

Carlos Ivorra Castillo

GEOMETRÍA

La geometría ilumina el intelecto y temple la mente. Todas sus pruebas son claras y ordenadas. Apenas caben errores en el razonamiento geométrico, pues está bien dispuesto y ordenado. Así, no es probable que la mente que se aplica a la geometría con regularidad cometa errores. De este modo, quien sabe geometría adquiere inteligencia.

IBN KHALDUN

Índice General

Preámbulo	vii
Introducción	ix
Capítulo I: La geometría absoluta	1
1.1 Axiomas de incidencia	2
1.2 Axiomas de ordenación	4
1.3 Dimensión	13
1.4 Ángulos y triángulos	21
1.5 Axiomas de congruencia	24
1.6 Suma de ángulos	29
1.7 Más sobre segmentos, ángulos y triángulos	32
1.8 Perpendiculares	33
1.9 Círculos y circunferencias	37
Capítulo II: La geometría euclídea	45
2.1 La geometría arquimediana	45
2.2 Longitud de segmentos. Los números reales	50
2.3 Amplitud de ángulos	59
2.4 Arcos y sectores circulares	65
2.5 El axioma de continuidad	67
2.6 El axioma de las paralelas	72
2.7 Los teoremas de Tales y Pitágoras	77
2.8 Las razones trigonométricas	82
Capítulo III: La geometría analítica	87
3.1 Vectores	87
3.2 Espacios afines	95
3.3 Espacios euclídeos	107
3.4 Circunferencias	118
3.5 Trigonometría	120

Capítulo IV: Números complejos y cuaternios	135
4.1 La geometría de los números complejos	135
4.2 Cuaternios	142
4.3 La geometría de los cuaternios	148
Capítulo V: Bijecciones afines	157
5.1 El grupo afín y el grupo lineal	157
5.2 Homotecias	159
5.3 El teorema fundamental de la geometría afín	167
5.4 Isometrías y semejanzas	171
5.5 Clasificación de isometrías	177
5.6 Orientación	185
Capítulo VI: Construcciones con regla y compás	193
6.1 Constructibilidad	193
6.2 Construcción de polígonos regulares	200
6.3 Construcciones con regla	207
6.4 Construcciones con compás	216
6.5 Construcciones con regla marcada	218
6.6 Construcciones con regla marcada y compás	227
6.7 Construcciones con regla y transportador	233
Capítulo VII: La geometría afín	239
7.1 Los axiomas de la geometría afín	240
7.2 Variedades afines	244
7.3 Introducción de coordenadas	250
7.4 Los axiomas de ordenación	266
7.5 Los axiomas de congruencia	269
7.6 Planos afines finitos	274
Capítulo VIII: La geometría proyectiva	281
8.1 Los axiomas de la geometría proyectiva	282
8.2 Espacios proyectivos analíticos	293
8.3 Geometría proyectiva analítica y sintética	310
8.4 Dualidad	314
8.5 Perspectividades	325
8.6 Cuadriláteros completos	328
8.7 Espacios sobre cuerpos ordenados	333
8.8 Planos proyectivos finitos	339
Capítulo IX: Cónicas	347
9.1 Cuádricas proyectivas	348
9.2 Cónicas proyectivas	359
9.3 Cónicas afines	366
9.4 Homografías entre haces de rectas	374
9.5 El teorema de Steiner	377

Capítulo X: La geometría parabólica	389
10.1 Espacios parabólicos	391
10.2 Espacios euclídeos	399
10.3 Cónicas euclídeas	406
10.4 Homografías de la esfera	419
Capítulo XI: La geometría hiperbólica	423
11.1 El plano hiperbólico	423
11.2 Medida de segmentos y ángulos	428
11.3 Trigonometría hiperbólica	433
11.4 El modelo de Poincaré	439
11.5 Las isometrías hiperbólicas	442
11.6 La geometría diferencial hiperbólica	447
Capítulo XII: La geometría elíptica	459
12.1 El plano elíptico	459
12.2 Biláteros y triángulos	462
12.3 Isometrías elípticas	466
12.4 Trigonometría elíptica	467
12.5 La geometría diferencial elíptica	473
Apéndice A: La geometría inversiva	483
A.1 La proyección estereográfica	484
A.2 Transformaciones esféricas	486
A.3 La geometría circular	490
Apéndice B: El código de Hamming	497
Índice de Materias	503

Preámbulo

Este libro forma parte de la “tercera edición” de mis libros de *Álgebra* [Al], *Geometría* [G] y *Análisis Matemático* [An], de los que he separado la *Teoría de grupos* en un nuevo volumen [TG]. La diferencia esencial es que he suprimido muchos ejemplos y aplicaciones que ahora se encuentran en la serie de libros “introdutorios” *Introducción a la teoría algebraica de números* [ITAl], *Introducción a la geometría euclídea* [IGE], *Introducción a la teoría analítica de números* [ITAn] e *Introducción al cálculo diferencial* [IC] y los he sustituido por contenidos nuevos.

Como en la edición anterior, los contenidos están distribuidos entre los cuatro libros de modo que pueden leerse simultáneamente siguiendo el orden que muestra el esquema de la página siguiente.

El primer capítulo de [Al] es una introducción a la teoría de conjuntos, cuyos aspectos más técnicos (los relacionados con el axioma de elección y la teoría de cardinales infinitos) se han relegado a dos apéndices. La teoría descrita es la teoría de Zermelo, que resulta más que suficiente para formalizar los contenidos de los cuatro libros. El único inconveniente es que “se queda corta” para desarrollar plenamente la teoría de cardinales infinitos, pero hemos preferido reducirla a lo imprescindible, aun al precio de no poder enunciar con total precisión algunos resultados sobre rango y dimensión de módulos y espacios vectoriales de dimensión infinita que, aunque resulta natural presentarlos al tratar estos conceptos, no son realmente necesarios en ningún momento.

El libro de *Álgebra* consta ahora (tras el capítulo [Al I] de fundamentos) de un primer bloque de cinco temas con los contenidos básicos del “álgebra abstracta” (incluyendo el álgebra lineal) y un segundo bloque de aplicaciones y resultados más avanzados.

El libro de *Geometría* empieza con dos capítulos que exponen la geometría euclídea a partir de unos axiomas al estilo de Hilbert, seguidos de un segundo bloque de cuatro capítulos dedicados a la geometría analítica y sus aplicaciones y un tercer bloque en el que analizamos más a fondo las geometrías afín y euclídea estudiadas en el bloque precedente e introducimos la geometría proyectiva y las geometrías no euclídeas.

El libro de *Análisis* consta de un primer bloque con los preliminares topológicos, un segundo bloque con los hechos básicos del cálculo diferencial e integral y un tercer bloque con temas más avanzados.

ÁLGEBRA	GEOMETRÍA	ANÁLISIS	GRUPOS
Al I Conjuntos			
Al II Anillos	G I G. absoluta		
Al III Aritmética	G II G. euclídea	An I Números reales	TG I Elementos
Al IV Módulos	G III G. analítica	An II Topología I	TG II Permutaciones I
Al V Cuerpos I	G IV Cuaternios	An III Topología II	TG III Estructura
Al VI Álgebra lineal	G V Bijecciones afines	An IV Tª de la medida I	TG IV Resolubles
Al VII Ecuaciones	G VI Regla y compás	An V Calc. diferencial	TG V Nilpotentes
Al VIII Enteros algebraicos	G VII G. afin	An VI Tª de la medida II	TG VI Caracteres
Al IX Cuerpos II	G VIII G. proyectiva	An VII Variedades	TG VII Permutaciones II
Al Ap A Ax. de elección	G IX Cónicas	An VIII Cál. vectorial	TG VIII Clásicos I
Al Ap B Ctos. infinitos	G X G. parabólica	An IX An. armónico	TG IX Clásicos II
Al Ap C Cuadrados latinos	G XI G. hiperbólica	An X Holomorfos	TG Ap A Banach-Tarski
	G XII G. elíptica	An Ap A Sólidos rígidos	
	G Ap A G. Inversiva	An Ap B Gravitación	
	G Ap B Hamming		

Finalmente, el libro de Teoría de grupos aparece dividido en el esquema anterior en un primer bloque con la teoría básica y un segundo bloque con contenidos más avanzados.

Salvo los dos apéndices sobre teoría de conjuntos de [Al], los demás apéndices contienen aplicaciones que, por motivos diversos, era preferible exponer unificadamente en lugar de dejarlas dispersas por el texto.

Remitimos a las introducciones de cada uno de los libros para una panorámica general más detallada de sus contenidos.

Introducción

En [IGE] presentamos los resultados básicos de la geometría euclídea y una selección de resultados de carácter más avanzado. Todos están demostrados a partir de hechos intuitivamente evidentes, pero, desde un punto de vista teórico, es interesante constatar que, de entre esos hechos intuitivamente evidentes que no requieren de ningún argumento para convencerse de su validez, es posible seleccionar unos pocos de forma que todos los demás pueden demostrarse formalmente a partir de ellos. Esto es lo que se llama “axiomatizar” la geometría.

La geometría axiomática Durante siglos, los matemáticos creyeron que los *Elementos* de Euclides contenían una axiomatización perfecta de la geometría, en el sentido de que todos los teoremas expuestos en ellos eran, en efecto, consecuencias lógicas de sus “postulados” y “nociones comunes”, presentados sin demostración. Sin embargo, a lo largo del siglo XIX diversos matemáticos fueron señalando defectos lógicos en los *Elementos*, es decir, señalaron hechos que Euclides usaba en sus demostraciones y que, aun siendo intuitivamente evidentes sin lugar a discusión, lo cierto era que no eran consecuencias lógicas de sus postulados.

La primera axiomatización de la geometría euclídea tridimensional rigurosa según los estándares modernos la publicó Hilbert en 1899. En los dos primeros capítulos de este libro estudiaremos una axiomática ligeramente distinta, cuyos axiomas son los recogidos en la página siguiente. Los discutiremos con detalle en su momento, pero los presentamos aquí para que el lector se haga una primera idea superficial del aspecto que tienen.

Desde un punto de vista lógico, pese a su similitud superficial, hay una diferencia sustancial entre la axiomática de Hilbert y la que estudiaremos aquí, y es que Hilbert planteó su axiomática como una teoría lógica independiente de cualquier otra. Sus axiomas debían concebirse como afirmaciones sobre varios *conceptos primitivos* no definidos, que en nuestra axiomática serían ocho:

- Punto, recta, plano.
- Las relaciones de incidencia “el punto A está en la recta r ” ($A \in r$) y “el punto A está en el plano P ” ($A \in P$).
- La relación $A - B - C$ (el punto B está entre los puntos A y C).
- La congruencia de segmentos y la congruencia de ángulos.

Axiomas de la geometría euclídea tridimensional

A1	Por cada par de puntos distintos P y Q pasa una única recta.
A2	Toda recta pasa al menos por dos puntos.
A3	Por cada tres puntos no colineales P, Q, R pasa un único plano.
A4	Si una recta tiene dos puntos en un plano, está contenida en él.
A5	Existen cuatro puntos no coplanares.
A6	Si dos planos tienen un punto en común, tienen dos puntos en común.
B1	Si se cumple $A - B - C$, entonces A, B y C son tres puntos colineales distintos dos a dos, y además se cumple $C - B - A$.
B2	Si A y B son puntos distintos, existe otro C tal que $A - B - C$.
B3	Dados tres puntos distintos en una misma recta, uno y sólo uno de ellos está entre los otros dos.
B4	Sean A, B, C tres puntos no colineales y sea r una recta contenida en el plano ABC , pero que no pase por ninguno de ellos. Si r pasa por un punto situado entre A y B , entonces r pasa por un punto entre A y C o bien por un punto entre B y C .
C1	Las congruencias de segmentos y ángulos son relaciones de equivalencia.
C2	Dados puntos $A \neq B$ y A' y una semirrecta s de origen A' , existe un único punto B' en s tal que $\overline{AB} \equiv \overline{A'B'}$.
C3	Sean A, B, C puntos colineales de modo que B esté entre A y C , sean A', B' y C' otros tres puntos en las mismas condiciones. Entonces, si $\overline{AB} \equiv \overline{A'B'}$ y $\overline{BC} \equiv \overline{B'C'}$, también $\overline{AC} \equiv \overline{A'C'}$.
C4	Sea L un ángulo, s una semirrecta y Π un semiplano cuya frontera sea la prolongación de s . Entonces existe un único ángulo L' contenido en Π , con un lado igual a s y tal que $L \equiv L'$.
C5	Dado un triángulo \widehat{ABC} , un segmento $\overline{A'B'} \equiv \overline{AB}$ y un semiplano Π de frontera la prolongación de $\overline{A'B'}$, existe un (único) triángulo $\widehat{A'B'C'}$ contenido en Π y congruente con \widehat{ABC} .
D	Si una recta r puede partirse como $r = s_1 \cup s_2$, donde s_1 y s_2 son dos conjuntos convexos disjuntos no vacíos, entonces existe un punto $P \in r$ tal que uno de los conjuntos s_i es una de las semirrectas determinadas por P en r y el otro es la semirrecta complementaria menos P .
E	Por cada punto exterior a una recta pasa una única paralela.

Los axiomas mencionan algunos conceptos más que pueden definirse a partir de estos ocho, por ejemplo, cuatro puntos son *coplanares* si están en un mismo plano, etc.

Todos los teoremas del sistema de Hilbert pueden deducirse *formalmente* a partir de sus axiomas en el sentido de que, desde un punto de vista lógico, es irrelevante el posible significado de los conceptos primitivos. No importa lo que sea un punto o una recta, etc., pues si los puntos, las rectas, etc. —sean lo que sean— cumplen cada uno de los axiomas de Hilbert, entonces, —sean lo que sean— cumplen también cada uno de los teoremas de la geometría euclídea.

Sin embargo, nuestro planteamiento no es exactamente éste, es decir, no vamos a presentar una axiomática “autosuficiente” de la geometría euclídea, sino que la vamos a sumergir en la teoría de conjuntos (la teoría determinada por los axiomas descritos en la sección [A1 1.2]). En estos términos, los axiomas de la tabla anterior no son axiomas lógicos en el sentido propio de la palabra, sino que técnicamente son parte de una definición conjuntista:

Un *espacio euclídeo tridimensional* es un conjunto E (a cuyos elementos llamaremos *puntos*) en el que hemos fijado dos familias de subconjuntos \mathcal{R} y \mathcal{P} , a cuyos elementos llamaremos respectivamente *rectas* y *planos*, así como una relación ternaria $R \subset E^3$ (de modo que escribiremos $A - B - C$ en lugar de $(A, B, C) \in R$) y dos relaciones binarias definidas, respectivamente sobre el conjuntos de todos los *segmentos* y el todos los *ángulos*¹ en E , de modo que se cumplen los axiomas de la tabla precedente.

Esta diferencia es esencial en la formulación del axioma D, que, tal cual lo hemos enunciado, no tendría sentido en el planteamiento de Hilbert, ya que habla de dos conjuntos arbitrarios s_1 y s_2 , lo cual presupone los axiomas de la teoría de conjuntos. En resumen: en la sección [A1 1.2] presentaremos unos axiomas que nos permiten hablar formalmente de conjuntos en general, es decir, que nos permiten probar teoremas sobre conjuntos sin necesidad de especificar a qué llamamos “conjunto”, sino que si unos objetos cualesquiera cumplen los axiomas de la teoría de conjuntos, entonces, sean lo que sean, cumplen también todos los teoremas de la teoría de conjuntos, y aquí definimos un espacio tridimensional euclídeo como una selección de conjuntos que cumplen ciertas propiedades expresadas en los 17 “axiomas” precedentes. Esto convierte a la geometría euclídea en una teoría integrada en el resto del *corpus* matemático y no como una teoría aislada.

En el capítulo I, estudiaremos las consecuencias de los tres primeros grupos de axiomas, mientras que en el capítulo II veremos que el axioma D nos permite medir segmentos y ángulos mediante números reales (definidos a partir de los axiomas de la teoría de conjuntos). El axioma E está separado del resto porque, tal y como explicamos en la introducción de [IGE], es equivalente al “quinto postulado de Euclides”, y tiene interés saber qué parte de la geometría euclídea puede demostrarse sin él. Dicha parte es lo que se conoce como *geometría absoluta*.

Sin embargo, si uno quiere entender la geometría euclídea sin preocuparse de su relación con las geometrías no euclídeas, lo natural es considerar al axioma E como un axioma más del grupo básico A, pues, como los demás axiomas de dicho grupo, habla únicamente de las relaciones de incidencia, que en nuestra formalización de la geometría es simplemente la pertenencia conjuntista de los puntos a las rectas y a los planos.

¹Los conceptos de “segmento” y “ángulo” pueden definirse a partir de los conceptos precedentes. Por ejemplo, el segmento de extremos $A, B \in E$ se define como el conjunto $\overline{AB} = \{P \in E \mid A - P - B\}$.

La geometría analítica En el apéndice A de [IGE] (y en el apéndice A de [IC] para el caso tridimensional) vimos que cada punto del plano (o del espacio) puede determinarse mediante un par (o una terna) de coordenadas reales, de modo que todos los conceptos geométricos pueden caracterizarse en términos algebraicos a través de las coordenadas de los puntos implicados.

En la sección 3.1 demostraremos esto mismo a partir de nuestros axiomas, y la sección 3.2 la dedicaremos a estudiar la estructura algebraica de *espacio afín*, que generaliza la idea de “espacio geométrico en el que cada punto está determinado por tres coordenadas reales” permitiendo que el número de coordenadas de cada punto (es decir, la dimensión del espacio) sea arbitrario, al igual que el cuerpo en el que varían las coordenadas, con lo que podemos hablar de un espacio afín de dimensión n sobre un cuerpo K . Así, en estos términos, podemos decir que el espacio tridimensional euclídeo es un espacio afín de dimensión 3 sobre el cuerpo \mathbb{R} de los números reales. Explícitamente, la definición de espacio afín es la siguiente:

Un *espacio afín* de dimensión n es una terna $(E, \vec{E}, +)$, donde E es un conjunto no vacío, a cuyos elementos se les llama *puntos*, \vec{E} es un espacio vectorial de dimensión n sobre un cuerpo K y $+$ es una aplicación $E \times \vec{E} \rightarrow E$ que cumple las propiedades siguientes:

1. Para cada par de puntos P, Q , existe un único vector \vec{v} tal que $Q = P + \vec{v}$, y se le representa por $\vec{v} = \overrightarrow{PQ}$.
2. $P + \vec{0} = P$,
3. $(P + \vec{v}) + \vec{w} = P + (\vec{v} + \vec{w})$.

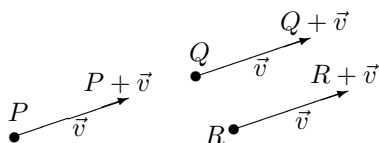
Esto presupone la definición de “espacio vectorial”, que es un conjunto \vec{E} , a cuyos elementos se les llama *vectores*, en el que hay definida una operación suma $+$: $\vec{E} \times \vec{E} \rightarrow \vec{E}$ y un producto de un *escalar* (un elemento de K) por un vector: \cdot : $K \times \vec{E} \rightarrow \vec{E}$, de modo que se cumplan todas las propiedades que uno puede esperar que se cumplan, y que recogemos en la tabla siguiente:

Axiomas de espacio vectorial

-
1. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ para todos los $\vec{u}, \vec{v}, \vec{w} \in \vec{E}$.
 2. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ para todos los $\vec{u}, \vec{v} \in \vec{E}$.
 3. Existe un elemento $\vec{0} \in \vec{E}$ tal que $\vec{v} + \vec{0} = \vec{v}$ para todo $\vec{v} \in \vec{E}$.
 4. Para todo $\vec{u} \in \vec{E}$ existe un $-\vec{u} \in \vec{E}$ tal que $\vec{u} + (-\vec{u}) = \vec{0}$.
 5. $\alpha(\vec{u} + \vec{v}) = \alpha\vec{u} + \alpha\vec{v}$, para todo $\alpha \in K$ y todos los $\vec{u}, \vec{v} \in \vec{E}$.
 6. $(\alpha + \beta)\vec{v} = \alpha\vec{v} + \beta\vec{v}$, para todos los $\alpha, \beta \in K$ y todo $\vec{v} \in \vec{E}$.
 7. $\alpha(\beta\vec{v}) = (\alpha\beta)\vec{v}$ para todos los $\alpha, \beta \in K$ y todo $\vec{v} \in \vec{E}$.
 8. $1\vec{v} = \vec{v}$, para todo $\vec{v} \in \vec{E}$.
-

Nuevamente, enumeramos estas propiedades para que el lector se haga una idea de su aspecto, pero la idea general con la que el lector debe quedarse ahora es simplemente que un espacio afín está determinado por un conjunto de puntos E y un conjunto de vectores \vec{E} , de modo que cada vector \vec{v} nos permite trasladar cada punto P a otro punto $Q = P + \vec{v}$, de modo que hay un único vector que traslada un punto dado a otro punto dado.

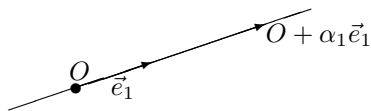
demostrar que un espacio euclídeo tridimensional E es un espacio afín de dimensión 3 sobre \mathbb{R} supone construir un espacio vectorial \vec{E} y definir la suma de un punto y un vector, y al hacerlo veremos cuál es la interpretación geométrica de esta suma:



Cuando sumamos un mismo vector \vec{v} no nulo a puntos distintos, los puntos P y $P + \vec{v}$ están siempre sobre una recta de un mismo haz de rectas paralelas determinado por \vec{v} (lo que podemos llamar la *dirección* de \vec{v}), de modo que $P + \vec{v}$ está a una distancia fija de P , también determinada por \vec{v} (lo que se llama el *módulo* de \vec{v}) y además $P + \vec{v}$ se obtiene moviendo P en la dirección correspondiente a \vec{v} siempre en un mismo sentido, también determinado por \vec{v} . Así pues, cada vector no nulo representa una traslación posible de los puntos de E a una distancia determinada y en una dirección y un sentido determinados.

Los elementos de K se llaman escalares porque aplican una “escala” los vectores. Por ejemplo (en el caso $K = \mathbb{R}$), el vector $3\vec{v}$ es el vector que tiene la misma dirección y el mismo sentido que \vec{v} , pero que traslada los puntos a una distancia tres veces mayor, mientras que $-3\vec{v}$ además invierte el sentido de la traslación. Por último, una suma $\vec{u} + \vec{v}$ de dos vectores representa la traslación que resulta de trasladar un punto primero según \vec{u} y, a continuación, trasladar el resultado según \vec{v} .

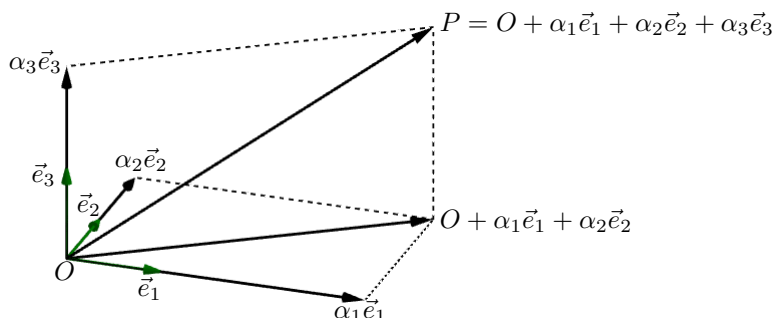
En estos términos, una recta está determinada por uno de sus puntos O y un *vector director* no nulo \vec{e}_1 , de forma que los puntos de la recta son los de la forma $O + \alpha_1 \vec{e}_1$, donde el escalar α_1 varía en \mathbb{R} . En otras palabras, los puntos de una recta son los que se pueden obtener trasladando un punto P cualquier distancia en una dirección fija, pero en cualquiera de los dos sentidos posibles:



E igualmente, un plano está determinado por uno de sus puntos O y dos *vectores directores* \vec{e}_1 y \vec{e}_2 no nulos que tengan direcciones distintas, de forma que sus puntos son los de la forma $O + \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2$, para escalares arbitrarios α_1 y α_2 , es decir, los que pueden obtenerse a partir de P mediante traslaciones en dos direcciones distintas (pero no en tres).

Si consideramos tres vectores $\vec{e}_1, \vec{e}_2, \vec{e}_3$ que tengan, no ya direcciones distintas, sino que sean *independientes*, en el sentido que de al movernos en la dirección de \vec{e}_3 nos salimos del plano determinado por \vec{e}_1 y \vec{e}_2 , entonces podemos llegar a todo punto del espacio, es decir, fijado un punto cualquiera O , todo punto del espacio se puede expresar de forma única como

$$P = O + \overrightarrow{OP} = O + \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \vec{e}_3,$$



es decir, que podemos llegar hasta cualquier punto del espacio partiendo de un punto arbitrario O y combinando traslaciones en tres direcciones prefijadas, y esto es precisamente lo que significa que el espacio E tiene tres dimensiones. En términos del espacio vectorial \vec{E} se puede expresar así:

Un espacio vectorial \vec{E} sobre un cuerpo K tiene *dimensión* 3 si tiene una *base* formada por 3 vectores $\vec{e}_1, \vec{e}_2, \vec{e}_3$, lo cual significa a su vez que cualquier vector $\vec{v} \in \vec{E}$ se expresa de forma única como

$$\vec{v} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \vec{e}_3,$$

para ciertos escalares $\alpha_1, \alpha_2, \alpha_3$.

En términos del espacio E , esto significa que, fijado un *origen de coordenadas* $O \in E$ y una base $\vec{e}_1, \vec{e}_2, \vec{e}_3$ de \vec{E} , cada punto $P \in E$ se expresa de forma única como

$$P = O + \overrightarrow{OP} = O + \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \vec{e}_3,$$

para ciertos escalares $(\alpha_1, \alpha_2, \alpha_3) \in K^3$ a los que podemos llamar *coordenadas* del punto P respecto del *sistema de referencia* $(O; \vec{e}_1, \vec{e}_2, \vec{e}_3)$.

La condición geométrica para que tres vectores de \vec{E} formen una base es la que ya hemos indicado: \vec{e}_1 puede ser cualquier vector no nulo, \vec{e}_2 tiene que ser un vector no nulo de dirección distinta, y \vec{e}_3 tiene que ser un vector que no pueda expresarse en la forma $\alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2$, es decir, que, al movernos en la dirección de \vec{e}_3 , salgamos del plano determinado por los dos vectores anteriores.

En este punto (o tal vez mejor cuando estudie el capítulo III), el lector familiarizado con la geometría analítica rudimentaria expuesta en los apéndices A de [IGE] e [IC], debería reflexionar sobre las enormes diferencias conceptuales

entre el concepto de “vector” considerado allí y el que estamos considerando ahora. Allí, un vector no era más que un par o una terna de números reales, de modo que, ante una terna $(\alpha_1, \alpha_2, \alpha_3)$, podíamos considerarla como un punto o como un vector según lo que hiciéramos con ella, mientras que, en un espacio afín, puntos y vectores son conceptos radicalmente distintos (elementos de E y elementos de \vec{E} , respectivamente). En particular, ni un punto ni un vector son ternas de números reales. A cada vector le podemos asociar una terna de números reales en cuanto fijamos una base de \vec{E} , pero no tenemos ningún criterio especial para seleccionar ninguna base en particular de \vec{E} , y según qué base elijamos, un mismo vector tendrá unas coordenadas u otras.

Por ejemplo, si $\vec{e}_1, \vec{e}_2, \vec{e}_3$ es una base de \vec{E} , las coordenadas de \vec{e}_1 en dicha base son $(1, 0, 0)$, pues, ciertamente, $\vec{e}_1 = 1 \cdot \vec{e}_1 + 0 \cdot \vec{e}_2 + 0 \cdot \vec{e}_3$, pero es fácil ver que $-\vec{e}_1, \vec{e}_2, \vec{e}_3$ también forman una base de \vec{E} y, respecto de esta segunda base, las coordenadas de \vec{e}_1 son $(-1, 0, 0)$.

Similarmente, no podemos hablar de coordenadas de un punto de E si no hemos fijado previamente un sistema de referencia, determinado por un origen O y una base $\vec{e}_1, \vec{e}_2, \vec{e}_3$.

Además, si P, Q son dos puntos de un espacio afín, la suma $P + Q$ no tiene sentido, no está definida, como tampoco está definido $3P$, mientras que la suma $\vec{u} + \vec{v}$ de dos vectores de \vec{E} siempre está definida, al igual que el producto $3\vec{u}$. Esto muestra que los conceptos “punto” y “vector” no son en modo alguno equiparables en este contexto. Otra cosa es que, una vez hemos fijado un sistema de referencia, tanto los puntos de E como los vectores de \vec{E} vienen determinados por ternas de números reales, pero una cosa es que una terna esté “identificando” a un punto o que esté “identificando” a un vector. A menudo será relevante tener claro si una definición geométrica depende de la expresión en coordenadas de unos puntos o vectores considerados o si, por el contrario, depende únicamente de los puntos y vectores en sí mismos, sin que sea relevante que sus coordenadas sean unas u otras en tal o cual sistema de referencia.

Hemos definido lo que significa que un espacio vectorial \vec{E} tenga dimensión 3, pero es evidente que en la definición podemos sustituir el 3 por cualquier otro número natural, de modo que la *geometría afín* es el estudio de los espacios afines de dimensión arbitraria n sobre un cuerpo arbitrario K .

Esto supone una doble generalización de la geometría tridimensional euclídea en dos sentidos distintos (dimensión arbitraria y cuerpo arbitrario), pero conviene enfatizar que incluso la geometría afín tridimensional sobre el cuerpo \mathbb{R} de los números reales es una versión débil —una abstracción— de la geometría tridimensional euclídea.

En efecto, si E es un espacio afín sobre \mathbb{R} , en él podemos hablar de puntos, rectas, planos, rectas paralelas, semirectas, semiplanos, segmentos, ángulos, etc., y demostrar que cumplen todo lo que cabe esperar que cumplan, pero no podemos hablar, por ejemplo, de “rectas perpendiculares” ni de “longitud de un segmento” ni de “amplitud de un ángulo”. Ni tampoco podemos hablar, por ejemplo, del módulo de un vector.

Por ejemplo, podemos decir que dos rectas de E son paralelas cuando admiten un mismo vector director (o, si fijamos en cada una un vector director arbitrario, si el vector director de una es un múltiplo del de la otra), pero, si tenemos dos rectas que pasan por un mismo punto P con vectores directores \vec{u} y \vec{v} , ¿qué significa que sean rectas perpendiculares? ¿Qué tienen que cumplir los vectores \vec{u} y \vec{v} para que sus direcciones sean perpendiculares?

Similarmente, cuando hemos interpretado el concepto de “vector” determinado por un espacio euclídeo tridimensional, hemos dicho que cada vector tiene asociado un módulo, que representa la distancia a la que traslada cada punto, pero eso tiene sentido porque estamos hablando de un espacio E que cumple los 17 axiomas de la geometría euclídea, a partir de los cuales se puede definir la longitud de un segmento, que es esencialmente lo mismo que la distancia entre dos puntos, pero si sólo sabemos que E es un espacio afín euclídeo sobre \mathbb{R} (tridimensional, si queremos, aunque esto no ayuda en nada), ¿qué es el módulo de un vector $\vec{v} \in \vec{E}$? o, equivalentemente, ¿qué es la distancia de un punto P a otro punto $P + \vec{v}$? Estas preguntas no tienen respuesta.

Lo que sucede es que, a partir de los 17 axiomas de la geometría euclídea tridimensional, no sólo se puede probar que el espacio E es un espacio afín tridimensional sobre \mathbb{R} , sino que podemos definir el módulo de un vector \vec{v} como la longitud del segmento de extremos P y $P + \vec{v}$ (que no depende del punto P elegido) y podemos establecer que dos vectores \vec{u} y \vec{v} son *ortogonales* si las rectas que pasan por un punto P con dichos vectores directores son perpendiculares, pero todo esto se define a partir de la estructura euclídea de E , no a partir de su estructura afín.

Más precisamente: si E es un espacio afín de dimensión 3 sobre un cuerpo K , podemos definir en él los conceptos de recta y plano de modo que cumplan los axiomas del grupo A y el axioma de las paralelas E , y si además el cuerpo K está ordenado es posible definir en E una relación “estar entre” de modo que se cumplan también los axiomas del grupo B, pero, aunque tomemos concretamente $K = \mathbb{R}$, la estructura afín no nos permite definir unas nociones de congruencia de segmentos y ángulos que nos permitan demostrar los axiomas del grupo C, de los cuales dependen las nociones de “longitud” o “perpendicularidad”.

Las relaciones de congruencia requieren fijar en \vec{E} una estructura algebraica adicional, determinada por lo que se llama un producto escalar:

Un *producto escalar* en un espacio vectorial \vec{E} sobre \mathbb{R} es una operación $\vec{E} \times \vec{E} \rightarrow \mathbb{R}$ que cumpla las propiedades siguientes:

1. $\vec{v}\vec{v} \geq 0$ y $\vec{v}\vec{v} = 0$ si y sólo si $\vec{v} = 0$,
2. $\vec{u}(\vec{v} + \vec{w}) = \vec{u}\vec{v} + \vec{u}\vec{w}$,
3. $(\alpha\vec{u})\vec{v} = \alpha(\vec{u}\vec{v}) = \vec{u}(\alpha\vec{v})$,
4. $\vec{u}\vec{v} = \vec{v}\vec{u}$.

Un *espacio vectorial euclídeo* es un espacio vectorial en el que hemos fijado un producto escalar, y un *espacio afín euclídeo* es un espacio afín E tal que en el espacio vectorial \vec{E} hemos fijado un producto escalar.

En la sección 3.3 veremos que si E es un espacio euclídeo tridimensional, no sólo es posible definir un espacio vectorial \vec{E} sobre \mathbb{R} con el cual E se convierte en un espacio afín tridimensional sobre \mathbb{R} , sino que veremos que la geometría de E también determina un producto escalar en \vec{E} , de modo que —ahora sí— todos los conceptos de la geometría euclídea pueden ser caracterizados algebraicamente mediante la estructura de espacio afín euclídeo tridimensional. Por ejemplo, dos vectores son ortogonales (es decir, definen rectas perpendiculares) si $\vec{u} \cdot \vec{v} = 0$ y el módulo de un vector se puede definir como $\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}}$.

En resumen:

- La geometría tridimensional euclídea es equivalente a todos los efectos a la geometría de los espacios afines euclídeos tridimensionales (sobre el cuerpo \mathbb{R} de los números reales).
- Una primera generalización consiste en eliminar la condición de que la dimensión del espacio sea precisamente 3, que equivale a considerar espacios afines euclídeos de dimensión arbitraria n (sobre el cuerpo \mathbb{R} de los números reales). Tenemos así la *geometría euclídea de dimensión arbitraria*.
- Una segunda generalización es la que resulta de eliminar el producto escalar y estudiar los espacios afines de dimensión arbitraria sobre el cuerpo \mathbb{R} de los números reales.

Tenemos así una geometría más débil en la que no están definidos los conceptos de congruencia, ortogonalidad, longitud de segmentos o medida de ángulos.

- Una tercera generalización consiste en considerar un cuerpo arbitrario K en lugar del cuerpo de los números reales. Tenemos así la *geometría afín* en general.

Por razones que enseguida explicaremos, conviene considerar una geometría afín un poco más general, en la cual K no sea necesariamente un cuerpo, sino meramente un *anillo de división*, que es lo mismo que un cuerpo salvo que no exigimos que el producto sea conmutativo.

En principio, el concepto de “espacio afín” es un concepto puramente algebraico, que en esencia no es más que un par de conjuntos en los que hay definidas unas operaciones que cumplen muchas propiedades asociativas, conmutativas, etc. Sin embargo, en el capítulo VII veremos que los espacios afines arbitrarios pueden ser definidos en términos puramente geométricos, pues la geometría afín (levemente generalizada como explicaremos enseguida) es equivalente a la geometría determinada por los axiomas de incidencia (los axiomas del grupo A más el axioma E de las paralelas). Más precisamente, en el caso tridimensional, podemos adoptar la definición siguiente:

Un *espacio afín tridimensional* es un conjunto E (a cuyos elementos llamaremos *puntos*) en el que hemos fijado dos familias de subconjuntos no vacíos \mathcal{R} y \mathcal{P} , a cuyos elementos llamaremos respectivamente *rectas* y *planos*, de modo que se cumplen los axiomas siguientes:

A1	Por cada par de puntos distintos P y Q pasa una única recta.
A2	Por cada punto pasa una única paralela a una recta dada.
A3	Por cada tres puntos no colineales P, Q, R pasa un único plano.
A4	Si una recta tiene dos puntos en un plano, está contenida en él.
A5	Existen cuatro puntos no coplanares.
A6	Si dos planos tienen un punto en común, tienen dos puntos en común.

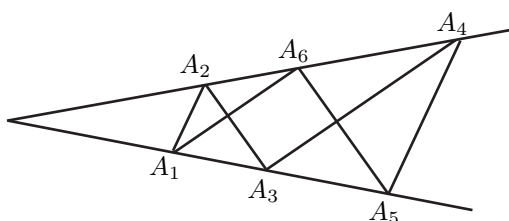
No es difícil ver que el axioma A2 original puede demostrarse a partir de los demás axiomas cuando incluimos entre ellos el axioma de las paralelas, por lo que simplemente hemos sustituido el axioma A2 por el axioma E.

En el capítulo VII probaremos que todo espacio afín en este sentido permite definir un anillo de división K de modo que E tiene estructura de espacio afín de dimensión 3 sobre K en el sentido algebraico.

Más aún, veremos que la tridimensionalidad del espacio viene determinada concretamente por el axioma A6, de modo que éste puede ser sustituido por otro que afirme que el espacio tiene cualquier dimensión n prefijada, es decir, que cada punto viene determinado por n coordenadas y no 3 (aunque para $n = 2$ debemos sustituir también A5 por la mera exigencia de que existan tres puntos no colineales).

No es posible demostrar que el producto en el anillo de división K tenga la propiedad conmutativa porque es fácil ver que si K es cualquier anillo de división, entonces K^3 , con las definiciones algebraicas usuales de recta y plano, satisface los seis axiomas de la geometría afín, pero sucede que la conmutatividad de K tiene una equivalencia puramente geométrica sorprendente: si E es un espacio afín en el sentido algebraico sobre un anillo de división K , se cumple que K es conmutativo si y sólo si E cumple el teorema de Pappus-Pascal:

Sean r y s dos rectas secantes y consideremos un hexágono tal que los vértices A_1, A_3, A_5 están sobre r y A_2, A_4, A_6 están sobre s (todos ellos distintos del punto de intersección). Si dos pares de lados opuestos son paralelos, el tercer par también lo es.



Veremos que este teorema se puede demostrar a partir de los axiomas de la geometría euclídea (en realidad sin necesidad del axioma D), pero no únicamente a partir de los axiomas de incidencia (incluyendo el axioma de las paralelas).

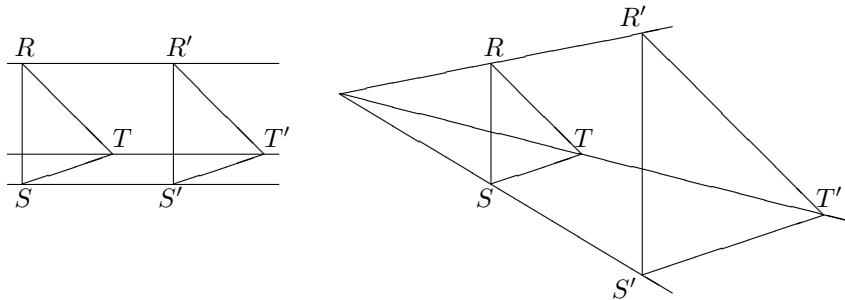
El hecho de que los axiomas de incidencia permitan dotar al espacio de una estructura de espacio afín en el sentido algebraico no es evidente en absoluto, hasta el punto de que esto no es necesariamente cierto en el caso de los planos

afines (es decir, en espacios afines de dimensión 2). Es fácil adaptar los axiomas de incidencia al caso bidimensional. Simplemente, A3, A4 y A6 se vuelven triviales, mientras que A5 hay que debilitarlo (y lo reenumeramos como A3), de modo que nos quedan los axiomas siguientes:

A1	Por cada par de puntos distintos P y Q pasa una única recta.
A2	Por cada punto pasa una única paralela a una recta dada.
A3	Existen tres puntos no colineales.

Lo que decimos es que no es cierto que un plano que cumpla estos axiomas admita necesariamente una estructura de plano afín en el sentido algebraico. Y la condición necesaria y suficiente para que esto suceda puede expresarse también en términos puramente geométricos. Resulta que un plano afín en el sentido geométrico es un plano afín algebraico sobre cierto anillo de división K si y sólo si satisface un caso particular del teorema de Desargues:

Sean r_1, r_2 y r_3 tres rectas distintas paralelas o concurrentes en un punto P . Sean R, R' puntos en r_1 , sean S, S' puntos en r_2 y T, T' puntos en r_3 distintos de P si las rectas se cortan. Si RS es paralela a $R'S'$ y RT es paralela a $R'T'$ entonces ST es paralela a $S'T'$.



Este teorema se puede demostrar en cualquier espacio afín en el sentido algebraico y en cualquier espacio afín de dimensión ≥ 3 en el sentido geométrico, pero hay planos afines en el sentido geométrico que no son *arguesianos*, es decir, que no satisfacen el teorema de Desargues y que, por consiguiente, no pueden ser dotados de una estructura algebraica de plano afín sobre ningún anillo de división.

La geometría proyectiva El teorema de Desargues que acabamos de enunciar no es exactamente el que demostramos en [IGE 4.59], sino una de las variantes demostradas en la nota posterior a dicho teorema. Dichas variantes son ejemplos de un fenómeno que nos apareció en [IGE] en varias ocasiones, y es que muchos resultados de la geometría euclídea válidos para rectas secantes siguen siendo ciertos para rectas paralelas (y viceversa) si entendemos que dos rectas paralelas se cortan en un “punto infinito”. En el capítulo VIII introducimos la geometría proyectiva, que permite plasmar y desarrollar esa idea con rigor.

Esencialmente, un espacio proyectivo P de dimensión n es un espacio afín E de la misma dimensión al que hemos añadido un punto nuevo por cada haz

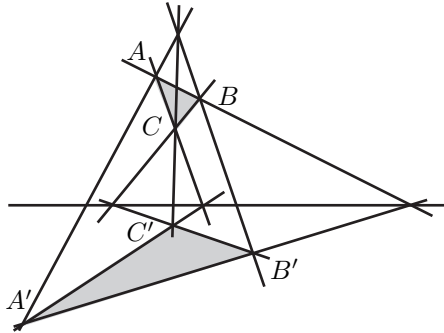
de rectas paralelas que hay en E , de modo que todas las rectas del haz se cortan ahora en su punto “infinito” correspondiente. Los espacios proyectivos, obtenidos de este modo, pueden caracterizarse por unos axiomas geométricos de incidencia que son variantes naturales de los axiomas de la geometría afín. Por ejemplo, un plano proyectivo puede definirse así:

Un *plano proyectivo* es un conjunto P , a cuyos elementos llamaremos *puntos*, en el que hemos fijado una familia de subconjuntos a los que llamaremos *rectas*, de modo que se cumplen los axiomas siguientes:

- P1** *Por cada par de puntos distintos pasa una única recta.*
P2 *Dos rectas cualesquiera tienen al menos un punto en común.*
P3 *Existen tres puntos no colineales.*
P4 *Toda recta tiene al menos tres puntos.*

Vemos que P2 niega la existencia de paralelas, pero esto no debe interpretarse como que la geometría proyectiva contradice a la geometría afín, sino que más bien la extiende: todo plano afín se convierte en un plano proyectivo si añadimos un punto “infinito” común a cada haz de rectas paralelas (de modo que los nuevos puntos infinitos forman una nueva “recta infinita”) y, recíprocamente, si a un plano proyectivo le quitamos los puntos de una recta cualquiera (y a cada recta le quitamos el punto en que corta a la recta eliminada) obtenemos un plano afín. El teorema de Desargues se enuncia de forma natural en el contexto de la geometría proyectiva:

Dados dos triángulos ABC y $A'B'C'$, las rectas AA' , BB' y CC' se cortan en un punto si y sólo si los puntos de corte de cada par de lados AB y $A'B'$, AC y $A'C'$, BC y $B'C'$ están alineados.

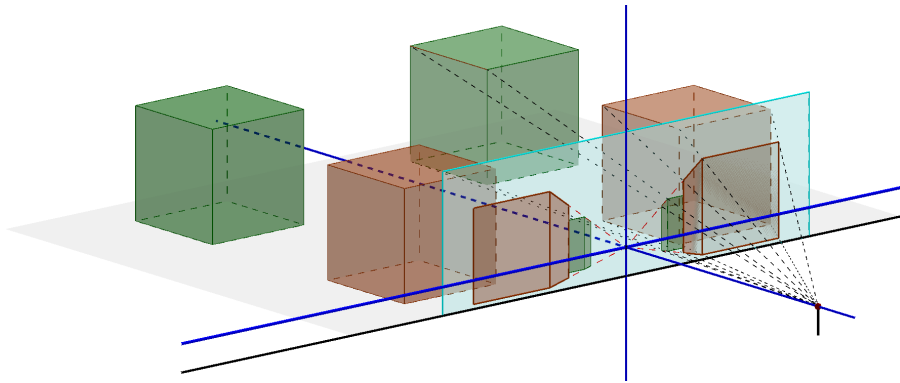


Este teorema puede demostrarse para espacios proyectivos de cualquier dimensión $n \geq 3$, mientras que los planos proyectivos pueden ser arguesianos o no arguesianos, según que cumplan o no el teorema de Desargues, y esto sucede si y sólo si el plano afín que resulta al eliminar una recta proyectiva es o no arguesiano. Como el plano (afín) de la geometría euclídea es arguesiano, también lo es el plano proyectivo que resulta de completarlo con puntos infinitos y, de hecho, la versión proyectiva del teorema de Desargues incluye en un único

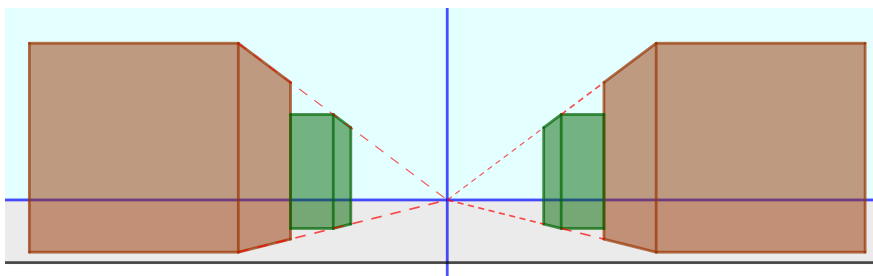
enunciado todas las variantes afines discutidas tras [IGE 4.59]. Por ejemplo, las rectas AA' , BB' y CC' pueden ser paralelas o concurrentes, según si se cortan en un punto finito o infinito, si la recta en la que se cortan los pares de lados es la recta infinita, es que los dos triángulos tienen sus lados paralelos dos a dos, pero también puede ser una recta infinita y que uno de los puntos de corte sea su punto infinito, en cuyo caso los triángulos sólo tienen un par de lados paralelos, etc., pero en todos los casos se cumple la equivalencia entre las dos afirmaciones.

La geometría proyectiva no surge únicamente como “compleción abstracta” de la geometría afín, sino que hay un contexto en el que aparece de forma natural, y así surgió históricamente: la pintura en perspectiva.

Imaginemos una persona situada ante varios edificios, que en la figura siguiente hemos representado esquemáticamente como cubos. Supongamos que entre los edificios y el observador se levanta un gran cristal transparente.



Podríamos considerar que el observador no está viendo los edificios, sino el cristal, de modo que si pintáramos cada punto del cristal con el color del rayo de luz que lo atraviesa, obtendríamos un cuadro opaco que proporcionaría la misma vista. La figura siguiente contiene dicho cuadro:



Las cuatro líneas discontinuas marcadas en el cuadro se corresponden con cuatro rectas paralelas en la imagen real, pero en el cuadro concurren en un punto. Ese punto de concurrencia no se corresponde con ningún punto real (pues las rectas no se cortan), sino que es una representación “tangible” del punto infinito común a esas cuatro rectas paralelas.

En el cuadro, el suelo y el cielo se cortan en una recta, la “línea del horizonte”. El cielo es algo un tanto “etéreo”, pero si por encima de los edificios se extendiera un techo paralelo al suelo que llegara hasta donde alcanza la vista, dicho plano paralelo al suelo se cortaría en el cuadro con el plano del suelo en la recta del horizonte. Dicha recta es la recta infinita en la que se cortan todos planos paralelos paralelos al suelo. Cualquier par de rectas paralelas contenidas en planos paralelos al suelo se cortan en un punto de la recta del horizonte, salvo si también son paralelas al cuadro, en cuyo caso en el cuadro se siguen viendo paralelas, porque su punto de corte es el punto infinito de la recta del horizonte, que no “cabe” en el cuadro.

La geometría necesaria para determinar el punto del cuadro en el que hay que representar un punto del espacio real no excede la expuesta en el apéndice A de [IC]:

Fijemos un sistema de referencia con origen O en el punto del cuadro que está justo enfrente del observador, a la altura de sus ojos.² Pongamos que el eje X es el eje horizontal contenido en el cuadro, el eje Y es el perpendicular al cuadro (y pasa por el ojo del observador) y el eje Z es el eje vertical. Así, el ojo del observador está en una posición $(0, -d, 0)$, donde d es la distancia del observador al cuadro.

Ahora consideremos un punto cualquiera $P = (x, y, z)$, con $y > 0$, para que se encuentre detrás del cuadro. La recta que une el punto P con el ojo del observador está formada por los puntos de coordenadas

$$(0, -d, 0) + \alpha(x, y + d, z)$$

(para $\alpha = 0$ tenemos el ojo, y para $\alpha = 1$ tenemos el punto P). Dicha recta corta al cuadro en el punto con segunda coordenada nula, de modo que el escalar α está determinado por la ecuación

$$-d + \alpha(y + d) = 0,$$

que nos da $\alpha = d/(y + d)$. Por lo tanto, las coordenadas (x', z') del punto del cuadro correspondiente a P son

$$(x', z') = \left(\frac{dx}{y + d}, \frac{dz}{y + d} \right)$$

Así pues, la correspondencia que transforma cada punto en su imagen en el cuadro es

$$\pi(x, y, z) = \left(\frac{dx}{y + d}, \frac{dz}{y + d} \right).$$

Ahora consideremos el problema siguiente: supongamos que el “cuadro” es una fotografía que hemos sacado de los edificios y que la tenemos ante nosotros, de modo que podemos medir sobre ella cualquier distancia que nos interese.

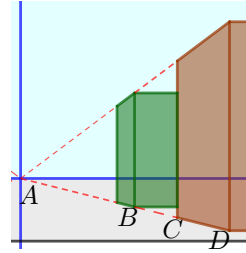
²Vamos a pasar por alto que el observador tiene dos ojos o, equivalentemente, vamos a despreciar la distancia entre sus dos ojos y considerar que ambos determinan un mismo punto.

Si sabemos que la fachada del primer edificio de la derecha (la arista del cubo) mide 10 metros, ¿podemos calcular la anchura de la calle que lo separa del edificio situado más al fondo? ¿Qué longitudes necesitamos medir sobre la fotografía para ello?

La geometría proyectiva nos da una respuesta muy simple a este problema.

Necesitamos medir las distancias AB , BC y CD indicadas en la figura. Notemos que el punto infinito A se determina con una regla prolongando las aristas de los cubos. Midiendo con una aplicación de dibujo sobre la figura original (no la versión reducida que aparece en esta página) hemos obtenido los valores:

$$AB = 9.15, \quad BC = 3.39, \quad CD = 4.1.$$



La geometría proyectiva nos permite afirmar que la anchura real L de la calle BC cumple la relación

$$\frac{CD}{AD} : \frac{BC}{AB} = \frac{10}{L}.$$

El miembro izquierdo da 0.665 (el valor exacto, calculado a partir de los datos con los que se ha construido la figura, es $2/3$) de donde concluimos que la anchura de la calle es $L = 15.03$ metros (el valor exacto es $L = 15$).

Ahora no estamos en condiciones de explicar cómo se obtiene la relación indicada, pero ponemos a continuación la explicación que el lector estará en condiciones de entender cuando haya estudiado el capítulo VIII:

Consideramos el espacio tridimensional contenido en un espacio proyectivo tridimensional. En él consideramos el plano H determinado por el cuadro y el plano H' determinado por el suelo. Sea O el punto donde se encuentra el ojo del observador. Entonces, la aplicación $\pi_O : H' \rightarrow H$ que a cada punto del suelo le hace corresponder el punto correspondiente en el cuadro es una proyección perspectiva (definición 8.50), luego es una homografía (teorema 8.51), luego conserva las razones dobles (está probado tras la definición 8.28). Según el teorema 8.30, tenemos que

$$|\mathcal{R}(D, B, C, A)| = \frac{CD}{AD} : \frac{BC}{AB},$$

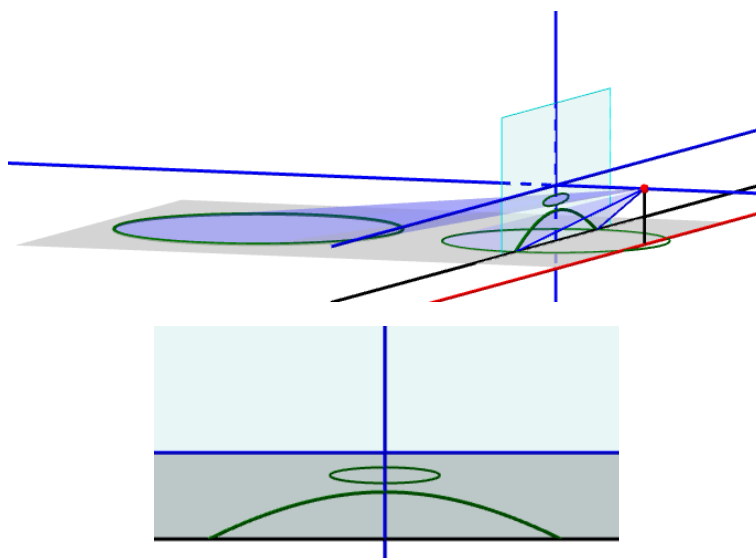
pero, si llamamos B', C', D' a los puntos del suelo H' cuyas imágenes en H son B, C, D , entonces A es la imagen por π del punto infinito de la recta que pasa por estos tres puntos, luego, por el mismo teorema,

$$|\mathcal{R}(D', B', C', \infty)| = \frac{D'C'}{B'C'} = \frac{10}{L}.$$

Como las razones dobles tienen que ser iguales, se tiene la igualdad que hemos usado. ■

Aunque el lector no esté en condiciones ahora de seguir el razonamiento, lo interesante es que observe cómo la geometría proyectiva nos permite operar con puntos infinitos en los cálculos. En el capítulo IX estudiaremos desde el punto de vista de la geometría proyectiva las secciones cónicas que estudiamos en el capítulo VII de [IGE], y veremos que en este contexto todas las cónicas son iguales, en el sentido de que las diferencias entre una y otra dependen únicamente de la elección del sistema de referencia. En particular, no hay ninguna diferencia intrínseca entre una elipse, una parábola y una hipérbola, sino que una misma cónica será considerada de uno u otro tipo en función de qué recta del plano tomemos como recta infinita. Si la recta infinita no corta a la cónica, la veremos como una elipse, si la corta en un punto como una parábola y si la corta en dos puntos como una hipérbola.

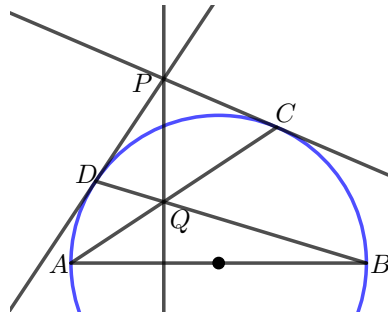
Por ejemplo, en la representación en perspectiva que hemos considerado, la línea del horizonte es la recta del cuadro que se corresponde con la recta infinita del suelo (o de cualquier otro plano paralelo), pero el cuadro tiene a su vez una recta infinita, que se corresponde con la recta r del suelo paralela al cuadro que pasa por los pies del observador pues, cuando unimos los puntos de esta recta con el ojo del observador, obtenemos una recta paralela al cuadro que, por lo tanto, no determina en el ningún punto (o, mejor dicho, corta al plano del cuadro en un punto de su recta infinita). Por ello, si una circunferencia (o cualquier cónica) pintada en el suelo tiene al observador en su exterior, su representación en el cuadro será una elipse, si toca a la recta r en un punto se verá como una parábola, y si la corta en dos puntos se verá como una hipérbola. La figura muestra el caso de la elipse y la hipérbola:



Otro contexto en el que la posibilidad de trabajar sistemáticamente con puntos infinitos resulta útil es al tratar con los conceptos de polo y polar respecto de una cónica, que en [IGE 5.25] introdujimos en el caso particular de una circunferencia.

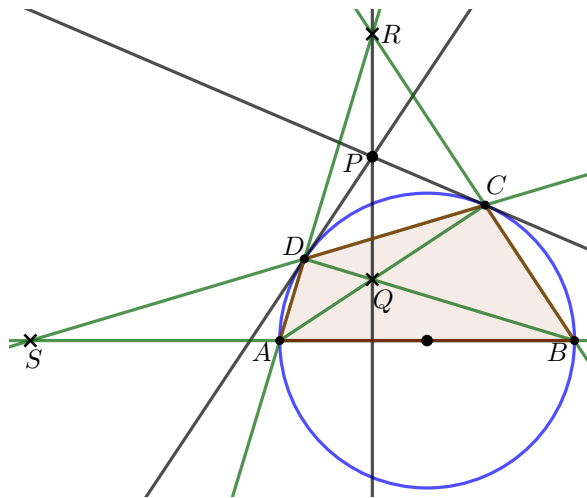
Por ejemplo, he aquí un problema que no es fácil resolver a partir de los resultados probados en [IGE], pero que tiene una solución muy simple basada en un teorema que demostraremos sobre la polaridad de una cónica:

En la figura siguiente, probar que la recta PQ es perpendicular al diámetro AB :



De nuevo exponemos aquí una solución que requiere algunos resultados de los capítulos VIII y IX, simplemente para que el lector aprecie la simplicidad formal del argumento:

Completamos la figura hasta formar un cuadrilátero completo:



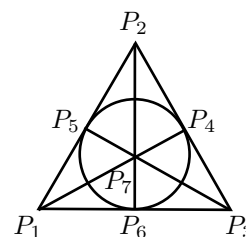
Por el teorema 9.10, tenemos que P es el polo de DC , luego está en la recta polar de S , que, por el teorema 9.11 es QR , luego concluimos que $PQ = RQ$ es la recta polar de S , que es obviamente perpendicular a AB , ya que puede calcularse uniendo los puntos donde las tangentes a la circunferencia por S tocan a la circunferencia (o, alternativamente, porque la recta polar de S pasa por el polo de la recta AB , que es el punto infinito común a las perpendiculares a AB (el punto de intersección de las tangentes a la circunferencia por A y B)). ■

Utilizando ahora resultados básicos de la geometría euclídea no es difícil probar que el triángulo DPQ (o CPQ) es isósceles, hecho que tampoco es evidente.

La geometría proyectiva es una extensión de la geometría afín, no de la geometría euclídea, en el sentido de que en un espacio proyectivo, al igual que en un espacio afín, no podemos medir longitudes ni ángulos. Esto hace que en una cónica proyectiva no podamos hablar de focos o directrices (que son conceptos métricos). Sin embargo, en el capítulo X estudiaremos la geometría parabólica, que es esencialmente la geometría de un espacio proyectivo dotado de una estructura adicional que determina una estructura de espacio euclídeo en cada espacio afín resultante de tomar un hiperplano como hiperplano infinito. En este contexto podremos definir todos los conceptos métricos asociados a las cónicas y demostrar que las cónicas que estamos estudiando son las mismas que las clásicas estudiadas en [IGE].

Geometrías finitas El tratamiento analítico de la geometría afín y la geometría proyectiva hace que cada cuerpo K determine un espacio afín y un espacio proyectivo (arguesianos) de cada dimensión d unívocamente determinados salvo isomorfismo. En particular, a partir de un cuerpo finito podemos construir espacios afines y proyectivos finitos de cualquier dimensión. Por ejemplo, el plano proyectivo más simple (correspondiente al cuerpo de 2 elementos) es el llamado *plano de Fano*, por el matemático italiano Gino Fano, que lo usó como ejemplo en un artículo de 1892.

Consta de 7 puntos y de otras tantas rectas, que en la figura están representadas por 6 segmentos y una circunferencia. Cada recta contiene tres puntos y por cada punto pasan tres rectas. Estas “simetrías y dualidades” hacen que las geometrías finitas no sean meras curiosidades, sino que tengan aplicaciones que se exploran en la llamada “teoría de diseños”, que surgió a partir del análisis de los problemas derivados de diseñar experimentos científicos o meramente estadísticos que permitan estudiar todas las combinaciones posibles de varios factores con el mínimo número de experimentos.



En el apéndice B veremos un ejemplo muy simple de aplicación del plano de Fano. Se trata de la construcción del código de Hamming para la detección y corrección de errores. Imaginemos que queremos transmitir un mensaje codificado como una sucesión de ceros y unos. Hay muchos contextos en los que no podemos descartar la posibilidad de que algunos de ellos se alteren durante la transmisión, es decir, que se envíe un 1 y se reciba un 0 o viceversa. Una medida contra este problema puede ser enviar el mensaje dos veces, de modo que si, por ejemplo, queremos transmitir el código 1011, podemos enviar 10111011 y así, si el receptor no recibe el mensaje exactamente duplicado sabrá que ha habido un error. Pero esta forma de detectar errores no es muy eficiente, pues, en primer lugar, si sólo se corrompe un dígito, sabremos que ha habido un error, pero no sabremos si el mensaje correcto es el formado por la primera parte de la sucesión

recibida o el formado por la segunda parte, y, en segundo lugar, si se corrompen dos dígitos correspondientes al mismo fragmento del mensaje, como por ejemplo el primero y el quinto, recibiremos un mensaje aparentemente correcto y no detectaremos el error.

Richard Wesley Hamming trabajaba para los *Bell Telephone Laboratories* y estaba molesto con los relativamente frecuentes errores que producían los lectores de tarjetas perforadas, así que en 1950 introdujo una forma de duplicar la longitud de un mensaje de cuatro dígitos binarios de modo que si en la transmisión se corrompe uno de los ocho dígitos enviados se puede reconstruir el mensaje original (se puede corregir el error) y que si se corrompen dos de ellos se puede detectar que ha habido un error, aunque no podamos saber cuál era el mensaje correcto.

Hay dos versiones del código de Hamming (y muchos otros códigos más que se conocen en general como “códigos de Hamming”, porque se basan en los mismos principios), que completan el mensaje que hay que transmitir con tres o cuatro dígitos de comprobación. El código de Hamming extendido usa cuatro. Por ejemplo, para transmitir el mensaje 1011, se completa hasta 1011|1000 y veremos que, tal y como hemos indicado, si este código se recibe con un dígito alterado, es posible saber cuál era el código original, mientras que si se producen dos alteraciones podemos saber que ha habido un error, aunque no podamos determinar cuál era el código emitido.

Geometrías no euclídeas En los dos últimos capítulos veremos cómo a partir de un plano proyectivo es posible definir dos geometrías planas no euclídeas: la geometría hiperbólica, que es la geometría no euclídea más parecida a la euclídea, en el sentido de que satisface todos los axiomas de la geometría euclídea (de incidencia, de ordenación y de congruencia, e incluso el axioma de continuidad D) menos el axioma de las paralelas, y la geometría elíptica, que es casi la misma que la geometría esférica, es decir, la geometría de la superficie de una esfera, conocida y aplicada por astrónomos y geógrafos desde la antigüedad. Como ejemplo resolveremos el problema de las dos elevaciones, cuya solución fue publicada por Gauss en 1812, consistente en determinar la longitud y la latitud de un punto de la superficie terrestre a partir de la observación de dos estrellas cuya posición en la esfera celeste es conocida. Concretamente resolveremos este caso particular (aunque el procedimiento de resolución es general y aplicable a cualquier otro caso):

El 15 de enero de 2021, en un punto del océano Atlántico, a la 1:15 hora de Greenwich, las altitudes de las estrellas Procyon y Beigelgeuse en la esfera celeste resultaron ser de 40° y 60° , respectivamente. Determinar las coordenadas y la hora local del punto de observación.

Con esto el lector puede formarse una idea superficial del contenido de este libro, aunque hay tres capítulos que no hemos mencionado: en el capítulo IV estudiamos la geometría de los números complejos y la de los cuaternios, que

son una estructura de anillo de división en \mathbb{R}^4 que, entre otras cosas, permite representar muy eficientemente los giros en el espacio tridimensional. El capítulo V aplicamos las técnicas del álgebra lineal expuestas en [A1] para obtener resultados fundamentales sobre las biyecciones afines en un espacio afín que, por una parte, se pueden definir como las biyecciones que conservan la estructura algebraica de espacio afín y, por otra parte —como demostraremos allí—, bajo hipótesis razonables sobre el cuerpo de escalares, resultan ser simplemente las biyecciones que transforman rectas en rectas. En el capítulo VI estudiaremos las posibilidades de construcción geométrica con regla y compás mucho más a fondo que en las secciones [IGE 6.4, 6.5], aprovechando la teoría algebraica sobre extensiones de cuerpos del capítulo V de [A1]. Finalmente, en el apéndice A profundizaremos en el concepto de inversión respecto de una circunferencia que presentamos en [IGE 5.12] y que usamos frecuentemente en [IGE].

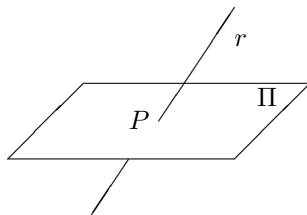
Capítulo I

La geometría absoluta

En este capítulo y los siguientes nos ocuparemos de formalizar en el seno de la teoría de conjuntos la geometría intuitiva, es decir, la geometría con la que interpretamos nuestras percepciones que, como se sabe, no coincide con la geometría que los físicos usan para describir el espacio físico. Podemos decir que lo que vamos a estudiar es el *espacio*. No podemos definir este concepto, pero todos tenemos una imagen intuitiva del mismo. Nuestra intención es introducir unos axiomas sobre el espacio que puedan considerarse intuitivamente evidentes y explorar sus consecuencias lógicas.

La primera aproximación a la caracterización matemática del espacio será considerar al espacio como un conjunto \mathbb{E} a cuyos elementos llamaremos *puntos*. Un punto es una posición en el espacio, carente de toda extensión. Un punto ortográfico es una buena imagen de un punto geométrico, si bien hemos de tener presente que la pequeña mancha de tinta tiene una cierta extensión, de la que debemos hacer abstracción.

Hay dos conceptos más que se encuentran al mismo nivel elemental que el concepto de punto. Se trata de los conceptos de *recta* y *plano*. De nuevo es imposible definir la característica que diferencia a una línea recta de una línea curva o a una superficie plana de una superficie curva, pero intuitivamente todos sabemos distinguir las rectas y los planos de las restantes curvas y superficies.



Una recta r y un plano Π que se cortan en un punto P .

Hemos de hacer una aclaración sobre la interpretación que vamos a dar a estas palabras: una hoja de papel proporciona una buena imagen de un plano, salvo por el hecho de que la hoja termina en unos bordes, mientras que para

nosotros un plano será una superficie ilimitada. Si apoyamos la hoja en una mesa, la superficie de la mesa representa una porción más amplia del mismo plano que representa la hoja. Similarmente, por una recta entenderemos una línea recta sin extremos, de modo que si trazamos una porción de recta con la ayuda de una regla, cualquier extensión de la misma por cualquiera de sus extremos será una porción mayor de la misma recta. Del mismo modo que hemos hecho con el espacio, podemos concebir las rectas y los planos como conjuntos de puntos, es decir, como *ciertos* subconjuntos¹ de \mathbb{E} que no podemos definir.

1.1 Axiomas de incidencia

Definición 1.1 Una *geometría de Hilbert* está formada por un conjunto \mathbb{E} al que llamaremos *espacio*, y a cuyos elementos llamaremos *puntos*, junto con dos familias de subconjuntos no vacíos de \mathbb{E} a cuyos elementos llamaremos respectivamente *rectas* y *planos*, de modo que se cumplan los cinco axiomas indicados a continuación.

Diremos que una recta o plano X *pasa* por un punto P , o que X *incide* en el punto P , si $P \in X$.

Axioma A1 *Por cada par de puntos distintos P y Q pasa una única recta, que representaremos por PQ .*

Axioma A2 *Toda recta pasa al menos por dos puntos.*

Diremos que tres o más puntos son *colineales* si hay una recta que pasa por todos ellos.

Axioma A3 *Por cada tres puntos no colineales P, Q, R pasa un único plano, al que representaremos por PQR .*

Axioma A4 *Si una recta tiene dos puntos en común con un plano Π , entonces está contenida en Π .*

Axioma A5 *Todo plano pasa al menos por tres puntos no colineales.*

Éstos no son los únicos axiomas de incidencia que vamos a considerar, sino que más adelante consideraremos algunos axiomas alternativos o adicionales, pero conviene distinguir los resultados que pueden probarse únicamente a partir de estos axiomas.

Veamos algunas consecuencias elementales:

Si una recta o un plano X no pasa por un punto P , diremos que P es un punto *exterior* a X .

¹Se puede cuestionar que intuitivamente el espacio sea un conjunto de puntos, ya que no podemos imaginarnos el espacio como formado por puntos, sino que más bien podemos imaginar algunos puntos en el espacio. En este sentido, el considerar el espacio como un conjunto de puntos es simplemente un convenio habitual en la teoría de conjuntos, en la cual todo es un conjunto. Pero es un convenio intrascendente, pues en la práctica, decir que \mathbb{E} es un conjunto de puntos sólo significa que expresaremos mediante la expresión $P \in \mathbb{E}$ la propiedad " P es un punto". Lo mismo se aplica a las rectas y los planos.

Teorema 1.2 *Toda recta tiene un punto exterior que esté en un plano dado que la contenga (Por A5)*

Teorema 1.3 *Si P y Q son puntos de una recta r y R es exterior a r , entonces P , Q y R no son colineales. (Por A1)*

Teorema 1.4 *Dos rectas distintas tienen como máximo un punto en común. (Por A1)*

Definición 1.5 Diremos que dos rectas son *paralelas* (gr. ‘una al lado de la otra’) si son iguales (en cuyo caso se dice también que son *coincidentes*) o bien ambas están contenidas en un mismo plano y no tienen puntos en común. Si dos rectas no tienen puntos en común, pero no están contenidas en el mismo plano, se dice que *se cruzan*. La única alternativa a estos casos es que las rectas tengan un único punto en común, en cuyo caso se dice que son *secantes* (lat. ‘que se cortan’).

Teorema 1.6 *Dados un plano y una recta, o bien no tienen puntos comunes, o bien tienen un único punto en común, o bien la recta está contenida en el plano. (Por A4)*

Diremos que una recta y un plano son *secantes* si se cortan en un punto y *paralelos* si no tienen puntos comunes.

Teorema 1.7 *Una recta y un punto exterior a ella están contenidos en un único plano. (Por A2, A3 y A4)*

Teorema 1.8 *Dos rectas secantes están contenidas en un único plano.*

El axioma de las paralelas El *axioma de las paralelas* es el siguiente axioma de incidencia:

Axioma A2b *Por cada punto pasa una única paralela a una recta dada.*

Observemos que esto equivale a exigir que por cada punto exterior a una recta dada pasa una única paralela, pues si el punto P está en la recta r entonces r es claramente la única recta paralela a r por P , y esto no requiere el axioma de las paralelas.

Hemos llamado A2b a este axioma porque podemos incorporarlo al grupo de axiomas de incidencia en sustitución del que hemos tomado como A2, ya que el axioma A2 original se puede demostrar a partir del axioma de las paralelas (y de los demás axiomas de incidencia). En efecto:

Dada una recta r , por definición es un conjunto no vacío, luego contiene un punto P . Tenemos que probar que contiene al menos otro punto. En caso contrario, sean Q y R dos puntos no colineales con P (por A5) y sea s la paralela a PQ por R . Entonces $r = \{P\}$ está contenida en el plano PQR , al igual que s , y de hecho r y s son paralelas, pues P no está en s . Pero entonces r y PQ son dos paralelas a s por el punto P , lo cual contradice al axioma de las paralelas. ■

La geometría que puede desarrollarse sin suponer el axioma de las paralelas se llama *geometría absoluta* o *geometría neutra*, y es la que vamos a estudiar en este primer capítulo.

1.2 Axiomas de ordenación

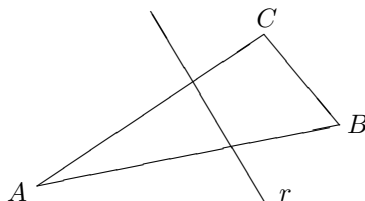
Definición 1.9 Una geometría de Hilbert está *ordenada* si hay definida una relación triádica entre los puntos del espacio que representaremos con la notación $A-B-C$ y leeremos “ B está entre A y C ”, de modo que se cumplen los axiomas siguientes:

Axioma B1 Si se cumple $A-B-C$, entonces A, B y C son tres puntos colineales distintos dos a dos, y además se cumple $C-B-A$.

Axioma B2 Si A y B son puntos distintos, existe otro C tal que $A-B-C$.

Axioma B3 Dados tres puntos distintos en una misma recta, uno y sólo uno de ellos está entre los otros dos.

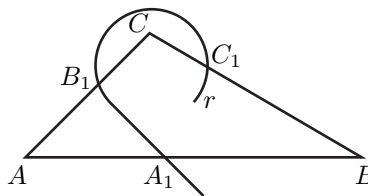
Axioma B4 (Axioma de Pasch)² Sean A, B, C tres puntos no colineales y sea r una recta contenida en el plano ABC , pero que no pase por ninguno de ellos. Si r pasa por un punto situado entre A y B , entonces r pasa por un punto entre A y C o bien por un punto entre B y C .



El axioma B4 admite una precisión:

Teorema 1.10 (Teorema de Pasch) En las condiciones del axioma B4, no puede suceder que r pase a la vez por un punto entre A y C y también por un punto entre B y C .

DEMOSTRACIÓN: Supongamos que r pasa por un punto A_1 entre A y B , por un punto B_1 entre A y C y por un punto C_1 entre B y C . Los tres puntos A_1, B_1, C_1 tienen que ser distintos entre sí, pues si, por ejemplo, $A_1 = B_1$, entonces tendríamos $AB = AC$, con lo que A, B, C serían colineales. Puesto que A_1, B_1, C_1 están sobre la recta r , el axioma B3 afirma que uno de ellos tiene que estar entre los otros dos. Por simetría podemos suponer que $A_1 - B_1 - C_1$.



²Moritz Pasch advirtió que Euclides usaba implícitamente este hecho, a pesar de que no se podía deducir de sus postulados.

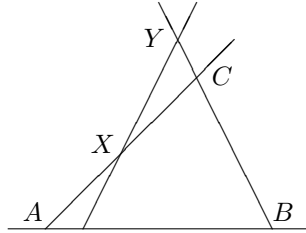
Como r no pasa por B , los puntos B, A_1, C_1 no son colineales, y la recta AC pasa por B_1 , que está entre A_1 y C_1 . Por el axioma B4, tiene que pasar por un punto entre B y A_1 o bien por un punto entre B y C_1 , pero dicho punto tiene que ser la intersección de AC con la recta $A_1B = AB$ o bien con $BC_1 = BC$, luego tiene que ser A o C , pero, por B3, ni A está entre A_1 y B , ni C está entre B y C_1 , con lo que tenemos una contradicción. ■

Definición 1.11 Dados dos puntos distintos A y B , llamaremos *segmento* (lat. 'corte') de extremos A y B al conjunto formado por A , B y todos los puntos situados entre A y B . Lo representaremos por \overline{AB} . Notemos que el axioma B1 implica que $\overline{AB} = \overline{BA}$, así como que $\overline{AC} \subset AC$. La recta AC (que es la única recta que contiene al segmento, se llama su *prolongación*.

Ahora probamos una variante del axioma B2:

Teorema 1.12 Entre dos puntos distintos siempre hay un tercero.

DEMOSTRACIÓN: Sean A y B dos puntos distintos. Sea X un punto no colineal con ellos, sea C tal que $A - X - C$. Entonces C no es colineal con A y B , pues la recta AC es distinta de AB (porque contiene a X) y por consiguiente sólo corta a AB en A . En particular $C \neq B$, luego podemos tomar Y tal que $Y - C - B$.



Es claro que todos los puntos están en el mismo plano AXB . Notemos que $X \neq Y$, pues la recta AX es distinta de CB , o de lo contrario A, X, B serían colineales, luego su único punto en común es C . Si fuera $Y = X$ sería otro punto común distinto de C .

Es fácil ver que la recta XY no pasa por A, B o C (de lo contrario se llega a que A, B, X son colineales). Pero XY pasa por X , que está entre A y C , y no puede pasar por ningún punto entre B y C porque la intersección de XY con BC es Y , que no está entre B y C (por B3). El axioma B4 nos da entonces que XY pasa por un punto situado entre A y B . ■

Seguidamente vamos a demostrar que toda recta divide a un plano en dos semiplanos y que todo punto divide a una recta en dos semirrectas. Para ello empezamos definiendo una relación de equivalencia:

Teorema 1.13 Sea Π un plano y r una recta contenida en Π . Entonces, la relación en $\Pi \setminus r$ dada por

$$P \sim_r Q \quad \text{si y sólo si} \quad P = Q \text{ o } \overline{PQ} \text{ no corta a } r$$

es una relación de equivalencia que determina exactamente dos clases de equivalencia.

DEMOSTRACIÓN: La relación es trivialmente reflexiva y simétrica. Veamos que también es transitiva. Supongamos, pues que $A \sim_r B$ y que $B \sim_r C$. Si dos de los puntos son iguales, es inmediato que $A \sim_r C$, así que podemos suponer que son distintos dos a dos. Ahora distinguimos dos casos:

a) Los puntos $A, B, C \in \Pi \setminus r$ no son colineales. Entonces tenemos que r no corta ni a \overline{AB} ni a \overline{BC} , luego el axioma B4 implica que tampoco corta a \overline{AC} , luego $A \sim_r C$.

b) Los puntos A, B, C están en una misma recta s . Como r contiene al menos dos puntos, podemos tomar $X \in r$ que sea distinto de la intersección de s con r (si es que existe tal intersección). Por B2 existe un punto D tal que $X - A - D$. Es claro entonces que $D \notin r$, porque la recta AD es distinta de r y ya corta a r en X . Similarmente, $D \notin s$, luego D no es colineal ni con A, B , ni con B, C , ni con A, C .

Por B3 tenemos que r no corta a \overline{AD} , ya que la intersección de r con AD es X , que no está en el segmento. Por lo tanto, $D \sim_r A$. Por el caso ya probado $D \sim_r B$ y $D \sim_r C$, de donde a su vez $A \sim_r C$.

Veamos ahora que hay dos clases de equivalencia. Dado $X \in r$ y $A \in \Pi \setminus r$, tomamos un punto B tal que $A - X - B$ (por B2), que claramente está en $\Pi \setminus r$ y r corta a \overline{AB} , luego A y B están en clases de equivalencia distintas. Así pues, hay al menos dos clases.

Veamos por último que cualquier otro punto $C \in \Pi \setminus r$ está relacionado con A o con B . Si A, B, C no son colineales, esto se debe a 1.10, pues r corta a \overline{AB} , luego no puede suceder que corte tanto a \overline{AC} como a \overline{BC} . Supongamos, pues, que A, B, C están en una recta s y tomemos $Y \in r$ distinto de la intersección X de r con s . Sea D tal que $Y - C - D$. Como antes, $D \in \Pi$ no está ni en r ni en s . Entonces $D \sim_r C$ y, por el caso ya probado, D está relacionado con A o con B , luego lo mismo le sucede a C . ■

Ahora podemos decir que dos puntos en $\Pi \setminus r$ están *al mismo lado* o *en lados opuestos* de r según que estén en la misma clase de equivalencia o en clases distintas respecto de la relación considerada en el teorema anterior.

Teorema 1.14 *Si A es un punto en una recta r , entonces la relación en $r \setminus \{A\}$ dada por*

$$P \sim_A Q \quad \text{si y sólo si} \quad P = Q \text{ o } A \notin \overline{PQ}$$

es una relación de equivalencia que determina exactamente dos clases de equivalencia.

DEMOSTRACIÓN: Sea X un punto que no esté en r , sea $s = AX$ y sea Π el plano que contiene a X y a r . Entonces $s \subset \Pi$ y se comprueba fácilmente que, para puntos de $r \setminus \{A\}$, se cumple

$$P \sim_A Q \quad \text{si y sólo si} \quad P \sim_s Q,$$

de donde se concluye que la relación es de equivalencia y que determina a lo sumo dos clases de equivalencia. Para probar que hay dos exactamente tomamos

cualquier punto $P \in r \setminus \{A\}$ y un punto tal que $P - A - Q$, con lo que P y Q no están relacionados. ■

Diremos que dos puntos $P, Q \in r \setminus \{A\}$ están *al mismo lado* o *en lados opuestos* de A según que estén en la misma clase de equivalencia o en clases de equivalencia distintas respecto de la relación considerada en el teorema anterior.

Nota Ahora ya podemos probar que un segmento determina sus extremos, es decir, que no puede ocurrir que un mismo conjunto pueda ser el segmento determinado por dos pares distintos de extremos.

Para ello basta observar que si $s = \overline{AB}$, entonces A y B son los dos únicos puntos de s con la propiedad de que no están entre otros dos puntos de s .

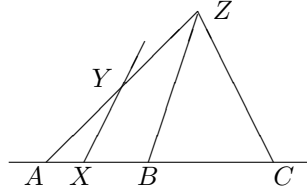
En efecto, si $P \in s$ es un punto distinto de A y B , existen puntos $X, Y \in s$ (concretamente, $X = A$ e $Y = B$) tales que $X - P - Y$. En cambio, si P es uno de los puntos A y B , esto es imposible, ya que si, por ejemplo, $P = A$, de $X, Y \in s$ se sigue que $X \sim_A B$, $Y \sim_A B$, luego $X \sim_A Y$, lo cual contradice a $X - A - Y$. ■

Teorema 1.15 *Si tres puntos cumplen $A - B - C$, entonces*

$$\overline{AB} \cap \overline{BC} = \{B\}, \quad \overline{AC} = \overline{AB} \cup \overline{BC}.$$

DEMOSTRACIÓN: Sea r la recta que contiene a A, B, C . Si se cumple $P \in \overline{AB} \cap \overline{BC}$, pero $P \neq B$, entonces $P \sim_B A$ y $P \sim_B C$, luego $A \sim_B C$, lo cual es falso. Esto nos da la primera igualdad.

Para demostrar la segunda tomamos $X \in \overline{AB}$. Si $X = A$ o $X = B$ es claro que $X \in \overline{AC}$. Supongamos, pues, que $X \neq A$ y $X \neq B$. Sea Y un punto no colineal con A, B, C y sea Z tal que $A - Y - Z$. Es claro que todos los puntos están en el mismo plano, así como que la recta XY no pasa por ninguno de los otros puntos.



La recta XY corta a la recta BC en el punto X , que no está entre B y C por la igualdad que ya hemos probado. Así pues, XY no pasa por ningún punto entre B y C . Si XY pasara por un punto entre Z y C , por el axioma B4 tendría que pasar por un punto entre Z y B , pero eso es imposible por 1.10 aplicado a A, B, Z , ya que XY pasa por un punto entre A y B . Concluimos que XY no pasa por ningún punto entre Z y C , luego por B4 tiene que pasar por un punto entre A y C , y dicho punto es necesariamente X . Así pues, $X \in \overline{AC}$.

Con esto hemos probado que $\overline{AB} \subset \overline{AC}$, e igualmente $\overline{BC} \subset \overline{AC}$, luego tenemos una inclusión.

Supongamos ahora que $X \in \overline{AC}$, y nuevamente podemos suponer que es distinto de A, B, C . Consideramos puntos Y, Z en las mismas condiciones anteriores. Ahora sabemos que XY pasa por un punto entre A y C y por un punto entre A y Z , luego no puede pasar por ningún punto entre Z y C .

Como pasa por un punto entre A y Z , o bien pasa por un punto entre A y B , en cuyo caso dicho punto es X y tenemos que $X \in \overline{AB}$, o bien pasa por un punto entre Z y B , en cuyo caso una nueva aplicación de B4 nos da que pasa por un punto entre B y C , que tiene que ser X , con lo que $X \in \overline{BC}$. ■

Definición 1.16 Un conjunto de puntos C es *convexo* si cuando $P, Q \in C$ son dos puntos distintos, entonces $\overline{PQ} \subset C$. Equivalentemente, si cuando $P-X-Q$ con $P, Q \in C$, entonces $X \in C$.

Es inmediato que el espacio \mathbb{E} , los planos y las rectas son conjuntos convexos, así como que la intersección de conjuntos convexos es convexa.

Del teorema anterior se sigue que los segmentos son conjuntos convexos. En efecto, si tenemos puntos distintos $P, Q \in \overline{AB}$, entonces, por el teorema anterior, $Q \in \overline{AP}$ o bien $Q \in \overline{PB}$. Intercambiando A y B si es preciso, no perdemos generalidad si suponemos el primer caso. Entonces, de nuevo por el teorema anterior vemos que $\overline{PQ} \subset \overline{AP} \subset \overline{AB}$.

Volvamos ahora al teorema 1.14:

Definición 1.17 Dado un punto O en una recta r , llamaremos *semirrectas* en r con *origen* O a los dos conjuntos que resultan de añadir el punto O a cada una de las dos clases de equivalencia dadas por el teorema 1.14.

Dados dos puntos distintos O y A , llamaremos \overrightarrow{OA} a la semirrecta que el punto O determina en la recta OA que contiene al punto A . Toda semirrecta puede expresarse de esta forma.

Una semirrecta está contenida en una única recta, a la que llamaremos su *prolongación*. Además, cada semirrecta $s = \overrightarrow{OA}$ determina su origen, pues O es el único punto de s que no está entre otros dos puntos de s .

En efecto, si $P \in s$ es un punto distinto de su origen O , entonces podemos tomar un punto tal que $O-P-Y$, de modo que $P \sim_O Y$, luego $Y \in s$, y así tenemos dos puntos $X, Y \in s$ (tomando $X = O$), de modo que $X-P-Y$. En cambio, esto no sucede si $P = O$, pues de $X, Y \in s$ se sigue que $X \sim_O Y$, y esto contradice a $X-P-Y$.

Dada una semirrecta s de origen O y prolongación r , la otra semirrecta s' que O determina en r se llama *semirrecta complementaria* de s . Es claro entonces que $s \cap s' = \{O\}$ y que $s \cup s' = r$.

Notemos por último que las semirrectas son convexas. En efecto, tomamos $P-X-Q$ con $P, Q \in \overrightarrow{OA}$. Si P y Q son distintos de O , entonces $X \neq O$ (o P y Q estarían en lados opuestos respecto de O) luego $X \sim_O Q$, porque de lo contrario $O \in \overline{XQ} \subset \overline{PQ}$, contradicción. Por lo tanto $X \in \overrightarrow{OA}$. Si, por ejemplo, $P = O$, entonces también $X \sim_O Q$, e igualmente $X \in \overrightarrow{OA}$.

Definición 1.18 Sea r una recta contenida en un plano Π . Llamaremos *semi-planos* en Π de *frontera* r a los conjuntos que resultan de unir con r cada una de las clases de equivalencia dadas por el teorema 1.13.

Cada semiplano está contenido en un único plano, al que llamaremos su *prolongación*. Vamos a probar que un semiplano determina su frontera. Para ello probamos un hecho general:

Sea S un semiplano en el plano Π de frontera r , sea $A \in r$ y sea $P \in S \setminus r$. Entonces $\overrightarrow{AP} \subset S$.

En efecto, si $X \in \overrightarrow{AP}$, entonces, o bien $X = A \in S$ o, en caso contrario, $X \sim_A P$ respecto de la recta $s = AP$, lo cual equivale a que A no está en \overline{XP} , lo cual equivale a que r no corte a \overline{XP} , porque el único punto de corte posible es A , lo cual equivale a que $X \sim_r P$, es decir, a que $X \in S$.

Por el contrario, la semirrecta complementaria de \overrightarrow{AP} no contiene más punto de S que el propio A , pues todo punto $Q \neq A$ de dicha semirrecta complementaria cumple $Q - A - P$, luego no $Q \sim_r P$, luego $Q \notin S$.

Veamos ahora que los puntos de r son los únicos puntos P de S que cumplen que existe una recta s que pasa por P tal que una de las semirrectas determinadas por P en s está contenida en S y la complementaria no tiene más puntos en común con S que el propio P .

Ya hemos probado que los puntos de r tienen esta propiedad. Tomemos ahora $P \in S \setminus r$ y consideremos cualquier recta s que pase por P con una semirrecta contenida en S . Eso implica en particular que $s \subset \Pi$. Si s no corta a r , entonces es claro que todos los puntos $Q \in s$ cumplen $Q \sim_r P$, luego las dos semirrectas de s están contenidas en S . Si s corta a r en un punto A , sabemos que $\overrightarrow{AP} \subset S$ y si tomamos un punto en s tal que $A - P - Q$, entonces A y Q están cada uno en una semirrecta distinta de P en s y ambos están en S (en el caso de Q porque $Q \sim_A P$, luego $Q \sim_r P$). Por lo tanto, s no cumple la propiedad considerada.

Así pues, si S es un semiplano, está determinada su frontera r , al igual que su prolongación Π , y el otro semiplano determinado por r en Π se llama el *semiplano complementario* de S .

Claramente, si S' es el semiplano complementario de S , entonces $S \cap S' = r$ es la frontera común a ambos y $S \cup S'$ es su prolongación común.

Ahora es fácil probar que los semiplanos son convexos: Si S es un semiplano de frontera r y $P - X - Q$ con $P, Q \in S$, distinguimos varios casos:

- 1) Si $P, Q \in r$, entonces $X \in PQ = r \subset S$.
- 2) Si $P \in r, Q \notin r$, entonces sabemos que $X \in \overrightarrow{PQ} \subset S$.
- 3) Si $P, Q \notin r$, entonces $X \notin r$ (o si no P y Q estarían en lados opuestos respecto de r) y $X \sim_r Q$, pues en caso contrario r cortaría a $\overline{XQ} \subset \overline{PQ}$ y de nuevo P y Q estarían separados por r , contradicción.

Ordenación de las rectas Seguidamente vamos a probar que los axiomas de ordenación permiten ordenar las rectas de forma natural. Concretamente, dados dos puntos distintos A y B en una recta, vamos a definir una relación de orden $X <_{AB} Y$ que se interpretará como que para llegar desde X hasta Y hay que desplazarse sobre la recta en el mismo sentido que para llegar desde A hasta B .

Definiremos en primer lugar la relación sobre las semirrectas y luego “pegaremos” las relaciones correspondientes a dos semirrectas complementarias.

Observemos que si dos puntos X, Y están en una semirrecta S de origen O (y son distintos de O), entonces $O - X - Y$ o bien $O - Y - X$, y el teorema 1.15 implica que $\overline{OX} \subsetneq \overline{OY}$ o bien $\overline{OY} \subsetneq \overline{OX}$. Como se ha de dar uno de los dos casos, tenemos de hecho una doble implicación:

$$O - X - Y \quad \text{si y sólo si} \quad \overline{OX} \subsetneq \overline{OY}.$$

Definición 1.19 Sea S una semirrecta de origen O . Definimos el orden $<_S^+$ en S como el dado por

$$X <_S^+ Y \quad \text{si y sólo si} \quad O - X - Y \quad \text{si y sólo si} \quad \overline{OX} \subsetneq \overline{OY},$$

para $X, Y \in S$ distintos de O , con el convenio adicional de que $O <_S^+ X$ para todo $X \in S$ distinto de O .

Es inmediato comprobar que se trata realmente de una relación de orden total (estricto) en S . Llamaremos \leq_S^+ a la relación no estricta correspondiente y llamaremos \leq_S^- a la relación inversa, de modo que

$$X <_S^- Y \quad \text{si y sólo si} \quad X - Y - O \quad \text{si y sólo si} \quad \overline{OY} \subsetneq \overline{OX},$$

para $X, Y \in S$ distintos de O , con el convenio adicional de que $X <_S^- O$ para todo $X \in S$ distinto de O .

Así O es el máximo para la relación \leq_S^- y el mínimo para la relación \leq_S^+ .

Dados tres puntos $A - O - B$, definimos el orden \leq_{AB} sobre la recta AB como el dado por

$$X \leq_{AB} Y \quad \text{si y sólo si} \quad \begin{cases} X \in \overrightarrow{OA}, Y \in \overrightarrow{OB} & \text{o} \\ X, Y \in \overrightarrow{OA} \text{ y } X \leq^- Y & \text{o} \\ X, Y \in \overrightarrow{OB} \text{ y } X \leq^+ Y, & \end{cases}$$

donde las relaciones \leq^- y \leq^+ son las correspondientes a las semirrectas \overrightarrow{OA} y \overrightarrow{OB} , respectivamente.

Observemos que lo que hace \leq_{AB} es ordenar la semirrecta \overrightarrow{OA} de modo que O sea el máximo y a continuación pone todos los puntos de \overrightarrow{OB} con O como punto mínimo.

Es claro que $A <_{AB} B$ y que \leq_{BA} es la relación inversa de \leq_{AB} .

En principio deberíamos escribir \leq_{AB}^O , porque la definición depende de los tres puntos, pero enseguida veremos que realmente no depende de la elección de O . Observemos de momento que, en realidad, A y B sólo intervienen en la definición para distinguir las dos semirrectas de origen O , de modo que si elegimos otros puntos $A' < O < B'$, es claro que $\leq_{A'B'}$ es la misma relación que \leq_{AB} . En otras palabras, para cada punto O de una recta r , tenemos definidas únicamente dos relaciones de orden en r mutuamente inversas.

Teorema 1.20 *Dada una recta r y en ella dos puntos distintos A y B , entonces:*

1. *La relación \leq_{AB} es una relación de orden total en r independiente de la elección del punto $A - O - B$ con la que se calcula.*
2. *Se cumple que $A <_{AB} B$*
3. *La relación \leq_{BA} es la relación inversa de \leq_{AB} .*
4. *C y D son puntos distintos en r , entonces \leq_{CD} coincide con \leq_{AB} o con \leq_{BA} (según si $A \leq_{CD} B$ o $B \leq_{CD} A$).*

DEMOSTRACIÓN: Consideremos ahora un punto $P \in AB$, $P \neq O$. Supongamos, por ejemplo, que $P \in \overrightarrow{OB}$, sea C tal que $O - P - C$ y llamemos $S = \overrightarrow{OB}$, $S' = \overrightarrow{PC}$.

Entonces $S' \subset S$, pues si $X \in S'$, o bien $X = P \in S$, o bien $X \sim_P C$ y, como no $O \sim_P C$, tampoco $X \sim_P O$, luego $O - P - X$, luego $X \in S$.

Si $X, Y \in S'$ son distintos de P , tenemos que $O - P - X$ y $O - P - Y$, luego

$$X <_{S'}^+ Y \leftrightarrow \overrightarrow{PX} \subsetneq \overrightarrow{PY} \leftrightarrow \overrightarrow{OP} \cup \overrightarrow{PX} \subsetneq \overrightarrow{OP} \cup \overrightarrow{PY} \leftrightarrow \overrightarrow{OX} \subsetneq \overrightarrow{OY} \leftrightarrow X <_{AB} Y.$$

Además, como $O - P - X$, se cumple que $P <_{AB} X$ (y también $P <_{S'}^+ X$), luego la restricción de \leq_{AB} a S' es $\leq_{S'}^+$.

Consideramos ahora la semirrecta complementaria de S' , que es $S'' = \overrightarrow{PO}$. Ahora se cumple que $S'' = \overrightarrow{OP} \cup \overrightarrow{OA}$. En efecto, si $X \in S''$ es distinto de P y de O , tenemos que $X \sim_P O$, luego $O - X - P$ o bien $X - O - P$. En el primer caso $X \in \overrightarrow{OP}$, y en el segundo, no $X \sim_O P$ y tampoco $A \sim_O P$, luego $X \sim_O A$, luego $X \in \overrightarrow{OA}$.

Vamos a probar que la restricción de \leq_{AB} a S'' es $\leq_{S''}^-$. Esto obliga a distinguir varios casos. En primer lugar veamos que P es el máximo de S'' para \leq_{AB} . En efecto, si $X \in S''$ es distinto de P , o bien $X = O$, en cuyo caso es claro que $O <_{AB} P$, porque O es el mínimo de S , o bien $O - X - P$, en cuyo caso llegamos a la misma conclusión, o bien $X \in \overrightarrow{OA}$, en cuyo caso también $X <_{AB} P$ por definición, pues X y P están en semirrectas opuestas respecto de O .

Ahora tomamos $X, Y \in S''$ distintos de P . Si ambos están en \overrightarrow{OP} y son distintos de O , entonces

$$X <_{AB} Y \leftrightarrow O - X - Y \leftrightarrow X - Y - P \leftrightarrow X <_{S''}^- Y.$$

La equivalencia central se debe a que $\overline{OP} = \overline{OX} \cup \overline{XP}$, pero $O - X - Y$ implica que Y no puede estar en el primer segmento, luego está en el segundo y, recíprocamente, $\overline{OP} = \overline{OY} \cup \overline{YP}$ y $X - Y - P$ implica que X no puede estar en el segundo segmento.

Supongamos ahora que X e Y están en \overrightarrow{OA} y son distintos de O . Entonces

$$X <_{AB} Y \leftrightarrow X - Y - O \leftrightarrow X - Y - P \leftrightarrow X <_{S''} Y.$$

En efecto, si $X - Y - O$, entonces no $X \sim_Y O$ y $O \sim_Y P$ (porque $Y - O - P$), luego no $X \sim_Y P$, luego $X - Y - P$. Recíprocamente, si $X - Y - P$, como $X - O - P$, es $\overline{XP} = \overline{XO} \cup \overline{OP}$, pero Y no está en el segundo segmento (porque $Y - O - P$), luego $X - Y - O$.

Ahora supongamos que $X \in \overrightarrow{OA}$ y que $Y \in \overline{OP}$ son puntos distintos y distintos de O . Entonces $X <_{AB} Y$ por definición y, por otra parte, como $X - O - P$, tenemos que $Y \in \overline{OP} \subset \overline{XP}$, luego $X - Y - P$, luego $X <_{S''} Y$.

Dejamos al lector los casos en los que uno de los puntos es igual a O , pues son más sencillos.

En total hemos probado que si $P \in \overrightarrow{OB}$ es distinto de O , entonces \leq_{AB}^O coincide con \leq_{OC}^P . Dicho de otro modo, la relación \leq_{AB}^O es una de las dos relaciones en r definidas a partir de las semirrectas con origen P . Esto es trivialmente cierto si $P = O$, y si $P \in \overrightarrow{OA}$ se cumple también aplicando lo que ya hemos probado a la relación \leq_{BA} (obtenemos que \leq_{BA} es una de las dos relaciones definidas a partir de P y, por consiguiente, \leq_{AB} es la otra). En particular, \leq_{AB} es independiente de la elección de O y esto completa la prueba del teorema. ■

Por último probamos que la relación de orden determina la relación “estar entre”:

Teorema 1.21 Sean A y B dos puntos distintos y sean X, Y, Z puntos de la recta AB . Entonces

$$X - Y - Z \quad \text{si y sólo si} \quad X <_{AB} Y <_{AB} Z \quad \text{o} \quad Z <_{AB} Y <_{AB} X.$$

DEMOSTRACIÓN: Si se cumple $X - Y - Z$, tomando $O = Y$ para definir las relaciones de orden, es inmediato que $X <_{XZ} Y <_{XZ} Z$. Como la relación \leq_{XZ} coincide con \leq_{AB} o con su inversa, la conclusión es inmediata.

Si se cumple $X <_{AB} Y <_{AB} Z$, entonces también $X <_{XZ} Y <_{XZ} Z$, y podemos calcular esta relación usando $O = Y$, y entonces es claro por la definición que X y Z están en semirrectas opuestas respecto de Y , luego se cumple $X - Y - Z$. ■

En particular vemos que

$$\overline{AB} = \{X \in AB \mid A \leq_{AB} X \leq_{AB} B\}, \quad \overrightarrow{AB} = \{X \in AB \mid A \leq_{AB} X\}.$$

1.3 Dimensión

Informalmente, podemos decir que una recta tiene una dimensión (gr. medida) en el sentido de que es necesario dar una coordenada para determinar un punto en ella (por ejemplo, indicar el km en el que se encuentra un vehículo en una carretera), mientras que un plano tiene dos dimensiones porque es necesario dar dos coordenadas para especificar la posición de un punto en él, y el espacio intuitivo tiene tres dimensiones porque, además de las coordenadas que sitúan un punto en un plano, hay que especificar su altura con una tercera coordenada. En este sentido podemos decir también que los puntos tienen dimensión cero, pues no hay que dar ninguna información para especificar un punto en un conjunto formado sólo por dicho punto.

Esta idea se formaliza en estos mismos términos a través de la geometría analítica, en la que a cada punto se le asocia unas coordenadas en un sistema de referencia. Este enfoque de la geometría lo presentaremos en el capítulo III, pero los axiomas de incidencia y ordenación nos bastan para distinguir los puntos, las rectas, los planos y el espacio según su dimensión.

Observemos que los axiomas de incidencia sólo aseguran que el espacio tiene dimensión mayor o igual que 2, concretamente a través del axioma A5, que afirma la existencia de tres puntos no alineados. Sin embargo, nada en ellos contradice la posibilidad de que exista un único plano Π que contenga todos los puntos del espacio \mathbb{E} , es decir, que el espacio sea un plano.

Si queremos estudiar exclusivamente la *geometría plana* podemos simplificar la axiomática para eliminar de ella el concepto de ‘plano’. Concretamente, podemos definir una *geometría plana de Hilbert* como un conjunto \mathbb{P} al que llamaremos “plano”, y a cuyos elementos llamaremos “puntos” junto con una familia no vacía de subconjuntos no vacíos de \mathbb{P} a cuyos elementos llamaremos “rectas”, de modo que se cumplan los tres axiomas siguientes:

Axioma A1 *Por cada par de puntos distintos pasa una única recta.*

Axioma A2 *Toda recta pasa al menos por dos puntos.*

Diremos que tres o más puntos son *colineales* si hay una recta que pasa por todos ellos.

Axioma A3 *Existen tres puntos no colineales.*

En este contexto dos rectas paralelas se definen como dos rectas iguales o sin puntos comunes (omitiendo la condición ahora trivial de que estén contenidas en el mismo plano). Si incorporamos el axioma de las paralelas, podemos hacerlo igualmente en sustitución de A2, pues la demostración de que el axioma de las paralelas implica A2 sólo se basa en lo que era el axioma A5, que aquí es A3.

A estos tres axiomas de la geometría plana podemos añadir los axiomas de ordenación sin más variación que eliminar en el axioma B4 el requisito ahora redundante de que la recta esté contenida en el plano determinado por los tres puntos.

Todos los resultados que hemos demostrado hasta ahora pueden probarse igualmente a partir de los axiomas de la geometría plana, sin más que eliminar toda referencia a planos en los enunciados o en las pruebas, aunque algunos de ellos se vuelven vacíos, como que dos rectas secantes están contenidas en un mismo plano.

Recíprocamente, todo resultado demostrable a partir de los axiomas de la geometría plana se traduce automáticamente en un resultado demostrable a partir de los axiomas generales sin más que observar que todo plano Π de una geometría de Hilbert (tomando como familia de rectas la de las rectas contenidas en él) satisface los axiomas de la geometría plana de Hilbert.

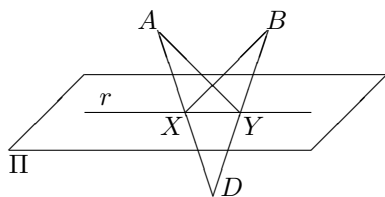
En efecto, el axioma A4 afirma que las rectas de una geometría de Hilbert que pasan por dos puntos de Π están contenidas en Π , luego son rectas de la geometría plana. Esto hace que Π cumpla el axioma A1, mientras que A2 se cumple en Π por el propio axioma A2 y el axioma A5 implica el axioma A3 para Π .

La mayor parte de los resultados que vamos a demostrar en los tres primeros capítulos son en realidad teoremas de la geometría plana. No obstante, ahora vamos a estudiar cómo completar los axiomas generales que hemos dado para que exijan que el espacio tenga exactamente tres dimensiones. Para ello vamos a describir con una notación homogénea algunas características que ya conocemos de puntos, rectas y planos:

- Diremos que un conjunto formado por un solo punto P_0 es siempre *afínmente independiente*, y definimos la *variedad afín* generada por P_0 como el conjunto $A^0(P_0) = \{P_0\}$.
- Diremos que un conjunto formado por dos puntos distintos P_0, P_1 es siempre *afínmente independiente*, y definimos la *variedad afín* generada por ellos como la recta $A^1(P_0, P_1) = P_0P_1$.
- Diremos que un conjunto formado por tres puntos P_0, P_1, P_2 es *afínmente independiente* si no son colineales, y definimos la *variedad afín* generada por ellos como el plano $A^2(P_0, P_1, P_2) = P_0P_1P_2$.
- Diremos que un conjunto formado por cuatro puntos P_0, P_1, P_2, P_3 es *afínmente independiente* si no son *coplanares*, es decir, si no están contenidos en un mismo plano.

A partir de aquí nos encaminamos a dar una definición razonable de la variedad afín $A^3(P_0, P_1, P_2, P_3)$ generada por cuatro puntos afínmente independientes, que será el análogo tridimensional de los conceptos de punto, recta y plano.

Definición 1.22 Diremos que un plano Π *separa* dos puntos A y B , y lo representaremos por $A - \Pi - B$, si $A, B \notin \Pi$ y Π corta al segmento \overline{AB} , es decir, si existe un punto $T \in \Pi$ (necesariamente único) tal que $A - T - B$.



Por lo tanto existe un punto Z en AY tal que $B - Z - X$. Similarmente, la recta XB tiene que cortar a \overline{AY} , pero la intersección tiene que ser el mismo Z , luego en total tenemos $B - Z - X$ y $A - Z - Y$.

Por el teorema anterior, $A - \Pi - C$ implica $Z - \Pi - C$, y una segunda aplicación

nos da $B - \Pi - C$. ■

Ahora ya es inmediato que la relación \sim_{Π} es una relación de equivalencia sobre los puntos exteriores al plano Π . La reflexividad y la simetría son obvias, mientras que la transitividad se sigue del teorema anterior: si $A \sim_{\Pi} B \sim_{\Pi} C$, por definición existe un D tal que $A - \Pi - D$ y $B - \Pi - D$, y por el teorema anterior $C - \Pi - D$, luego $A \sim_{\Pi} C$.

Definición 1.26 Si Π es un plano y P es un punto exterior, definimos

$$A^3(\Pi, P) = \{X \in \mathbb{E} \mid X \sim_{\Pi} P, \text{ o bien } X \in \Pi, \text{ o bien } X - \Pi - P\}.$$

Equivalentemente, si definimos

$$SA^3(\Pi, P) = \{X \in \mathbb{E} \mid X \sim_{\Pi} P\},$$

y $P - \Pi - Q$, entonces $A^3(\Pi, P) = SA^3(\Pi, P) \cup \Pi \cup SA^3(\Pi, Q)$.

En efecto, sólo hay que tener en cuenta que $X - \Pi - P$ equivale a $X \sim_{\Pi} Q$, por el teorema anterior.

Con esto tenemos definido el “espacio” y los “semiespacios” definidos por un plano y un punto exterior. Nuestra intención es definir

$$A^3(P_0, P_1, P_2, P_3) = A^3(P_0P_1P_2, P_3),$$

pero necesitamos justificar que la definición no depende del orden de los puntos, y algunos hechos más. Empezamos demostrando lo siguiente:

Teorema 1.27 Si Π es un plano y P, Q son dos puntos exteriores de modo que $Q \in A^3(\Pi, P)$, entonces $A^3(\Pi, P) = A^3(\Pi, Q)$.

DEMOSTRACIÓN: Distinguimos dos casos:

1) Si $P - \Pi - Q$ entonces directamente

$$A^3(\Pi, P) = SA^3(\Pi, P) \cup \Pi \cup SA^3(\Pi, Q) = A^3(\Pi, Q).$$

2) Si $Q \sim_{\Pi} P$, entonces tomamos $R - \Pi - P$ (es claro que siempre podemos tomar un punto así, por el axioma B2), con lo que también $R - \Pi - Q$, y así

$$\begin{aligned} A^3(\Pi, P) &= SA^3(\Pi, P) \cup \Pi \cup SA^3(\Pi, R) \\ &= SA^3(\Pi, Q) \cup \Pi \cup SA^3(\Pi, R) = A^3(\Pi, Q), \end{aligned}$$

donde la igualdad $SA^3(\Pi, P) = SA^3(\Pi, Q)$ se sigue de que dos puntos relacionados tienen la misma clase de equivalencia. ■

Teorema 1.28 *Sea Π un plano, sea P un punto exterior, sea r una recta contenida en Π y sea Π' el plano que contiene a r y a P . Entonces $\Pi' \subset A^3(\Pi, P)$.*

DEMOSTRACIÓN: El plano Π' es la unión disjunta de la recta r y los dos semiplanos en que ésta lo divide (desprovistos de su frontera r). Trivialmente $r \subset \Pi \subset A^3(\Pi, P)$. Si un punto Q está en el semiplano de Π' opuesto a P , entonces existe un $T \in r \subset \Pi$ tal que $Q - T - P$, luego $Q - \Pi - P$, luego $Q \in A^3(\Pi, P)$. Por último, si Q' está en el mismo semiplano que P , entonces $Q' - \Pi - Q$, luego $Q' \sim_{\Pi} P$, luego $Q' \in A^3(\Pi, P)$. ■

Supongamos ahora que Π y Π' son dos planos cuya intersección sea una recta r . Entonces podemos definir $A^3(\Pi, \Pi') = A^3(\Pi, P)$, donde P es cualquier punto de $\Pi' \setminus \Pi = \Pi' \setminus r$ sin que importe la elección de P .

En efecto, si tomamos otro punto $P' \in \Pi' \setminus \Pi$, entonces Π' es el plano que contiene a r y a P' , luego el teorema anterior nos da que $P' \in \Pi' \subset A^3(\Pi, P)$, y el teorema 1.27 nos da a su vez que $A^3(\Pi, P) = A^3(\Pi, P')$.

Teorema 1.29 *Sean Π y Π' dos planos cuya intersección sea una recta r . Entonces $\Pi \subset A^3(\Pi, \Pi')$, $\Pi' \subset A^3(\Pi, \Pi')$ y $A^3(\Pi, \Pi') = A^3(\Pi', \Pi)$.*

DEMOSTRACIÓN: Las dos inclusiones son obvias por la definición. Por simetría basta probar que $A^3(\Pi, \Pi') \subset A^3(\Pi', \Pi)$. Pongamos que $A^3(\Pi, \Pi') = A^3(\Pi, P)$, para cierto $P \in \Pi' \setminus \Pi$. Tomemos $X \in r$ y un punto Q tal que $P - X - Q$. Notemos que $P - \Pi - Q$ y $Q \in \Pi' \setminus r$.

Tomemos cualquier punto $Z \in A^3(\Pi, \Pi')$. Si $Z \in \Pi$ o $Z \in \Pi'$, ya sabemos que $Z \in A^3(\Pi', \Pi)$, luego podemos suponer lo contrario. Entonces $Z \in SA^3(\Pi, P)$ o $Z \in SA^3(\Pi, Q)$. Por simetría podemos suponer el primer caso.

Entonces $Z \sim_{\Pi} P$, luego $Z - \Pi - Q$, luego existe un $T \in \Pi$ tal que $Z - T - Q$. No puede ser que $T \in \Pi'$, pues entonces también $Z \in \Pi'$, luego $T \neq Q$. Ahora consideramos cualquier recta contenida en Π' que pase por Q y consideramos el plano Π^* que contiene a ésta y a T . Entonces el teorema anterior nos da que $Z \in \Pi^* \subset A^3(\Pi', T) = A^3(\Pi', \Pi)$. ■

Finalmente podemos definir:

Definición 1.30 *Dados cuatro puntos P_0, P_1, P_2, P_3 afínmente independientes, es decir, no coplanares, la variedad afín (tridimensional) que generan es*

$$A^3(P_0, P_1, P_2, P_3) = A^3(P_0P_1P_2, P_3).$$

Ahora es inmediato que la definición no depende del orden de los puntos. En efecto, observemos que los planos $P_0P_1P_2$ y $P_0P_1P_3$ se cortan en la recta P_0P_1 (la intersección no puede contener ningún otro punto X , pues en tal caso $P_0P_1P_2 = P_0P_1X = P_0P_1P_3$, luego $P_3 \in P_0P_1P_2$, y los cuatro puntos serían coplanares). Por lo tanto

$$\begin{aligned} A^3(P_0, P_1, P_2, P_3) &= A^3(P_0P_1P_2, P_3) = A^3(P_0P_1P_2, P_0P_1P_3) \\ &= A^3(P_0P_1P_3, P_0P_1P_2) = A^3(P_0P_1P_3, P_2) = A^3(P_0, P_1, P_3, P_2), \end{aligned}$$

e igual que hemos intercambiado P_2 con P_3 , podemos poner en cuarto lugar

cualquiera de los puntos. Ahora probamos el teorema fundamental sobre variedades tridimensionales:

Teorema 1.31 *Si E es una variedad afín tridimensional y $P_0, P_1, P_2, P_3 \in E$ son puntos no coplanares, entonces $E = A^3(P_0, P_1, P_2, P_3)$.*

DEMOSTRACIÓN: En principio $E = A^3(Q_0, Q_1, Q_2, Q_3) = A^3(Q_1Q_2Q_3, Q_0)$, para ciertos puntos afinmente independientes Q_i . Como los P_i también son afinmente independientes, alguno de ellos no está en el plano $Q_1Q_2Q_3$. No perdemos generalidad si suponemos que es P_0 , y entonces el teorema 1.27 nos da que

$$E = A^3(Q_1Q_2Q_3, P_0) = A^3(P_0, Q_1, Q_2, Q_3) = A^3(P_0Q_2Q_3, Q_1).$$

Nuevamente, uno de los puntos P_1, P_2, P_3 no puede estar en el plano $P_0Q_2Q_3$, y no perdemos generalidad si suponemos que es P_1 , con lo que

$$E = A^3(P_0Q_2Q_3, P_1) = A^3(P_0, P_1, Q_2, Q_3).$$

Continuando de este modo llegamos a que $E = A^3(P_0, P_1, P_2, P_3)$. ■

Por consiguiente:

Teorema 1.32 *Por cuatro puntos no coplanares P_0, P_1, P_2, P_3 pasa una única variedad afín tridimensional, que es la que representamos por $A^3(P_0, P_1, P_2, P_3)$.*

Notemos que, por definición, toda variedad afín tridimensional contiene cuatro puntos no coplanares. Más aún, si $X \subset E = A^3(P_0, P_1, P_2, P_3)$ está formado por uno, dos o tres puntos afinmente independientes, entonces X se puede extender hasta un generador de E .

En efecto, si $Q_0 \in E$, algún P_i tiene que ser distinto de Q_0 , luego existe un $Q_1 \in E$ distinto de Q_0 . A su vez, si $Q_0, Q_1 \in E$ son puntos distintos, algún P_i tiene que ser exterior a la recta Q_0Q_1 , luego existe un $Q_2 \in E$ tal que Q_0, Q_1, Q_2 no son colineales. Por último, dados tres puntos no colineales $Q_0, Q_1, Q_2 \in E$, algún P_i tiene que ser exterior al plano $Q_0Q_1Q_2$, luego existe un $Q_3 \in E$ tal que $Q_0, Q_1, Q_2, Q_3 \in E$ son puntos no coplanares, y entonces $E = A^3(Q_0, Q_1, Q_2, Q_3)$.

De aquí se sigue a su vez que si una variedad afín tridimensional E contiene a dos puntos, también contiene a la recta que los une, y si contiene tres puntos no colineales, también contiene al plano que determinan. Esto es la versión tridimensional del axioma A4. También es claro que si E contiene una recta r y un punto exterior P , entonces contiene al plano que determinan y, en particular, a la paralela a r por P .

Necesitamos una última propiedad de las variedades afines tridimensionales. Notemos que, en general, si Π y Π' son dos planos distintos, su intersección puede ser vacía, puede ser un punto o puede ser una recta, pero no hay más opciones, pues si la intersección contiene dos puntos, entonces contiene la recta r que los une (por A4), y no puede contener otro punto P exterior a r , ya que entonces $\Pi = \Pi'$ serían ambos el único plano que contiene a P y a dos puntos de r . En una variedad afín tridimensional uno de los casos es imposible:

Teorema 1.33 *Si E es una variedad afín tridimensional y $\Pi, \Pi' \subset E$ son dos planos distintos con intersección no vacía, entonces su intersección es una recta.*

DEMOSTRACIÓN: Supongamos que la intersección fuera sólo un punto P . Podemos tomar puntos P' y P'' en Π' tales que P, P', P'' no sean colineales. En particular, $P', P'' \notin \Pi$ (pues en caso contrario tendrían que ser P). Por lo tanto, $E = A^3(\Pi, P')$ y, como $P'' \in \Pi' \subset E$, tiene que darse uno de los dos casos: $P' - \Pi - P''$ o bien $P' \sim_{\Pi} P''$.

Si se da el primer caso existe un punto $T \in \Pi$ tal que $P' - T - P''$, pero entonces $T \in \Pi \cap \Pi'$, luego $T = P$, y llegamos a que P, P', P'' son colineales, en contra de lo supuesto.

Si se da el segundo caso tomamos P''' tal que $P''' - P - P''$ y entonces $P''' \in \Pi'$ y $P''' - \Pi - P''$, luego $P''' - \Pi - P'$. Más aún, P, P' y P''' siguen siendo no colineales, por lo que están en el caso anterior y tenemos igualmente una contradicción. ■

Ahora tenemos varias caracterizaciones de la tridimensionalidad del espacio:

Teorema 1.34 *Supongamos que existen cuatro puntos no coplanares. Las afirmaciones siguientes son equivalentes:*

1. *El espacio \mathbb{E} es una variedad afín de dimensión 3.*
2. *Si P_0, P_1, P_2, P_3 son puntos no coplanares, $\mathbb{E} = A^3(P_0, P_1, P_2, P_3)$.*
3. *Todo plano Π tiene un punto exterior y la relación \sim_{Π} en $\mathbb{E} \setminus \Pi$ determina dos clases de equivalencia, es decir, Π divide a $\mathbb{E} \setminus \Pi$ en dos semiespacios, de forma que dos puntos están en el mismo semiespacio si y sólo si Π no corta el segmento que los une.*
4. *Si dos planos tienen un punto en común, entonces tienen dos puntos en común (y, por consiguiente, su intersección es una recta).*

DEMOSTRACIÓN: 1) \Rightarrow 2) es el teorema 1.31.

2) \Leftrightarrow 3) Si P_0, P_1, P_2 son tres puntos no colineales en Π , sabemos que existe un cuarto punto P de modo que $\mathbb{E} = A^3(P_0, P_1, P_2, P) = A^3(\Pi, P)$. En particular P es un punto exterior a Π y 3) se cumple por definición de $A^3(\Pi, P)$. Recíprocamente, si se cumple 3), dados puntos P_i , tomamos $\Pi = P_0P_1P_2$, y entonces 3) implica que $\mathbb{E} = A^3(\Pi, P_3)$.

2) \Rightarrow 4) es el teorema anterior.

4) \Rightarrow 1). Por hipótesis existen cuatro puntos no coplanares P_0, P_1, P_2, P_3 . Si $\mathbb{E} \neq A^3(P_0, P_1, P_2, P_3)$, entonces existe un punto $P_4 \in \mathbb{E} \setminus A^3(P_0, P_1, P_2, P_3)$. Entonces P_0, P_3, P_4 no son colineales, pues si lo fueran $P_4 \in A^3(P_0, P_1, P_2, P_3)$. Por consiguiente, podemos considerar los planos $\Pi = P_0P_1P_2$ y $\Pi' = P_0P_3P_4$. Claramente $P_0 \in \Pi \cap \Pi'$. Por 4) tenemos que $\Pi \cap \Pi'$ es una recta r , en la cual no puede estar P_3 (pues en tal caso $P_3 \in \Pi$ y los cuatro puntos dados serían coplanares). Por lo tanto Π' es el plano determinado por dos puntos de $r \subset \Pi$ y P_3 , luego $P_4 \in \Pi' \subset A^3(P_0, P_1, P_2, P_3)$, contradicción. ■

Con esto tenemos ya una forma (o varias equivalentes) de completar los axiomas de incidencia para asegurar que el espacio sea tridimensional:

Definición 1.35 Una *geometría de Hilbert tridimensional* es una geometría de Hilbert que incluya los axiomas de incidencia siguientes:

Axioma A5b *Existen cuatro puntos no coplanares.*

Axioma A6 *Si dos planos tienen un punto en común, entonces tienen dos puntos en común.*

Hemos llamado A5b al primero de los axiomas porque con estos dos nuevos axiomas se puede demostrar el antiguo axioma A5. En efecto:

DEMOSTRACIÓN: Sea Π un plano y vamos a ver que contiene tres puntos no colineales. Por definición Π es no vacío, luego podemos tomar un punto $P \in \Pi$. Por el axioma A5b existen puntos Q y R no colineales con P . Si ambos están en Π ya hemos terminado. En caso contrario el plano PQR corta a Π en P y por A6 en otro punto más. No perdemos generalidad si suponemos que es Q , con lo que $R \notin \Pi$. De nuevo por A5b existe un punto S no coplanar con P, Q, R . El plano PRS corta a Π en P y en otro punto P' que no puede estar en PQ , pues PQ está contenida en PQR (por A4) y en tal caso los planos PRS y PQR tendrían en común los puntos P, R y P' , luego serían iguales (por A3) y P, Q, R, S serían coplanares. Por lo tanto, los puntos P, Q, P' son tres puntos no colineales en Π . ■

Del mismo modo que los planos cumplen los axiomas de la geometría plana, es fácil ver que las variedades tridimensionales cumplen los axiomas de la geometría tridimensional, por lo que los teoremas de la geometría tridimensional son aplicables a geometrías de dimensiones superiores particularizados a subvariedades tridimensionales.

En la geometría tridimensional dos planos se dicen *coincidentes* si son iguales, *secantes* si su intersección es una recta y *paralelos* si son iguales o no tienen puntos comunes. No obstante, del mismo modo que para que dos rectas sean paralelas se exige que estén contenidas en un mismo plano (aunque esto es trivial en el caso de la geometría plana), para que dos planos se consideren paralelos tienen que estar contenidos en la misma variedad afín tridimensional (aunque esto es trivial en la geometría tridimensional).

Observemos que un punto intermedio entre no suponer los axiomas de tridimensionalidad y suponerlos consiste en suponer únicamente A5b (sin reemplazar entonces a A5), lo cual significa postular que el espacio tiene al menos dimensión 3, pero sin excluir que pueda tener una dimensión mayor. Conviene observar una consecuencia de A5b:

Teorema 1.36 *Toda recta es intersección de dos planos.*

DEMOSTRACIÓN: Sea $r = AB$ una recta. Como existen puntos no colineales, tiene que haber un punto C exterior a r . Sea Π el plano ABC . Como existen puntos no coplanares, tiene que haber un punto D exterior a Π . Sea $\Pi' = ABD$. Es claro entonces que $r = \Pi \cap \Pi'$. ■

De momento estamos interesados en describir la geometría intuitiva, es decir, tridimensional, pero es conveniente no usar los axiomas que exigen la tridimensionalidad del espacio si no son necesarios, de modo que los resultados que obtengamos valdrán igualmente para geometrías de dimensiones superiores. Vamos a ver un ejemplo de cómo el concepto de variedad afín tridimensional nos permite evitar estos axiomas en una demostración.

Teorema 1.37 *Admitiendo el axioma de las paralelas, si dos rectas son paralelas a una tercera, entonces son paralelas entre sí.*

DEMOSTRACIÓN: Sean r_1 y r_2 paralelas a r_3 . Supongamos que r_1 y r_2 no son paralelas (en particular que son distintas). Sea Π_1 el plano que contiene a r_1 , r_3 y sea Π_2 el plano que contiene a r_2 y r_3 . Si $\Pi_1 = \Pi_2$ entonces r_1 y r_2 están en el mismo plano, y como no son paralelas se cortan en un punto P , con lo que r_3 tiene dos paralelas distintas por P , contradicción.

Si $\Pi_1 \neq \Pi_2$ entonces la intersección de ambos planos es la recta r_3 , luego podemos formar la variedad afín $E = A^3(\Pi_1, \Pi_2)$. Sea Π el plano que contiene a $r_1 \subset \Pi_1$ y a un punto Q de $r_2 \subset \Pi_2$. Entonces $E = A^3(\Pi_1, Q)$ y $\Pi \subset E$ por el teorema 1.28. Por consiguiente, el teorema 1.33 nos da que Π corta a Π_2 en una recta r' disjunta de r_3 (pues si R está en $r' \cap r_3$ entonces no puede estar en r_1 , ya que entonces estaría en $r_1 \cap r_3$. De aquí que $\Pi = \Pi_1$, pues tienen a r_1 y a R en común, luego $r' = r_3$ y Q estaría en $r_2 \cap r_3$). Por lo tanto r' es paralela a r_3 y pasa por un punto de r_2 , luego $r' = r_2$. De nuevo tenemos que r_1 y r_2 están en un mismo plano, y concluimos como antes. ■

Nota Conviene observar que el teorema anterior se demuestra a partir de los axiomas de incidencia y ordenación sin los axiomas de tridimensionalidad (los de ordenación intervienen en la construcción de la variedad afín $A^3(\Pi_1, \Pi_2)$), pero si suponemos los axiomas de incidencia de la geometría tridimensional, entonces no es necesario considerar dicha variedad, ya que la hemos usado únicamente para probar que los planos Π y Π_2 se cortan en una recta, y esto se obtiene directamente de A6. ■

1.4 Ángulos y triángulos

Definición 1.38 Sean l_1 y l_2 dos semirrectas con origen común O y no contenidas en la misma recta. Sean r_1 y r_2 sus respectivas prolongaciones. Sea Π el plano que las contiene. Hemos visto que l_1 está contenido en uno de los semiplanos en que r_2 divide a Π y l_2 está contenido en uno de los semiplanos en que r_1 divide a Π . Llamaremos *ángulo* (lat. ‘rincón’) de *vértice* (lat. ‘cumbre’) O y lados l_1 y l_2 a la intersección del semiplano de Π respecto a r_2 que contiene a l_1 con el semiplano de Π respecto a r_1 que contiene a l_2 . Lo representaremos $\widehat{l_1 l_2}$. Los puntos de l_1 y l_2 constituyen la *frontera* del ángulo.

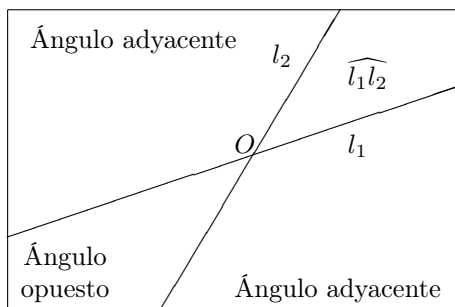
Observemos que $\widehat{l_1 l_2}$ contiene más puntos, aparte de los de sus lados. De hecho es un conjunto convexo, pues es la intersección de dos conjuntos convexos.

Por lo tanto, si A y B son puntos en l_1 y l_2 respectivamente, entonces todos los puntos entre ellos están en el ángulo.

Si tres puntos A , O y B no son colineales, llamaremos \widehat{AOB} al ángulo de vértice O y lados \overrightarrow{OA} y \overrightarrow{OB} .

Un ángulo está contenido en un único plano, llamado su *soporte*. Se comprueba que un ángulo determina su vértice y sus lados.

Dos rectas secantes dividen el plano que las contiene en cuatro ángulos con vértice común. Dos ángulos con el mismo vértice, un lado en común y los otros lados formados por semirrectas complementarias se llaman ángulos *adyacentes*. Dos ángulos con el mismo vértice y cuyos lados son semirrectas complementarias se llaman ángulos *opuestos por el vértice*. Cada ángulo tiene exactamente dos ángulos adyacentes y un ángulo opuesto por el vértice.



Teorema 1.39 Sean A , O , B puntos no colineales en un plano Π . Entonces una semirrecta de origen O y contenida en Π está contenida en el ángulo \widehat{AOB} si y sólo si corta al segmento \overline{AB} .

DEMOSTRACIÓN: Sea s una semirrecta de origen O y contenida en Π . Podemos suponer que no es uno de los lados de \widehat{AOB} . Si s corta a \overline{AB} en un punto X y P es cualquier otro punto de s (distinto de O), entonces la prolongación del segmento \overline{PX} es la prolongación de s , que corta a los lados del ángulo en el punto O , y éste no está en \overline{PX} . Por lo tanto P y X están en los mismos semiplanos respecto a los lados del ángulo, y como X está en el ángulo, P también. Esto prueba que s está contenida en \widehat{AOB} .

Recíprocamente, supongamos que s está contenida en \widehat{AOB} . Sea \bar{s} la prolongación de s y consideremos un punto C tal que O esté entre C y B . Los puntos A , B y C no son colineales, y \bar{s} no pasa por ninguno de ellos. Como \bar{s} pasa por O , que es un punto entre C y B , por el axioma B4 ha de pasar por un punto entre A y B o bien por un punto entre A y C . Ahora bien, s está contenida en \widehat{AOB} y puesto que al pasar por O cruza a la vez sus dos lados, es claro que la semirrecta complementaria de s está contenida en el ángulo opuesto por el vértice a \widehat{AOB} . Sin embargo el segmento \overline{AC} está contenido en el ángulo \widehat{AOC} , que es uno de los adyacentes a \widehat{AOB} . Por lo tanto \bar{s} no puede cortar a \overline{AC} . Así pues, \bar{s} corta a \overline{AB} . Sin embargo, la semirrecta complementaria de s

no puede cortar a este segmento, pues está fuera de \widehat{AOB} . Concluimos que es s quien corta al segmento. ■

Observemos que si un punto P está en un ángulo \widehat{AOB} entonces la semirrecta \overrightarrow{OP} está contenida en \widehat{AOB} .

Teorema 1.40 *Si dos ángulos tienen un lado en común y están contenidos en el mismo semiplano respecto a él, uno de ellos está contenido en el otro.*

DEMOSTRACIÓN: Los ángulos serán de la forma \widehat{AOB}_1 y \widehat{AOB}_2 . Supongamos que el segmento \overline{AB}_1 no corta a la recta OB_2 . Entonces todos los puntos de \overline{AB}_1 están en el mismo semiplano que A respecto a OB_2 , y por hipótesis también están en el mismo semiplano que B_2 respecto a OA . Por lo tanto \overline{AB}_1 está contenido en \widehat{AOB}_2 . Si P es cualquier punto de \widehat{AOB}_1 (que no esté en OA) la semirrecta \overrightarrow{OP} corta a \overline{AB}_1 en un punto X . Si $X = P$ ya tenemos que P está en \widehat{AOB}_2 . En caso contrario, el segmento \overline{XP} no contiene a O , por lo que X y P están en los mismos semiplanos respecto a los lados de \widehat{AOB}_2 , luego P está en \widehat{AOB}_2 .

Supongamos ahora que \overline{AB}_1 corta a la recta OB_2 en un punto X . Entonces X está en \widehat{AOB}_1 , luego está en el mismo semiplano que B_2 respecto a OA , luego en realidad X está en \overrightarrow{OB}_2 . Esto implica que $\widehat{AOX} = \widehat{AOB}_2$, luego no perdemos generalidad si suponemos que $X = B_2$. Ahora, si P está en \widehat{AOB}_2 la semirrecta \overrightarrow{OP} corta a \overline{AB}_2 , luego corta a \overline{AB}_1 , luego está contenida en \widehat{AOB}_1 , luego P está en \widehat{AOB}_1 . ■

De la prueba del teorema anterior y del teorema 1.39 se deduce el hecho siguiente:

Teorema 1.41 *Sean l_1, l_2 y l_3 semirrectas de origen O tales que dos de ellas no estén sobre una misma recta y de modo que l_2 y l_3 estén contenidas en un mismo semiplano respecto de l_1 . Entonces $\widehat{l_1l_2}$ está contenido en $\widehat{l_1l_3}$ si y sólo si l_2 está contenida en $\widehat{l_1l_3}$.*

También tenemos esta equivalencia:

Teorema 1.42 *Sean l_1, l_2 y l_3 semirrectas de origen O tales que dos de ellas no estén sobre una misma recta y de modo que l_2 y l_3 estén contenidas en un mismo semiplano respecto de l_1 . Entonces l_2 está contenida en $\widehat{l_1l_3}$ si y sólo si l_1 y l_3 están en semiplanos opuestos respecto de l_2 .*

DEMOSTRACIÓN: Pongamos que $l_1 = \overrightarrow{OA}$, $l_3 = \overrightarrow{OC}$. Notemos que l_1 está contenida en uno de los semiplanos determinados por l_2 en el plano que contiene a ambas, pues si P está en l_1 y es distinto de O , entonces \overline{AP} no contiene a O , luego no corta a la prolongación de l_2 , luego l_1 está en el semiplano que contiene a A . Lo mismo vale para l_3 .

Si l_2 está en $\widehat{l_1 l_3}$, por el teorema 1.39 tenemos que $l_2 = \overrightarrow{OB}$, con $A - B - C$. Entonces A y C están en semiplanos opuestos respecto de l_2 , luego lo mismo vale para l_1 y l_3 , pues l_1 está en el semiplano de A y l_3 en el de C .

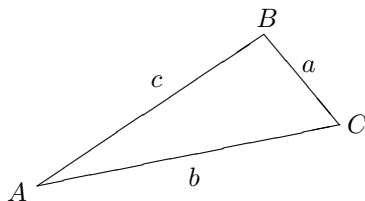
Recíprocamente, si l_1 y l_3 están en semiplanos opuestos respecto a l_2 , el segmento AC corta a la prolongación de l_2 en un punto $B \neq O$ (pues si fuera $B = O$ entonces l_1 y l_3 estarían alineadas). Por el teorema 1.39 tenemos que la semirrecta \overrightarrow{OB} está contenida en $\widehat{l_1 l_3}$, luego en particular l_3 y \overrightarrow{OB} están en el mismo semiplano respecto de l_1 . Basta probar que es l_2 , pero en caso contrario sería su semirrecta opuesta, y así l_2 y l_3 estarían en semiplanos opuestos respecto de l_1 , en contra de lo supuesto. ■

Definición 1.43 Sean A , B y C tres puntos no colineales. Llamaremos *triángulo de vértices A , B y C* a la intersección de los ángulos \widehat{BAC} , \widehat{ABC} y \widehat{ACB} . Lo representaremos por \widehat{ABC} .

Los ángulos \widehat{BAC} , \widehat{ABC} y \widehat{ACB} se llaman *ángulos* del triángulo \widehat{ABC} . Cuando no haya ambigüedad, nos referiremos a ellos como \hat{A} , \hat{B} y \hat{C} , respectivamente (es decir, los nombraremos por sus vértices). Los segmentos \overline{AB} , \overline{AC} y \overline{BC} se llaman *lados* de \widehat{ABC} . Los tres lados de un triángulo forman su *frontera*.

Los lados \overline{AB} y \overline{AC} se llaman lados *contiguos* al ángulo \hat{A} , mientras que el lado \overline{BC} es el lado *opuesto* al ángulo \hat{A} (similarmente con los otros dos ángulos).

Normalmente llamaremos a , b y c a los lados de un triángulo \widehat{ABC} , de modo que a será el lado opuesto al ángulo \hat{A} , b será el lado opuesto a \hat{B} y c será el lado opuesto a \hat{C} .



1.5 Axiomas de congruencia

Continuamos introduciendo conceptos geométricos básicos estudiando ahora la congruencia de figuras. La idea subyacente es que dos figuras son congruentes si se diferencian a lo sumo en su posición en el espacio, es decir, si una puede convertirse en la otra mediante un movimiento. Aunque en principio el concepto de congruencia es aplicable a cualquier figura, de momento sólo necesitamos considerar congruencias de segmentos, ángulos y triángulos. Además, la congruencia de triángulos puede definirse en términos de las otras dos.

Definición 1.44 Una *geometría métrica* es una geometría ordenada junto con dos relaciones, que llamaremos de *congruencia* y las representaremos por \equiv , definidas respectivamente sobre los conjuntos de los segmentos y ángulos, y que cumplen los axiomas siguientes:

Axioma C1 Las dos congruencias son relaciones de equivalencia, es decir, son reflexivas, simétricas y transitivas.

Axioma C2 Dados puntos $A \neq B$ y A' y una semirrecta s de origen A' , existe un único punto B' en s tal que $\overline{AB} \equiv \overline{A'B'}$.

Axioma C3 Sean A, B, C puntos colineales de modo que B esté entre A y C , sean A', B' y C' otros tres puntos en las mismas condiciones. Entonces, si $\overline{AB} \equiv \overline{A'B'}$ y $\overline{BC} \equiv \overline{B'C'}$, también $\overline{AC} \equiv \overline{A'C'}$.

Axioma C4 Sea L un ángulo, s una semirrecta y Π un semiplano cuya frontera sea la prolongación de s . Entonces existe un único ángulo L' contenido en Π , con un lado igual a s y tal que $L \equiv L'$.

Diremos que dos triángulos T y T' son *congruentes* si existe una correspondencia entre sus vértices para la cual cada par de lados y ángulos correspondientes son congruentes. En lo sucesivo, cuando digamos que dos triángulos \overline{ABC} y $\overline{A'B'C'}$ son congruentes, se sobrentenderá que cumplen la definición para la correspondencia $A \rightarrow A', B \rightarrow B', C \rightarrow C'$, es decir, que $\overline{AB} \equiv \overline{A'B'}$, etc.

Axioma C5 Dado un triángulo \overline{ABC} , un segmento $\overline{A'B'} \equiv \overline{AB}$ y un semiplano Π de frontera la prolongación de $\overline{A'B'}$, existe un (único) triángulo $\overline{A'B'C'}$ contenido en Π y congruente con \overline{ABC} .

Observemos que la unicidad del triángulo se sigue del axioma C4, pues las rectas $A'C'$ y $B'C'$ son únicas, y C' ha de ser su intersección. En el lenguaje tradicional de la geometría es costumbre hablar de ángulos, segmentos y triángulos ‘iguales’ en el sentido que aquí hemos dado a la palabra ‘congruentes’, mientras que para indicar que dos segmentos, ángulos o triángulos son iguales en el sentido conjuntista, es decir, que contienen los mismos puntos, se suele decir que son ‘coincidentes’. Nosotros usaremos estos términos excepto cuando pueda dar lugar a confusión.

Comencemos estudiando las propiedades de la congruencia de segmentos. El axioma C3 nos permite definir una suma:

Teorema 1.45 Dados dos segmentos \overline{AB} y \overline{CD} existe un segmento \overline{PQ} con la propiedad de que existe un punto R entre P y Q de modo que $\overline{PR} \equiv \overline{AB}$ y $\overline{RQ} \equiv \overline{CD}$. La clase de congruencia de \overline{PQ} sólo depende de las clases de congruencia de \overline{AB} y \overline{CD} .

DEMOSTRACIÓN: Tomamos por ejemplo $P = A$ y $R = B$. Ahora consideramos la semirrecta de AB con origen B y que no contiene a A . Por el axioma C2 existe en ella un punto Q tal que $\overline{BQ} \equiv \overline{CD}$. Es claro que P, Q y R cumplen lo pedido. El resto es consecuencia del axioma C3. ■

Definición 1.46 En las condiciones del teorema anterior, escribiremos

$$\overline{PQ} \equiv \overline{AB} + \overline{CD},$$

entendiendo la expresión como una igualdad entre clases de congruencia.³

De este modo tenemos definida la suma de dos (clases de) segmentos cualesquiera. Es obvio que esta suma es asociativa y conmutativa. El hecho siguiente no es exactamente una consecuencia inmediata de la definición de suma:

Teorema 1.47 *Si $\overline{PQ} \equiv u + v$ entonces existe un único punto R entre P y Q tal que $\overline{PR} \equiv u$ y $\overline{RQ} \equiv v$.*

DEMOSTRACIÓN: Por el axioma C2 existe un único punto R en la semirrecta \overrightarrow{PQ} tal que $\overline{PR} \equiv u$. Del mismo modo, existe un punto R' en la semirrecta de PQ de origen R y que no contiene a P de modo que $\overline{RR'} \equiv v$. Por definición de suma tenemos que $\overline{PR'} \equiv u + v \equiv \overline{PQ}$, luego la unicidad del axioma C2 implica que $R' = Q$. ■

De aquí se sigue que la suma de segmentos es simplificable:

Teorema 1.48 *Dados tres segmentos u, v y w , si $u + v \equiv u + w$ entonces $v \equiv w$.*

DEMOSTRACIÓN: Sea $u + v \equiv u + w \equiv \overline{PQ}$. Entonces existe un punto R entre P y Q tal que $\overline{PR} \equiv u$ y $\overline{RQ} \equiv v$. También tiene que existir un punto R' tal que $\overline{PR'} \equiv u$ y $\overline{R'Q} \equiv w$, y por la unicidad de C2 ha de ser $R = R'$, luego $v \equiv \overline{RQ} = \overline{R'Q} \equiv w$. ■

Definición 1.49 Diremos que un segmento u es menor que un segmento v (y lo representaremos por $u < v$) si existe un segmento w tal que $v \equiv u + w$. En tal caso el teorema anterior afirma que w es único (salvo congruencia) y lo llamaremos *resta* o *diferencia* de u y v , y lo representaremos $w \equiv v - u$.

De las propiedades de la suma se sigue inmediatamente que la desigualdad de segmentos depende sólo de las clases de congruencia y es una relación de orden estricto. También es fácil probar lo siguiente:

Teorema 1.50 *Si $\overline{AB} < \overline{AC}$ y ambos segmentos están situados sobre una semirrecta de origen A , entonces B está entre A y C .*

Nos ocupamos ahora de la congruencia de ángulos y triángulos. Comenzamos con dos criterios de igualdad de triángulos.

Teorema 1.51 (Criterio lado-ángulo-lado) *Si dos triángulos $T = \widehat{ABC}$ y $T' = \widehat{A'B'C'}$ cumplen $\overline{AB} \equiv \overline{A'B'}$, $\widehat{A} \equiv \widehat{A'}$ y $\overline{AC} \equiv \overline{A'C'}$ entonces $T \equiv T'$.*

³Esto significa que, dados dos segmentos u y v no tenemos definido ningún segmento en concreto al que llamar $u + v$, sino una única clase de congruencia de segmentos w que cumplen $w \equiv u + v$.

DEMOSTRACIÓN: Por el axioma C5 el triángulo T es igual a un triángulo $T'' = \widehat{A'B'C''}$ contenido en el mismo semiplano que T' respecto a la recta $A'B'$ y con todos sus lados y ángulos congruentes con los de T . En particular, usando el axioma C1 resulta que $\widehat{A'C'} \equiv \widehat{A'C''}$ y $\widehat{B'A'C'} \equiv \widehat{B'A'C''}$.

Estos últimos ángulos tienen un lado en común y están contenidos en el mismo semiplano respecto a dicho lado, luego por el axioma C4 son coincidentes. En particular la semirrecta $\overrightarrow{A'C'}$ coincide con $\overrightarrow{A'C''}$. Ahora el axioma C2 implica que $C' = C''$, por lo que T' coincide con T'' , luego T y T' son iguales. ■

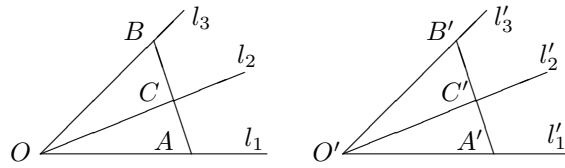
Teorema 1.52 (Criterio ángulo-lado-ángulo) *Si dos triángulos $T = \widehat{ABC}$ y $T' = \widehat{A'B'C'}$ cumplen $\overline{AB} \equiv \overline{A'B'}$, $\hat{A} \equiv \hat{A}'$ y $\hat{B} \equiv \hat{B}'$ entonces $T \equiv T'$.*

DEMOSTRACIÓN: Razonamos igual que en el teorema anterior, con lo que obtenemos un triángulo $T'' = \widehat{A'B'C''}$ igual a T y que comparte con T' el lado $\overline{A'B'}$ y los ángulos \hat{A}' y \hat{B}' . Esto implica que $\overrightarrow{A'C'} = \overrightarrow{A'C''}$ y $\overrightarrow{B'C'} = \overrightarrow{B'C''}$, pero C' y C'' son los respectivos puntos donde se cortan estas semirrectas, luego $C' = C''$ y concluimos igualmente. ■

Más adelante probaremos que si dos triángulos tienen dos ángulos iguales entonces tienen los tres ángulos iguales, con lo que el criterio anterior cubrirá cualquier caso en que dos triángulos tengan iguales dos ángulos y un lado. Más delicado es probar el criterio lado-lado-lado. Necesitamos algunos resultados previos:

Teorema 1.53 *Sean l_1, l_2 y l_3 semirrectas de origen O de modo que no haya dos sobre una misma recta y con l_2 y l_3 en el mismo semiplano respecto de l_1 . Sean l'_1, l'_2 y l'_3 semirrectas de origen O' en las mismas condiciones. Si l_2 está contenida en $\widehat{l_1 l_3}$, $\widehat{l_1 l_2} \equiv \widehat{l'_1 l'_2}$ y $\widehat{l_1 l_3} \equiv \widehat{l'_1 l'_3}$, entonces l'_2 está en $\widehat{l'_1 l'_3}$ y $\widehat{l_2 l_3} \equiv \widehat{l'_2 l'_3}$.*

DEMOSTRACIÓN: Tomemos puntos $A \in l_1$ y $B \in l_3$. Por el teorema 1.39 sabemos que la semirrecta l_2 corta a \overline{AB} en un punto C . Consideremos puntos $A' \in l'_1$ y $B' \in l'_3$ tales que $\overline{O'A'} \equiv \overline{OA}$ y $\overline{O'B'} \equiv \overline{OB}$.



Por el criterio LAL de igualdad de triángulos tenemos que $\widehat{AOB} \equiv \widehat{A'O'B'}$, luego $\overline{AB} \equiv \overline{A'B'}$. Sea C' el punto de $\overline{A'B'}$ que cumple $\overline{AC} \equiv \overline{A'C'}$. Como $\overline{AC} < \overline{AB}$ resulta que C' está entre A' y B' . Veamos que está en l'_2 . Esto se debe a que los triángulos \widehat{AOC} y $\widehat{A'O'C'}$ son iguales, porque tienen iguales el ángulo $\hat{A} \equiv \hat{A}'$ y los lados $\overline{OA} \equiv \overline{O'A'}$ y $\overline{AC} \equiv \overline{A'C'}$. Por lo tanto $\widehat{AOC} \equiv \widehat{A'O'C'}$, y por la unicidad del axioma C4 ha de ser $l'_2 = \overrightarrow{O'C'}$. (Notemos que $\overrightarrow{O'C'}$ está en $\widehat{l'_1 l'_3}$ por 1.39, luego está en el mismo semiplano que l'_3 respecto de l'_1 , luego por hipótesis también en el mismo que l'_2 .) En particular l'_2 está en $\widehat{l'_1 l'_3}$.

El teorema 1.48 implica que $\overline{BC} \equiv \overline{B'C'}$, y el criterio LAL (teniendo en cuenta que $\hat{B} \equiv \hat{B}'$) implica que $\widehat{BOC} \equiv \widehat{B'O'C'}$, luego en particular $\widehat{l_2l_3} \equiv \widehat{l'_2l'_3}$. ■

A su vez:

Teorema 1.54 Sean l_1, l_2 y l_3 semirrectas de origen O tales que l_2 esté contenida en $\widehat{l_1l_3}$. Sean l'_1, l'_2 y l'_3 semirrectas de origen O' en las mismas condiciones. Si $\widehat{l_1l_2} \equiv \widehat{l'_1l'_2}$ y $\widehat{l_2l_3} \equiv \widehat{l'_2l'_3}$ entonces $\widehat{l_1l_3} \equiv \widehat{l'_1l'_3}$.

DEMOSTRACIÓN: Por el axioma C4 existe una semirrecta l''_3 de origen O' y contenida en el mismo semiplano respecto a l'_1 que l'_3 y de modo que $\widehat{l_1l_3} \equiv \widehat{l'_1l''_3}$. Hemos de probar que $l'_3 = l''_3$, pero esto es consecuencia de la unicidad del axioma C4, ya que, por una parte, el teorema anterior implica que l'_2 está en $\widehat{l'_1l''_3}$ y que $\widehat{l'_2l''_3} \equiv \widehat{l_2l_3} \equiv \widehat{l'_2l'_3}$, mientras que el teorema 1.42 nos da que l'_3 y l''_3 están en el mismo semiplano respecto de l'_2 , ya que ambas están en el semiplano opuesto a l'_1 . ■

Definición 1.55 Un triángulo es *equilátero* (lat. 'de lados iguales') si sus tres lados son iguales. Un triángulo es *isósceles* (gr. 'de piernas iguales') si tiene al menos dos lados iguales. Un triángulo es *escaleno* (gr. 'oblicuo') si sus lados son desiguales dos a dos.

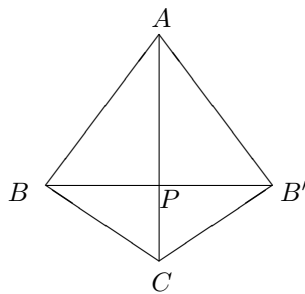
Probamos ahora un resultado con una prueba mucho más elemental de lo que podría preverse:

Teorema 1.56 (Criterio del triángulo isósceles) Si en un triángulo \widehat{ABC} se cumple $\overline{CA} \equiv \overline{CB}$ entonces $\hat{A} \equiv \hat{B}$.

DEMOSTRACIÓN: Basta aplicar el criterio LAL a los triángulos (coincidentes) \widehat{ACB} y \widehat{BCA} (es decir, tomando $A' = B, B' = A$ y $C' = C$). ■

Observemos que el teorema anterior implica que los triángulos equiláteros tienen también sus tres ángulos iguales. Finalmente podemos probar:

Teorema 1.57 (Criterio lado-lado-lado) Si $T = \widehat{ABC}$ y $T' = \widehat{A'B'C'}$ cumplen $\overline{AB} \equiv \overline{A'B'}$, $\overline{AC} \equiv \overline{A'C'}$ y $\overline{BC} \equiv \overline{B'C'}$, entonces $T \equiv T'$.



DEMOSTRACIÓN: Trasladando uno de los triángulos (por C5) podemos suponer que $\overline{AC} = \overline{A'C'}$ y que los vértices B y B' se encuentran en semiplanos distintos respecto a este lado común.

En estos términos $\overline{AB} \equiv \overline{A'B'}$, $\overline{BC} \equiv \overline{B'C'}$ y hemos de probar que $\widehat{A'B'C'} \equiv \widehat{ABC}$. Para ello basta probar que $\widehat{AB'C'} \equiv \widehat{ABC}$, pues entonces el criterio LAL nos da el resultado.

Tenemos que el segmento $\overline{BB'}$ corta a la recta AC en un punto P . Si es $P = C$ concluimos por el teorema anterior para

$\widehat{ABB'}$, e igualmente si $P = A$. Supongamos, pues, que A, P, C son distintos dos a dos. Si es $A - P - C$, aplicamos el teorema anterior a los triángulos $\widehat{ABB'}$ y $\widehat{CBB'}$, con lo que obtenemos $\widehat{ABB'} \equiv \widehat{AB'B}$ y $\widehat{CBB'} \equiv \widehat{CB'B}$. Entonces el teorema 1.54 implica que $\widehat{AB'C} \equiv \widehat{ABC}$. Si es $P - C - A$ o $P - A - C$ concluimos análogamente usando ahora el teorema 1.53. ■

1.6 Suma de ángulos

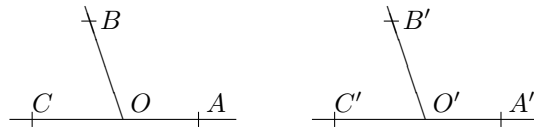
El teorema 1.54 nos permite definir una suma de ángulos de forma similar a como hemos definido la suma de segmentos. Conviene definir primero la ordenación de los ángulos.

Definición 1.58 Diremos que un ángulo A es menor que un ángulo B (y lo representaremos por $A < B$) si existen ángulos A' y B' congruentes con A y B respectivamente, con un lado en común, situados en un mismo semiplano respecto a dicho lado y de modo que A' está (estrictamente) contenido en B' .

Si A y B son dos ángulos no congruentes, por el axioma C4 existen ángulos A' y B' en las condiciones de la definición, por el teorema 1.40 uno de ellos estará contenido en el otro y por el teorema 1.54 el resultado de la comparación depende sólo de las clases de congruencia de A y B . Ahora es fácil probar que la relación que acabamos de definir es ciertamente una relación de orden total estricto sobre (las clases de congruencia de) todos los ángulos, así como que si dos ángulos comparten un lado y están contenidos en un mismo semiplano respecto a dicho lado, entonces el menor estará contenido en el mayor.

Teorema 1.59 Si L y L' son ángulos iguales, S es un ángulo adyacente a L y S' es un ángulo adyacente a L' , entonces S y S' también son iguales.

DEMOSTRACIÓN: Digamos que $L = \widehat{AOB}$ y que $S = \widehat{COB}$, donde los puntos C, O, A están alineados. Sea O' el vértice de L' y S' , tomemos B' en el lado común entre ambos y de modo que $\overline{OB} \equiv \overline{O'B'}$. Sea A' en el otro lado de L' de modo que $\overline{OA} \equiv \overline{O'A'}$ y sea C' en el otro lado de S' de modo que $\overline{OC} \equiv \overline{O'C'}$.



Es claro entonces que $\overline{CA} \equiv \overline{C'A'}$, $\widehat{AOB} \equiv \widehat{A'O'B'}$ (por el criterio LAL), de donde $\overline{AB} \equiv \overline{A'B'}$, de donde $\widehat{CAB} \equiv \widehat{C'A'B'}$ (por el mismo criterio), de donde $\overline{BC} \equiv \overline{B'C'}$, de donde $\widehat{COB} \equiv \widehat{C'O'B'}$ (por el criterio LLL). Esto implica que $\widehat{COB} \equiv \widehat{C'O'B'}$, es decir, $S \equiv S'$. ■

Como consecuencia inmediata tenemos:

Teorema 1.60 *Los dos ángulos adyacentes a un ángulo dado son iguales entre sí. Dos ángulos opuestos por el vértice son iguales entre sí.*

(Notemos que dos ángulos opuestos por el vértice son adyacentes a un mismo ángulo.)

Definición 1.61 Dos ángulos son *suplementarios* si uno es congruente con un ángulo adyacente al otro.

Es obvio que la relación de ser suplementarios depende sólo de las clases de congruencia de los ángulos, así como que es simétrica. Si dos ángulos suplementarios tienen un lado en común y están en semiplanos opuestos respecto a éste, entonces son adyacentes.

Teorema 1.62 *Si un ángulo A es menor que un ángulo B , entonces el suplementario de B es menor que el suplementario de A .*

DEMOSTRACIÓN: Podemos suponer que A y B tienen un lado en común y están contenidos en un mismo semiplano respecto a éste. Digamos $A = \widehat{l_1 l_2}$ y $B = \widehat{l_1 l_3}$. Sea l'_1 la semirrecta complementaria de l_1 . Entonces los suplementarios de A y B son $\widehat{l'_1 l_2}$ y $\widehat{l'_1 l_3}$. Basta probar que l_3 está contenida en $\widehat{l'_1 l_2}$. Por el teorema 1.39 esto equivale a que l'_1 y l_2 estén en semiplanos distintos respecto a l_3 , lo que a su vez equivale a que l_1 y l_2 estén en el mismo semiplano respecto a l_3 , y por definición de ángulo esto equivale a que l_2 esté contenida en $\widehat{l_1 l_3}$, lo cual es cierto por hipótesis. ■

Definición 1.63 Un *semihaz de semirrectas* (lat. *fascem* = ‘manejo’) es el conjunto de todas las semirrectas con un origen común O contenidas en un semiplano S que tenga a O en su frontera.

Notemos que un semihaz determina su origen y su semiplano. El origen O divide a la frontera de S en dos semirrectas s y t a las que llamaremos *extremos* del semihaz. Cada semirrecta l en un semihaz que sea distinta de sus extremos determina dos ángulos suplementarios \widehat{sl} y \widehat{tl} . Esto nos permite definir dos ordenaciones totales en el semihaz: una dada por

$$l \leq_{st} l' \quad \text{si y sólo si} \quad \widehat{sl} \leq \widehat{sl'}$$

(con el convenio adicional de que $s \leq_{st} l \leq_{st} t$, para cualquier semirrecta l distinta de s y t), y otra \leq_{ts} definida análogamente. El teorema anterior prueba que ambas ordenaciones son mutuamente inversas.

En particular, si l_1 , l_2 y l_3 son tres semirrectas de un semihaz, diremos que l_2 está *entre* l_1 y l_3 si $l_1 \leq_{st} l_2 \leq_{st} l_3$, donde el orden de los extremos s y t es irrelevante. Es fácil ver que esto sucede si y sólo si l_1 y l_3 están en semiplanos distintos respecto a l_2 .

Ahora conviene adoptar el convenio siguiente:

Definición 1.64 Llamaremos *ángulos llanos* a los semiplanos. Extendemos la congruencia de ángulos a los ángulos llanos estipulando que todos ellos son congruentes entre sí y no son congruentes con ningún ángulo en sentido estricto. Extendemos la relación de orden estipulando que un ángulo llano es mayor que cualquier ángulo en sentido estricto.

Notemos que un ángulo llano no tiene definidos un vértice y unos lados. Podemos considerar como tales a un punto cualquiera de su frontera y las semirrectas que éste determina, pero hay infinitos ángulos llanos con el mismo vértice y los mismos lados (todos los semiplanos con una misma recta como frontera). Los ángulos llanos no tienen suplementario. Si s y t son los extremos de un semihaz de semirrectas contenidas en el semiplano S , entonces convendremos en que \widehat{st} representa precisamente a S (pero esto sólo tiene sentido con respecto a un semihaz prefijado). Con este convenio se cumple el teorema siguiente (la demostración es muy simple):

Teorema 1.65 Sean l_1 y l_2 dos semirrectas distintas en un semihaz prefijado. Entonces $\widehat{l_1 l_2}$ es la unión de todas las semirrectas del semihaz que están entre l_1 y l_2 .

Todas estas consideraciones nos permiten estudiar cómodamente la suma de ángulos.

Definición 1.66 Diremos que un ángulo $A = \widehat{l_1 l_2}$ es la *suma* de dos ángulos B y C si existe una semirrecta l_3 de origen el vértice de A y contenida en A tal que $B \equiv \widehat{l_1 l_3}$ y $C \equiv \widehat{l_3 l_2}$. Lo representaremos por $A \equiv B + C$.

Como en el caso de los segmentos, es claro que la relación $A \equiv B + C$ puede verse como una igualdad entre clases de congruencia de los ángulos. Por ejemplo, notemos que si A , B y C cumplen la definición anterior y $A' \equiv A$, entonces $B < A$, luego existe un ángulo B' congruente con B con un lado en común con A' y contenido en A' . Por el teorema 1.53, los otros lados de B' y A' forman un ángulo C' congruente con C , luego A' también es una suma de B y C . Similarmente, el teorema 1.54 implica que todas las sumas de B y C son congruentes.

Convenimos en que un ángulo llano A es la suma de dos ángulos B y C si y sólo si éstos son suplementarios. Observemos que la definición general de suma es aplicable a este caso tomando como vértice de A cualquier punto de su frontera.

La suma de ángulos presenta una diferencia importante con la de segmentos, y es que no todo par de ángulos tiene una suma. Concretamente:

Teorema 1.67 Dos ángulos B y C admiten una suma si y sólo si C es menor o igual que el suplementario de B .

DEMOSTRACIÓN: Sea A una suma de B y C . Es claro que existe un semihaz de semirrectas contenidas en el semiplano de A y con un extremo l_1 igual a un lado de A (incluso si A es un semiplano). Sea l_3 el otro lado de A . Por definición de suma, existe una semirrecta l_2 entre l_1 y l_3 de modo que $\widehat{l_1 l_2} \equiv B$ y $\widehat{l_2 l_3} \equiv C$.

Si llamamos l_4 a la semirrecta complementaria de l_1 tenemos

$$l_1 \leq_{l_1 l_4} l_2 \leq_{l_1 l_4} l_3 \leq_{l_1 l_4} l_4,$$

luego $C \equiv \widehat{l_2 l_3} \leq \widehat{l_2 l_4}$, y este último ángulo es el suplementario de B .

Recíprocamente, si C es menor o igual que el suplementario de B , entonces podemos tomar $B = \widehat{l_1 l_2}$, donde l_1 es un extremo de un semihaz de semirrectas al cual pertenece l_2 . Por hipótesis $C \leq \widehat{l_4 l_2}$, donde l_4 es la semirrecta complementaria de l_1 , luego existe una semirrecta l_3 entre l_2 y l_4 tal que $C \equiv \widehat{l_2 l_3}$. Es claro entonces que $\widehat{l_1 l_3}$ es una suma de B y C . ■

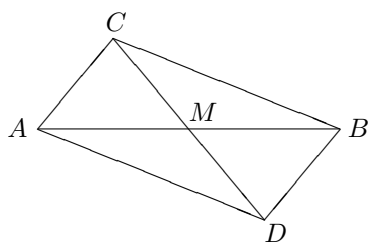
Se comprueba sin dificultad que la suma de ángulos es asociativa, es decir, si $A + B$ es sumable con C , entonces A también es sumable con $B + C$ y $(A + B) + C \equiv A + (B + C)$. Así mismo es conmutativa y simplificable. Si $B < C$ entonces existe un único ángulo D (salvo congruencia) tal que $C \equiv B + D$. Lo representaremos por $D \equiv C - B$.

1.7 Más sobre segmentos, ángulos y triángulos

Con los resultados de que disponemos hasta el momento ya podemos probar con cierta agilidad muchas propiedades intuitivamente evidentes acerca de segmentos, ángulos y triángulos. Recogemos aquí las que nos harán falta más adelante para estudiar la perpendicularidad, el paralelismo y las circunferencias, entre otras nociones.

Teorema 1.68 *Todo segmento \overline{AB} contiene un único punto M que cumple $\overline{AM} \equiv \overline{MB}$. Se le llama punto medio del segmento.*

DEMOSTRACIÓN: Sea C un punto cualquiera fuera de la recta AB . El triángulo \widehat{ACB} es congruente con un único triángulo \widehat{BDA} contenido en el semiplano complementario del que contiene al primero.



Los puntos C , A y D no pueden ser colineales, pues entonces los ángulos \widehat{CAB} y \widehat{BAD} serían suplementarios, luego también lo serían los ángulos (iguales a éstos) \widehat{ABD} y \widehat{CBA} , luego C , B y D serían también colineales, y así los cuatro puntos estarían alineados, en contra de la elección de C .

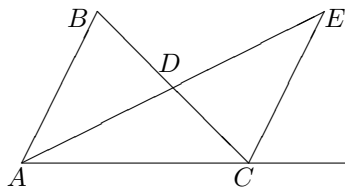
Tenemos, pues, dos triángulos \widehat{ACD} y \widehat{BDC} , que tienen sus lados iguales, luego también sus ángulos. Además \overline{CD} corta a la recta AB en un punto M (porque C y D están en semiplanos distintos). Los triángulos \widehat{ACM} y \widehat{BDM} son iguales por el criterio ALA, luego $\overline{AM} \equiv \overline{MB}$. El punto M ha de estar entre A y B , pues en caso contrario habría dos segmentos iguales con extremo M y el otro extremo al mismo lado de M .

La unicidad es fácil de probar: Si hubiera dos puntos medios M_1 y M_2 , podemos suponer que M_1 está entre A y M_2 . Entonces $\overline{AM_2} = \overline{AM_1} + \overline{M_1M_2}$, luego $\overline{AB} = 2\overline{AM_2} = 2\overline{AM_1} + 2\overline{M_1M_2} = \overline{AB} + 2\overline{M_1M_2}$, lo cual es absurdo. ■

Teorema 1.69 *Dado un ángulo $\widehat{l_1l_2}$, existe una única semirrecta l contenida en él, tal que $\widehat{l_1l} \equiv \widehat{l}l_2$. Se la llama bisectriz del ángulo.*

DEMOSTRACIÓN: Sea O el vértice del ángulo. Tomemos un punto A en l_1 y sea B en l_2 tal que $\overline{OA} \equiv \overline{OB}$. Sea M el punto medio de \overline{AB} y l la semirrecta de origen O y que pasa por M . Claramente l está contenida en el ángulo. Los triángulos \overline{OAM} y \overline{OBM} tienen los tres lados iguales, luego también los ángulos. En particular $\widehat{l_1l} \equiv \widehat{l}l_2$. La unicidad se prueba como en el caso de los segmentos o, alternativamente, se prueba que una bisectriz ha de pasar por el punto medio de AB . ■

Teorema 1.70 *Todo ángulo de un triángulo es menor que el suplementario de cualquier otro de los ángulos.*



DEMOSTRACIÓN: Sea el triángulo \widehat{ABC} . Vamos a probar que el suplementario de \hat{C} es mayor que \hat{B} . Sea D el punto medio del lado \overline{BC} . Consideremos la semirrecta \overrightarrow{AD} y, sobre ella, sea E el punto que cumple $\overline{AD} \equiv \overline{DE}$. Los triángulos \overline{ABD} y \overline{EDC} son iguales por el criterio LAL , luego $\widehat{DCE} \equiv \hat{B}$ está contenido en el ángulo adyacente a \hat{C} . ■

El teorema siguiente generaliza al criterio del triángulo isósceles.

Teorema 1.71 *Los ángulos de un triángulo satisfacen las mismas desigualdades que sus respectivos lados opuestos.*

DEMOSTRACIÓN: Sea \widehat{ABC} un triángulo y supongamos, por ejemplo, que $\overline{BC} < \overline{AB}$. Entonces existe un punto D en \overline{AB} tal que $\overline{BD} \equiv \overline{BC}$.

Entonces $\hat{A} = \widehat{CAD} < \widehat{CDB}$, por el teorema anterior, pues el segundo es el suplementario de un ángulo de \widehat{ADC} ; a su vez $\widehat{CDB} = \widehat{DCB}$, porque el triángulo \widehat{DCB} es isósceles, y por último $\widehat{DCB} < \widehat{ACB} = \hat{C}$. El recíproco es obvio: Si $\hat{A} < \hat{C}$ no puede ser $\overline{AB} < \overline{BC}$ por la parte ya probada, y tampoco puede darse la igualdad por el criterio del triángulo isósceles. ■

En particular tenemos que un triángulo es equilátero si y sólo si tiene sus tres ángulos iguales, es isósceles si y sólo si tiene dos ángulos iguales y es escaleno si y sólo si tiene sus tres ángulos desiguales.

1.8 Perpendiculares

Ahora estamos en condiciones de estudiar el concepto de perpendicularidad. La definición básica es la siguiente:

Definición 1.72 Un ángulo es *recto* si es su propio suplementario.

Teorema 1.73 *Existen ángulos rectos.*

DEMOSTRACIÓN: Sea $\widehat{l_1 l_2}$ un ángulo cualquiera. Sea O su vértice, sea A un punto en l_1 y sea B el punto de l_2 que cumple $\overline{OA} \equiv \overline{OB}$. Sea M el punto medio de \overline{AB} . Entonces los ángulos \widehat{OMA} y \widehat{OMB} son adyacentes y por otra parte son iguales, pues los triángulos correspondientes tienen los lados iguales. Por consiguiente ambos son ángulos rectos. ■

Dos ángulos rectos cualesquiera son iguales, pues si fuera $R_1 < R_2$, sus suplementarios, por una parte, deberían cumplir la relación inversa, pero por otra parte deberían cumplir la misma. Por otro lado, es obvio que todo ángulo congruente con un ángulo recto es recto. La existencia de ángulos rectos generaliza el teorema 1.69 al caso de ángulos llanos.

Definición 1.74 Dos rectas son *perpendiculares* (lat. *perpendicularum* = ‘plomada’) si son secantes y uno de los ángulos que forman—y por consiguiente los cuatro— es recto.

Dos semirrectas o segmentos son perpendiculares si lo son sus prolongaciones. Un ángulo es *agudo* (lat. ‘con punta’) si es menor que un ángulo recto. Un ángulo es *obtuso* (lat. ‘sin punta’) si es mayor que un ángulo recto.

Es claro que el suplementario de un ángulo agudo es un ángulo obtuso y viceversa. El teorema 1.70 implica que todo triángulo tiene al menos dos ángulos agudos, pues si tiene uno obtuso su suplementario es agudo, y los otros dos son menores que éste. Esto nos permite clasificar los triángulos en *obtusángulos*, *rectángulos* y *acutángulos* según si tienen, respectivamente, un ángulo obtuso, un ángulo recto o si todos sus ángulos son agudos. En un triángulo rectángulo, los lados perpendiculares se llaman *catetos* (gr. ‘perpendiculares’) y el lado situado bajo el ángulo recto se llama *hipotenusa* (gr. ‘tendido por debajo’). El teorema 1.71 implica que la hipotenusa de un triángulo rectángulo es mayor que los catetos.

Teorema 1.75 *Dada una recta r y un punto P contenidos en un plano Π , existe una única recta perpendicular a r que pasa por P y está contenida en Π .*

DEMOSTRACIÓN: Si el punto P está en r es evidente, pues basta transportar un ángulo recto sobre una de las semirrectas que P determina en r . La unicidad también es clara. Supongamos ahora que P no está en r .

Sea A un punto de r , sea P' el único punto del semiplano de Π opuesto al que contiene a P y que cumple que el ángulo que r forma con $\overrightarrow{AP'}$ es igual al que forma con \overrightarrow{AP} así como que $\overline{AP} \equiv \overline{AP'}$.

Si P , A y P' están alineados entonces la recta PP' forma con r dos ángulos adyacentes iguales, luego es perpendicular a r y pasa por P . Si no están alineados entonces la recta PP' corta a r en un punto B distinto de A . Los triángulos \widehat{ABP} y $\widehat{ABP'}$ son iguales, luego también lo son los ángulos \widehat{ABP} y $\widehat{ABP'}$, que

además son adyacentes. Por lo tanto la recta PP' es perpendicular a r y pasa por P . Si hubiera dos perpendiculares a r que pasaran por P , formarían un triángulo con dos ángulos rectos, lo cual es imposible. ■

El punto donde la perpendicular a una recta r por un punto P corta a r se llama *pie* de la perpendicular a r por P .

Dados dos puntos A y B en un plano Π , la perpendicular (en Π) al segmento \overline{AB} por su punto medio se llama *mediatriz* de \overline{AB} . Es inmediato comprobar que la mediatriz de un segmento \overline{AB} contiene exactamente a los puntos que equidistan de sus extremos, es decir, que cumplen $\overline{AX} \equiv \overline{BX}$.

Observemos que si P es un punto exterior a una recta r y Q es el pie de su perpendicular por P , entonces \overline{PQ} es menor que \overline{PR} para cualquier otro punto R de r . En efecto, el triángulo \widehat{PQR} es rectángulo, y su lado mayor es la hipotenusa \overline{PR} .

Teorema 1.76 *Todo lado de un triángulo es menor que la suma de los otros dos y mayor que su diferencia.*

DEMOSTRACIÓN: Consideremos un triángulo \widehat{ABC} y tracemos la perpendicular por A a la recta BC . Distinguimos tres casos.

1) Si el pie de la perpendicular es B o C (por ejemplo B), entonces el triángulo es rectángulo y \overline{BC} es menor que la hipotenusa \overline{AC} , y en particular es menor que $\overline{AC} + \overline{AB}$.

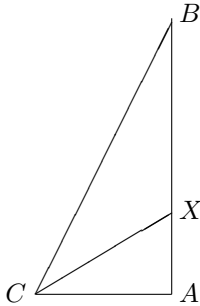
2) Si el pie de la perpendicular está fuera del segmento \overline{BC} , digamos que es un punto P de modo que B está entre P y C , entonces $\overline{BC} < \overline{PC}$, que es el cateto de un triángulo rectángulo de hipotenusa \overline{AC} , luego tenemos que $\overline{BC} < \overline{AC} < \overline{AC} + \overline{AB}$.

3) Si el pie de la perpendicular es un punto X entre B y C , entonces tenemos dos triángulos rectángulos \widehat{AXB} y \widehat{AXC} , de hipotenusas \overline{AB} y \overline{AC} , luego

$$\overline{BC} \equiv \overline{BX} + \overline{XC} < \overline{AB} + \overline{AC}.$$

La segunda propiedad es consecuencia de la primera. Si llamamos a, b y c a los tres lados y, por ejemplo, $b \leq c$, entonces $a + b > c$ implica $a > c - b$. ■

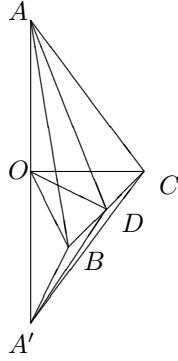
Teorema 1.77 *Sea \widehat{ABC} un triángulo rectángulo, cuyo ángulo recto es \hat{A} y sea X un punto entre A y B . Entonces $\overline{CX} < \overline{CB}$.*



DEMOSTRACIÓN: Como X está entre A y B , los ángulos \widehat{CXA} y \widehat{CXB} son suplementarios, y el primero es agudo, porque \widehat{CAX} es rectángulo y su ángulo recto es \hat{A} , luego el segundo es obtuso. Por lo tanto, aplicando el teorema 1.71 al triángulo \widehat{CXB} , concluimos que $\overline{CX} < \overline{CB}$. ■

Ahora pasamos a ocuparnos de la perpendicularidad entre rectas y planos. El resultado básico es el siguiente:

Teorema 1.78 *Si una recta es perpendicular a dos rectas contenidas en un plano, entonces es perpendicular a todas las rectas contenidas en dicho plano y que pasan por el punto de corte.*



DEMOSTRACIÓN: Sea O el punto de corte entre la recta y el plano. Sea A otro punto de la recta y A' el simétrico de A respecto a O (es decir, el que cumple $\overline{AO} \equiv \overline{OA'}$). Consideremos una recta cualquiera contenida en el plano que pase por O . Fijemos en particular una de sus semirrectas de origen O . Ésta estará contenida en uno de los cuatro ángulos en que las rectas de la hipótesis dividen al plano. Digamos que este ángulo es \widehat{BOC} . Entonces la semirrecta corta al segmento \overline{BC} en un punto D . En estos términos tenemos que los ángulos \widehat{AOB} y \widehat{AOC} son rectos, y queremos probar que \widehat{AOD} también lo es.

Los triángulos \widehat{AOC} y $\widehat{A'OC}$ son iguales por el criterio LAL y lo mismo sucede con \widehat{AOB} y $\widehat{A'OB}$. Esto implica que los triángulos \widehat{ABC} y $\widehat{A'BC}$ también son iguales, por el criterio LLL. De aquí pasamos a que los triángulos \widehat{ADB} y $\widehat{A'DB}$ también son iguales, esta vez por el criterio LAL, lo que nos da finalmente la igualdad de los triángulos \widehat{AOD} y $\widehat{A'OD}$, pues sus lados son iguales. Esto implica que los ángulos \widehat{AOD} y $\widehat{A'OD}$ son iguales, a la vez que adyacentes, luego son rectos. ■

Definición 1.79 Diremos que una recta es *perpendicular* a un plano Π si lo corta en un punto P y es perpendicular a todas las rectas contenidas en Π que pasan por P .

Los teoremas siguientes valen únicamente en un espacio tridimensional:

Teorema 1.80 *La unión de todas las rectas perpendiculares a una recta r que pasan por uno de sus puntos es un plano.*

DEMOSTRACIÓN: Sea P un punto de r . Consideremos dos planos que pasen por r (teorema 1.36), tracemos en ellos sendas perpendiculares a r por P y sea Π el plano que las contiene. Entonces r es perpendicular a dos rectas de Π , luego por el teorema anterior es perpendicular a todas las rectas que pasan por P y están contenidas en Π . Falta probar que toda recta perpendicular a r por P está contenida en Π . Sea s una de estas rectas. Entonces, por el axioma A6, el plano que contiene a r y a s corta a Π en una recta que, según sabemos, es perpendicular a r . Como en un mismo plano sólo puede haber una perpendicular, dicha intersección es s , luego s está contenida en Π . ■

Teorema 1.81 *Dado un plano Π y un punto A , existe una única recta perpendicular a Π que pasa por A .*

DEMOSTRACIÓN: Supongamos primero que A está en Π . Entonces consideramos dos rectas perpendiculares contenidas en Π que se corten en A . Sus respectivos planos perpendiculares se cortarán en una recta, que será perpendicular a las dos elegidas, luego a todas las de Π .

Supongamos ahora que A no está en Π . Tomemos una recta cualquiera r contenida en Π , sea $D \in \Pi$ el pie de la perpendicular a r por A , sean B y C puntos en r situados en semirrectas opuestas respecto a D , sea s la perpendicular a r por D contenida en Π , sea A' el punto situado en el semiplano de frontera s complementario del que contiene a A y que hace que $\overline{DA'}$ forme el mismo ángulo con s que \overline{DA} y además $\overline{DA'} \equiv \overline{DA}$ y sea O el punto en que $\overline{AA'}$ corta a s (la figura de la prueba del teorema 1.78 ilustra también la situación actual).

Por construcción $\widehat{ODA} \equiv \widehat{ODA'}$, de donde $\widehat{ADB} \equiv \widehat{A'DB}$ y $\widehat{ADC} \equiv \widehat{AD'C}$ (recordar que ambos son rectángulos). Por lo tanto tenemos que $\widehat{ABO} \equiv \widehat{A'BO}$ y $\widehat{ACO} \equiv \widehat{A'CO}$, luego ambos son rectángulos, y así la recta AA' es perpendicular a OB y OC , luego a Π . ■

Como en el caso de las rectas, el punto donde la perpendicular a un plano Π por un punto A corta a Π se llama *pie* de la perpendicular a Π por A .

1.9 Círculos y circunferencias

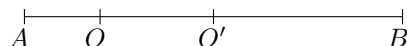
Definición 1.82 Dado un plano Π , un punto O en Π y un segmento r , llamaremos *círculo* de *centro* O y *radio* r al conjunto de todos los puntos P de Π tales que $\overline{OP} \leq r$. Llamaremos *circunferencia* (lat. *circumferre* = ‘llevar alrededor’) de *centro* O y *radio* r en Π al conjunto de todos los puntos P de Π tales que $\overline{OP} \equiv r$.

Cada círculo tiene asociada una circunferencia (la del mismo centro y radio), a la que se llama también su *frontera*. Los puntos del círculo que no pertenecen a la circunferencia se llaman *interiores*, mientras que los puntos que no pertenecen al círculo se llaman puntos *exteriores* a él.

También se llama *radio* de un círculo o circunferencia a cualquier segmento que una su centro con uno de los puntos de la circunferencia. Es claro que todos los radios son congruentes entre sí. Un segmento que une dos puntos de una circunferencia se llama *cuerda* de la misma. Una cuerda que pase por el centro se llama *diámetro* (gr. ‘medida transversal’). Es fácil ver que un diámetro es igual a dos veces el radio.

Se comprueba fácilmente que cada círculo o circunferencia contiene al menos tres puntos no colineales, con lo que determina el plano que lo contiene, al que llamaremos su *soporte*. Veamos que también determina su centro y su radio (éste último salvo congruencia):

Supongamos que un mismo conjunto de puntos fuera a la vez un círculo (o circunferencia) de centro O y radio r y de centro O' y radio r' . Entonces consideramos los puntos que indica la figura:



Tomamos A de modo que $\overline{OA} \equiv r$. Así, A está en la circunferencia (y en el círculo), pero como $\overline{O'A} > r$, tiene que ser $r' > r$. Pero razonando con B que cumple $\overline{O'B} \equiv r'$ concluimos igualmente que $r' < r$, contradicción.

Esto prueba que el centro de un círculo o circunferencia es único. En el caso de la circunferencia es claro entonces que el radio es único salvo congruencia, pues cualquier segmento que una un punto de la circunferencia con su centro es un radio. En el caso del círculo un radio está determinado por la propiedad de que es un segmento que une el centro con un punto del círculo y que no hay ninguno mayor que cumpla lo mismo.

En particular, esto implica que cada círculo determina su frontera y viceversa, pues cualquiera de los dos determina el centro y el radio y éstos a su vez determinan el otro. Esto justifica que hablemos, por ejemplo, de puntos interiores y exteriores a una circunferencia cuando propiamente son interiores o exteriores al círculo que la tiene por frontera.

Teorema 1.83 *Los círculos son conjuntos convexos.*

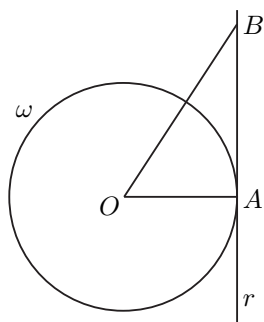
DEMOSTRACIÓN: Sean A y B dos puntos en un círculo de centro O y radio r . Si O está en la recta AB es fácil ver que todo X tal que $A - X - B$ cumple $\overline{OX} \leq r$. En caso contrario sea C el pie de la perpendicular a AB por O . Entonces \widehat{OCA} es un triángulo rectángulo, luego $\overline{OC} < \overline{OA} \leq r$, luego C está en el círculo, y el teorema 1.77 implica que todo X que cumpla $A - X - C$ o bien $B - X - C$ cumple $\overline{OX} < \overline{OA}$, $\overline{OB} \leq r$, y esto incluye a todos los puntos de $A - X - B$ distintos de C . ■

Definición 1.84 Diremos que una recta r es *tangente* (lat. ‘que toca’) a una circunferencia ω si r y ω están contenidas en el mismo plano y no tienen más que un punto en común. Igualmente diremos que dos circunferencias ω y ω' son *tangentes* si están contenidas en el mismo plano y no tienen más que un punto en común.

Veamos que toda circunferencia tiene rectas tangentes:

Teorema 1.85 *Sea ω una circunferencia de centro O y radio \overline{OA} . Entonces la recta perpendicular a OA por el punto A (en el plano de ω) es tangente a ω , excepto por el punto A , todos sus puntos son exteriores al círculo delimitado por ω . Recíprocamente, si r es una recta tangente a ω por el punto A , entonces r es perpendicular a OA . En particular, por cada punto de una circunferencia pasa una única recta tangente.*

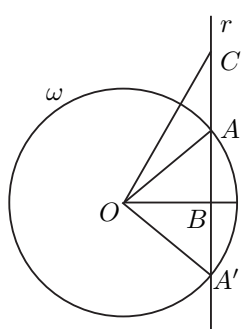
DEMOSTRACIÓN: Si r es la perpendicular a OA por A y B es cualquier otro de sus puntos, entonces el triángulo \widehat{AOB} es rectángulo y su hipotenusa es mayor que sus catetos: $\overline{OB} > \overline{OA}$, luego B es exterior al círculo. En particular, ningún otro punto de r aparte de A está en ω , luego r es tangente.



Recíprocamente, si r es una recta tangente a ω que pasa por el punto A , no puede ser OA , porque esta recta corta a ω en otro punto A' tal que $A'-O-A$. Sea B el pie de la perpendicular a r por O . Basta probar que $B = A$. En caso contrario, sea $A-B-A'$ de modo que $\overline{AB} \equiv \overline{BA'}$. Los triángulos \widehat{OAB} y $\widehat{OA'B}$ son rectángulos y tienen dos lados iguales (uno coincidente y otro congruente). Por lo tanto son semejantes, luego $\overline{OA} \equiv \overline{OA'}$, luego $A' \in \omega$ y es un segundo punto donde r corta a ω , contradicción. ■

Teorema 1.86 Sea ω una circunferencia y r una recta contenida en su mismo plano. Si r corta a ω pero no es tangente a ella, entonces se cortan exactamente en dos puntos.

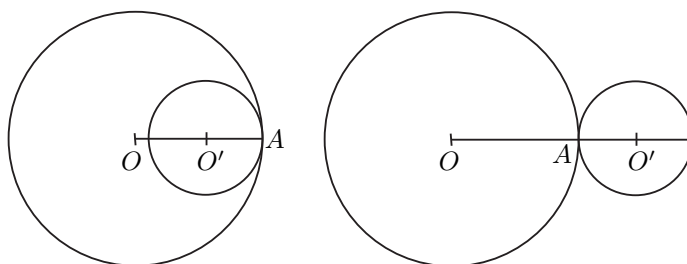
DEMOSTRACIÓN: Sea A un punto de corte entre r y ω . Si r no es tangente a ω , entonces no es perpendicular a OA , por el teorema anterior.



Sea B el pie de la perpendicular a r por O y sea $A-B-A'$ tal que $\overline{AB} \equiv \overline{BA'}$. Como en la prueba del teorema anterior razonamos que A' también está en ω , y ahora basta aplicar el teorema 1.77 para probar que cualquier otro punto C de r no está en ω . (El caso en que $B = O$ hay que tratarlo aparte.) De paso podemos concluir que la intersección de r con el círculo determinado por ω es AA' . ■

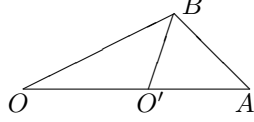
Ahora estudiamos la tangencia entre dos circunferencias:

Teorema 1.87 Sean O, O', A tres puntos colineales distintos dos a dos. Entonces, la circunferencia ω de centro O y radio \overline{OA} (en un plano dado) es tangente a la circunferencia ω' de centro O' y radio $\overline{O'A}$ (en el mismo plano). Recíprocamente, si dos circunferencias son tangentes en un punto A , entonces sus centros son colineales con A .

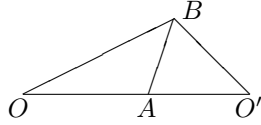


DEMOSTRACIÓN: Es fácil ver que ω y ω' no pueden tener otro punto en común sobre la recta r que contiene a los tres puntos (distinguiendo dos casos sobre su posición relativa). Supongamos ahora que hay un segundo punto de

intersección B fuera de dicha recta. Distinguiamos también los dos casos $O - O' - A$ y $O - A - O'$.



Si se cumple $O - O' - A$, como $\overline{OB} \equiv \overline{OA}$, se cumple $\widehat{OAB} \equiv \widehat{OBA}$, pero por el mismo motivo $\widehat{OAB} = \widehat{O'AB} \equiv \widehat{O'BA}$. Así pues $\widehat{OBA} \equiv \widehat{O'BA}$, lo cual contradice la unicidad del axioma C4.



En el caso en que $O - A - O'$ los dos triángulos \widehat{OAB} , $\widehat{O'AB}$, son isósceles, luego sus ángulos en A y en B deberían ser iguales, luego agudos, pero los ángulos \widehat{OAB} y $\widehat{O'AB}$ son suplementarios, luego no pueden ser los dos agudos, contradicción.

Recíprocamente, supongamos que las dos circunferencias son tangentes en A , pero que O , O' y A no son colineales. Entonces podemos formar un triángulo $\widehat{OO'B} \equiv \widehat{OO'A}$ en el semiplano opuesto al que contiene al punto A , y así tenemos otro punto B en la intersección de las dos circunferencias, contradicción. ■

Teorema 1.88 *Si dos círculos tienen circunferencias tangentes, o bien uno está contenido en el otro o bien no tienen más punto en común que el punto de tangencia.*

DEMOSTRACIÓN: Sean O y O' los centros y sea A el punto de tangencia. Por el teorema anterior son colineales. Distingamos los casos $O - O' - A$ y $O - A - O'$.

Supongamos $O - O' - A$, con lo que $\overline{O'A} < \overline{OA}$. Sean B y B' los puntos de la recta OO' tales que $\overline{BO} \equiv \overline{OA}$ y $\overline{B'O'} \equiv \overline{O'A}$. Es fácil ver que

$$\overline{B'A} \equiv \overline{B'O'} + \overline{O'A} < \overline{BO} + \overline{OA} \equiv \overline{BA}.$$

Como B y B' están al mismo lado respecto de A (el lado al que están O y O'), podemos concluir que $\overline{B'A} \subset \overline{BA}$. Pero la intersección de OO' con el círculo de centro O' es $\overline{B'A}$, y la intersección con el círculo de centro O es \overline{BA} . Por lo tanto, si X es un punto del círculo de centro O' contenido en OO' , podemos afirmar que está en el círculo de centro O .

Supongamos ahora que X no está en la recta OO' . Entonces podemos aplicar el teorema 1.76, según el cual

$$\overline{OX} < \overline{OO'} + \overline{O'X} \leq \overline{OO'} + \overline{O'A} \equiv \overline{OA},$$

lo que prueba que X está en el círculo de centro O .

Ahora supongamos que $O - A - O'$. Tomamos puntos $B - O - A - O' - B'$ tales que $\overline{BO} \equiv \overline{OA}$ y $\overline{AO'} \equiv \overline{O'B'}$. Ahora sucede que las intersecciones de los círculos con la recta OO' son los segmentos \overline{BA} y $\overline{AB'}$, que sólo tienen en común el punto A . Así pues, un hipotético punto común alternativo X no puede estar en la recta OO' , pero si existiera tal punto, el teorema 1.76 nos daría que

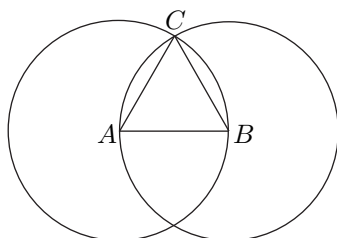
$$\overline{OO'} < \overline{OX} + \overline{XO'} < \overline{OA} + \overline{AO'} \equiv \overline{OO'},$$

contradicción. ■

Teorema 1.89 *Si dos circunferencias (coplanares) distintas tienen un punto en común y no son tangentes, entonces tienen exactamente dos puntos en común.*

DEMOSTRACIÓN: Sean O y O' los centros respectivos y sea A un punto común. Por el teorema 1.87 O , O' y A no son colineales, y el argumento de la parte final de la prueba muestra que existe un segundo punto de corte B situado al otro lado de la recta OO' respecto de A . Falta probar que no hay más. Ahora bien, si C es un punto de corte, tampoco puede ser colineal con O y O' , porque entonces las circunferencias serían tangentes, y el triángulo $\widehat{OO'C}$ sólo puede ser uno de los dos triángulos $\widehat{OO'A}$, $\widehat{OO'B}$, por el axioma C5, luego C tiene que ser A o B . ■

El primer resultado que presenta Euclides en sus *Elementos* es la construcción de un triángulo equilátero. Consiste en tomar un segmento arbitrario \overline{AB} , trazar con el compás las circunferencias de centros A y B y radio \overline{AB} y tomar como tercer vértice del triángulo uno de los dos puntos en los que éstas se cortan.



Sin embargo, la existencia de tales puntos de corte no se deduce de los postulados de Euclides! Tampoco se deduce de nuestros axiomas, sino que de momento necesitamos introducir un axioma específico para esta situación:

Axioma de las circunferencias *Sean ω y ω' dos circunferencias coplanares. Si ω' pasa por un punto interior y otro exterior a ω , entonces ambas circunferencias se cortan.*

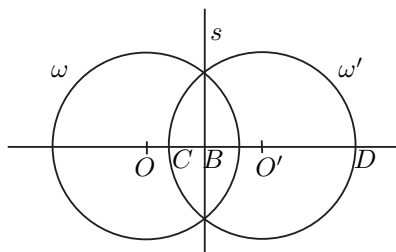
Notemos que en las hipótesis del axioma las circunferencias no pueden ser tangentes, por el teorema 1.88, luego la conclusión es, más concretamente, que se cortan en dos puntos.

Con este axioma ya se puede justificar fácilmente la existencia de triángulos equiláteros mediante la construcción de Euclides. A su vez deducimos una propiedad análoga sobre existencia de cortes entre rectas y circunferencias:

Teorema 1.90 *Si una recta está contenida en el plano de una circunferencia y pasa por un punto interior a ésta, entonces corta a la circunferencia en dos puntos.*

DEMOSTRACIÓN: Sean O y r el centro y el radio de la circunferencia ω y sea s la recta. Si s pasa por O es inmediato que corta a la circunferencia en dos puntos. En caso contrario sea B el punto de s por donde pasa la perpendicular

a s por O . Sea $O - B - O'$ de modo que $\overline{OB} \equiv \overline{BO'}$. Sea ω' la circunferencia de centro O' y radio r (contenida en el mismo plano). Ésta corta a OO' en dos puntos C y D . Concretamente, tomamos C en el mismo lado que O respecto de O' y D en el opuesto, como muestra la figura.



Que s contenga un punto A interior a ω se traduce en que $\overline{OB} \leq \overline{OA} < r$, luego B es también interior a ω . Como $\overline{OB} \equiv \overline{BO'}$, resulta que $\overline{O'B} < r \equiv \overline{O'C}$. Pero B y C están en el mismo lado de O' , luego tiene que ser $C - B - O'$, luego C y O' están en lados opuestos respecto de s . Lo mismo pasa con O y O' , luego C y O están al mismo lado de s . Vamos a probar que C es

interior a ω , para lo cual distinguimos dos casos:

Si $O - C - B$, entonces $\overline{OC} < \overline{OB} < r$, luego C es interior.

Si $C - O - B$, como $\overline{CO} < \overline{CO'}$, también $C - O - O'$, luego $\overline{CO} < \overline{CO'} \equiv r$, y llegamos a la misma conclusión.

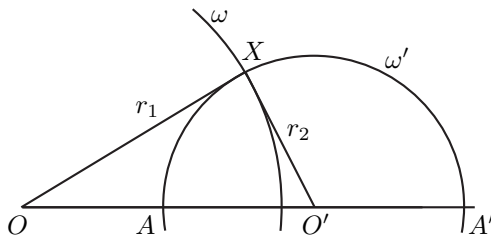
Si $C = O$ la conclusión es trivial.

Por otra parte, $O - O' - D$, luego $\overline{OD} > \overline{O'D} \equiv r$, luego D es exterior a ω .

Esto nos permite aplicar el axioma de las circunferencias, que nos da que ω y ω' se cortan en dos puntos E y E' . Sólo queda observar que ambos están en s , pues equidistan de O y O' y s es la mediatriz del segmento OO' . ■

Ahora podemos demostrar que las condiciones necesarias dadas por el teorema 1.76 para que exista un triángulo con unos lados dados son también suficientes:

Teorema 1.91 Sean r_1, r_2 y r_3 tres segmentos que cumplan las desigualdades $r_2 \leq r_3$ y $r_3 - r_2 < r_1 < r_2 + r_3$ (o simplemente $r_1 < r_2 + r_3$ si $r_2 = r_3$). Entonces existe un triángulo de lados r_1, r_2 y r_3 .



DEMOSTRACIÓN: Tomamos un segmento $\overline{OO'} \equiv r_3$ y consideramos la circunferencia ω de centro O y radio r_1 y la circunferencia ω' de centro O' y radio r_2 . Sean A y A' los puntos en que ω' corta a la recta OO' . Concretamente, sea A el que está al mismo lado de O' que O .

Entonces $\overline{OA} \equiv r_3 - r_2$ (o bien $O = A$), luego por hipótesis A es interior a la circunferencia ω . Por otra parte, $\overline{OA'} \equiv r_2 + r_3 > r_1$, luego A' es exterior a la circunferencia ω .

El axioma de las circunferencias nos da que ω y ω' se cortan en un punto X de modo que el triángulo $\widehat{OO'X}$ cumple lo requerido. ■

Otra aplicación es la siguiente:

Teorema 1.92 *Dados dos segmentos $c < h$ existe un triángulo rectángulo cuya hipotenusa es congruente con h y uno de sus catetos es congruente con c .*

DEMOSTRACIÓN: Sea $c = \overline{OA}$ y sea ω la circunferencia de centro O y radio h . La perpendicular a OA por A pasa por A , que es interior a ω , luego corta a ω en un punto B . El triángulo \widehat{OAB} cumple lo requerido. ■

Ejercicio: Deducir del teorema anterior que por un punto exterior a una circunferencia pasan dos de sus rectas tangentes.

Capítulo II

La geometría euclídea

Los axiomas de la geometría absoluta que hemos presentado en el capítulo anterior no son suficientes para demostrar todos los resultados geométricos intuitivamente verdaderos. Sucede que son necesarios dos axiomas más: el axioma de continuidad, que nos permitirá identificar los puntos de una recta con los números reales, y el axioma de las paralelas, que lo hemos enunciado en el capítulo anterior, pero del que todavía no hemos extraído ninguna consecuencia. Antes de introducir el axioma de continuidad trabajaremos con otro más débil, la propiedad de Arquímedes, que basta para formalizar la teoría de las proporciones de Eudoxo discutida en el capítulo II de [IGE].

2.1 La geometría arquimediana

Si reflexionamos en lo que supone medir un segmento, tenemos que entender que, en sentido estricto, es imposible determinar la longitud de un segmento mediante un número. Lo único que podemos hacer es encontrar un número que determine la longitud de un segmento en comparación con la de otro conocido que tomemos arbitrariamente como unidad. Por ejemplo, decimos que la razón de dos segmentos v y u es, por ejemplo, $4 : 7$ si al dividir el segundo en siete partes iguales y sumar cuatro de estas partes obtenemos un segmento igual al primero. En términos de medidas esto significa que si tomamos a u como unidad, entonces v mide $4/7$ unidades.

Los primeros geómetras griegos no concebían otra forma de comparar longitudes entre segmentos que encontrar una ‘razón’ entre ellas, y por ello daban por hecho que, como era inconcebible que un segmento no pudiera ser medido, debía guardar siempre una cierta razón con cualquier unidad prefijada. Sin embargo, pronto se dieron cuenta de que no es así, sino que, fijado un segmento unitario, siempre es posible encontrar otro cuya longitud sea ‘irracional’ respecto a él, pero nunca llegaron a asimilar completamente este hecho. No obstante, Eudoxo desarrolló una teoría sobre proporciones entre segmentos diseñada especialmente para ser compatible con la existencia de tales longitudes irracionales.

En el capítulo anterior hemos definido la suma de segmentos, de modo que tres segmentos cumplen $w \equiv u + v$ si $w = \overline{PQ}$ y existe un punto R entre P y Q tal que $u \equiv \overline{PR}$ y $v \equiv \overline{RQ}$. Hay que tener presente que no hemos definido un único segmento $u + v$, sino que existe una única clase de congruencia de segmentos w que cumplen $w \equiv u + v$.

En particular, dado un número natural $m \geq 1$, podemos definir $v \equiv mu$ como que v resulta de sumar u consigo mismo m veces, y nuevamente no tenemos definido un segmento mu , sino una única clase de congruencia de segmentos v tales que $v \equiv mu$. Esta relación se interpreta como que el segmento v puede dividirse en m segmentos, todos ellos congruentes con u (y en particular congruentes entre sí).

El teorema 1.48, que afirma que la suma de segmentos es simplificable, nos da que si $v \equiv mu$ y $v' \equiv mu'$ y $v \equiv v'$, entonces $u \equiv u'$. Esto significa que si un segmento v se puede dividir en m partes iguales, la longitud de dichas partes está completamente determinada por la longitud de v y por el número m .

Hay un hecho elemental que no podemos demostrar con los axiomas que hemos dado hasta el momento, aunque veremos (teorema 2.49) que se deduce del axioma de las paralelas:

Propiedad de la división de segmentos *Dado un segmento v y un número natural $n \geq 1$, existe un segmento u tal que $v \equiv nu$.*

En otras palabras: todo segmento se puede dividir en cualquier número de partes iguales. El teorema 1.68 sobre la existencia del punto medio implica que para todo segmento v existe otro segmento u tal que $v \equiv 2u$, de donde se sigue claramente que v puede dividirse en 2^n partes iguales, para todo natural n , pero, con los axiomas que hemos dado, no podemos demostrar ni siquiera que todo segmento se pueda dividir en tres partes iguales.

Para que la exposición sea más natural vamos a suponer la propiedad de división de segmentos, pero al mismo tiempo explicaremos mediante notas al pie cómo se puede sustituir en todo momento por su restricción a potencias de 2, que tenemos demostrada.

Según hemos indicado, si se da la relación $u \equiv nv$, la clase de congruencia de u determina la de v , por lo que podemos escribir $v \equiv u/n$, donde hay que entender, como siempre, que no tenemos definido ningún segmento u/n en particular, sino que v es cualquiera de los segmentos que cumplen $u \equiv nv$, que son todos congruentes entre sí (si es que existen).

Es fácil ver que $m(u/mn) \equiv u/n$, es decir, que si dividimos un segmento u en mn partes iguales y tomamos m de ellas, obtenemos un segmento congruente con el que resulta de dividir u en n partes iguales. Esto hace que si $r = m/n = m'/n'$ son dos fracciones correspondientes a un mismo número racional positivo r (con m, n, m', n' números naturales no nulos) entonces

$$m(u/n) \equiv mn'(u/(nn')) \equiv nm'(u/nn') \equiv m'(u/n').$$

A partir de aquí ya podemos definir el concepto de ‘razón’ entre dos segmentos en el sentido que le daban los griegos:

Definición 2.1 Si u, v son dos segmentos y $r = m/n$ es un número racional positivo, definimos¹ $v \equiv ru$ como $v \equiv m(u/n)$.

Acabamos de probar que esta definición depende de r y no de la expresión que elijamos para r como cociente de números naturales. Una vez más, no tenemos definido ningún segmento ru en particular, sino que tenemos una única clase de congruencia de segmentos v que cumplen $v \equiv ru$, y dicha clase sólo depende de r y de la clase de congruencia de u .

Así, ru es el segmento que resulta de dividir u en n partes iguales y sumar m de ellas. Si $v \equiv ru$ podemos decir que el número racional r es la ‘razón’ entre los segmentos u y v .

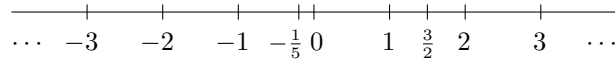
Una comprobación rutinaria nos da las propiedades siguientes:

1. $(rs)u \equiv r(su), \quad r(u+v) \equiv ru+rv, \quad (r+s)u \equiv ru+su, \quad 1u \equiv u,$
2. Si $ru \equiv rv$ entonces $u \equiv v$,
3. Si $ru \equiv su$, entonces $r = s$,
4. Si $r < s$ entonces $ru < su$,
5. Si $u < v$ entonces $ru < rv$,

para todos los números racionales positivos r, s y todos los segmentos u y v .

Podemos decir que al establecer una relación $v \equiv ru$ entre dos segmentos hemos “medido” v tomando a u como unidad, en el sentido de que el número r determina el tamaño de v en función del tamaño de u . Pero, según indicábamos en la introducción de esta sección, este concepto de ‘medida’ no va a ser suficiente, y vamos a introducir otro más general. Antes conviene considerar el concepto siguiente:

Una *recta graduada* es una recta en la que hemos seleccionado arbitrariamente dos puntos distintos, P_0 y P_1 . Entonces a cada número racional positivo r le podemos asignar unívocamente un punto P_r de la semirrecta $\overrightarrow{P_0P_1}$, a saber, el único que cumple $\overline{P_0P_r} \equiv r \overline{P_0P_1}$. Convenimos también en asignar a los números racionales negativos puntos en la semirrecta complementaria, de modo que si $r < 0$ entonces P_r está determinado por la relación $\overline{P_0P_r} \equiv (-r) \overline{P_0P_1}$.



¹Dado un segmento u y un número racional $r > 0$, la propiedad de división de segmentos implica que existe otro segmento v que cumple $v \equiv ru$. Si no suponemos esta propiedad, podemos, no obstante, demostrar la existencia de tal v cuando r es un *número racional diádico* (positivo) es decir, un número de la forma $r = m/2^n$. Observemos que la suma, la resta y el producto de números diádicos es de nuevo un número diádico. Dicho en otras palabras, los números diádicos forman un subanillo del cuerpo \mathbb{Q} de los números racionales.

Es importante notar que esta asignación de números racionales a algunos puntos de una recta depende de la elección arbitraria de los puntos P_0 y P_1 o, en otros términos, de la elección de P_0 , de la unidad de medida $u = \overline{P_0P_1}$ y de la orientación de la recta (es decir, de la semirrecta determinada por P_0 en la que situamos P_1). Cambiando estas elecciones podemos obtener distintas graduaciones de una misma recta. En una recta graduada consideramos siempre el orden para el cual $P_0 < P_1$, y así el orden en \mathbb{Q} se corresponde con el orden de la recta.²

En estos términos, el problema que tenemos planteado es si todo punto de una recta graduada es de la forma P_r , para cierto número racional r o si, por el contrario, existen puntos de la recta que no son de esta forma, es decir, puntos *incommensurables* con respecto a la unidad prefijada, puntos que no guardan ninguna razón $m : n$ con respecto a dicha unidad.

En este punto se vuelve esencial otra propiedad que tampoco puede demostrarse a partir de los axiomas que hemos dado hasta ahora (ni siquiera contando el axioma de las paralelas):

Propiedad de Arquímedes *Para todo par de segmentos u y v existe un número natural n tal que $nu > v$.*

En términos de una recta graduada, lo que expresa esta propiedad es que si vamos avanzando por una recta con pasos iguales: P_0, P_1, P_2, \dots tarde o temprano rebasamos cualquier punto de la recta, es decir, que no hay puntos “infinitamente alejados” a los que nunca podremos llegar por muchos pasos que demos.

Naturalmente, esto vale tanto si avanzamos en un sentido como en otro, de modo que, para todo punto P de la recta, existe un número natural n tal que $P_{-n} < P < P_n$.

Tal y como acabamos de indicar, la propiedad de Arquímedes va a ser fundamental en todos los resultados de este capítulo, por lo que podemos decir que aquí vamos a estudiar la *geometría arquimediana*.

A partir de la propiedad de Arquímedes podemos demostrar un resultado más preciso:

Teorema 2.2 *Si $P < Q$ son dos puntos distintos de una recta graduada, existe un número racional r tal que $P < P_r < Q$.*

DEMOSTRACIÓN: Si P y Q están en semirrectas distintas respecto a P_0 es obvio. Podemos suponer que $P_0 < P < Q$. El caso contrario es análogo. Sea $u = \overline{P_0P_1}$ y $v = \overline{PQ}$. Por la propiedad de Arquímedes existe un número natural³ n tal que $u < nv$ o, equivalentemente, $(1/n)u < v$. De nuevo por la

²Si no suponemos la propiedad de la división de segmentos tenemos definido igualmente P_r cuando r es un número racional diádico.

³Si no suponemos la propiedad de la división de segmentos, podemos sustituir n por 2^n , pues también $u < nv \leq 2^n v$ y la prueba vale igualmente, de modo que obtenemos un número racional diádico r , para el cual está definido P_r , que cumple igualmente $P < P_r < Q$.

propiedad de Arquímedes existe un número m tal que $(m/n)u > \overline{P_0P}$. Podemos tomar el mínimo que cumpla esto. Entonces

$$\frac{m-1}{n}u \leq \overline{P_0P}, \quad \text{luego} \quad \frac{m}{n}u \leq \overline{P_0P} + \frac{1}{n}u < \overline{P_0P} + \overline{PQ} = \overline{P_0Q}.$$

Por consiguiente si llamamos $r = m/n$ tenemos que $\overline{P_0P} < \overline{P_0P_r} < \overline{P_0Q}$, es decir, $P < P_r < Q$. ■

Este teorema nos da una forma de determinar el tamaño de *cualquier* segmento en relación a otro, si no puede ser con un número racional, al menos con un conjunto infinito de ellos:

Definición 2.3 Si P es un punto de una recta graduada, llamaremos⁴

$$\alpha_P = \{r \in \mathbb{Q} \mid P_r < P\}.$$

La idea es que cada punto P de una recta graduada divide en dos partes disjuntas a los números racionales: por una parte están los números r que cumplen $P_r < P$ y por otra los que cumplen $P \leq P_r$. Una y otra se determinan mutuamente, así que hemos asociado a P arbitrariamente una de las dos. El hecho es que α_P determina completamente la posición de P en la recta:

Teorema 2.4 Si $P < Q$ son dos puntos distintos en una recta graduada, entonces $\alpha_P \subsetneq \alpha_Q$. En particular, si dos puntos P y Q de una recta graduada cumplen $\alpha_P = \alpha_Q$, entonces $P = Q$.

DEMOSTRACIÓN: Es claro que $\alpha_P \subset \alpha_Q$, y por el teorema anterior existe un número racional r tal que $P < P_r < Q$, luego $r \in \alpha_Q \setminus \alpha_P$, luego la inclusión es estricta. ■

Definición 2.5 Dados dos segmentos u y v , consideramos una recta graduada con unidad u y llamamos P al único punto de la semirrecta positiva que cumple $\overline{P_0P} \equiv v$. Entonces podemos definir la *proporción* de v respecto a u como

$$\frac{v}{u} = \alpha_P.$$

Es fácil ver que esta proporción sólo depende de las clases de congruencia de u y v (en particular, que no depende de la elección del origen o de la orientación de la recta graduada), así como que

$$v \equiv v' \quad \text{si y sólo si} \quad \frac{v}{u} = \frac{v'}{u},$$

de modo que el tamaño de un segmento v (es decir, su clase de congruencia) está determinado por la unidad u y la proporción v/u .

⁴Si no suponemos la propiedad de la división de segmentos podemos definir igualmente

$$\alpha_P = \{r \in \mathbb{Q} \mid \text{existen } m, n \in \mathbb{N} \text{ tales que } r \leq \frac{m}{2^n}, \quad P_{m/2^n} < P\}.$$

Es fácil ver que si los segmentos son divisibles ambas definiciones son equivalentes.

En particular, para todo punto P de la semirrecta positiva de una recta graduada, tenemos que

$$\frac{\overline{P_0P}}{\overline{P_0P_1}} = \alpha_P.$$

Así pues, como no tenemos garantías de que vaya a existir un número racional que cumpla precisamente $v \equiv ru$, hemos encontrado una forma de determinar la longitud de v a partir de u mediante un conjunto infinito de números racionales. En la sección siguiente vamos a ver cómo podemos determinar la proporción entre dos segmentos con un único número, pero eso no significa en realidad que vayamos a *mejorar* lo que acabamos de obtener, sino simplemente que lo vamos a *asumir* y vamos a adoptar el convenio de considerar como un único número a un conjunto infinito de números racionales.

2.2 Longitud de segmentos. Los números reales

A cada punto P de una recta graduada le hemos asignado un conjunto α_P de números racionales que determina completamente su posición en la recta. Estos conjuntos α_P no son arbitrarios, sino que tienen que cumplir ciertas propiedades:

Definición 2.6 Una *sección inicial abierta* de \mathbb{Q} es un subconjunto $\alpha \subset \mathbb{Q}$ que cumpla:

1. Si $r \in \alpha$ y $s \leq r$ entonces $s \in \alpha$.
2. Si $r \in \alpha$ existe un $s \in \alpha$ tal que $r < s$.

Llamaremos $\overline{\mathbb{R}}$ al conjunto de todas las secciones iniciales abiertas en \mathbb{Q} .

Es fácil probar que los conjuntos α_P son secciones iniciales abiertas que cumplen dos propiedades más: no son vacías (pues hay puntos anteriores a P) y no contienen a todos los números racionales (pues hay puntos posteriores).

Definimos $\pm\infty \in \overline{\mathbb{R}}$ como $-\infty = \emptyset$ y $+\infty = \mathbb{Q}$. Llamaremos conjunto de los *números reales* a $\mathbb{R} = \overline{\mathbb{R}} \setminus \{\pm\infty\}$.

Teorema 2.7 $\overline{\mathbb{R}}$ es un conjunto totalmente ordenado por la inclusión con mínimo $-\infty$ y máximo $+\infty$. A su vez, \mathbb{R} es un conjunto totalmente ordenado por la inclusión sin máximo ni mínimo.

DEMOSTRACIÓN: La inclusión es claramente un orden parcial. Sólo falta ver que en este caso es total.

Sean $\alpha, \beta \in \overline{\mathbb{R}}$. Supongamos $\alpha \neq \beta$. Por ejemplo, pongamos que existe $b \in \beta \setminus \alpha$.

Si $a \in \alpha$ entonces $a < b$ (o si no $b \in \alpha$), de donde $a \in \beta$. Esto prueba que $\alpha \subset \beta$, es decir, $\alpha \leq \beta$. Por lo tanto $\overline{\mathbb{R}}$ está totalmente ordenado.

Es claro que $-\infty = \emptyset$ está contenido en todo elemento de $\overline{\mathbb{R}}$, luego es el mínimo, y todo elemento de $\overline{\mathbb{R}}$ está contenido en $+\infty = \mathbb{Q}$, luego éste es el máximo.

El hecho de que $\overline{\mathbb{R}}$ esté totalmente ordenado por la inclusión implica que lo mismo vale para \mathbb{R} . Falta probar que \mathbb{R} no tiene máximo ni mínimo.

Si $\alpha \in \mathbb{R}$ entonces $\alpha \neq +\infty$, luego existe un número racional $r \in \mathbb{Q} \setminus \alpha$. Es fácil ver que si $r < s \in \mathbb{Q}$, el conjunto $\beta = \{t \in \mathbb{Q} \mid t < s\}$ es un número real tal que $\alpha < \beta$. Por lo tanto \mathbb{R} no tiene máximo.

Si $\alpha \in \mathbb{R}$, entonces $\alpha \neq -\infty$, luego existe un número racional $r \in \alpha$. Si $s \in \mathbb{Q}$ y $s < r$, entonces el conjunto $\beta = \{t \in \mathbb{Q} \mid t < s\}$ es un número real tal que $\beta < \alpha$. Por lo tanto \mathbb{R} no tiene mínimo. ■

La relación de orden en \mathbb{R} y en $\overline{\mathbb{R}}$ tiene una propiedad fundamental, pero para enunciarla necesitamos los conceptos de supremo e ínfimo:

Definición 2.8 Sea A un conjunto parcialmente ordenado y $X \subset A$.

Un elemento $s \in A$ es un *supremo* de X si es una cota superior de X y, para cualquier otra cota superior c de X , se cumple que $s \leq c$.

Un elemento $i \in A$ es un *ínfimo* de X si es una cota inferior de X y, para cualquier otra cota inferior c de X , se cumple que $c \leq i$.

Notemos que si un conjunto X tiene supremo éste es único, pues dos supremos s y s' son ambas cotas superiores de X , y la definición implica entonces que $s \leq s'$ y $s' \leq s$, luego $s = s'$. Lo mismo vale para los ínfimos, por lo que podemos usar las notaciones $\sup X$ e $\inf X$ para referirnos al supremo y al ínfimo de un conjunto, aunque teniendo en cuenta de que pueden no existir.

En suma, el supremo de un conjunto es la menor de sus cotas superiores, mientras que el ínfimo es la mayor de sus cotas inferiores. Ahora podemos probar la propiedad fundamental del orden de los números reales:

Teorema 2.9 Se cumplen las propiedades siguientes:

1. Todo subconjunto de $\overline{\mathbb{R}}$ tiene supremo e ínfimo.
2. Si un subconjunto no vacío de \mathbb{R} está acotado superiormente tiene supremo, y si está acotado inferiormente tiene ínfimo.

DEMOSTRACIÓN: 1) Sea S un subconjunto de $\overline{\mathbb{R}}$. Es inmediato comprobar que

$$\bigcup_{\alpha \in S} \alpha \in \overline{\mathbb{R}}$$

y es obviamente el supremo de S . El ínfimo de S no es sino el supremo del conjunto de sus cotas inferiores.

2) Un subconjunto de \mathbb{R} no vacío y acotado superiormente tiene supremo en $\overline{\mathbb{R}}$, como no es vacío el supremo no es $-\infty$, como tiene una cota en \mathbb{R} tampoco es $+\infty$, luego tiene supremo en \mathbb{R} . Análogamente con ínfimos. ■

El teorema 2.4 tiene ahora la traducción siguiente:

Teorema 2.10 Si r es una recta graduada, la aplicación $r \rightarrow \mathbb{R}$ determinada por $P \mapsto \alpha_P$ cumple que si $P < Q$, entonces $\alpha_P < \alpha_Q$. En particular es inyectiva.

Así, con el “truco” de considerar “números” a ciertos conjuntos infinitos de números, podemos decir que cada punto de una recta graduada está completamente determinado por un “único” número.

Observemos ahora que, en una recta graduada, para cada número racional r , el punto P_r está determinado, por una parte, por el número racional r y, por otra, por el número real α_{P_r} . El teorema siguiente nos permite eliminar esta duplicidad:

Teorema 2.11 *La aplicación $i : \mathbb{Q} \rightarrow \mathbb{R}$ dada por*

$$i(r) = \{s \in \mathbb{Q} \mid s < r\}$$

es inyectiva y conserva el orden. Si identificamos \mathbb{Q} con su imagen en \mathbb{R} , entre dos números reales hay siempre un número racional.

DEMOSTRACIÓN: Es inmediato que $i(r) \in \mathbb{R}$. Si $r, s \in \mathbb{Q}$ con $r < s$, entonces existe un $t \in \mathbb{Q}$ tal que $r < t < s$, de donde resulta que $t \in i(s) \setminus i(r)$. Como obviamente $i(r) \subset i(s)$, concluimos que $i(r) < i(s)$. Esto prueba que i es inyectiva y que conserva el orden.

Si $\alpha < \beta$, existirá $b \in \beta \setminus \alpha$. Sean $r, s \in \beta$ tales que $b < r < s$. Entonces todo $a \in \alpha$ es menor que r , luego $\alpha \leq i(r)$. Así mismo todo $t < r$ está en β , luego $i(r) \leq \beta$. Las desigualdades son estrictas pues $b \in i(r) \setminus \alpha$ y $s \in \beta \setminus i(r)$. ■

En lo sucesivo consideraremos a los números racionales como parte de los números reales a través de la aplicación que acabamos de definir. Para “olvidarnos” definitivamente de que los números reales son, por definición, conjuntos de números racionales, sólo tenemos que observar cómo se puede expresar la pertenencia a un número real en términos de la relación de orden:

Si r es un número racional y α es un número real, entonces $r \in \alpha$ si y sólo si $i(r) < \alpha$, es decir, si y sólo si r se identifica con un número real $r < \alpha$.

En efecto, si $r < \alpha$, entonces $i(r) \subsetneq \alpha$, luego existe un $s \in \alpha \setminus i(r)$, luego $r \leq s \in \alpha$, luego $r \in \alpha$, por definición de sección inicial abierta. Recíprocamente, si $r \in \alpha$ entonces todo $s < r$ cumple $s \in \alpha$, luego $i(r) \subset \alpha$, pero existe un $s \in \alpha$ tal que $r < s$, luego $s \in \alpha \setminus i(r)$, y así $i(r) \subsetneq \alpha$, es decir, $r < \alpha$.

Podemos expresar esto diciendo que hemos definido un número real como el conjunto de los números racionales menores que él mismo. Dicho así es circular, pero la idea es que definimos un número real como el conjunto de los números racionales “que queremos que sean” menores que él, y luego hacemos que lo sean identificando los números racionales con parte de los números reales.

Nota Una idea fundamental en todo lo que estamos viendo es que un número real α está determinado por su posición con respecto a los números racionales, en el sentido de que si dos números reales α y β cumplen que todo número

racional $r < \alpha$ es también $r < \beta$ y viceversa (o, equivalentemente, que todo $r < \alpha$ cumple $r < \beta$ y todo $r > \alpha$ cumple $r > \beta$), entonces $\alpha = \beta$. La razón es que si fuera, por ejemplo, $\alpha < \beta$, existiría un número racional $\alpha < r < \beta$.

Y la razón por la que podemos sustituir los números racionales por los números racionales diádicos es que también es cierto que entre dos números reales hay siempre un número racional diádico. En efecto, si $\alpha < \beta$, podemos tomar números racionales $\alpha < r < r' < \beta$, luego basta probar que entre dos números racionales hay un número racional diádico. Ahora bien, podemos tomar un n tal que $1/(r' - r) < n < 2^n$, con lo que $0 < 1/2^n < r' - r$. Claramente existe un mínimo número natural $m > 2^n r$, con lo que

$$\frac{m-1}{2^n} \leq r < \frac{m}{2^n} = \frac{m-1}{2^n} + \frac{1}{2^n} < r + r' - r = r'.$$

Por lo tanto, $m/2^n$ cumple lo requerido. ■

Notemos que ahora tenemos con los números reales el problema opuesto al que teníamos planteado con los números racionales: el teorema 2.10 afirma que todo punto de una recta graduada tiene asignado un número real, pero no sabemos si a todo número real le corresponde un punto en la recta. Podemos pensar que los números reales representan todas las medidas posibles de un segmento (admitiendo la propiedad arquimediana). Veremos que los axiomas que hemos dado —incluyendo al de las paralelas, que de momento no estamos usando— implican que necesariamente tiene que haber segmentos de longitud irracional, pero dichos axiomas no obligan a que todo número irracional sea necesariamente la longitud de un segmento. Nos ocuparemos de esto más adelante. De momento vamos a extender la graduación de las rectas:

Definición 2.12 En una recta graduada, si β es un número real llamaremos P_β al único punto (si existe) que cumple $\alpha_{P_\beta} = \beta$, y que está determinado por que, para todo número racional r , se cumple $P_r < P_\beta$ si y sólo si $r < \beta$.

Notemos que ahora, si r es un número racional, tenemos definidos dos puntos P_r , pero en realidad son el mismo, pues, para P_r en el sentido original,

$$\alpha_{P_r} = \{s \in \mathbb{Q} \mid P_s < P_r\} = \{s \in \mathbb{Q} \mid s < r\} = i(r),$$

luego con la identificación de r con $i(r)$ tenemos que $\alpha_{P_r} = r$ y P_r en el sentido que acabamos de definir es precisamente el P_r que ya teníamos definido.

Observemos que la proporción entre dos segmentos definida en 2.5 es un número real positivo. En lugar de $v/u = \alpha$ podemos escribir $v \equiv \alpha u$, donde hemos cambiado el signo $=$ por el de congruencia porque dos segmentos u y v cualesquiera determinan un único número real positivo $\alpha = v/u$, pero α y u no determinan un segmento v en concreto, sino que a lo sumo existe una clase de congruencia de segmentos v que cumplen $v \equiv \alpha u$ (aunque en principio puede no existir ninguna).

Observemos que, según hemos visto en 2.5, si $\alpha > 0$ tiene asociado el punto P_α en una recta graduada, entonces

$$\frac{\overline{P_0P_\alpha}}{\overline{P_0P_1}} = \alpha,$$

o equivalentemente, $\overline{P_0P_\alpha} \equiv \alpha \overline{P_0P_1}$.

El teorema siguiente precisa el significado de estas proporciones:

Teorema 2.13 *Dados dos segmentos u, v y un número real $\alpha > 0$, la relación $v \equiv \alpha u$ equivale a que, para todo número racional $0 < r < \alpha$, se cumple $ru < v$ y, para todo número racional $\alpha < r$, se cumple $v < ru$.*

DEMOSTRACIÓN: Fijamos una recta graduada con u como unidad, y en ella consideramos el punto P de la semirrecta positiva tal que $\overline{P_0P} \equiv v$. La relación $v \equiv \alpha u$ es equivalente a $v/u = \alpha$, es decir, a que $P = P_\alpha$, y esto equivale a que un número racional r cumple $r < \alpha$ si y sólo si $P_r < P$. A su vez, esto equivale a que si $0 < r < \alpha$ entonces $P_r < P$ y si $\alpha < r$ entonces $P < P_r$ (no necesitamos considerar los números racionales $r \leq 0$ porque cumplen trivialmente $P_r < P$). Teniendo en cuenta que $\overline{P_0P_r} \equiv ru$ y que $\overline{P_0P} \equiv v$, esto equivale a que $ru < v$ cuando $0 < r < \alpha$ y $v < ru$ cuando $\alpha < r$. ■

Ahora es claro que $v \equiv ru$ en el sentido que ya teníamos definido para números racionales r si y sólo si $v \equiv ru$, en el sentido que hemos definido para números reales. En efecto, si llamamos $\alpha = i(r)$ y $v \equiv ru$, tenemos que si $0 < s < \alpha$ entonces $s < r$, luego $su < v$ y si $\alpha < s$ entonces $r < s$, luego $v < su$, luego el teorema anterior nos da que $v \equiv \alpha u$, es decir, $ru \equiv \alpha u$. El recíproco es trivial: si $v \equiv ru$ en el sentido de números reales y $w \equiv ru$ en el sentido de números racionales, hemos probado que también $w \equiv ru$ en el sentido de números reales, luego $v \equiv w \equiv ru$ en el sentido de números racionales.

Definición 2.14 Llamaremos *longitud* de un segmento v respecto de una unidad u a la proporción $l_u(v) = v/u \in \mathbb{R}$.

El teorema anterior expresa que el número real $\alpha = v/u$ parte el conjunto de los números racionales en dos mitades, de modo que si $0 < r < \alpha$ el segmento ru “se queda corto” al compararlo con v y si $\alpha < r$ el segmento ru “se pasa” al compararlo con v , luego el número α representa la longitud exacta de v .

Hemos visto en 2.5 que dos segmentos son congruentes si y sólo si tienen la misma longitud α (respecto de una misma unidad u). Más aún, razonamos ahora que segmentos mayores tienen mayor longitud:

Si $0 < \alpha < \beta$ entonces $\alpha u < \beta u$ (si es que están definidos), porque podemos tomar un número racional $\alpha < r < \beta$ y el teorema 2.13 nos da que $\alpha u < ru < \beta u$. Esto implica a su vez el recíproco: si $\alpha u < \beta u$, es que $\alpha < \beta$, pues si fuera $\beta \leq \alpha$ tendríamos que $\beta u < \alpha u$ o bien $\beta u = \alpha u$.

Vamos a definir una suma de números reales de manera que la longitud de una suma de segmentos será la suma de sus longitudes.

Definición 2.15 Dados $\alpha, \beta \in \mathbb{R}$, definimos

$$\alpha + \beta = \sup\{r + s \mid r, s \in \mathbb{Q}, r < \alpha, s < \beta\}.$$

Ciertamente tenemos que $\alpha + \beta \in \mathbb{R}$, pues el conjunto está acotado superiormente por la suma de dos números racionales mayores que α y β , respectivamente.

Teorema 2.16 *Se cumple:*

1. La suma de números reales es asociativa y conmutativa, tiene a 0 por elemento neutro y todo número real tiene un opuesto.
2. La suma $+$ restringida a \mathbb{Q} es la suma usual en \mathbb{Q} .
3. Si $\alpha \leq \beta$ entonces $\alpha + \gamma \leq \beta + \gamma$.

DEMOSTRACIÓN: Notemos en general que para probar que dos números reales α y β son iguales es suficiente probar que todo número racional x que cumple $x < \alpha$ cumple también $x \leq \beta$ y viceversa.

Si $x \in \mathbb{Q}$, $x < (\alpha + \beta) + \gamma$, entonces existen $y, t \in \mathbb{Q}$ tales que $x < y + t$, $y < \alpha + \beta$, $t < \gamma$, luego existen $r, s \in \mathbb{Q}$, tales que $y < r + s$, $r < \alpha$, $s < \beta$, luego $s + t < \beta + \gamma$ y $x < r + (s + t) < \alpha + (\beta + \gamma)$. Similarmente se recorre el camino contrario, luego $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

Es obvio que la suma es conmutativa. Veamos ahora que $\alpha + 0 = \alpha$. Si $x \in \mathbb{Q}$, $x < \alpha + 0$ existen $r, s \in \mathbb{Q}$ tales que $x < r + s$, $r < \alpha$, $s < 0$, luego $x < r + s < r < \alpha$. Así mismo, si $r \in \mathbb{Q}$, $r < \alpha$ entonces existe un $s \in \mathbb{Q}$ tal que $r < s < \alpha$, con lo que $r = s + (r - s) < \alpha + 0$.

Dado un número real α , definimos

$$-\alpha = \sup\{-r \mid r \in \mathbb{Q}, \alpha < r\}.$$

Una cota superior del conjunto es $-s$, donde $s \in \mathbb{Q}$, $s < \alpha$, luego $-\alpha$ es un número real. Veamos que $\alpha + (-\alpha) = 0$.

Si $x \in \mathbb{Q}$, $x < \alpha + (-\alpha)$ entonces existen $r, s \in \mathbb{Q}$ tales que $x < r + s$, $r < \alpha$, $-s > \alpha$, luego $x < r + s < 0$. Recíprocamente, si $x \in \mathbb{Q}$, $x < 0$, tomamos números racionales $x < -u < 0$ y $v < \alpha$. Entonces la sucesión $v + nu$ sobrepasará (un número racional mayor que) α para algún natural n , que podemos tomar mínimo. Así obtenemos un número $r < \alpha$ tal que $s = r + u > \alpha$ (si $r + u = \alpha$ cambiamos u por un número mayor que siga cumpliendo $x < -u < 0$). Entonces $x < -u = r - s < \alpha + (-\alpha)$.

Dados $u, v \in \mathbb{Q}$, si $x \in \mathbb{Q}$ cumple $x < i(u) + i(v)$ entonces existen $r, s \in \mathbb{Q}$ tales que $x < r + s$, $r < u$, $s < v$, luego $x < u + v$, luego $x < i(u + v)$. Si $x < i(u + v)$ tomamos $r \in \mathbb{Q}$ tal que $0 < r < (u + v - x)/2$, de modo que $x < (u - r) + (v - r) < i(u) + i(v)$. Por lo tanto $i(u + v) = i(u) + i(v)$.

Si $r < \alpha$ y $s < \gamma$ son números racionales y $\alpha \leq \beta$, entonces $r + s \leq \beta + \gamma$ por definición de suma, luego tomando el supremo, $\alpha + \gamma \leq \beta + \gamma$. ■

La interpretación geométrica de la suma de números reales está expresada en la relación siguiente, según la cual, como habíamos anticipado, la longitud de una suma de segmentos es la suma de las longitudes.

Teorema 2.17 *Si α y β son números reales positivos y u es un segmento, entonces⁵ $(\alpha + \beta)u \equiv \alpha u + \beta u$.*

DEMOSTRACIÓN: Sea $\alpha u + \beta u \equiv \gamma u$. Hay que ver que $\gamma = \alpha + \beta$. Si $r < \alpha$ y $s < \beta$ son números racionales positivos, entonces

$$(r + s)u \equiv ru + su < \alpha u + \beta u \equiv \gamma u,$$

luego $r + s < \gamma$, lo que prueba que $\alpha + \beta \leq \gamma$. Si fuera $\alpha + \beta < \gamma$ sea $\delta > 0$ tal que $\gamma = \alpha + \beta + \delta$ (existe por la propiedad 3 del teorema anterior). Sea $r \in \mathbb{Q}$ tal que $0 < r < \delta$. Tomemos números racionales s y t tales que $\alpha < s < \alpha + r/2$ y $\beta < t < \beta + r/2$. Entonces $s + t < \alpha + \beta + r < \alpha + \beta + \delta = \gamma$, luego

$$\alpha u + \beta u < su + tu \equiv (s + t)u < \gamma u \equiv \alpha u + \beta u,$$

con lo que tenemos una contradicción. Por consiguiente se da la igualdad. ■

Observaciones El teorema anterior implica también que si $\alpha, \beta > 0$ y existen αu y $(\alpha + \beta)u$, entonces existe βu . En efecto, como $\alpha < \alpha + \beta$, se cumple que $\alpha u < (\alpha + \beta)u$, luego existe un segmento (que podemos expresar como γu) tal que $\alpha u + \gamma u \equiv (\alpha + \beta)u$ pero por el teorema anterior esto equivale a $(\alpha + \gamma)u \equiv (\alpha + \beta)u$, de donde $\alpha + \gamma = \alpha + \beta$, luego $\gamma = \beta$.

Si a un número real $\alpha > 0$ le corresponde un punto P_α en una recta graduada, entonces a $-\alpha$ le corresponde el punto simétrico respecto a P_0 , es decir, el punto de la semirrecta negativa que cumple $\overline{P_{-\alpha}P_0} \equiv \overline{P_0P_\alpha}$. En efecto, si llamamos P a este punto, entonces, para todo número racional r , se cumple $P_r < P$ si y sólo si $P_\alpha < P_{-r}$ (por la definición de los puntos P_r para números racionales), si y sólo si $\alpha < -r$, si y sólo si $r < -\alpha$, luego $P = P_{-\alpha}$.

De esto se desprende que si α y β tienen asignados puntos en una recta graduada, lo mismo sucede con $\alpha + \beta$ y $\alpha - \beta$. ■

Teorema 2.18 *Si α y β son dos números reales distintos que se corresponden con puntos P_α y P_β de una recta graduada, entonces la longitud del segmento $\overline{P_\alpha P_\beta}$ (respecto de la unidad de la recta) es $|\alpha - \beta|$.*

DEMOSTRACIÓN: Supongamos que $\alpha < \beta$, y veamos que $\overline{P_\alpha P_\beta} \equiv \overline{P_0 P_{\beta - \alpha}}$.

En efecto, si $0 < \alpha < \beta$, tenemos que $\overline{P_0 P_\beta} \equiv \overline{P_0 P_\alpha} + \overline{P_\alpha P_\beta}$, es decir, $\beta u \equiv \alpha u + \overline{P_\alpha P_\beta}$, donde $u = \overline{P_0 P_1}$ es la unidad de la recta graduada, pero por el teorema anterior y las observaciones posteriores, $\beta u \equiv \alpha u + (\beta - \alpha)u$, luego concluimos que $\overline{P_\alpha P_\beta} \equiv (\beta - \alpha)u \equiv \overline{P_0 P_{\beta - \alpha}}$.

⁵Aquí hay que entender que si están definidos αu y βu , entonces también está definido $(\alpha + \beta)u$ y cumple la relación indicada.

Si $\alpha < \beta < 0$ sabemos que $\overline{P_\alpha P_0} \equiv \overline{P_0 P_{-\alpha}}$ y $\overline{P_\beta P_0} \equiv \overline{P_0 P_{-\beta}}$, de donde $\overline{P_\alpha P_\beta} \equiv \overline{P_{-\beta} P_{-\alpha}}$ y el problema se reduce al caso ya probado.

Similarmente, si $\alpha < 0 < \beta$ partimos de que $\overline{P_\alpha P_\beta} \equiv \overline{P_\alpha P_0} + \overline{P_0 P_\beta}$ y usamos los dos casos anteriores. Sólo faltan los casos en que $\alpha = 0$ o $\beta = 0$, que son inmediatos.

Así pues, $\overline{P_\alpha P_\beta} \equiv (\beta - \alpha)u$, lo que significa que la longitud de $\overline{P_\alpha P_\beta}$ respecto de u es $\beta - \alpha$. Si $\beta < \alpha$ usamos que $\overline{P_\alpha P_\beta} = \overline{P_\beta P_\alpha}$ y, por la parte ya probada, la longitud es $\alpha - \beta = |\beta - \alpha|$. ■

La *distancia* entre dos puntos P y Q se define como la longitud del segmento \overline{PQ} (respecto de una unidad de medida prefijada u) si $P \neq Q$ y como 0 si $P = Q$. En estos términos, el teorema anterior afirma que la distancia entre dos puntos P_α y P_β de una recta graduada es $|\alpha - \beta|$.

Si tomamos dos unidades de longitud, por ejemplo el metro y el centímetro, para expresar en centímetros una longitud dada en metros hemos de multiplicar por 100, debido a que $1\text{m} = 100\text{cm}$. En general, si tenemos dos unidades u y v , para expresar respecto de v la longitud de un segmento dada respecto de u necesitamos conocer la longitud de u respecto de v , es decir, el número β que cumple $u \equiv \beta v$. Entonces, si $s \equiv \alpha u$, tendremos también $s \equiv \alpha\beta v$, de modo que la longitud de s en términos de v se obtiene multiplicando por β su longitud en términos de u . Todo esto es fácil de probar si las longitudes son todas racionales, pero, si no es así, ni siquiera tenemos definido el producto $\alpha\beta$. Vamos a definir un producto de números reales que dé sentido a estas fórmulas.

Definición 2.19 Si α y β son números reales positivos definimos

$$\alpha\beta = \sup\{rs \mid r, s \in \mathbb{Q}, 0 < r < \alpha, 0 < s < \beta\}.$$

El producto de dos números reales no nulos se define por las relaciones:

$$\alpha\beta = \begin{cases} -((-\alpha)\beta) & \text{si } \alpha < 0, \beta > 0, \\ -(\alpha(-\beta)) & \text{si } \alpha > 0, \beta < 0, \\ (-\alpha)(-\beta) & \text{si } \alpha < 0, \beta < 0. \end{cases}$$

Finalmente, si $\alpha = 0$ o $\beta = 0$ definimos $\alpha\beta = 0$.

Teorema 2.20 $(\mathbb{R}, +, \cdot, \leq)$ es un cuerpo ordenado que contiene a \mathbb{Q} como subcuerpo.

DEMOSTRACIÓN: La prueba de que el producto de números reales positivos es asociativo, conmutativo, tiene por neutro a 1 y de que todo número real positivo α tiene un inverso α^{-1} se obtiene cambiando sumas por productos en la prueba de que la suma de números reales tiene estas propiedades. Después las propiedades se trasladan formalmente a números reales arbitrarios a partir de la definición de producto. Para probar que \mathbb{R} es un cuerpo sólo queda comprobar que la suma distribuye al producto.

Tomemos en primer lugar $\alpha, \beta, \gamma > 0$ y veamos que $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

Si $r \in \mathbb{Q}$ cumple $0 < r < \alpha(\beta + \gamma)$, entonces existen $u, v \in \mathbb{Q}$ positivos y tales que $r < uv$, $u < \alpha$, $v < \beta + \gamma$, luego existen $x, y \in \mathbb{Q}$ positivos tales que $v < x + y$, $x < \beta$, $y < \gamma$. Entonces $r < u(x + y) = ux + uy \leq \alpha\beta + \alpha\gamma$.

Si $0 < r < \alpha\beta + \alpha\gamma$ entonces existen $u, v \in \mathbb{Q}$ positivos tales que $r < u + v$, $u < \alpha\beta$, $v < \alpha\gamma$. A su vez existen $a, b, c, d \in \mathbb{Q}$ positivos de modo que $r < ab + cd$, $a < \alpha$, $b < \beta$, $c < \alpha$, $d < \gamma$. Sea $e = \max\{a, c\}$. Entonces $e < \alpha$ y $r < eb + ed = e(b + d) < \alpha(\beta + \gamma)$. Esto prueba la igualdad.

Si $\beta + \gamma \geq 0$, $\beta \geq 0$, $\gamma < 0$, entonces $\alpha\beta = \alpha((\beta + \gamma) - \gamma) = \alpha(\beta + \gamma) + \alpha(-\beta)$ (puesto que $\beta + \gamma \geq 0$ y $-\gamma \geq 0$), de donde $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$. Los demás casos se siguen formalmente de éstos dos.

Se cumple que \mathbb{R} es un cuerpo ordenado, pues esto significa que la suma cumple la tercera propiedad del teorema 2.16 y que el producto de números positivos es positivo, lo cual es inmediato por la definición.

El hecho de que la suma en \mathbb{R} extienda a la suma en \mathbb{Q} implica sucesivamente que el producto en \mathbb{R} extiende al producto de \mathbb{N} , al de \mathbb{Z} y al de \mathbb{Q} . ■

A partir de aquí podemos usar en \mathbb{R} todas las propiedades generales de los cuerpos ordenados, así como los conceptos definibles en ellos, como el signo, el valor absoluto, etc.

La interpretación geométrica del producto de números reales es la que ya habíamos indicado:

Teorema 2.21 *Si α y β son números reales positivos y u es un segmento, entonces $\alpha(\beta u) \equiv (\alpha\beta)u$.*

DEMOSTRACIÓN: Sea $\alpha(\beta u) \equiv \gamma u$. Tenemos que probar que $\gamma = \alpha\beta$. Si $r < \gamma$, entonces⁶ $ru < \alpha(\beta u)$, luego existe $0 < r' < \alpha$ tal que $ru < r'(\beta u)$, luego $(r/r')u < \beta u$, luego $r/r' < \beta$, luego $r < r'\beta < \alpha\beta$. Esto prueba que $\gamma \leq \alpha\beta$.

Si $r < \alpha\beta$ existen números racionales $0 < a < \alpha$, $0 < b < \beta$ tales que $r < ab$, luego $ru < abu$, luego $(r/a)u < bu < \beta u$, luego $ru < a(\beta u) < \alpha(\beta u) = \gamma u$, luego $r < \gamma$, luego $\gamma = \alpha\beta$. ■

Observemos que en la prueba del teorema anterior suponemos que están definidos βu y $\alpha(\beta u)$, y demostramos que también lo está $(\alpha\beta)u$. Esto no significa que si están definidos αu y βu , también tenga que estarlo $(\alpha\beta)u$. Veremos que esto puede probarse a partir del axioma de las paralelas.

Una consecuencia que sí que podemos extraer del teorema anterior es que si $u < v$ y $\alpha > 0$, entonces $\alpha u < \alpha v$ (si están definidos).

En efecto, expresamos $v = \beta u > 1u$, luego $\beta > 1$ y por el teorema anterior $\alpha v = (\alpha\beta)u > \alpha u$.

Terminamos con una caracterización de la longitud de los segmentos que nos será útil más adelante:

⁶Aquí usamos que, en general, si $s < \alpha u$, entonces existe un número racional $0 < r < \alpha$ tal que $s < ru$. En efecto, será $s = \beta u$, con $0 < \beta < \alpha$, y basta tomar $\beta < r < \alpha$ para que $s = \beta u < ru$.

Teorema 2.22 *Sea m una aplicación que a cada segmento le asigna un número real positivo y que cumpla las propiedades siguientes:*

1. *Si $u \equiv v$ entonces $m(u) = m(v)$,*
2. *Para todo par de segmentos u, v se cumple $m(u + v) = m(u) + m(v)$.*

Entonces, para todo par de segmentos u, v se cumple

$$\frac{m(v)}{m(u)} = \frac{v}{u}.$$

DEMOSTRACIÓN: Para todo número natural q se cumple por hipótesis que $m(qx) = qm(x)$, y si aplicamos esto a $x = u/q$ tenemos que $m(u) = qm(u/q)$, o equivalentemente

$$\frac{m(u/q)}{m(u)} = \frac{1}{q}.$$

Multiplicando esta igualdad por un número natural p obtenemos, para cualquier número racional positivo $r = p/q$, que

$$\frac{m(ru)}{m(u)} = r.$$

Si dos segmentos cumplen $x < y$, entonces existe un segmento z tal que $y \equiv x + z$, luego $m(x) < m(x) + m(z) = m(y)$. Por lo tanto, si $\alpha > 0$ es un número real para el que está definido αu y r, s son números racionales tales que $r < \alpha < s$ tenemos

$$r = \frac{m(ru)}{m(u)} < \frac{m(\alpha u)}{m(u)} < \frac{m(su)}{m(u)} = s.$$

Como esto es válido para todo r y todo s , ha de ser

$$\frac{m(\alpha u)}{m(u)} = \alpha.$$

Expresando $v \equiv \alpha u$ tenemos la relación buscada. ■

Lo que afirma este teorema es que si una aplicación m cumple las propiedades indicadas, tomamos una unidad de longitud u y llamamos $\alpha = m(u)$, entonces m asigna a cada segmento su longitud multiplicada por α .

2.3 Amplitud de ángulos

Vamos a asignar un número real a cada ángulo tal y como hemos hecho con los segmentos. En general, dados dos ángulos L y M y un número natural m , la relación $M \equiv mL$ indicará que M se obtiene sumando L consigo mismo m veces. Sin embargo, a diferencia de lo que sucede con la suma de segmentos, no tiene por qué haber ningún ángulo M que cumpla esto, incluso puede ocurrir que no lo haya para $m = 2$. Por ejemplo, si L es un ángulo recto, está definido $2L$, pero no $3L$, si L es la mitad de un ángulo recto, está definido $4L$, pero no $5L$, etc.

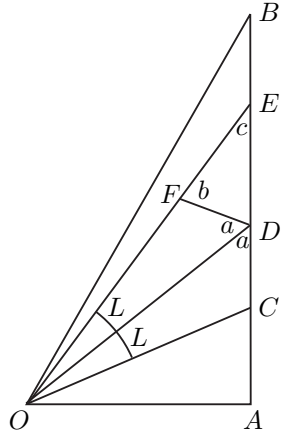
En primer lugar vamos a probar que la propiedad de Arquímedes implica una propiedad análoga para ángulos:

Teorema 2.23 *Dados dos ángulos L y M , existe un número natural n tal que nL no está definido o bien $nL > M$.*

DEMOSTRACIÓN: Pongamos que $M = \widehat{AOB}$, con $\overline{OA} \equiv \overline{OB}$. Sea C el punto medio de \overline{AB} , de modo que la recta OC es la bisectriz de M . Sea $M' = \widehat{AOC}$. Entonces $M = 2M'$ y basta probar el teorema para M' en vez de M , por lo que podemos suponer que M es uno de los ángulos agudos de un triángulo rectángulo.

Sea \widehat{OAB} un triángulo rectángulo con $\hat{O} = M$ y el ángulo recto \hat{A} . Si $L \geq M$ no hay nada que probar. En caso contrario, suponemos por reducción al absurdo que nL está definido para todo $n > 0$ y que $nL < M$. Podemos tomar A_n en \overline{AB} de modo que $\widehat{AOA_n} \equiv nL$.

Vamos a probar que $\overline{AA_1} < \overline{A_1A_2} < \overline{A_2A_3} < \dots$. Para ello llamamos C, D, E a tres A_i 's consecutivos, de modo que tenemos la situación que muestra la figura:



Observamos que el ángulo $a = \widehat{ODA}$ es agudo, pues forma parte del triángulo rectángulo \widehat{ODA} . Por lo tanto, su suplementario \widehat{ODE} es obtuso, luego podemos trazar una semirrecta \overline{DF} de modo que $a \equiv \widehat{ODF}$. Además el punto F podemos tomarlo en \overline{OE} .

Ahora, $\widehat{ODF} \equiv \widehat{ODC}$, porque comparten un lado y los dos ángulos de sus extremos, luego se cumple $\overline{FD} \equiv \overline{DC}$.

Por otra parte, $b = \widehat{DFE}$ es un ángulo exterior del triángulo \widehat{OFD} , luego $a < b$ (teorema 1.70). Igualmente, $a = \widehat{CDO}$ es un ángulo exterior de \widehat{EDO} , luego $c < a$, y concluimos que $c < b$. El teorema 1.71 aplicado a \widehat{DEF} implica entonces que

$$\overline{DC} \equiv \overline{FD} < \overline{DE},$$

como queríamos probar.

Por la propiedad de Arquímedes existe un n tal que $\overline{AB} < n\overline{AA_1}$, y como $\overline{AA_1} < \overline{A_iA_{i+1}}$, esto implica que $\overline{AB} < \overline{AA_n}$, lo cual es imposible, porque A_n está en \overline{AB} . ■

Nota Si aplicamos el teorema anterior a un ángulo recto R , obtenemos que para todo ángulo L existe un n tal que nL no está definido o bien $nL > R$, en cuyo caso $2nL$ ya no está definido, luego siempre existe un natural n tal que nL no está definido. ■

No es posible demostrar que todo ángulo se puede dividir en n partes iguales ni siquiera suponiendo el axioma de las paralelas. Por ello trabajaremos con bisecciones y números racionales diádicos, igual que podríamos haber hecho en el caso de los segmentos.

Si $n > 0$ es un número natural, definimos la relación $M \equiv L/n$ como que M puede sumarse consigo mismo n veces y el resultado es un ángulo congruente con L . El teorema 1.69 nos asegura que siempre existen ángulos $M \equiv L/2^n$, para todo número natural n . Además, en tal caso $M \equiv 2^{n'}(L/2^{n+n'})$, es decir, si dividimos un ángulo en $2^{n+n'}$ partes y sumamos $2^{n'}$ de ellas, el ángulo resultante es congruente con el que habríamos obtenido dividiéndolo en 2^n partes.

Como consecuencia, si están definidos $m(L/2^n)$ y $m'/(2^{n'})$ y $m/2^n = m'/2^{n'}$, entonces

$$m(L/2^n) \equiv m2^{n'}(L/2^{n+n'}) \equiv m'2^n(L/2^{n+n'}) \equiv m'(L/2^{n'}).$$

En otras palabras, si $r = m/2^n$ es un número racional diádico positivo, podemos definir la relación $M \equiv rL$ como $M \equiv m(L/2^n)$. Tal vez no exista ningún ángulo M que cumpla esto, pero si existe no depende de la representación de r como fracción, y la clase de congruencia de L determina la clase de congruencia de M .

Claramente, si está definido $(m/2^n)L$, también lo está $(m'/2^{n'})L$, para todo número natural $0 < m' \leq m$, y también es claro que si $r < r'$, entonces $rL < r'L$ (si están definidos).

Teorema 2.24 *Sea L un ángulo. Entonces*

1. *Existe un número real $\pi \geq 1$ tal que, para todo número racional diádico $r > 0$, se cumple que rL está definido si y sólo si $r \leq \pi$.*
2. *Dados dos ángulos $U < V$, hay un número racional diádico $r > 0$ tal que $U < rL < V$.*

DEMOSTRACIÓN: Por la nota siguiente al teorema 2.23 sabemos que existe un natural, que podemos tomar de la forma 2^n , tal que $2^n L$ no está definido.

Si $r = m/2^k > 2^n$ es un número racional diádico, podemos suponer que $k > n$ y entonces $2^{n+k} \leq m$. Si estuviera definido rL , también lo estaría $(2^{n+k}/2^k)L$, es decir, $2^n L$, lo cual es falso. Por lo tanto, 2^n es una cota superior del conjunto de los números diádicos r para los que rL está definido. Basta tomar como π el supremo de dicho conjunto.

Dados dos ángulos L y L' existe un número natural n tal que $(1/2^n)L < L'$. En efecto, si fuera $L' \leq (1/2^n)L$ para todo n , entonces es claro que $2^n L'$ estaría definido para todo n , y nL' también.

Sean ahora dos ángulos $U < V$. Existe un número natural n de manera que $(1/2^n)L < V - U < V$. Por el teorema 2.23 existe un mínimo número natural

m tal que $(m/2^n)L$ no está definido o es mayor que U . No puede ser $m = 1$, luego $((m - 1)/2^n)L$ está definido y es menor o igual que U , pero entonces

$$\frac{m}{2^n}L \equiv \frac{m-1}{2^n}L + \frac{1}{2^n}L < U + (V - U) \equiv V,$$

luego $(m/2^n)L$ sí que está definido luego tiene que ser

$$U < \frac{m}{2^n}L < V,$$

como había que probar. ■

Ahora ya podemos definir la medida de ángulos:

Definición 2.25 Dados dos ángulos L y U , definimos

$$L/U = \sup\{r \in \mathbb{R} \mid r \text{ es un número racional diádico positivo y } rU < L\}.$$

Así $\alpha = L/U$ es un número real porque el conjunto considerado está acotado superiormente por el número π dado por el teorema anterior. De hecho, $\alpha \leq \pi$. Además $\alpha > 0$, porque dicho teorema nos da también que existe un $r > 0$ tal que $L/2 < rU < L$, luego $\alpha \geq r > 0$.

Es evidente que α sólo depende de las clases de congruencia de U y L y, más aún, si $L/U = L'/U$, entonces $L \equiv L'$, pues en caso contrario sería $L < L'$ o $L' < L$. No perdemos generalidad si suponemos que $L < L'$, entonces, aplicando dos veces el teorema anterior, existen r y r' tales que $L < rU < r'U < L'$, y entonces, tomando supremos, $L/U \leq r < r' \leq L'/U$, contradicción.

Observemos que la definición de L/U tiene sentido incluso si L es un ángulo llano, en cuyo caso L/U es simplemente el supremo de todos los números diádicos r para los que rU está definido, que claramente es el número π dado por el teorema anterior. De hecho, $L/U = \pi$ si y sólo si L es un ángulo llano.

Si $\alpha = L/U$ escribiremos también $L \equiv \alpha U$, donde nuevamente hay que entender que no hemos definido un ángulo αU en particular, sino que, a lo sumo, existe una clase de congruencia de ángulos L que cumplen la relación $L \equiv \alpha U$. Se cumple el análogo al teorema 2.13:

Teorema 2.26 *Dados dos ángulos L y U y $\alpha > 0$ un número real, la relación $L \equiv \alpha U$ equivale a que para todo número racional $r > 0$ para el que esté definido rU se cumple $rU < L$ si $r < \alpha$ y $L < rU$ si $r > \alpha$.*

DEMOSTRACIÓN: Si $r < \alpha$, existe un número diádico $r < r' < \alpha$ tal que $r'U < L$, por definición de αU , luego $rU < r'U < L$. Si $r > \alpha$, es fácil ver que existe un número diádico tal que $\alpha < r' < r$, y entonces $L \leq r'U < rU$. ■

De este teorema se sigue en particular que si $r > 0$ es un número racional, rU en el sentido que acabamos de definir coincide con el definido previamente en términos de subdivisiones de U .

Como en el caso del teorema 2.13, este resultado justifica que consideremos al único número real $\alpha > 0$ que cumple $L = \alpha U$ como la *amplitud* del ángulo L medida respecto del ángulo unitario U , en el sentido de que para todo número racional $r > 0$ para el que esté definido rU se cumple que rU “se queda corto” al compararlo con L si $r\alpha$, mientras que “se pasa” si $r > \alpha$, luego α está indicando la amplitud exacta de L respecto de U .

Las propiedades siguientes se demuestran sin dificultad de modo totalmente análogo a las correspondientes para segmentos:

Teorema 2.27 Sean α, β números reales y L, L' dos ángulos. Entonces se cumplen las propiedades siguientes (donde suponemos que las sumas y productos están definidos):

1. Si $\alpha < \beta$, entonces $\alpha L < \beta L$,
2. Si $L < L'$ entonces $\alpha L < \alpha L'$,
3. $(\alpha + \beta)L \equiv \alpha L + \beta L$,
4. $\alpha(L + L') \equiv \alpha L + \alpha L'$,
5. $(\alpha(\beta L)) \equiv (\alpha\beta)L$.

Por ejemplo, la propiedad 3 se interpreta como que la amplitud de la suma de dos ángulos es la suma de sus amplitudes (de donde se sigue a su vez que la amplitud de una resta es la resta de las amplitudes), y la propiedad 5 nos dice que para pasar de la amplitud α de un ángulo medida respecto de un ángulo L' a la medida respecto de un ángulo L tal que $L' = \beta L$ debemos multiplicar $\alpha\beta$.

Nota Hay una diferencia destacable entre la medida de segmentos y la medida de ángulos, y es que para medir segmentos es necesario fijar unidad de medida y no tenemos ningún criterio para seleccionar ninguna. En cambio, sí que tenemos criterios para seleccionar intrínsecamente una unidad de medida de ángulos. Por ejemplo, podemos tomar como unidad los ángulos rectos, o los ángulos llanos.

Hemos visto que cada ángulo U que fijemos como unidad determina un número real π , que es la medida de un ángulo llano tomando U como unidad. Por ejemplo, si fijamos como U un ángulo recto tenemos que $\pi = 2$, es decir, un ángulo llano mide dos ángulos rectos.

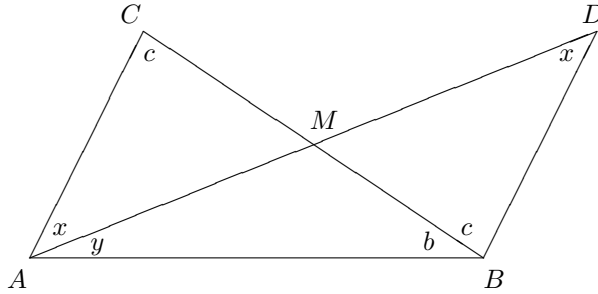
Por razones históricas, es habitual tomar como unidad de medida la clase de los ángulos $U \equiv (1/90)R$, donde R es un ángulo recto (la existencia de este ángulo no puede demostrarse con los axiomas que conocemos hasta ahora, pero sí con el axioma de continuidad). Esta unidad recibe el nombre de *grado sexagesimal*, de modo que un ángulo recto tiene una amplitud de 90° y un ángulo llano de 180° .

De momento no vamos a fijar ninguna unidad de medida de ángulos, y usaremos la letra π para referirnos a la medida de un ángulo llano respecto de una unidad de ángulos arbitraria. ■

El teorema 1.70 puede reenunciarse ahora como que la suma de dos ángulos de un triángulo es siempre menor que π . En realidad podemos probar un poco más:

Teorema 2.28 (Lagrange) *La suma de los tres ángulos de un triángulo es menor o igual que π .*

DEMOSTRACIÓN: Dado un triángulo \widehat{ABC} cuyos ángulos midan a , b y c , consideramos el punto medio M del lado \overline{BC} y prolongamos el segmento \overline{AM} hasta un segmento $\overline{AD} \equiv 2\overline{AM}$. Es claro entonces que $\widehat{AMC} \equiv \widehat{DMB}$, con lo que el triángulo \widehat{ABD} resulta tener un ángulo igual a $b + c$ y otros dos ángulos x e y tales que $a = x + y$. Uno de estos dos será menor o igual que $a/2$.



De este modo, dado un triángulo arbitrario hemos obtenido otro cuyos ángulos suman la misma cantidad pero uno de ellos es menor o igual que la mitad de uno de los ángulos originales. Repitiendo el proceso podemos obtener triángulos con la misma suma de ángulos y donde uno de ellos sea menor o igual que $a/2^n$. Si la suma de los ángulos del triángulo original fuera $\pi + \epsilon$, entonces podemos llegar a un triángulo cuyos ángulos sumen $\pi + \epsilon$ con uno de ellos menor que ϵ . Por lo tanto, los otros dos sumarán al menos π , lo cual contradice al teorema 1.70. ■

Terminamos lo referente a la medida de ángulos con un resultado análogo al teorema 2.22. Omitimos la prueba por ser completamente análoga.

Teorema 2.29 *Sea m una aplicación que a cada ángulo le asigna un número real positivo y que cumpla las propiedades siguientes:*

1. Si $L \equiv L'$ entonces $m(L) = m(L')$,
2. Si L , y L' se pueden sumar, entonces $m(L + L') = m(L) + m(L')$.

Entonces, para todo par de ángulos L , L' se cumple

$$\frac{m(L)}{m(L')} = \frac{L}{L'}.$$

2.4 Arcos y sectores circulares

Fijemos una circunferencia ω de centro O y radio r . Si A y B son dos puntos de ω no diametralmente opuestos (es decir, no alineados con O), llamaremos *arco menor* de extremos A y B a la intersección de ω con \widehat{AOB} . El *arco mayor* de extremos A y B es el conjunto formado por los puntos de ω que no están en el arco menor, más los puntos A y B . Diremos que estos dos arcos son mutuamente *complementarios*.

Un arco determina a sus extremos, pues son los únicos puntos P con la propiedad de que todo ángulo de vértice O y que contenga a P en su interior (es decir, no en su frontera) contiene puntos de ω externos al arco.

Si A y B son puntos de ω diametralmente opuestos llamaremos *arcos* de extremos A y B a las intersecciones de ω con los dos semiplanos determinados por la recta AB . Así tenemos dos pares de arcos complementarios, sin que podamos decir que uno es mayor que otro. Estos arcos se llaman también *semicircunferencias*. Las *semicircunferencias* determinan sus extremos del mismo modo que los demás arcos. Usaremos la notación \widehat{AB} para referirnos a un arco de extremos A y B en una circunferencia dada, pero hemos de tener presente que \widehat{AB} no determina el arco, sino que éste depende de la circunferencia y además hay dos arcos distintos con los mismos extremos. Cuando no sean *semicircunferencias* los distinguiremos si conviene como \widehat{AB}_M (arco mayor) y \widehat{AB}_m (arco menor).

El segmento que une los extremos de un arco se llaman *cuerdas* del arco. Dos arcos complementarios comparten la misma cuerda.

Si en lugar de partir de una circunferencia partimos de un círculo obtenemos las definiciones de *sector circular* y *semicírculo*. La frontera de un sector circular está formada por dos radios \overline{OA} , \overline{OB} y un arco de extremos A y B , de modo que los arcos se corresponden biunívocamente con los sectores, por lo que es equivalente trabajar con unos u otros.

Los arcos admiten la caracterización siguiente:

Teorema 2.30 *Los arcos de extremos A y B en una circunferencia ω son las intersecciones con ω de los semiplanos determinados por la recta AB .*

DEMOSTRACIÓN: Si AB pasa por el centro O de ω los arcos de extremos A y B son *semicircunferencias*, y el resultado es cierto por definición. Supongamos el caso contrario. Si P es un punto del arco menor \widehat{AB}_m , entonces P está en el ángulo \widehat{AOB} , luego la semirrecta \overrightarrow{OP} corta al segmento \overline{AB} . Como este segmento está contenido en el círculo, el punto de corte ha de estar en \overline{OP} , y esto significa que O y P están en semiplanos distintos respecto a AB . Recíprocamente, si P es un punto de ω en el semiplano opuesto a O respecto a AB , entonces \overline{OP} ha de cortar a AB . Como \overline{OP} está contenido en el círculo, el punto de corte ha de estar en \overline{AB} , luego la semirrecta \overrightarrow{OP} está contenida en \widehat{AOB} , lo que se traduce en que P está en \widehat{AB}_m .

Con esto hemos probado que \widehat{AB}_m es la intersección de ω con el semiplano de AB que no contiene a O . Obviamente entonces el arco mayor ha de ser la intersección de ω con el semiplano que contiene a O . ■

Definición 2.31 La *amplitud* de un arco de extremos AB (en un círculo de centro O) es la del ángulo \widehat{AOB} si el arco es menor, 2π menos esta amplitud si es mayor y π si es una semicircunferencia.

De este modo, los arcos menores son los arcos de amplitud menor que π y los mayores son los de amplitud mayor que π . Podemos pensar en las circunferencias como arcos de amplitud igual 2π .

Teorema 2.32 Sean \widehat{AB} y \widehat{BC} dos arcos de una circunferencia con el extremo B como único punto común. Entonces $\widehat{AB} \cup \widehat{BC}$ es un arco de extremos A y C , y su amplitud es la suma de las amplitudes de los arcos dados.

DEMOSTRACIÓN: Sea O el centro de la circunferencia que contiene a los arcos. Representemos por $m(X)$ la amplitud de un ángulo o arco X . Supongamos que la amplitud de uno de los arcos es mayor que π , por ejemplo $m(\widehat{AB}) > \pi$. Entonces \widehat{AB} es un arco mayor, luego está contenido en el complementario del ángulo \widehat{AOB} . Como \widehat{BC} no tiene puntos en común con \widehat{AB} (salvo B) ha de estar contenido en \widehat{AOB} . En particular C es un punto interior de este ángulo, luego $\widehat{AOB} = \widehat{AOC} + \widehat{COB}$. Es claro entonces que $\widehat{AB} \cup \widehat{BC} = \widehat{AC}_M$. Además

$$m(\widehat{AC}_M) = 2\pi - m(\widehat{AOC}) = 2\pi - m(\widehat{AOB}) + m(\widehat{COB}) = m(\widehat{AB}) + m(\widehat{BC}).$$

Si uno de los arcos tiene amplitud π la prueba es casi idéntica. Supongamos ahora que ambos arcos son menores que π . Esto significa que ambos arcos son menores, luego están contenidos en los ángulos \widehat{AOB} y \widehat{BOC} . Llamemos A' y B' a los puntos diametralmente opuestos a A y B .

Notemos que A y C han de estar en semiplanos opuestos respecto a BB' , pues en caso contrario los ángulos \widehat{AOB} y \widehat{BOC} estarían uno contenido en otro, y los arcos también. Por lo tanto C está en uno de los dos ángulos $\widehat{BOA'}$ o $\widehat{A'OB'}$.

Si C está en $\widehat{BOA'}$ (incluyendo el caso $C = A'$) entonces tenemos la relación $\widehat{AOC} = \widehat{AOB} + \widehat{BOC}$, luego $\widehat{AB} \cup \widehat{BC} = \widehat{AC}_m$ y la relación entre las amplitudes es clara.

Si C está en $\widehat{A'OB'}$ entonces $\widehat{BOC} = \widehat{BOA'} + \widehat{A'OC}$, luego

$$\widehat{AB} \cup \widehat{BC} = \widehat{AB} \cup \widehat{BA'} \cup \widehat{A'C} = \widehat{AA'} \cup \widehat{A'C}_m = \widehat{AC}_M,$$

donde $\widehat{AA'}$ es la semicircunferencia que contiene a B . Las amplitudes cumplen la relación indicada:

$$\begin{aligned} m(\widehat{AC}_M) &= 2\pi - m(\widehat{AOC}) = 2\pi - \pi + m(\widehat{A'OC}) = \pi + m(\widehat{A'OC}) \\ &= m(\widehat{AOB}) + m(\widehat{BOA'}) + m(\widehat{A'OC}) = m(\widehat{AOB}) + m(\widehat{BOC}) \\ &= m(\widehat{AB}) + m(\widehat{BC}). \end{aligned}$$

■

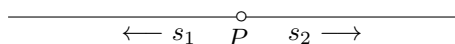
2.5 El axioma de continuidad

Hemos llamado geometría métrica (tridimensional o bidimensional) a todo conjunto de puntos \mathbb{E} en el que hemos seleccionado una familia de rectas, otra de planos (el el caso tridimensional), una relación triádica “estar entre” y dos relaciones de congruencia de segmentos y ángulos, de modo que se cumplen los tres bloques de axiomas de incidencia (A), de ordenación (B) y de congruencia (C).

A este cuerpo básico de axiomas hemos añadido tres axiomas más, el axioma de las circunferencias, la propiedad de división de segmentos y la propiedad de Arquímedes. Ahora presentamos un axioma que implica estos tres hechos:

Axioma de continuidad *Si una recta r puede partirse como $r = s_1 \cup s_2$, donde s_1 y s_2 son dos conjuntos convexos disjuntos no vacíos, entonces existe un punto $P \in r$ tal que uno de los conjuntos s_i es una de las semirrectas determinadas por P en r y el otro es la semirrecta complementaria menos P .*

Si tenemos una descomposición $r = s_1 \cup s_2$ en las condiciones del axioma de continuidad y $U \in s_1$, $V \in s_2$, entonces el requisito de que s_1 y s_2 sean convexos se traduce en que todos los puntos de s_1 tienen que ser $<_{UV}$ que todos los puntos de s_2 . Si s_1 no tuviera máximo y s_2 no tuviera mínimo, entonces ambos conjuntos estarían poniendo en evidencia la presencia de un “agujero” en la recta, y eso es lo que niega el axioma de continuidad al afirmar que tiene que haber un punto P que sea el máximo de s_1 o bien el mínimo de s_2 .



Por lo tanto, el axioma de continuidad puede interpretarse como que las rectas “no tienen huecos”. Quizá se entienda mejor esta idea al ver las aplicaciones.

En la práctica suele ser más conveniente aplicar una versión equivalente para segmentos del axioma de continuidad:

Si $\overline{AB} = s_1 \cup s_2$ es una descomposición de un segmento en dos conjuntos convexos disjuntos con $A \in s_1$ y $B \in s_2$. Supongamos que s_1 y s_2 contienen otros puntos además de A y B . Entonces existe un punto $A - P - B$ tal que s_1 y s_2 son los segmentos \overline{AP} y \overline{PB} salvo que a uno de ellos le falta P .

Para demostrar esta versión tomamos puntos $U - A - B - V$ y consideramos $s'_1 = \overrightarrow{AU} \cup s_1$, $s'_2 = \overrightarrow{BV} \cup s_2$. Se comprueba fácilmente que cumplen los requisitos del axioma de continuidad, y el punto P que éste proporciona cumple también la conclusión de la variante. Para probar el recíproco, dados s_1 y s_2 en las condiciones del axioma de continuidad, tomamos $A \in s_1$ y $B \in s_2$ distintos de sus posibles máximo o mínimo y aplicamos la variante a $\overline{AB} \cap s_1$ y $\overline{AB} \cap s_2$.

Demostración de la propiedad de Arquímedes Podemos suponer que $u = \overline{AB}$ y $v = \overline{AC}$, así como que ambos están sobre una misma semirrecta s de origen A . Supongamos que el resultado es falso, es decir, que $nu < v$ para todo número natural n (observemos que si se diera la igualdad para un n , entonces $n + 1$ cumpliría el teorema).

Para cada $n > 0$, sea A_n el punto de s que cumple que $\overline{AA_n} \equiv nu$. Llamemos X al conjunto de todos los puntos de la semirrecta complementaria a s más los puntos P de s que cumplen $P <_{AB} A_n$ para algún n . Sea Y el conjunto de los puntos de s que no están en X (por ejemplo C). Es evidente que se cumplen las hipótesis del axioma de continuidad, luego existe un punto D en s de modo que X e Y son las semirrectas de origen D , salvo que uno de los dos no contendrá a D , pero esto no afecta a que podemos tomar un punto E de X tal que $\overline{ED} \equiv u$. Si fuera $E - A - D$ o bien $E = A$, cambiamos E por otro punto $A - E' - D$, con lo que se cumple $\overline{E'D} \leq \overline{ED} \equiv u$. Así podemos suponer que $A - E - D$. Como $E \in X$ existe un número n tal que $E <_{AB} A_n$, lo que significa que $\overline{AE} < \overline{AA_n} \equiv nu$, de donde

$$\overline{AD} \equiv \overline{AE} + \overline{ED} < nu + u = (n + 1)u \equiv \overline{AA_{n+1}},$$

pero entonces A_{n+1} es un punto de X posterior a D , contradicción. ■

Demostración de la propiedad de división de segmentos Sea \overline{AB} un segmento y $n \geq 2$ un número natural. Llamamos X al conjunto formado por A y los puntos $P \neq A$ de \overline{AB} tales que $n\overline{AP} < \overline{AB}$. Notemos que X contiene otros puntos aparte de A , pues si $n < 2^m$ y $2^{-m}\overline{AB} \equiv \overline{AP}$, entonces también $P \in X$. Sea Y el conjunto de los puntos de \overline{AB} que no están en X (entonces $B \in Y$, y también pertenece a Y el punto medio del segmento \overline{AB}).

Es obvio que se cumplen las hipótesis del axioma de continuidad (de la variante para segmentos), por lo que existe un punto $A - C - B$ de manera que $\overline{AC} \setminus \{C\} \subset X \subset \overline{AC}$. Veamos que $n\overline{AC} \equiv \overline{AB}$.

Supongamos que $n\overline{AC} < \overline{AB}$. Sea entonces D un punto entre A y B tal que $n\overline{AC} \equiv \overline{AD}$. Sea m un número natural tal que $2^m > n$. Sea $u \equiv 2^{-m}\overline{DB}$. Entonces tenemos $nu < 2^m u \equiv \overline{DB}$. Sea $A - C - E$ de modo que $\overline{CE} \equiv u$. Así,

$$\overline{AE} < n\overline{AE} \equiv n\overline{AC} + n\overline{CE} < \overline{AD} + \overline{DB} \equiv \overline{AB}.$$

Esto implica en particular que $A - E - B$, luego $E \in X$, pero por otra parte $E \notin \overline{AC}$, contradicción.

Supongamos ahora que $\overline{AB} < n\overline{AC}$. Entonces tomamos $D \in \overline{AB}$ tal que $n\overline{AC} \equiv \overline{AD}$, con lo que necesariamente $A - B - D$. Como antes, tomamos $u \equiv 2^{-m}\overline{BD}$ y así $nu < \overline{BD}$. Tiene que ser $u < \overline{AC}$, pues si $\overline{AC} \leq u$, entonces

$$\overline{AD} \equiv n\overline{AC} \leq nu < \overline{BC} < \overline{AB} + \overline{BD} = \overline{AD},$$

y tenemos una contradicción. Por lo tanto podemos tomar $A - E - C$ de modo que $\overline{EC} \equiv u$, y así

$$\overline{AB} + \overline{BD} \equiv \overline{AD} \equiv n\overline{AC} \equiv n\overline{AE} + n\overline{EC} < n\overline{AE} + \overline{BD},$$

luego $\overline{AB} < n\overline{AE}$. Esto implica que $E \in Y \subset \overline{CB}$, pero $E \in \overline{AC}$, y ambos segmentos no tienen más que a C como punto común, contradicción. ■

Nota Hemos dado esta prueba para mostrar que si suponemos el axioma de continuidad podemos definir la longitud de segmentos contando desde el principio con la propiedad de la división de segmentos, sin necesidad de considerar números diádicos, pero obtenemos una demostración mucho más sencilla como consecuencia del teorema siguiente:

Teorema 2.33 *Dado un segmento u y un número real $\alpha > 0$, existe un segmento v tal que $v \equiv \alpha u$.*

DEMOSTRACIÓN: Consideremos una recta graduada r con unidad u y sean

$$s_1 = \{P \in r \mid \alpha_P < \alpha\}, \quad s_2 = \{P \in r \mid \alpha_P \geq \alpha\}.$$

Es inmediato comprobar que estos conjuntos cumplen las hipótesis del axioma de continuidad, luego existe un punto $P \in r$ tal que s_2 es una semirrecta de origen P y $s_1 \cup \{P\}$ es la semirrecta complementaria (notemos que la definición de s_1 implica que no tiene máximo). Sea $\beta = \alpha_P$. Entonces $\beta \geq \alpha$ por definición de s_2 , pero si fuera $\beta > \alpha$, podríamos tomar un número racional diádico $\alpha < r < \beta$, y entonces $P_r \in s_2$ por definición de s_2 , pero entonces es un elemento de s_2 menor que P_β , que es el mínimo de s_2 , contradicción. ■

Así pues, el axioma de continuidad garantiza que todo número real es la longitud de un segmento respecto de cualquier unidad prefijada, de donde se sigue inmediatamente que la aplicación $P \mapsto \alpha_P$ considerada en el teorema 2.10 es biyectiva o, dicho de otro modo, que cada graduación de una recta r establece una biyección entre los puntos de r y los números reales, cuya inversa es $\alpha \mapsto P_\alpha$.

En otros términos, el axioma de continuidad (que sería más razonable llamar “axioma de completitud”) afirma que cada recta está “completa”, en el sentido de que tiene todos los puntos que puede tener admitiendo la propiedad de Arquímedes (pues si admitimos que una recta pueda tener puntos infinitamente alejados, entonces le “cabén” aún más puntos).

Teniendo esto en cuenta, la propiedad de división de segmentos es inmediata, pues, dado cualquier segmento u y un número natural $n \geq 1$, está definido $v \equiv (1/n)u$, que cumple $u \equiv nv$.

Teorema 2.34 *Sea L un ángulo arbitrario y sea $\pi \geq 1$ el número real dado por el teorema 2.24. Entonces, para todo número real $0 < \alpha \leq \pi$ existe un ángulo L' tal que $L' \equiv \alpha L$.*

DEMOSTRACIÓN: Podemos suponer que L no es un ángulo llano, pues si el teorema vale para un ángulo recto R (en cuyo caso $\pi = 2$) vale también para un ángulo llano (con $\pi = 1$), pues si $0 < \alpha \leq 1$, el ángulo $2\alpha R \equiv 2(\alpha R)$ cumple lo requerido.

También podemos suponer que $0 < \alpha < \pi$, pues un ángulo llano L' cumple $L' \equiv \pi L$.

Fijemos un número racional diádico $\alpha < r_0 < \pi$ y sea $L' = r_0 L \equiv \widehat{AOB}$. Entonces, a cada punto $P \in \overline{AB}$ distinto de A le podemos asignar el ángulo $L_P = \widehat{AOP} \leq L'$. En efecto, la semirrecta \overrightarrow{OP} está contenida en L' por 1.39, el teorema 1.41 nos da que $L_P \subset L'$ y entonces $L_P \leq L'$ por definición.

Más aún, si $0 < r \leq r_0$ es un número diádico, entonces $rL \leq r_0 L = L'$, luego existe un $P_r \in \overline{AB}$ tal que $L_{P_r} \equiv rL$.

Llamamos X al conjunto formado por A y los puntos $P \neq A$ de \overline{AB} tales que $L_P \leq rL$, para cierto número diádico $0 < r < \alpha$. En particular, esto incluye a todos los puntos P_r con $0 < r < \alpha$. Sea Y el conjunto de los puntos P en \overline{AB} que no están en X , lo cual incluye a todos los puntos P_r con $\alpha \leq r \leq r_0$.

Es fácil probar que X e Y cumplen las hipótesis de la variante para intervalos del axioma de continuidad, luego existe un punto $A - C - B$ de modo que $\overline{AC} \setminus \{C\} \subset X \subset \overline{AC}$. Sea $L_C \equiv \beta L \leq L'$, luego $0 < \beta \leq r_0$. Vamos a probar que $\beta = \alpha$.

Si $\beta < \alpha$, podemos tomar un número diádico $\beta < r < \alpha$, con lo que tenemos $L_C \equiv \beta L < rL \equiv L_{P_r}$. Esto implica que la semirrecta \overrightarrow{OC} esté contenida en L_{P_r} , lo que a su vez implica que $A - C - P_r$, pero $P_r \in X$, y esto es una contradicción.

Similarmente, si $\alpha < \beta$ podemos tomar un número diádico $\alpha < r < \beta$, y entonces $L_{P_r} \equiv rL < \beta L \equiv L_C$, de donde $A - P_r - C$, pero $P_r \in Y$, y de nuevo tenemos una contradicción. ■

Así pues, fijada una unidad de ángulos U para la que π sea la amplitud del ángulo llano, existen ángulos de todas las amplitudes posibles $0 < \alpha \leq \pi$ respecto de U como unidad de medida. En particular:

Teorema 2.35 (Propiedad de la división de ángulos) *Para todo ángulo L y todo número natural $n \geq 1$ existe un ángulo L' tal que $L \equiv nL'$.*

DEMOSTRACIÓN: Basta tomar $L' \equiv (1/n)L$, que está definido, pues se cumple $1/n \leq 1 \leq \pi$. ■

Vamos a probar que el axioma de continuidad implica también el axioma de las circunferencias, para lo cual demostraremos primero el teorema 1.90:

Demostración del teorema 1.90 Sea ω una circunferencia de centro O y radio r y sea s una recta en su mismo plano que pase por un punto interior A . Si $A = O$ la conclusión es trivial. En caso contrario sea B el pie de la perpendicular a s por O . El teorema 1.77 implica que B también es interior a ω , luego podemos suponer que $B = A$.

Fijemos una semirrecta s_0 de s respecto de A y llamemos X al conjunto de los puntos P de s que están en la semirrecta complementaria a s_0 o bien son interiores a ω . Sea Y el conjunto de los puntos de s que no están en X (es decir, que están en la semirrecta s_0 y son exteriores a ω o están en ω). El teorema 1.77 permite probar fácilmente que estos conjuntos están en las hipótesis del axioma

de continuidad, por lo que son semirrectas de origen un punto C . Basta probar que $C \in \omega$. Supongamos que $\overline{OC} < r$. Entonces podemos tomar $A - C - C'$ de modo que $\overline{CC'} < r - \overline{OC}$. El teorema 1.76 aplicado al triángulo $\overline{OCC'}$ nos da que $\overline{OC'} < \overline{OC} + \overline{CC'} < r$, luego $C' \in X$, lo cual contradice que C sea el extremo de X .

Si $\overline{OC} > r$ la situación es similar. Ahora podemos tomar $A - C' - C$ de modo que $\overline{C'C} < \overline{OC} - r$, y entonces $\overline{OC'} > \overline{OC} - \overline{C'C} > r$, con lo que $C' \in Y$, lo que contradice que C sea el extremo de Y . Por lo tanto $\overline{OC} \equiv r$. ■

Demostración del axioma de las circunferencias Sea ω una circunferencia de centro O y radio r , y sea ω' una circunferencia (en el mismo plano) de centro O' y radio r' . Supongamos que ω' pasa por un punto interior y por un punto exterior a ω .

Sea $r = \overline{OO'}$, sean P, Q los puntos de r que cumplen $\overline{OP} \equiv \overline{OQ} \equiv r$ y sean P', Q' los que cumplen $\overline{O'P'} \equiv \overline{O'Q'} \equiv r'$. Fijando en r el orden $\leq_{OO'}$, podemos suponer que $Q < O < P, P' < O' < Q'$.

Es claro que la intersección con r del círculo de frontera ω es \overline{QP} y el del círculo de frontera ω' es $\overline{P'Q'}$. Si estos segmentos son disjuntos, es fácil ver que los círculos son disjuntos, por lo que ω' no tiene puntos interiores a ω . Esto vale incluso si $P = P'$, pues entonces las circunferencias son tangentes en P y todos los puntos de ω' distintos de P son exteriores a ω .

Por lo tanto, los segmentos \overline{QP} y $\overline{P'Q'}$ se cortan, y no sólo en P . Tampoco puede suceder que uno esté contenido en el otro. Por ejemplo, si se cumple $P' \leq Q < O < P \leq Q'$, es claro que ω' no tiene puntos interiores de ω en r , y si $A \in \omega'$ fuera un punto interior de ω fuera de r , entonces

$$\overline{O'A} < \overline{O'O} + \overline{OA} < \overline{O'O} + \overline{OQ} \equiv \overline{O'Q} \leq \overline{O'P'} \equiv r',$$

en contradicción con que $A \in \omega'$. La otra opción es $Q \leq P' < O' < Q' \leq P$. Entonces ω' no tiene puntos exteriores a ω en r , y si $A \in \omega'$ es exterior a ω fuera de r , entonces

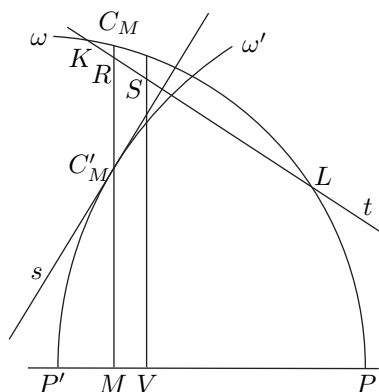
$$\overline{OA} < \overline{OO'} + \overline{O'A} < \overline{OO'} + \overline{O'Q'} \equiv \overline{OQ'} \leq \overline{OP} \equiv r,$$

en contradicción con que A es exterior a ω .

Así pues, tiene que ser $Q < P' < P < Q'$, con lo que P' es interior a ω y P es interior a ω' . Es claro entonces que todos los puntos de $\overline{P'P}$ salvo los extremos son interiores a ambas circunferencias. Fijemos un semiplano Σ del plano de ω respecto de la recta OO' . Para cada $A \in \overline{P'P}$ que no sea un extremo, la semirrecta perpendicular a $P'P$ por A contenida en Σ tiene por origen un punto interior a ambas circunferencias, luego las corta en puntos $C_A \in \omega$ y $C'_A \in \omega'$. Esto vale también para los extremos si entendemos que $C_P = P$ y $C'_{P'} = P'$. Basta ver que existe un punto $A \in \overline{P'P}$ tal que $C_A = C'_A$. Supongamos lo contrario.

Llamamos X al conjunto de todos los puntos U de $\overline{P'P}$ para los que existe un punto V tal que $P' \leq U < V < P$ con $\overline{VC'_V} < \overline{VC_V}$, más los puntos de la

semirrecta de origen P' que no contiene a P . Llamamos Y al conjunto de los puntos de PP' que no están en X . En particular contiene a todos los puntos de la semirrecta de origen P que no contiene a P' . Es inmediato que se cumplen las hipótesis del axioma de continuidad, por lo que X e Y son semirrectas de origen un punto M , que claramente tiene que estar en $\overline{P'P}$. Supongamos que cumple $\overline{MC'_M} < \overline{MC_M}$.



Sea s la tangente a ω' por C'_M , sea R un punto entre C'_M y C_M , sea t una recta que pase por R y corte a la recta $P'P$ en cualquier punto distinto de M . Entonces t corta a ω en dos puntos K y L , de modo que los puntos del segmento \overline{KL} distintos de sus extremos son interiores a ω . Sea S un punto de este segmento que esté en el mismo semiplano que P respecto a MR y que sea anterior a un eventual corte entre KL y la tangente s . Tracemos desde S la perpendicular a $P'P$, que cortará a esta recta en un punto V .

Como las rectas MR y VS tienen una perpendicular común ($P'P$) no pueden cortarse (o formarían un triángulo con dos ángulos rectos). Esto implica que V está entre M y P . Ahora notamos que el segmento \overline{RS} no corta a s , luego S está en el mismo semiplano que R respecto a s , es decir, en el semiplano opuesto a M , o sea, en el que no está ω' . Por lo tanto S es exterior a ω' y por consiguiente \overline{VS} corta a ω' antes de llegar a S , pero como S es interior a ω , resulta que \overline{VS} corta a ω después de llegar a S , o sea, $\overline{VC'_V} < \overline{VC_V}$. Por consiguiente cualquier punto entre M y V está en X , lo cual es imposible.

Notemos que, de hecho, si $M - U - V$, la perpendicular a $P'P$ por U no puede cortar a MR ni a VS (o formaría triángulos con dos ángulos rectos), luego por B4 tiene que cortar a \overline{MS} y a su vez a \overline{RS} en un punto S' que cumple el mismo criterio con el que hemos elegido a S , por lo que el mismo argumento muestra que $\overline{UC'_V} < \overline{UC_V}$.

Esto es importante porque ahora, invirtiendo los papeles de las dos circunferencias, llegamos a que si $\overline{MC_M} < \overline{MC'_M}$ existe un punto U entre P' y M de modo que todos los puntos entre U y M cumplen la misma desigualdad, pero entonces $U \in X$ (por ser anterior a M) y tendría que existir un V tal que $P' < U < V$ de modo que $\overline{VC'_V} < \overline{VC_V}$, pero necesariamente entonces $P' < U < M < V$, y aplicando a V el razonamiento inicial que hemos aplicado a M llegamos a que $V \in X$, lo que de nuevo es una contradicción. ■

2.6 El axioma de las paralelas

Vamos finalmente a explorar las consecuencias que resultan de añadir el axioma de las paralelas a los axiomas restantes que hemos tomado para la geometría absoluta.

Recordemos que dos rectas son paralelas si son iguales o bien están contenidas en un mismo plano y no tienen puntos en común, así como que dos planos son paralelos si son iguales o bien están contenidos en la misma variedad afín de dimensión 3 y no tienen puntos en común. Muy poco es lo que podemos decir sobre rectas y planos paralelos sin recurrir al axioma de las paralelas. No obstante, podemos demostrar su existencia:

Teorema 2.36 *Por cada punto pasa una paralela a una recta dada.*

DEMOSTRACIÓN: Dada una recta r y un punto P , podemos suponer que P es exterior a r , ya que en caso contrario sirve r como paralela. Sea s la perpendicular a r por P y sea t la perpendicular a s por P contenida en el plano de r y s . Las rectas r y t han de ser paralelas, pues si se cortaran formarían un triángulo con dos ángulos rectos. ■

Teorema 2.37 *Por cada punto pasa un plano paralelo a un plano dado.*

DEMOSTRACIÓN: Sea Π un plano y P un punto que podemos suponer exterior. Supongamos en primer lugar los axiomas de tridimensionalidad. Sea r la recta perpendicular a Π por P y sea Π_1 el plano perpendicular a r por P . Entonces Π y Π_1 son planos paralelos, pues si tuvieran un punto en común Q , el triángulo formado por P , Q y la intersección de r con Π tendría dos ángulos rectos. Sin los axiomas de tridimensionalidad aplicamos el razonamiento precedente a la variedad afín $A^3(\Pi, P)$, y así obtenemos un plano paralelo. ■

Definición 2.38 Una *geometría de Hilbert euclídea* es una geometría de Hilbert que satisfaga los axiomas A, B, C, junto con el axioma siguiente:

Axioma de las paralelas *Por cada punto exterior a una recta pasa una única paralela.*

Ya hemos observado que la existencia de paralelas (únicas) por puntos de una recta es trivial por definición, por lo que al admitir este axioma tenemos de hecho la existencia de una única paralela a cada recta por cada punto.

En principio no incluimos la propiedad de Arquímedes en la definición de geometría euclídea, y los resultados de esta sección no la requieren. Tampoco suponemos los axiomas de tridimensionalidad.

En el capítulo I hemos visto que el axioma de las paralelas puede reemplazar al axioma de incidencia A2, y hemos probado el teorema 1.37, que afirma que el paralelismo es una relación de equivalencia. El análogo para planos del axioma de las paralelas se demuestra fácilmente:

Teorema 2.39 *Por cada punto pasa un único plano paralelo a un plano dado.*

DEMOSTRACIÓN: Sea Π un plano y P un punto que podemos suponer exterior. Ya hemos probado que existe un plano paralelo. Si hay dos, digamos Π_1 y Π_2 , ambos tienen que estar contenidos en la variedad tridimensional $A^3(\Pi, P)$, luego se cortan en una recta s . Consideremos un plano Π' que pase por P , por un punto $P' \in \Pi_1 \setminus s$ y por un punto de Π . Entonces $\Pi' \cap \Pi$ es una recta r paralela tanto a $r_1 = \Pi' \cap \Pi_1$ como a $r_2 = \Pi' \cap \Pi_2$, que son dos rectas distintas (ya que $P' \in r_1 \setminus r_2$) y ambas pasan por P , contradicción. ■

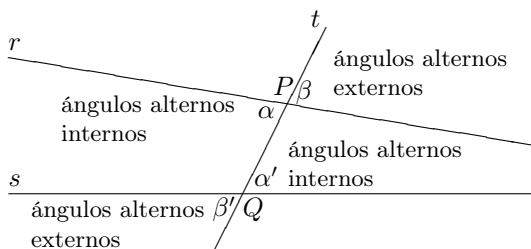
Más fácilmente se demuestra:

Teorema 2.40 *Dos planos paralelos a un tercero son paralelos.*

El teorema 1.37 equivale a que si dos rectas son paralelas, toda recta contenida en su plano común que corte a una de ellas tiene que cortar también a la otra.

Consideremos una figura formada por dos rectas r y s que cortan a una recta t en puntos distintos P y Q . De los ocho ángulos determinados por los dos cortes, llamaremos *ángulos internos* de la figura a los dos que tienen vértice P y que están contenidos en el semiplano de r que contiene a Q , así como a los dos que tienen vértice Q y están contenidos en el semiplano de s que contiene a P . Hay, pues, cuatro ángulos internos, dos con vértice P , uno en cada semiplano respecto a t , y dos con vértice Q , también uno en cada semiplano respecto a t . Los cuatro ángulos restantes son *externos*. También hay dos en cada vértice y cada uno en un semiplano distinto respecto a t . Diremos que dos ángulos son *alternos* si tienen vértices distintos y están contenidos en semiplanos distintos respecto a t .

En la figura siguiente α y α' son ángulos alternos internos, mientras que β y β' son ángulos alternos externos.



Teorema 2.41 *Sean r y s dos rectas que corten a una recta t en dos puntos P y Q . Entonces r y s son paralelas si y sólo si los ángulos alternos internos que determinan son iguales.*

DEMOSTRACIÓN: Si los ángulos alternos internos son iguales, entonces dos ángulos internos no alternos son suplementarios, luego no pueden ser los ángulos de un triángulo, cosa que ocurriría si las rectas r y s se cortaran.

Supongamos ahora que r y s son paralelas. Tomamos uno de los ángulos internos de vértice P y lo transportamos a un ángulo de vértice Q , con un lado

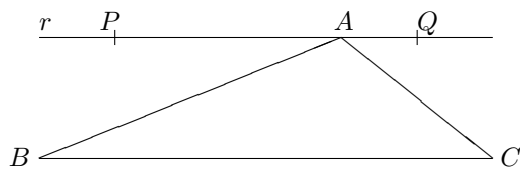
igual a \overrightarrow{QP} y con el otro lado en el semiplano opuesto al ángulo de partida respecto de t . La prolongación de dicho lado es una recta s' de modo que t determina ángulos alternos internos iguales en r y s' . Por la parte ya probada r y s' son rectas paralelas, y como s' pasa por Q , por el axioma de las paralelas ha de ser $s = s'$, luego ciertamente los ángulos alternos internos son iguales. ■

Así pues cuando una recta t corta a dos rectas paralelas r y s se forman ocho ángulos divididos en dos grupos de cuatro que son iguales entre sí (porque cada ángulo interno es opuesto por el vértice a uno externo). Los de un grupo son los suplementarios de los del otro grupo. En particular, si una recta corta a dos paralelas y es perpendicular a una, también lo es a la otra (pues en tal caso los ocho ángulos son iguales, luego son rectos).

Como aplicación inmediata tenemos el hecho siguiente:

Teorema 2.42 *La suma de los ángulos de un triángulo es igual a un ángulo llano.*

DEMOSTRACIÓN: Dado un triángulo cualquiera \widehat{ABC} , trazamos la paralela r al lado BC por el punto A . Obviamente r deja a B y C en un mismo semiplano.



Las semirrectas \overrightarrow{AB} y \overrightarrow{AC} dividen a dicho semiplano en tres ángulos cuya suma es π . Uno de ellos es \hat{A} , y los otros dos son iguales a \hat{B} y \hat{C} por el teorema anterior.

Con más detalle: tomemos puntos P y Q a ambos lados de A en r . Entonces los ángulos \widehat{PAB} y \widehat{PAC} están contenidos en el mismo semiplano respecto a r (pues r no corta a \overline{BC}) luego uno está contenido en el otro. Supongamos por ejemplo que \widehat{PAB} es el menor. Entonces \overrightarrow{AB} está contenida en \widehat{PAC} , luego corta a \overline{PC} , luego P y C están en semiplanos distintos respecto a AB , de donde se sigue que \widehat{PAB} y \widehat{ABC} son alternos internos. Así pues, $\widehat{PAB} \equiv \hat{B}$, luego $\widehat{PAC} \equiv \widehat{PAB} + \widehat{BAC} \equiv \hat{A} + \hat{B}$. Por otra parte, \widehat{QAC} es adyacente a \widehat{PAC} y de aquí que B y Q estén en semiplanos distintos respecto a AC , con lo que \widehat{QAC} y \widehat{ACB} son alternos internos. En total queda que $\hat{A} + \hat{B} + \hat{C} = \widehat{PAC} + \widehat{CAQ}$, que es un ángulo llano. ■

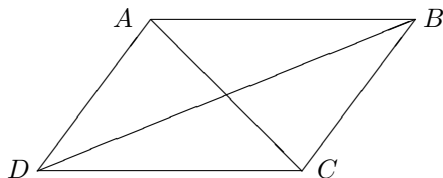
En particular, si dos triángulos tienen dos ángulos iguales, de hecho tienen los tres ángulos iguales, pues el tercero es el suplementario de la suma de los otros dos, luego el criterio de igualdad ALA puede aplicarse con dos ángulos cualesquiera, no necesariamente los contiguos al lado.

Ejercicio: Probar que por tres puntos no alineados pasa una única circunferencia.

Definición 2.43 Sean A, B, C, D cuatro puntos coplanares no colineales tres a tres tales que la recta AB sea paralela a CD y la recta AD sea paralela a BC . Entonces C y D están en el mismo semiplano respecto a AB , A y B están en el mismo semiplano respecto a CD , A y D están en el mismo semiplano respecto a BC y B y C están en el mismo semiplano respecto a AD . A la intersección de estos cuatro semiplanos se le llama *paralelogramo* (gr. ‘de líneas paralelas’) de vértices A, B, C y D .

Los segmentos \overline{AB} , \overline{BC} , \overline{CD} y \overline{DA} se llaman *lados* del paralelogramo. Dos lados son *contiguos* si tienen un vértice en común y son *opuestos* en caso contrario. La unión de todos ellos constituye su *frontera*.

Los ángulos \widehat{DAB} , \widehat{ABC} , \widehat{BCD} y \widehat{CDA} se llaman *ángulos* del paralelogramo. Los segmentos \overline{AC} y \overline{BD} se llaman *diagonales* del paralelogramo. Dos vértices son *contiguos* u *opuestos* según si son extremos de un lado o de una diagonal.



Los paralelogramos son intersecciones de semiplanos, luego son convexos. En particular las diagonales de un paralelogramo están contenidas en él. No es difícil probar que un paralelogramo determina sus vértices.

Teorema 2.44 *Los lados y los ángulos opuestos de un paralelogramo son iguales, los ángulos contiguos son suplementarios, las diagonales se cortan por su punto medio.*

DEMOSTRACIÓN: Con la notación de la definición anterior, tenemos que A y D están en el mismo semiplano respecto a BC , luego si D y B estuvieran en el mismo semiplano respecto a AC tendríamos que D pertenecería al ángulo \widehat{CAB} , luego AD cortaría al segmento BC , cuando en realidad son paralelos. Así pues, B y D están en semiplanos distintos respecto a AC , por lo que \overline{BD} corta a AC . Similarmente llegamos a que \overline{AC} ha de cortar a la recta BD , y como el punto de corte ha de ser el mismo en ambos casos, concluimos que las diagonales \overline{AC} y \overline{BD} se cortan en un punto P .

Es claro que dos ángulos contiguos, por ejemplo \hat{A} y \hat{B} son ángulos internos respecto a la figura formada por las rectas paralelas AD y BC cortadas por AB , pero no son alternos, sino que ambos están en el mismo semiplano respecto a AB . Por lo tanto \hat{A} y un ángulo adyacente a \hat{B} sí son ángulos alternos internos, luego \hat{A} y \hat{B} son suplementarios. De aquí que dos ángulos opuestos sean iguales, pues tienen un suplementario común.

Puesto que A y C están en semiplanos distintos respecto a BD , los ángulos \widehat{ABD} y \widehat{BDC} son alternos internos, luego son iguales. Igualmente se cumple que $\widehat{ADB} \equiv \widehat{DBC}$, luego $\widehat{ADB} \equiv \widehat{CBD}$, lo que en particular implica que los lados opuestos son iguales.

Es fácil ver también que $\widehat{PAB} \equiv \widehat{PCD}$, de donde se sigue que $\overline{PB} \equiv \overline{PD}$ y $\overline{PA} \equiv \overline{PC}$, luego P es el punto medio de las diagonales. ■

Definición 2.45 Un *rombo* (gr. ‘alternado’) es un paralelogramo cuyos cuatro lados son iguales. Un *rectángulo* es un paralelogramo cuyos cuatro ángulos son iguales (luego los cuatro son rectos). Un *cuadrado* es un paralelogramo con los cuatro lados y los cuatro ángulos iguales.

Ejercicio: Probar que las diagonales de un rombo son perpendiculares, y son las bisectrices de los ángulos que unen.

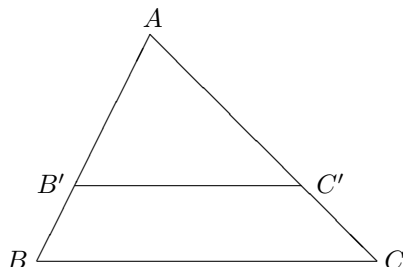
2.7 Los teoremas de Tales y Pitágoras

En esta sección trabajamos en una geometría euclídea con la propiedad de Arquímedes.

Definición 2.46 Dos triángulos \widehat{ABC} y $\widehat{A'B'C'}$ son *semejantes* si sus lados son proporcionales dos a dos, es decir, si

$$\frac{\overline{AB}}{\overline{A'B'}} = \frac{\overline{AC}}{\overline{A'C'}} = \frac{\overline{BC}}{\overline{B'C'}}.$$

Diremos que dos triángulos \widehat{ABC} y $\widehat{A'B'C'}$ están en *posición de Tales* si, nombrando los vértices adecuadamente, $A = A'$, B' está en \overline{AB} y $B'C'$ es paralela a BC (con lo que también C' está en \overline{AC}).



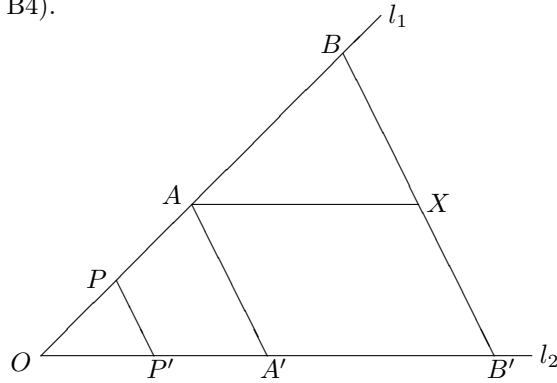
Es fácil probar el teorema siguiente:

Teorema 2.47 Si dos triángulos \widehat{ABC} y $\widehat{A'B'C'}$ están en posición de Tales entonces $\hat{A} = \hat{A}'$, $\hat{B} = \hat{B}'$ y $\hat{C} = \hat{C}'$. Recíprocamente, si \widehat{ABC} y $\widehat{A'B'C'}$ cumplen estas igualdades existe un triángulo $\widehat{A''B''C''}$ semejante a $\widehat{A'B'C'}$ de modo que \widehat{ABC} y $\widehat{A''B''C''}$ están en posición de Tales.

Con ayuda de este hecho ya podemos probar el resultado básico:

Teorema 2.48 Sean l_1 y l_2 dos semirrectas no alineadas de origen O . Sean P y P' puntos de l_1 y l_2 respectivamente distintos de O . Sean A y B dos puntos cualesquiera de l_1 distintos entre sí y distintos de O . Sean A' y B' los puntos donde las rectas paralelas a PP' por A y B , respectivamente, cortan a l_2 . Entonces la clase de congruencia de $\overline{A'B'}$ depende de la clase de congruencia de \overline{AB} , pero no de la elección de los puntos A y B en l_1 .

DEMOSTRACIÓN: Podemos suponer que A está más cerca de O que B . Notemos que la recta paralela a PP' por A ha de cortar efectivamente a l_2 , o de lo contrario serían paralelas, luego PP' también sería paralela a l_2 , lo cual es absurdo. Tracemos por A la paralela a l_2 , que cortará a $\overline{BB'}$ en un punto X (por el axioma B4).



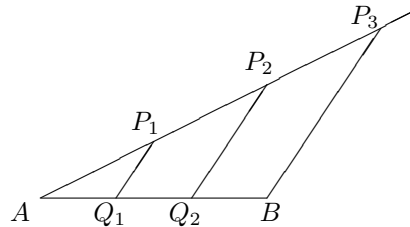
Los triángulos $\widehat{BOB'}$ y \widehat{BAX} están en posición de Tales, luego se cumple $\widehat{BAX} = \widehat{BOB'} = \widehat{POP'}$. Lo mismo sucede con $\widehat{OPP'}$ y $\widehat{OBB'}$, luego también $\widehat{ABX} = \widehat{OBB'} = \widehat{OPP'}$.

Esto significa que los ángulos del triángulo \widehat{ABX} son independientes de la elección de A y B . Por lo tanto, si hacemos otras elecciones distintas conservando la clase de congruencia de \overline{AB} obtendremos triángulos iguales, luego los lados \overline{AX} también serán iguales, pero $\overline{AX} = \overline{A'B'}$ porque son los lados opuestos de un paralelogramo. Así pues, la clase de congruencia de $\overline{A'B'}$ sólo depende de la clase de congruencia de \overline{AB} . ■

Como primera aplicación demostramos lo siguiente:

Teorema 2.49 (Propiedad de la división de segmentos) *Todo segmento se puede dividir en cualquier número prefijado de partes iguales.*

DEMOSTRACIÓN: La figura muestra el proceso de división de un segmento en tres partes iguales, pero el método es general:



Consideremos cualquier semirrecta de origen A no colineal con AB , elegimos cualquier punto P_1 en ella y lo transportamos las veces que queramos hasta determinar (en este caso dos) puntos P_2 y P_3 de modo que $\overline{AP_1} \equiv \overline{P_1P_2} \equiv \overline{P_2P_3}$. Seguidamente consideramos la recta BP_3 y trazamos las paralelas a esta recta que pasan por P_1 y P_2 , que cortarán a \overline{AB} en puntos Q_1 y Q_2 , respectivamente. Por el teorema anterior $\overline{AQ_1} \equiv \overline{Q_1Q_2} \equiv \overline{Q_2B}$. ■

Ya estamos en condiciones de demostrar los teoremas fundamentales de la geometría euclídea:

Teorema 2.50 (Teorema de Tales) *Si dos triángulos están en posición de Tales entonces son semejantes.*

DEMOSTRACIÓN: Sean \widehat{ABC} y $\widehat{A'B'C'}$ triángulos en posición de Tales, digamos con el vértice A en común. Para cada segmento s podemos tomar dos puntos P y Q en \overline{AB} tales que $\overline{PQ} \equiv s$, transportarlos a \overline{AC} mediante rectas paralelas a BC de acuerdo con el teorema 2.48 y obtener así un segmento cuya longitud sólo depende de la longitud de s . Llamemos $m(s)$ a esta longitud. Es obvio que se cumplen las hipótesis del teorema 2.22, luego en particular, y teniendo en cuenta que \overline{AB} se transforma en \overline{AC} y $\overline{A'B'}$ se transforma en $\overline{A'C'}$, concluimos que

$$\frac{\overline{AB}}{\overline{A'B'}} = \frac{\overline{AC}}{\overline{A'C'}}.$$

Ahora bien, los dos triángulos dados tienen los ángulos iguales, luego (pasando a otros congruentes) podemos ponerlos en posición de Tales con otro vértice en común, por ejemplo B , y entonces obtenemos la igualdad que nos falta para concluir que son semejantes. ■

Teorema 2.51 *Dos triángulos son semejantes si y sólo si tienen los ángulos iguales.*

DEMOSTRACIÓN: Si dos triángulos tienen los ángulos iguales, entonces pueden ponerse en posición de Tales, luego son semejantes. Supongamos ahora que \widehat{ABC} y $\widehat{A'B'C'}$ son dos triángulos semejantes. Es claro que podemos construir un triángulo $\widehat{A''B''C''}$ en posición de Tales respecto a \widehat{ABC} y de modo que $\overline{A'B'} \equiv \overline{A''B''}$. Entonces los triángulos $\widehat{A'B'C'}$ y $\widehat{A''B''C''}$ son semejantes y tienen un lado igual. De la definición de semejanza se sigue entonces que tienen los tres lados iguales, luego son iguales. ■

De aquí se sigue una variante:

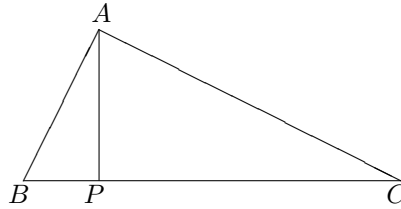
Teorema 2.52 *Dos triángulos son semejantes si sólo si tienen un ángulo igual y sus lados adyacentes proporcionales.*

DEMOSTRACIÓN: Sustituyendo uno de los triángulos por otro congruente (y por lo tanto semejante) podemos suponer que ambos tienen un ángulo coincidente. Sean, pues \widehat{ABC} y $\widehat{A'B'C'}$, donde las semirrectas \overline{AB} y $\overline{A'C'}$ coinciden

respectivamente con $\overrightarrow{AB'}$ y $\overrightarrow{AC''}$. Trazamos la paralela a BC que pasa por B' , la cual cortará a AC en un punto C''' . Por el teorema de Tales, el triángulo $\widehat{AB'C''}$ es semejante a \widehat{ABC} luego a $\widehat{AB'C'}$, con lo que sus lados son proporcionales, pero tienen un lado coincidente, de hecho todos sus lados han de ser iguales. En particular $\overline{AC'} \equiv \overline{AC''}$, luego $C' = C''$ y así $\widehat{AB'C'}$ es semejante a \widehat{ABC} . El recíproco es obvio. ■

Teorema 2.53 (Teorema de Pitágoras) *En un triángulo rectángulo, el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los catetos.*

DEMOSTRACIÓN: Sea \widehat{ABC} un triángulo rectángulo, donde \hat{A} es el ángulo recto. Sea P el pie de la perpendicular a BC por A . No puede ocurrir que P coincida con B o con C , o de lo contrario el triángulo tendría dos ángulos rectos. Además P ha de estar en el segmento \overline{BC} , pues si, por ejemplo, B estuviera entre P y C los ángulos \widehat{PBA} y \widehat{CBA} serían adyacentes, pero el primero forma parte del triángulo \widehat{PBA} , que tiene un ángulo recto en \hat{P} , luego \widehat{PBA} sería agudo, y \widehat{CBA} sería obtuso, lo cual es imposible. Por lo tanto P está ciertamente en \overline{BC} .



El triángulo rectángulo \widehat{ABP} comparte un ángulo agudo con el triángulo también rectángulo \widehat{CBA} , luego tienen dos—y por consiguiente tres—ángulos iguales, es decir, son semejantes. Así pues

$$\frac{\overline{AB}}{\overline{BP}} = \frac{\overline{BC}}{\overline{AB}},$$

y por consideraciones análogas sobre el triángulo \widehat{ACP} concluimos que

$$\frac{\overline{AC}}{\overline{CP}} = \frac{\overline{BC}}{\overline{AC}}.$$

Por lo tanto

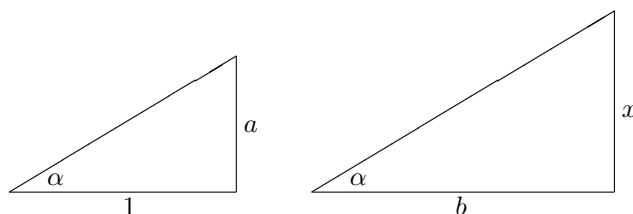
$$\overline{AB}^2 + \overline{AC}^2 = \overline{BC} \overline{BP} + \overline{BC} \overline{CP} = \overline{BC} (\overline{BP} + \overline{CP}) = \overline{BC}^2. \quad \blacksquare$$

De aquí se sigue fácilmente que la diagonal de un cuadrado mide $\sqrt{2}$ veces su lado, pero no existe ningún número racional cuyo cuadrado sea 2, luego la diagonal no guarda una proporción racional con el lado. Así fue como los griegos

descubrieron que una de las figuras geométricas más sencillas, el cuadrado, determina dos segmentos mutuamente inconmensurables, es decir, la existencia de números irracionales, y esto explica por qué hemos tenido que introducir los números reales a la hora de medir segmentos. Ahora bien, en ausencia del axioma de continuidad no necesitamos todos los números reales para medir segmentos. Ahora podemos decir algo más al respecto:

Teorema 2.54 *El conjunto \mathbb{R}_G de los números reales que tienen asociados puntos en una recta graduada es un subcuerpo de \mathbb{R} , el mismo para todas las rectas.*

DEMOSTRACIÓN: Fijemos una recta graduada. Por el teorema 2.17 y las observaciones posteriores sabemos que la suma y la resta de números de \mathbb{R}_G está en \mathbb{R}_G . Ahora, dado cualquier par de segmentos de longitudes a y b , podemos construir dos triángulos como indica la figura:



El primer triángulo rectángulo lo construimos fijando los catetos, lo cual determina la hipotenusa y, en particular, el ángulo α . El segundo lo construimos transportando el ángulo α y fijando el cateto de longitud b . Entonces ambos triángulos tienen los mismos ángulos, luego son semejantes, luego $b/1 = x/a$, es decir, $x = ab$. Esto prueba que el producto de elementos positivos de \mathbb{R}_G está en \mathbb{R}_G , y de aquí se generaliza trivialmente a elementos arbitrarios.

Ahora repetimos la construcción intercambiando la a y el 1 en el primer triángulo y poniendo $b = 1$ en el segundo. El resultado es $x/1 = 1/a$, luego $x = a^{-1}$. Esto prueba que el inverso de un número positivo de \mathbb{R}_G está en \mathbb{R}_G , y claramente lo mismo vale para números negativos.

Con esto tenemos que \mathbb{R}_G es un subcuerpo de \mathbb{R} . Falta probar que no depende de la elección de la recta graduada. Ahora bien, si tomamos otra recta graduada cuyo cuerpo asociado es \mathbb{R}'_G y su unidad mide $u \in \mathbb{R}_G$ respecto de la primera, entonces cada segmento que mida $a \in \mathbb{R}'_G$ respecto a la segunda recta medirá $b = ua \in \mathbb{R}_G$ respecto a la primera, pero entonces $a = b/u \in \mathbb{R}_G$. Esto prueba que $\mathbb{R}'_G \subset \mathbb{R}_G$, y por simetría se da también la inclusión opuesta. ■

Es importante tener claro que \mathbb{R}_G es el mismo cuerpo para todas las rectas graduadas de una misma geometría euclídea con la propiedad de Arquímedes, pero nada impide en principio que existan distintas geometrías euclídeas (es decir, distintos espacios que cumplan los axiomas de la geometría euclídea) que determinen distintos subcuerpos \mathbb{R}_G de \mathbb{R} .

No obstante, \mathbb{R}_G no puede ser cualquier subcuerpo de \mathbb{R} , pues el teorema de Pitágoras implica que debe cumplir una condición adicional:

Definición 2.55 Un cuerpo K es *pitagórico* si la suma de dos cuadrados de K es de nuevo un cuadrado en K .

Por ejemplo, \mathbb{R} es pitagórico porque toda suma de cuadrados es ≥ 0 y todo número real ≥ 0 tiene raíz cuadrada (véase [An 1.15]). En cambio, \mathbb{Q} no es pitagórico, pues $2 = 1^2 + 1^2$ no tiene raíz cuadrada en \mathbb{Q} . Es inmediato que si K es un cuerpo pitagórico, entonces toda suma finita de cuadrados es un cuadrado.

Teorema 2.56 *El cuerpo \mathbb{R}_G es pitagórico.*

DEMOSTRACIÓN: Tomemos $\alpha, \beta \in \mathbb{R}_G$ y veamos que $\gamma = \alpha^2 + \beta^2$ tiene raíz cuadrada en \mathbb{R}_G . No perdemos generalidad si suponemos $\alpha, \beta > 0$. Entonces podemos construir un triángulo rectángulo cuyos catetos midan α y β , con lo que su hipotenusa medirá $\sqrt{\gamma} \in \mathbb{R}_G$. ■

Por ejemplo, ahora podemos concluir que si n es un número natural, entonces $\sqrt{n} \in \mathbb{R}_G$. Obviamente es cierto para $n = 0$ y si $\sqrt{n} \in \mathbb{R}_G$, entonces tenemos que $n + 1 = \sqrt{n}^2 + 1^2$ es una suma de cuadrados en \mathbb{R}_G , luego $\sqrt{n+1} \in \mathbb{R}_G$.

2.8 Las razones trigonométricas

Sabemos que los ángulos y los lados de un triángulo están sometidos a muchas relaciones. Por ejemplo, si conocemos los ángulos de un triángulo y uno de sus lados, los otros dos lados están completamente determinados, pero en principio no sabemos cómo calcularlos a partir de estos datos. Este tipo de problemas constituye el objeto de la *trigonometría* (gr. ‘medida de triángulos’), aunque ahora no profundizaremos mucho en ella, pues más adelante estaremos en condiciones de hacerlo con más fluidez.

Definición 2.57 Sea L un ángulo agudo de vértice O , sea A un punto arbitrario en uno de sus lados y sea B el pie de la perpendicular al lado opuesto por A .

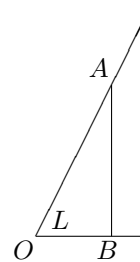
Entonces el triángulo \widehat{AOB} tiene un ángulo recto y otro igual a L , luego sus ángulos son independientes de la elección de A e incluso del lado en que lo tomamos (más aún, sólo dependen de la clase de congruencia de L).

Si tomamos puntos distintos obtendremos triángulos semejantes, luego podemos definir el *seno* y el *coseno* de L como las razones

$$\operatorname{sen} L = \frac{\overline{AB}}{\overline{OA}}, \quad \operatorname{cos} L = \frac{\overline{OB}}{\overline{OA}}.$$

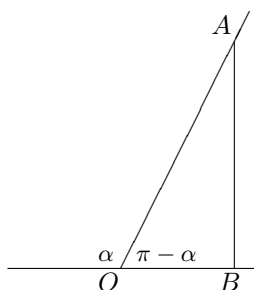
Puesto que el seno y el coseno dependen sólo de las clases de congruencia de los ángulos, podemos definir el seno y el coseno de un número real positivo $0 < \alpha < \pi/2$ como el seno y el coseno de los ángulos de amplitud correspondiente.

Volviendo a la figura anterior, observamos que podemos tomar el punto A a una unidad de distancia de O , y entonces el seno y el coseno de L son



simplemente las longitudes de \overline{AB} y \overline{OB} . El teorema de Pitágoras nos da entonces la que llamaremos *relación fundamental*:

$$\operatorname{sen}^2 \alpha + \operatorname{cos}^2 \alpha = 1.$$



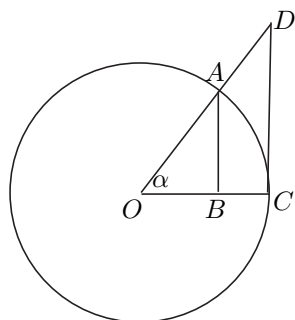
Las definiciones de seno y coseno valen tal cual para ángulos obtusos, pero en este caso convendremos en que el coseno tiene signo negativo. Es claro entonces que se cumplen las relaciones:

$$\operatorname{sen} \alpha = \operatorname{sen}(\pi - \alpha), \quad \operatorname{cos} \alpha = -\operatorname{cos}(\pi - \alpha).$$

Además se sigue cumpliendo la relación fundamental.

Observemos que el seno y el coseno pueden caracterizarse del modo siguiente: dado un ángulo L , si tomamos un punto en uno de sus lados a una distancia r del vértice, su proyección sobre el otro se encuentra a una distancia $r \operatorname{cos} L$ del vértice (donde el signo es negativo si el punto de corte está fuera del lado opuesto) y la distancia entre el punto y su proyección es $r \operatorname{sen} L$. Ahora bien, todo esto tiene sentido también si el ángulo es recto, en cuyo caso la proyección del punto escogido es el propio vértice del ángulo, de modo que todo lo dicho se sigue cumpliendo si definimos

$$\operatorname{sen} \frac{\pi}{2} = 1, \quad \operatorname{cos} \frac{\pi}{2} = 0.$$



Es útil considerar también la *tangente* de un ángulo α dada por

$$\tan \alpha = \frac{\operatorname{sen} \alpha}{\operatorname{cos} \alpha},$$

que está definida para todos los ángulos excepto para $\alpha = \pi/2$, porque el coseno de un ángulo recto es nulo. Para un ángulo agudo, si la circunferencia de la figura tiene radio 1, tenemos que

$$\operatorname{sen} \alpha = \overline{AB} = \frac{\overline{CD}}{\overline{OD}}, \quad \operatorname{cos} \alpha = \overline{OB} = \frac{\overline{OC}}{\overline{OD}}, \quad \tan \alpha = \frac{\overline{AB}}{\overline{OB}} = \overline{CD}.$$

Para ángulos obtusos se cumple trivialmente la relación

$$\tan(\pi - \alpha) = -\tan \alpha,$$

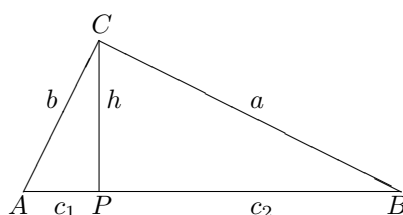
porque el coseno cambia de signo.

La relación entre las razones trigonométricas y los problemas sobre triángulos viene dada por el teorema del coseno y el teorema de los senos, que demostramos a continuación. El teorema del coseno es una generalización del teorema de Pitágoras:

Teorema 2.58 (Teorema del coseno) *Todo triángulo \widehat{ABC} verifica la relación*

$$a^2 = b^2 + c^2 - 2bc \cos \hat{A}.$$

DEMOSTRACIÓN: Sea P el pie de la altura del triángulo desde C (que puede o no estar en \overline{AB}). Podemos suponer que el triángulo no es rectángulo, pues si no el teorema se reduce al de Pitágoras. Entonces P es distinto de B y C . Sean c_1 , c_2 y h las longitudes indicadas en la figura.



Por el teorema de Pitágoras se cumple

$$a^2 = h^2 + c_2^2, \quad b^2 = h^2 + c_1^2,$$

luego $a^2 = b^2 - c_1^2 + c_2^2$.

Por otra parte, $c = c_2 \pm c_1$, donde el signo es positivo si A es agudo y negativo si es obtuso. Así, $c_2^2 = (c \mp c_1)^2 = c^2 + c_1^2 \mp 2cc_1$, con lo que $a^2 = b^2 + c^2 \mp 2cc_1$. Ahora bien, $c_1 = \pm b \cos \hat{A}$, luego concluimos que $a^2 = b^2 + c^2 - 2bc \cos \hat{A}$. ■

Para probar el teorema de los senos recordamos que la mediatriz de un segmento es la perpendicular por su punto medio, así como que está formada por los puntos que equidistan de sus extremos. Las mediatrices de dos lados de un triángulo \widehat{ABC} , por ejemplo, \overline{AB} y \overline{BC} se cortan en un punto O que equidista de A y B por una parte y de B y C por otra, luego equidista de los tres vértices, lo que implica que también se encuentra sobre la tercera mediatriz.

Definición 2.59 Se llama *circuncentro* de un triángulo a la intersección O de las mediatrices de sus lados. La distancia R del punto O a cualquiera de los vértices se llama *circunradio* del triángulo, la circunferencia de centro O y radio R se llama *circunferencia circunscrita* al triángulo.

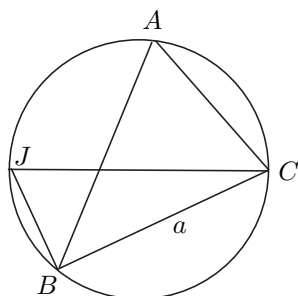
Claramente la circunferencia circunscrita a un triángulo es la única que pasa por sus vértices.

Teorema 2.60 (Teorema de los senos) *Si R es el circunradio de un triángulo \widehat{ABC} , entonces se cumple*

$$\frac{a}{\sin \hat{A}} = \frac{b}{\sin \hat{B}} = \frac{c}{\sin \hat{C}} = 2R.$$

DEMOSTRACIÓN: Sea ω la circunferencia circunscrita del triángulo. Sea J el punto diametralmente opuesto a C . Si $J = B$ entonces el ángulo \hat{A} abarca un semicírculo, luego es un ángulo recto, y trivialmente se cumple

$$\frac{a}{\text{sen } \hat{A}} = \frac{2R}{1} = 2R.$$



Si $B \neq J$ entonces los ángulos \hat{J} y \hat{A} abarcan arcos que comparten la cuerda \overline{BC} . Esto implica que los ángulos son iguales o suplementarios, luego tienen el mismo seno. Por otra parte el ángulo \widehat{JBC} es recto, pues abarca un semicírculo. Así pues

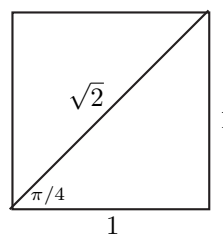
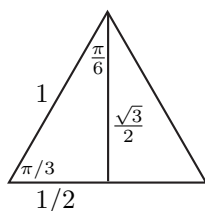
$$\frac{a}{\text{sen } \hat{A}} = \frac{a}{\text{sen } \hat{J}} = \frac{a}{a/2R} = 2R.$$

■

Notemos que el teorema de los senos refina al teorema de Tales, pues dos triángulos en posición de Tales tienen sus ángulos iguales, y el teorema de los senos nos da entonces que sus lados son proporcionales.

Algunas razones trigonométricas Analizando un triángulo equilátero y un cuadrado podemos deducir fácilmente las razones trigonométricas de algunos ángulos:

α	$\text{sen } \alpha$	$\text{cos } \alpha$	$\text{tan } \alpha$
$\pi/6$	$1/2$	$\sqrt{3}/2$	$1/\sqrt{3}$
$\pi/4$	$\sqrt{2}/2$	$\sqrt{2}/2$	1
$\pi/3$	$\sqrt{3}/2$	$1/2$	$\sqrt{3}$
$\pi/2$	1	0	$-$



■

Capítulo III

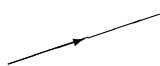
La geometría analítica

En este capítulo mostraremos cómo la geometría euclídea que hemos presentado en los dos capítulos precedentes se puede expresar en términos puramente algebraicos. Este enfoque algebraico de la geometría es lo que se conoce como *geometría analítica*. En el apéndice A de [IGE] presentamos una versión rudimentaria de la geometría analítica plana (y en el apéndice A de [IC] la generalizamos al caso tridimensional). En ambos casos introdujimos la distinción entre puntos y vectores de forma muy superficial, de modo que un par (o una terna) de números reales podía ser considerada como un punto o como un vector según el contexto. Aquí, en cambio, vamos a establecer una diferencia muy precisa entre los conceptos de punto y vector. En la primera sección los introduciremos en el contexto de la geometría euclídea tal y como la hemos presentado en los capítulos anteriores y en la sección siguiente introduciremos la estructura algebraica de espacio afín que toma los conceptos de “punto” y “vector” como conceptos fundamentales para una axiomatización algebraica de la geometría.

3.1 Vectores

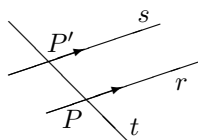
En esta sección consideramos una geometría euclídea arquimediana en el sentido de la definición 2.38. No necesitamos suponer el axioma de continuidad, de modo que llamaremos R al cuerpo \mathbb{R}_G dado por el teorema 2.54, es decir, el cuerpo de todos los números reales que tienen asignado un punto en cualquier recta graduada de un espacio euclídeo, cuyos elementos positivos son las longitudes de los segmentos. El axioma de continuidad equivale a que $R = \mathbb{R}$.

Definición 3.1 Una *orientación* en una recta es cualquiera de las dos relaciones de orden $<_{AB}$ mutuamente inversas determinadas por dos cualesquiera de sus puntos. Una *recta orientada* es un par $(r, <)$ formado por una recta y una de sus dos orientaciones posibles.



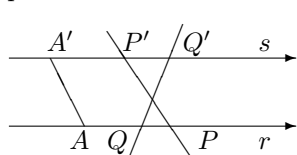
Podemos representar una recta orientada marcando una flecha en ella, como muestra la figura, de modo que al movernos en la dirección de la flecha pasamos a puntos mayores respecto

de la orientación seleccionada. Si fijamos un punto P en una recta orientada r , entonces P divide a r en dos semirrectas, una formada por los puntos menores o iguales que P y otra por los puntos mayores o iguales que P . Las llamaremos, respectivamente, semirrecta menor y semirrecta mayor.



Diremos que dos rectas paralelas orientadas $r \neq s$ tienen *la misma orientación* si existe una recta t que las corta en puntos P y P' de modo que las semirrectas menores determinadas por estos puntos estén contenidas en el mismo semiplano respecto de t (luego las mayores también).

Veamos que esta definición no depende de la recta t escogida. En efecto, supongamos que esto se cumple con una recta t y que t' es cualquier otra recta que corte a r y a s en dos puntos Q y Q' . Tomemos puntos A y A' en r y s que sean menores que P, Q, P', Q' respectivamente. Basta probar que t' no corta al segmento $\overline{AA'}$, pues entonces A y A' están en el mismo semiplano respecto a t' , y como están ambos en las semirrectas menores respecto a P' y Q' , concluimos que ambas están en el mismo semiplano respecto a t' .



Supongamos que $\overline{AA'}$ corta a t' . Es fácil ver que entonces de hecho corta al segmento $\overline{QQ'}$. (El primer segmento está contenido en la *banda* determinada por r y s , esto es, en la intersección del semiplano de frontera r que contiene a s con el semiplano de frontera s que contiene a r . Por otra parte la intersección de t' con esta banda es $\overline{QQ'}$, luego si la recta corta a $\overline{AA'}$ el punto de corte ha de estar en dicho segmento.)

Si $\overline{AA'}$ corta a $\overline{QQ'}$ y $P \neq Q$, entonces el axioma B4 aplicado al triángulo $\overline{PQQ'}$ obtenemos que AA' corta a $\overline{PQ'}$ (no puede cortar al otro lado \overline{PQ} porque entonces A estaría entre P y Q). Si $P = Q$ tenemos directamente que AA' corta a $\overline{PQ'}$. Por el mismo argumento aplicado al triángulo $\overline{PQ'P'}$ llegamos a que AA' corta a $\overline{PP'}$, pero el punto de corte ha de estar en $\overline{AA'}$, con lo que A y A' están en semiplanos distintos respecto de t' , en contradicción con la hipótesis. ■

Así pues, siempre que una recta t corta a dos rectas paralelas igualmente orientadas r y s , las semirrectas menores quedan en el mismo semiplano (y las mayores también). Como consecuencia:

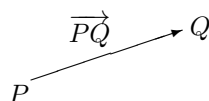
Teorema 3.2 Sean r_1, r_2 y r_3 tres rectas orientadas paralelas dos a dos. Si r_1 tiene la misma orientación que r_2 y r_2 tiene la misma orientación que r_3 , entonces r_1 tiene la misma orientación que r_3 .

DEMOSTRACIÓN: El resultado es trivial si dos de las rectas coinciden, así que suponemos lo contrario. Si las tres están contenidas en un mismo plano, una recta t que corte a r_1 y a r_2 tiene que cortar a r_3 , y entonces tenemos que la semirrecta menor de r_1 está en el mismo semiplano que la de r_2 , y ésta está en el mismo semiplano que la de r_3 , luego lo mismo vale para r_1 y r_3 , luego su orientación es la misma.

Si las tres rectas no están contenidas en un mismo plano, tomamos un plano Π que pase por un punto de cada una de ellas. Sea Π_{12} el plano que contiene a r_1 y r_2 , e igualmente tenemos Π_{23} y Π_{13} . La recta $\Pi \cap \Pi_{12}$ corta a r_1 y r_2 en dos puntos de modo que las semirrectas menores están en el mismo semiplano de Π_{12} , pero esto equivale a que estén en el mismo semiespacio respecto de Π . Similarmente, las semirrectas menores de r_2 y r_3 están en el mismo semiespacio respecto de Π , luego lo mismo vale para r_1 y r_3 , pero esto significa que las semirrectas menores de estas rectas respecto de los puntos de corte con la recta $\Pi_{13} \cap \Pi$ están en el mismo semiplano respecto de Π_{13} , luego r_1 y r_3 tienen la misma orientación. ■

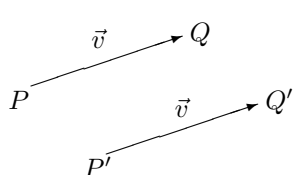
Con esto ya estamos en condiciones de iniciar el tránsito hacia la geometría analítica:

Definición 3.3 Un *vector fijo* (lat ‘transportador’) es un par ordenado de puntos del plano. Lo representaremos \overrightarrow{PQ} (no confundir con la notación para las semirrectas). El punto P se llama *origen* del vector \overrightarrow{PQ} , mientras que el punto Q es el *extremo*. Si el origen es igual al extremo diremos que el vector es *nulo*.



Gráficamente, los vectores fijos no nulos los podemos representar mediante flechas. La figura muestra un vector \overrightarrow{PQ} . El vector \overrightarrow{QP} estaría representado por la flecha opuesta, con la punta en P en lugar de en Q .

Ahora queremos hacer que los vectores fijos dejen de “estar fijos”, de modo que podamos trasladarlos libremente por el espacio y afirmar, por ejemplo, que los dos vectores de la figura son un mismo vector \vec{v} .



Lo cierto es que \overrightarrow{PQ} y $\overrightarrow{P'Q'}$ son dos vectores fijos distintos, y cuando queremos que dos cosas distintas puedan pasar a verse como “la misma cosa”, la estrategia es definir la relación de equivalencia que queremos que las identifique y pasar a considerar las clases de equivalencia correspondientes.

Nos ocupamos de ello seguidamente:

Llamaremos *norma* de un vector fijo no nulo \overrightarrow{PQ} a la longitud $\|\overrightarrow{PQ}\|$ del segmento \overline{PQ} (respecto a una unidad de longitud prefijada). Convenimos que la norma de los vectores nulos es 0. Notemos que la norma depende de la unidad de medida elegida, pero la relación de equivalencia “tener la misma norma” es independiente de dicha unidad.

Diremos que dos vectores fijos no nulos \overrightarrow{PQ} y $\overrightarrow{P'Q'}$ tienen la misma dirección si las rectas PQ y $P'Q'$ son paralelas.

En tal caso, diremos que los dos vectores tienen el mismo sentido si las relaciones \leq_{PQ} en PQ y $\leq_{P'Q'}$ en $P'Q'$ determinan la misma orientación. En caso contrario diremos que tienen sentidos opuestos.

Es claro que si $P \neq Q$, los vectores \overrightarrow{PQ} y \overrightarrow{QP} tienen la misma norma y la misma dirección, pero sentidos opuestos.

Diremos que dos vectores fijos \overrightarrow{PQ} y $\overrightarrow{P'Q'}$ son *equipolentes* (lat. de igual fuerza) si ambos son nulos o si ambos son no nulos y tienen la misma norma, la misma dirección y el mismo sentido. Es claro que la equipolencia es una relación de equivalencia sobre el conjunto de los vectores fijos.

Un *vector libre* es una clase de equipolencia de vectores fijos. Si \vec{v} es un vector libre, a los vectores fijos que lo componen los llamaremos *trasladados* de \vec{v} . Si \overrightarrow{PQ} es un trasladado de \vec{v} escribiremos simplemente $\vec{v} = \overrightarrow{PQ}$. Por definición de equipolencia, todos los vectores nulos determinan un mismo vector libre al que representaremos por $\vec{0}$.

Ahora ya podemos decir que los dos vectores de la figura anterior son dos trasladados del mismo vector libre \vec{v} .

En general, dado un punto P y un vector libre \vec{v} , existe un único punto Q tal que $\vec{v} = \overrightarrow{PQ}$. En efecto, si $\vec{v} = \vec{0}$ es claro. Si $\vec{v} = \overrightarrow{AB}$, un vector con origen P y la misma dirección que \vec{v} ha de tener el extremo en la recta paralela a AB por P . En dicha recta hay dos puntos Q tales que \overrightarrow{PQ} tiene la misma norma que \vec{v} , y cada uno de ellos determina una ordenación $<_{PQ}$ diferente, luego sólo uno de ellos hace que \overrightarrow{PQ} tenga el mismo sentido que \vec{v} .

Al único punto Q que cumple $\overrightarrow{PQ} = \vec{v}$ lo llamaremos *trasladado* de P por \vec{v} y lo representaremos por $Q = P + \vec{v}$. También diremos que \overrightarrow{PQ} es el *trasladado* de \vec{v} de origen P . En particular tenemos $P + \overrightarrow{PQ} = Q$.

Así, podemos pensar en los vectores libres como en reglas para trasladar un punto cualquiera P hasta otro punto Q , de modo que un mismo vector traslada a todos los puntos en una misma dirección, en un mismo sentido y recorriendo una misma distancia. El vector nulo es el único que no mueve los puntos.

Podemos decir que dos vectores tienen la misma norma si dos de sus trasladados respectivos tienen la misma norma, lo cual no depende de la elección de los trasladados. Si fijamos una unidad de longitud, podemos definir incluso la norma de un vector $\|\vec{v}\|$ como la norma de cualquiera de sus trasladados.

Similarmenete, podemos decir que dos vectores no nulos \vec{v} y \vec{w} tienen la misma dirección si esto les sucede a dos cualesquiera de sus trasladados. Si los tomamos con origen común $\vec{v} = \overrightarrow{PQ}$ y $\vec{w} = \overrightarrow{PQ'}$, esto equivale a que los puntos P , Q y Q' sean colineales (pues las rectas PQ y PQ' tienen que ser paralelas, luego iguales).

Diremos que dos vectores con la misma dirección tienen el mismo sentido si esto lo cumplen dos cualesquiera de sus trasladados, y en el caso en que $\vec{v} = \overrightarrow{PQ}$ y $\vec{w} = \overrightarrow{PQ'}$, esto equivale a que Q y Q' , además de ser colineales con P , estén en la misma semirrecta respecto a P .

Dado un vector libre $\vec{v} \neq \vec{0}$ y un número $\alpha \in \mathbb{R}$ no nulo, definimos el vector libre $\alpha\vec{v}$ como el único vector con norma $|\alpha|\|\vec{v}\|$, con la misma dirección que \vec{v} y cuyo sentido es el mismo que el de \vec{v} si $\alpha > 0$ o el opuesto si $\alpha < 0$. Definimos también $0\vec{v} = \alpha\vec{0} = \vec{0}$.

Observemos que, en efecto, existe un único vector libre en tales condiciones y, más aún, que no depende de la unidad de longitud con la que calculamos las normas. En efecto, si $\vec{v} = \overrightarrow{PQ}$, sabemos que $|\alpha|\|\vec{v}\| \in R$ es la longitud de un segmento s , y que en la recta PQ hay exactamente dos puntos A_1 y A_2 tales que $\overrightarrow{PA_1} \equiv \overrightarrow{PA_2} \equiv s$, de los cuales, uno estará en la misma semirrecta que Q respecto de P y el otro en la opuesta. Entonces $\alpha\vec{v} = \overrightarrow{PA_i}$, donde el índice i se elige según el signo de α . La unicidad se debe a que la definición determina la norma, la dirección y el sentido de $\alpha\vec{v}$.

Si hemos hecho el cálculo con una unidad de medida u y tomamos otro segmento u' , entonces $u \equiv \delta u'$, para cierto $\delta \in R$. Por consiguiente

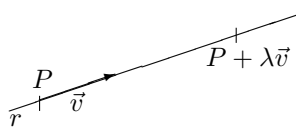
$$\begin{aligned} \overrightarrow{PQ} \equiv \|\vec{v}\|_u u &\equiv \|\vec{v}\|_{u'} \delta u' &\Rightarrow &\quad \|\vec{v}\|_{u'} = \|\vec{v}\|_u \delta. \\ s \equiv |\alpha|\|\vec{v}\|_u u &\equiv |\alpha|\|\vec{v}\|_{u'} \delta u' = \|\alpha\|\|\vec{v}\|_{u'} u', \end{aligned}$$

luego el segmento s es independiente de la unidad elegida, y por tanto $\alpha\vec{v}$ también.

Es fácil comprobar que $(\alpha\beta)\vec{v} = \alpha(\beta\vec{v})$, así como que $1\vec{v} = \vec{v}$.

Dada una recta r , un vector director de r es cualquier vector $\vec{v} = \overrightarrow{PQ}$, donde P y Q son dos puntos distintos de r .

Teorema 3.4 Si P es un punto de una recta r y \vec{v} es un vector director, entonces cada punto de r se expresa de forma única como $Q = P + \lambda\vec{v}$, con $\lambda \in R$.



DEMOSTRACIÓN: En principio $\vec{v} = \overrightarrow{AB}$, para cierto par de puntos A, B en r , pero si lo trasladamos para que tenga origen en P , entonces $\vec{v} = \overrightarrow{PP'}$, donde P' está en r , porque las rectas PP' y $r = PA$ son paralelas, luego coinciden.

Los vectores \vec{v} y \overrightarrow{PQ} tienen la misma dirección. Por lo tanto existe un $\lambda \in R$ tal que $\overrightarrow{PQ} = \lambda\vec{v}$, luego $Q = P + \lambda\vec{v}$.

Si $Q = P + \lambda\vec{v} = P + \lambda'\vec{v}$, entonces $\lambda\vec{v} = \lambda'\vec{v} = \overrightarrow{PQ}$. Obviamente $\lambda = 0$ si y sólo si $\lambda' = 0$. Si no son nulos, comparando las normas y los sentidos concluimos que $\lambda = \lambda'$. ■

Si r es una recta graduada, $\vec{v} = \overrightarrow{P_0P_1}$ y $\alpha \in R$, entonces $P_\alpha = P_0 + \alpha\vec{v}$ es un punto de la recta cuya distancia a P_0 (medida respecto de la unidad de longitud de la graduación) es α y que está en la semirrecta de P_1 o en la contraria según que α sea positivo o negativo. En definitiva, $P_\alpha = P_0 + \alpha\overrightarrow{P_0P_1}$ es el punto de la recta que se corresponde con α a través de la graduación (el que ya veníamos llamando P_α).

Más en general, se cumple que $P_\alpha + \overrightarrow{P_0P_\beta} = P_{\alpha+\beta}$.

En efecto, si llamamos $Q = P_\alpha + \overrightarrow{P_0P_\beta}$, entonces $\overrightarrow{P_\alpha Q} = \overrightarrow{P_0P_\beta} = \beta\overrightarrow{P_0P_1}$, luego Q está en r y dista $|\beta|$ unidades de P_α . Teniendo en cuenta que la distancia entre dos puntos P_α y P_β es $|\alpha - \beta|$, en este caso tiene que ser $Q = P_{\alpha+\beta}$.

Si $\beta > 0$ entonces $P_0 < P_\beta$, luego $P_\alpha < Q$ (pues ambos pares de puntos han de inducir el mismo orden en r), luego $Q = P_{\alpha+\beta}$. Similarmente, si $\beta < 0$ ha de ser $Q < P_\alpha$, luego también $Q = P_{\alpha+\beta}$.

En particular, si \vec{v} es cualquier vector no nulo, P es cualquier punto, llamamos $P_0 = P$, $P_1 = P + \vec{v}$ y consideramos la recta graduada P_0P_1 , de modo que $P_\alpha = P + \alpha\vec{v}$, $P_\beta = P + \beta\vec{v}$ y $\overrightarrow{P_0P_\beta} = \beta\vec{v}$, queda

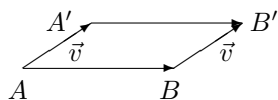
$$P + (\alpha + \beta)\vec{v} = (P + \alpha\vec{v}) + \beta\vec{v}.$$

Las propiedades básicas de los vectores se siguen del teorema siguiente:

Teorema 3.5 Para todo par de puntos A y B y todo vector libre \vec{v} , si $A' = A + \vec{v}$ y $B' = B + \vec{v}$, entonces $\overrightarrow{AB} = \overrightarrow{A'B'}$.

DEMOSTRACIÓN: Podemos suponer que $\vec{v} \neq \vec{0}$. Supongamos ahora que A , B y A' están alineados. Graduemos la recta que los contiene tomando $P_0 = A$ y $P_1 = B$. Sea $\vec{v} = \overrightarrow{P_0P_\alpha}$. Entonces $A + \vec{v} = P_\alpha$ y $B + \vec{v} = P_{1+\alpha}$, y claramente $\overrightarrow{P_0P_1} = \overrightarrow{P_\alpha P_{\alpha+1}}$ (ambos miden una unidad de longitud e inducen la ordenación de la graduación). Así pues, $\overrightarrow{AB} = \overrightarrow{A'B'}$.

Supongamos ahora que A' no está en AB . Entonces las rectas AA' y BB' son paralelas. Si las ordenamos con las relaciones $<_{AA'}$ y $<_{BB'}$ la orientación es la misma, luego A' y B' están en el mismo semiplano respecto de AB .



La paralela a AB que pasa por A' corta a BB' en un punto X que forma un paralelogramo con A , A' , B , luego X dista de B lo mismo que A' dista de A , luego lo mismo que B' dista de B . Como además B' y X están en el mismo semiplano respecto de AB (el que contiene a B), de hecho están en la misma semirrecta de origen B , luego $X = B'$. Así pues, $\overrightarrow{A'B'}$ y \overrightarrow{AB} tienen la misma dirección y la misma norma. Como la recta AA' deja a B y a B' en el mismo semiplano, el sentido también es el mismo (B' y B están en las semirrectas mayores para las ordenaciones que inducen $\overrightarrow{A'B'}$ y \overrightarrow{AB}). Esto prueba que los dos vectores son iguales. ■

Teorema 3.6 Para todos los puntos A y B y todos los vectores \vec{v} y \vec{w} se cumple:

1. Si $A + \vec{v} = B + \vec{v}$ entonces $A = B$.
2. Si $A + \vec{v} = A + \vec{w}$ entonces $\vec{v} = \vec{w}$.
3. $(A + \vec{v}) + \vec{w} = (A + \vec{w}) + \vec{v}$.
4. Si $A' = (A + \vec{v}) + \vec{w}$ y $B' = (B + \vec{v}) + \vec{w}$, entonces $\overrightarrow{AA'} = \overrightarrow{BB'}$.

DEMOSTRACIÓN: 1) Si llamamos $C = A + \vec{v} = B + \vec{v}$, el teorema anterior nos da que \overrightarrow{AB} es equipolente a \overrightarrow{CC} , luego ha de ser $A = B$.

2) Si $C = A + \vec{v} = A + \vec{w}$, obviamente $\vec{v} = \overrightarrow{AC} = \vec{w}$.

3) Sean $B = A + \vec{v}$, $A' = A + \vec{w}$ y $B' = B + \vec{w}$. Por el teorema anterior $\overrightarrow{A'B'} = \overrightarrow{AB} = \vec{v}$. Así

$$(A + \vec{v}) + \vec{w} = B + \vec{w} = B', \quad (A + \vec{w}) + \vec{v} = A' + \overrightarrow{A'B'} = B'.$$

4) Sea $\vec{x} = \overrightarrow{AB}$. Entonces

$$B' = (B + \vec{v}) + \vec{w} = ((A + \vec{x}) + \vec{v}) + \vec{w} = ((A + \vec{v}) + \vec{w}) + \vec{x} = A' + \vec{x},$$

y como también $B = A + \vec{x}$, el teorema anterior implica que $\overrightarrow{AB} = \overrightarrow{A'B'}$. ■

La propiedad 4) del teorema anterior nos permite definir la suma de dos vectores \vec{v} y \vec{w} como el único vector $\vec{v} + \vec{w}$ que cumple

$$A + (\vec{v} + \vec{w}) = (A + \vec{v}) + \vec{w},$$

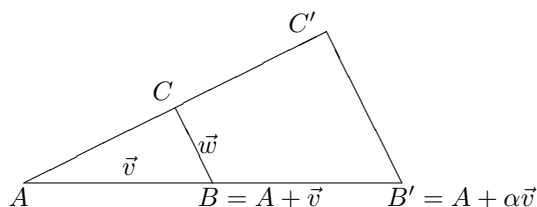
para todo punto A .

Ahora ya estamos en condiciones de probar que los vectores libres que hemos construido forman un ejemplo de espacio vectorial en el sentido algebraico definido en [Al 4.1]:

Teorema 3.7 *El conjunto V de los vectores libres con la suma y el producto que hemos definido forma un espacio vectorial sobre R en el sentido algebraico del término, es decir, se cumplen las propiedades siguientes:*

1. $(\vec{v} + \vec{w}) + \vec{x} = \vec{v} + (\vec{w} + \vec{x})$,
2. $\vec{v} + \vec{w} = \vec{w} + \vec{v}$,
3. $\vec{v} + \vec{0} = \vec{v}$,
4. Para cada vector \vec{v} existe un vector $-\vec{v}$ tal que $\vec{v} + (-\vec{v}) = \vec{0}$,
5. $\alpha(\beta\vec{v}) = (\alpha\beta)\vec{v}$,
6. $(\alpha + \beta)\vec{v} = \alpha\vec{v} + \beta\vec{v}$,
7. $\alpha(\vec{v} + \vec{w}) = \alpha\vec{v} + \alpha\vec{w}$,
8. $1\vec{v} = \vec{v}$.

DEMOSTRACIÓN: Todas las propiedades son inmediatas a partir de lo ya probado salvo la séptima. Las propiedades restantes permiten reducir la prueba al caso en que $\alpha > 0$ y los vectores \vec{v} y \vec{w} tienen direcciones distintas.



Sea A un punto cualquiera, $B = A + \vec{v}$, $C = A + \vec{v} + \vec{w}$ y $B' = A + \alpha\vec{v}$. Sea C' el punto donde la paralela a BC por B' corta a AC .

Entonces $\overrightarrow{B'C'}$ tiene la misma dirección y sentido que \vec{w} , y por el teorema de Tales su norma es $\alpha\|\vec{w}\|$, luego $\overrightarrow{B'C'} = \alpha\vec{w}$ y por consiguiente $\overrightarrow{AC'} = \alpha\vec{v} + \alpha\vec{w}$.

Por otra parte $\overrightarrow{AC'}$ tiene la misma dirección y sentido que $\vec{v} + \vec{w}$ y también por el teorema de Tales, su norma es $\alpha\|\vec{v} + \vec{w}\|$, luego $\overrightarrow{AC'} = \alpha(\vec{v} + \vec{w})$. ■

El teorema 3.4 se interpreta ahora como que una recta está formada por los trasladados de uno cualquiera de sus puntos mediante los vectores de un subespacio vectorial de V de dimensión 1. Vamos a probar un resultado análogo para los planos:

Teorema 3.8 *Sea Π un plano, sean P , Q y R tres puntos de π no colineales. Sean $\vec{u} = \overrightarrow{PQ}$ y $\vec{v} = \overrightarrow{PR}$. Entonces cada punto de Π se expresa de forma única como*

$$X = P + \lambda\vec{u} + \mu\vec{v}, \quad \lambda, \mu \in R.$$

DEMOSTRACIÓN: Dado un punto X en Π , consideramos la recta paralela a PR que pasa por X . Ésta cortará a PQ en un punto de la forma $Y = P + \lambda\vec{u}$. El vector \overrightarrow{YX} tiene la misma dirección que $\overrightarrow{PR} = \vec{v}$, luego será de la forma $\overrightarrow{YX} = \mu\vec{v}$. Así

$$X = P + \overrightarrow{PY} + \overrightarrow{YX} = P + \lambda\vec{u} + \mu\vec{v}.$$

La unicidad equivale a que los vectores \vec{u} y \vec{v} sean linealmente independientes, pero esto se sigue de que P , Q y R no son colineales. Es fácil ver de modo similar que todo punto de la forma indicada está en Π . ■

Similarmente:

Teorema 3.9 *Sean P_0, P_1, P_2, P_3 cuatro puntos no coplanares y consideremos la variedad afín tridimensional $E = A^3(P_0, P_1, P_2, P_3)$. Consideremos los vectores $\vec{u} = \overrightarrow{P_0P_1}$, $\vec{v} = \overrightarrow{P_0P_2}$ y $\vec{w} = \overrightarrow{P_0P_3}$. Entonces todo punto $X \in E$ se expresa de forma única como*

$$X = P_0 + \lambda\vec{u} + \mu\vec{v} + \nu\vec{w}, \quad \lambda, \mu, \nu \in R.$$

DEMOSTRACIÓN: Sea $\Pi = P_0P_1P_2$. El plano $\Pi' = P_0P_3X \subset E$ corta a Π en una recta r que corta a P_0P_3 en P_0 . Además contiene a la recta paralela a P_0P_3 que pasa por X , la cual tiene que cortar también a r en un punto $Y \in \Pi$ que, por el teorema anterior, será de la forma

$$Y = P_0 + \lambda\vec{u} + \mu\vec{v}.$$

El vector \overrightarrow{YX} tiene la misma dirección que $\overrightarrow{P_0P_3} = \vec{w}$, luego $\overrightarrow{YX} = \nu\vec{w}$. Como en el teorema anterior, concluimos que X tiene la forma indicada. La expresión es única porque los vectores \vec{u} , \vec{v} y \vec{w} son linealmente independientes (o de lo contrario E estaría contenida en un plano). ■

En particular, los axiomas de tridimensionalidad implican (y, de hecho, equivalen a) que el espacio V de los vectores libres tiene dimensión 3, pues entonces $\vec{E} = E$ y todo vector libre es de la forma $\overrightarrow{P_0X}$, para cierto punto X , luego $V = \langle \vec{u}, \vec{v}, \vec{w} \rangle$.

En otras palabras, lo que hemos probado es que toda recta se obtiene sumando a uno de sus puntos todos los vectores de un subespacio vectorial de V de dimensión 1, que todo plano se obtiene sumando a uno de sus puntos todos los vectores de un subespacio vectorial de dimensión 2, y que todos los puntos de una variedad tridimensional (que son todos los puntos del espacio si suponemos los axiomas de tridimensionalidad) se obtienen sumando a un punto cualquiera todos los vectores de V , que tiene dimensión 3.

En particular, cada punto de una recta está determinado por un parámetro real, cada punto de un plano está determinado por dos parámetros reales y cada punto de un espacio tridimensional está determinado por tres parámetros reales. Eso precisa el sentido en que podemos decir que las rectas son objetos de una dimensión, los planos son objetos de dos dimensiones y que el espacio tiene tres dimensiones, a la vez que da una interpretación geométrica al concepto abstracto de dimensión de un espacio vectorial.

3.2 Espacios afines

Observemos que, en los términos de la sección anterior, los conceptos geométricos de “punto” y “vector” corresponden a dos clases de objetos completamente distintos. Los puntos son simplemente “los puntos”, es decir, los elementos del espacio, mientras que un vector es una construcción abstracta que determina un modo “uniforme” de movernos de un punto a otro (mediante un movimiento en línea recta, de la misma longitud y en la misma dirección y sentido, sea cual sea el punto del que partamos). Estos conceptos de “punto” y “vector” contrastan con los considerados en el apéndice A de [IGE], donde, tanto los puntos como los vectores, no eran sino pares de \mathbb{R}^2 . Ahora vamos a introducir una estructura algebraica que axiomatiza las propiedades de los puntos y los vectores, pero manteniendo una diferencia nítida entre ambos conceptos, y luego veremos en qué sentido ambos pueden “confundirse” en la práctica.

Definición 3.10 Un *espacio afín* de dimensión n es una terna $(E, \vec{E}, +)$, donde E es un conjunto no vacío, a cuyos elementos se les llama *puntos*, \vec{E} es un espacio vectorial de dimensión n sobre un cuerpo K y $+$ es una aplicación $E \times \vec{E} \rightarrow E$ que cumple las propiedades siguientes:

1. Para cada par de puntos P, Q , existe un único vector \vec{v} tal que $Q = P + \vec{v}$, y se le representa por $\vec{v} = \overrightarrow{PQ}$.
2. $P + \vec{0} = P$,
3. $(P + \vec{v}) + \vec{w} = P + (\vec{v} + \vec{w})$.

En la sección anterior hemos probado que el espacio de una geometría (tridimensional) euclídea es un espacio afín tridimensional sobre un cuerpo de escalares R arquimediano (que será concretamente el cuerpo \mathbb{R} de los números reales si suponemos el axioma de continuidad). Es fácil demostrar que todas las propiedades de vectores que hemos probado allí son válidas en cualquier espacio afín. Por ejemplo,

$$P + \overrightarrow{PQ} + \overrightarrow{QP} = Q + \overrightarrow{QP} = P = P + \vec{0}, \quad \text{luego} \quad \overrightarrow{QP} = -\overrightarrow{PQ}.$$

Del mismo modo se prueba que $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$.

Una *variedad lineal*¹ de dimensión m en un espacio afín es un conjunto de puntos de la forma

$$L = P + \vec{L} = \{P + \vec{w} \mid \vec{w} \in \vec{L}\},$$

donde \vec{L} es un subespacio vectorial de \vec{E} de dimensión m , llamado *espacio director* de la variedad.

Notemos que $\vec{L} = \{\vec{v} \in \vec{E} \mid P + \vec{v} \in L\}$, luego \vec{L} está determinado por L . Recíprocamente, una variedad lineal L está determinada por su espacio director \vec{L} y uno cualquiera de sus puntos, pues si $Q \in P + \vec{L}$, entonces $\overrightarrow{PQ} \in \vec{L}$, luego $Q + \vec{L} = P + \overrightarrow{PQ} + \vec{L} = P + \vec{L} = L$.

En la sección anterior hemos probado que las rectas son las variedades lineales de dimensión 1 del espacio de la geometría euclídea, mientras que los planos son las variedades de dimensión 2. Por ello las variedades lineales de dimensión 1 y 2 de cualquier espacio afín reciben el nombre de *rectas* y *planos*. Las variedades lineales de dimensión 0 son los puntos (con rigor, los conjuntos con un solo punto). En un espacio afín de dimensión n , las variedades de dimensión $n - 1$ se llaman *hiperplanos*.

Toda variedad lineal $L = P + \vec{L}$ de dimensión m se puede considerar en sí misma como un espacio afín de dimensión m con \vec{L} como espacio vectorial asociado.

Dado un conjunto de puntos $A \neq \emptyset$, existe una mínima variedad lineal que lo contiene, la llamaremos la variedad lineal generada por A y la representaremos por $\langle A \rangle$. En efecto, si $P \in A$ es fácil ver que

$$\langle A \rangle = P + \langle \overrightarrow{PQ} \mid Q \in A \rangle.$$

es una variedad que contiene a A y está contenida en cualquier otra que cumpla lo mismo.

Diremos que $n + 1$ puntos A_0, \dots, A_n son *afínmente independientes* si la variedad lineal que generan tiene dimensión n (en caso contrario se dice que son *afínmente dependientes*). Según lo que acabamos de ver, esto equivale a que los vectores $\overrightarrow{A_0A_1}, \dots, \overrightarrow{A_0A_n}$ sean linealmente independientes.

¹El nombre hace referencia a que los elementos de una variedad están determinados por *varias* coordenadas.

En particular tres puntos P, Q, R son colineales si y sólo si son afinmente dependientes, si y sólo si los vectores \overrightarrow{PQ} y \overrightarrow{PR} son linealmente dependientes. Análogamente, cuatro puntos P, Q, R, S son coplanares si y sólo si $\overrightarrow{PQ}, \overrightarrow{PR}$ y \overrightarrow{PS} son linealmente dependientes, etc.

Es claro que en un espacio afín de dimensión n hay conjuntos afinmente independientes con $n + 1$ puntos, pero no con más.

Diremos que dos variedades lineales son *paralelas* si el espacio director de una está contenido en el de la otra.

El teorema siguiente prueba que esta noción general de paralelismo coincide con la que conocemos para rectas y planos en los espacios tridimensionales.

Teorema 3.11 *Dos variedades lineales de dimensiones k y l con $k \leq l$ son paralelas si y sólo si están contenidas en una variedad de dimensión $l + 1$ y una está contenida en la otra o bien no tienen puntos en común.*

DEMOSTRACIÓN: Sean $P + \langle \vec{v}_1, \dots, \vec{v}_k \rangle$ y $Q + \langle \vec{w}_1, \dots, \vec{w}_l \rangle$ las dos variedades. Si $\langle \vec{v}_1, \dots, \vec{v}_k \rangle \subset \langle \vec{w}_1, \dots, \vec{w}_l \rangle$, entonces ambas están contenidas en la variedad $P + \langle \overrightarrow{PQ}, \vec{w}_1, \dots, \vec{w}_l \rangle$ de dimensión $l + 1$ y, si tienen un punto en común, podemos tomar $P = Q$ y es claro que una está contenida en la otra.

Supongamos ahora que no se cortan y están contenidas en una variedad de dimensión $l + 1$. Veamos en primer lugar que \overrightarrow{PQ} no es combinación lineal de $\vec{v}_1, \dots, \vec{v}_k, \vec{w}_1, \dots, \vec{w}_l$. Si lo fuera tendríamos

$$\alpha_1 \vec{v}_1 + \dots + \alpha_k \vec{v}_k = \overrightarrow{PQ} + \beta_1 \vec{w}_1 + \dots + \beta_l \vec{w}_l,$$

con lo que el punto

$$P + \alpha_1 \vec{v}_1 + \dots + \alpha_k \vec{v}_k = Q + \beta_1 \vec{w}_1 + \dots + \beta_l \vec{w}_l$$

estaría en las dos variedades.

Una variedad de dimensión $l + 1$ que contenga a ambas variedades contiene a los puntos $P, Q, P + \vec{v}_i, Q + \vec{w}_i$, luego su espacio director contiene a los vectores $\overrightarrow{PQ}, \vec{v}_i, \vec{w}_i$. Como la dimensión es $l + 1$ dicho espacio ha de ser $\langle \overrightarrow{PQ}, \vec{w}_1, \dots, \vec{w}_l \rangle$, luego

$$\langle \vec{v}_1, \dots, \vec{v}_k \rangle \subset \langle \overrightarrow{PQ}, \vec{w}_1, \dots, \vec{w}_l \rangle,$$

y como \overrightarrow{PQ} no es combinación lineal de ninguno de los demás vectores,

$$\langle \vec{v}_1, \dots, \vec{v}_k \rangle \subset \langle \vec{w}_1, \dots, \vec{w}_l \rangle,$$

lo que prueba que las variedades son paralelas. ■

Así, dos rectas son paralelas si y sólo si están contenidas en un plano y son iguales o no tienen puntos comunes, dos hiperplanos (en particular dos planos en un espacio tridimensional) son paralelos si y sólo si no se cortan.

Ejercicio: Probar que si dos variedades lineales tienen intersección no vacía entonces ésta es una variedad lineal y que su espacio director es la intersección de los espacios directores.

Hemos necesitado todos los axiomas de la geometría euclídea (menos el de continuidad² para dotar al espacio de estructura de espacio afín (sobre un cuerpo ordenado arquimediano). Sin embargo, si partimos de la estructura de espacio afín, podemos obtener espacios que cumplan más o menos axiomas de la geometría euclídea. Para empezar, ahora es fácil demostrar el teorema siguiente:

Teorema 3.12 *Cualquier espacio afín de dimensión ≥ 2 (sobre un cuerpo arbitrario) cumple el grupo de axiomas A de la sección 1.1, así como el axioma de las paralelas. Si la dimensión es ≥ 3 se cumple además el axioma A5b, y si la dimensión es exactamente 3 se cumple el axioma A6.*

Por ejemplo, para probar A6 observamos que si dos planos en un espacio tridimensional tienen un punto en común P , la intersección es una variedad lineal cuyo espacio director es la intersección de dos subespacios de dimensión 2, luego la relación

$$\dim(V + W) = \dim V + \dim W - \dim(V \cap W)$$

implica que la dimensión de la intersección ha de ser al menos 1, luego la intersección es un plano o una recta.

Así pues, todos los conceptos y propiedades geométricas que puedan definirse / demostrarse a partir de los axiomas de incidencia más el axioma de las paralelas de la geometría euclídea son aplicables en realidad a cualquier espacio afín sobre un cuerpo arbitrario. Ahora vamos a probar que en cualquier espacio afín sobre un cuerpo ordenado se cumplen también los axiomas de ordenación, luego podemos contar con todos los hechos que se deducen de éstos:

Teorema 3.13 *En un espacio afín sobre un cuerpo ordenado se cumplen los axiomas de ordenación si definimos*

$$A - B - C \quad \text{si y sólo si} \quad B = A + \lambda \overrightarrow{AC}, \quad 0 < \lambda < 1.$$

DEMOSTRACIÓN: Veamos únicamente el axioma B4, que es el más técnico.

Sean A , B y C tres puntos no colineales y sea r una recta en su mismo plano que no pase por ninguno de ellos, pero sí por un punto X entre A y B . Sea \vec{v} un vector director de r , es decir, $r = X + \langle \vec{v} \rangle$. La hipótesis es que $X = A + \alpha \overrightarrow{AB}$, con $0 < \alpha < 1$.

Los vectores \vec{v} y \overrightarrow{AB} han de ser linealmente independientes, pues en caso contrario $A = X - \alpha \overrightarrow{AB} = X + \delta \vec{v}$ estaría en r .

Es claro que ABC es el plano $X + \langle \vec{v}, \overrightarrow{AB} \rangle$, luego

$$C = X + \beta \vec{v} + \gamma \overrightarrow{AB},$$

para ciertos β , γ , donde $\gamma \neq 0$, o de lo contrario C estaría en r .

²No es trivial, pero también podríamos haber prescindido de la propiedad arquimediana.

Supongamos que r corta a la recta AC . Entonces lo hace en un punto de la forma

$$X + \lambda\vec{v} = A + \mu\overrightarrow{AC}.$$

Desarrollando:

$$\begin{aligned} X + \lambda\vec{v} &= X + \overrightarrow{XA} + \mu(\overrightarrow{AX} + \overrightarrow{XC}) = X - \alpha\overrightarrow{AB} + \mu(\alpha\overrightarrow{AB} + \beta\vec{v} + \gamma\overrightarrow{AB}) \\ &= X + (-\alpha + \mu(\alpha + \gamma))\overrightarrow{AB} + \mu\beta\vec{v}. \end{aligned}$$

Esto sucede si y sólo si $\gamma \neq -\alpha$ y $\mu = \frac{\alpha}{\alpha + \gamma}$, y esto sucede con $0 < \mu < 1$ si y sólo si $\gamma > 0$.

Supongamos ahora que r corta a BC . Entonces lo hace en un punto de la forma

$$X + \lambda\vec{v} = B + \mu\overrightarrow{BC}.$$

Desarrollando:

$$\begin{aligned} X + \lambda\vec{v} &= X + \overrightarrow{XA} + \overrightarrow{AB} + \mu(\overrightarrow{BA} + \overrightarrow{AX} + \overrightarrow{XC}) \\ &= X - \alpha\overrightarrow{AB} + \overrightarrow{AB} + \mu(-\overrightarrow{AB} + \alpha\overrightarrow{AB} + \beta\vec{v} + \gamma\overrightarrow{AB}) \\ &= X + (1 - \alpha + \mu(\alpha + \gamma - 1))\overrightarrow{AB} + \mu\beta\vec{v}. \end{aligned}$$

Esto sucede si y sólo si $\alpha + \gamma \neq 1$ y $\mu = \frac{1 - \alpha}{1 - \alpha - \gamma}$, y se cumple con $0 < \mu < 1$ si y sólo si $\gamma < 0$.

Así pues, r corta a \overline{AC} si y sólo si $\gamma > 0$ y corta a \overline{BC} si y sólo si $\gamma < 0$. Por lo tanto, se da uno y sólo uno de los dos casos. ■

Nota El argumento precedente puede reinterpretarse como sigue: tenemos un plano arbitrario, formado por los puntos de la forma $C = X + \beta\vec{v} + \gamma\vec{w}$. Entonces, C está a un lado u otro de la recta r formada por los puntos $X + \beta\vec{v}$ según que sea $\gamma > 0$ o bien $\gamma < 0$.

En efecto, hemos visto que si $\gamma > 0$ entonces r corta a \overline{AC} , luego C está en el mismo lado que B , y si $\gamma < 0$ entonces r corta a \overline{BC} , luego C está en el mismo lado que A .

A su vez, dados tres puntos no colineales O, A, B , el ángulo \widehat{AOB} está formado por los puntos de la forma $O + \lambda\overrightarrow{OA} + \mu\overrightarrow{OB}$ que están a la vez en el semiplano respecto a la recta $O + \lambda\overrightarrow{OA}$ que contiene a B y en el semiplano respecto a la recta $O + \mu\overrightarrow{OB}$ que contiene a A . Concluimos que

$$\widehat{AOB} = \{O + \lambda\overrightarrow{OA} + \mu\overrightarrow{OB} \mid \lambda, \mu \geq 0\}.$$

(Los semiplanos correctos son los correspondientes a los signos positivos, pues A se obtiene con $\lambda = 1$ y B con $\mu = 1$). Equivalentemente, todo ángulo es de la forma

$$A(O; \vec{v}, \vec{w}) = \{O + \lambda\vec{v} + \mu\vec{w} \mid \lambda, \mu \geq 0\},$$

donde \vec{v} y \vec{w} son dos vectores linealmente independientes. ■

Observemos ahora que

$$A - (A + \lambda\vec{v}) - (A + \mu\vec{v}) \quad \text{si y sólo si} \quad \mu\vec{v} = \alpha\lambda\vec{v}, \quad \text{con} \quad 0 < \alpha < 1,$$

es decir, si y sólo si $|\lambda| < |\mu|$. De aquí se sigue fácilmente que las dos semirrectas que A determina en la recta formada por los puntos $A + \lambda\vec{v}$ son

$$\{A + \lambda\vec{v} \mid \lambda > 0\} \quad \text{y} \quad \{A + \lambda\vec{v} \mid \lambda < 0\},$$

así como que las dos ordenaciones posibles de la recta son simplemente la dada por

$$A + \lambda\vec{v} < A + \mu\vec{v} \leftrightarrow \lambda < \mu$$

y su inversa.

Resumimos lo que hemos obtenido junto con otros hechos adicionales evidentes:

- En todo espacio afín de dimensión ≥ 2 podemos definir los conceptos de recta y plano de modo que se satisfacen los axiomas de incidencia básicos del grupo A y el axioma de las paralelas. Si la dimensión es 3 se cumplen también los axiomas de tridimensionalidad.
- Si el cuerpo K del espacio afín es un cuerpo ordenado, podemos definir además una relación $A - B - C$ que satisface los axiomas de ordenación.
- Un espacio afín sobre un cuerpo ordenado R satisface la propiedad de Arquímedes si y sólo si R es arquimediano en el sentido de [An 1.6], lo que equivale [An 1.14] a que R sea un subcuerpo de \mathbb{R} .
- Un espacio afín sobre un cuerpo ordenado R satisface el axioma de continuidad si y sólo si $R = \mathbb{R}$.

Nos falta determinar qué más le hemos de exigir a un espacio afín para que podamos definir en él unas relaciones de congruencia con las que se cumplan los axiomas de congruencia. Pospondremos este problema hasta la sección siguiente, pero, para entender cuál es exactamente el problema al que nos enfrentamos conviene apurar al máximo lo que podemos hacer en un espacio afín arbitrario:

Definición 3.14 Sean P, Q, R y S cuatro puntos colineales en un espacio afín sobre un cuerpo arbitrario K y $P \neq Q$, entonces $\overrightarrow{RS} = \alpha \overrightarrow{PQ}$, para un único $\alpha \in K$. Por lo tanto podemos definir

$$\frac{\overrightarrow{RS}}{\overrightarrow{PQ}} = \alpha.$$

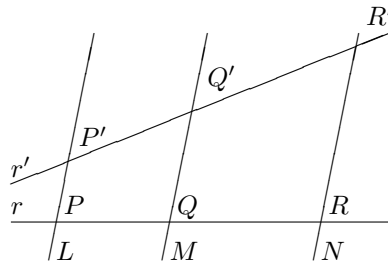
Podemos pensar que esta relación expresa (en un contexto muy abstracto), que el segmento \overline{RS} mide α si tomamos \overline{PQ} como unidad de longitud. No obstante, hay que tener presente que si K es un cuerpo ordenado, entonces α puede ser positivo o negativo, por lo que en realidad contiene algo más de información que la mera comparación de longitudes. Más concretamente, esta definición generaliza a la que dimos en la sección 3.6 de [IGE], y que empleamos, por ejemplo, para enunciar los teoremas de Menelao o Ceva.

El problema es que así sólo podemos comparar las longitudes (y orientaciones) de segmentos colineales (o, un poco más en general, paralelos), pero si tenemos tres puntos no colineales P, Q, R en un espacio afín, no tenemos ningún criterio para comparar la longitud de \overline{PR} respecto a la de \overline{PQ} . Pese a ello, podemos demostrar el teorema de Tales:

Teorema 3.15 (Teorema de Tales) Sean L, M, N tres hiperplanos paralelos en un espacio afín E sobre un cuerpo K . Si r es cualquier recta no paralela a estos hiperplanos, la razón

$$\frac{\overrightarrow{PQ}}{\overrightarrow{PR}}, \quad \text{donde } r \cap L = \{P\}, \quad r \cap M = \{Q\}, \quad r \cap N = \{R\},$$

es independiente de r .



DEMOSTRACIÓN: Sea r' otra recta no paralela a L, M, N que los corte en los puntos P', Q', R' . Vamos a suponer que $P \neq P'$ y $R \neq R'$. El argumento que sigue se simplifica en caso contrario. Entonces las rectas PP' y RR' son paralelas, luego $\overrightarrow{RR'} = \lambda \overrightarrow{PP'}$, para cierto $\lambda \in K$. Por otra parte, llamemos $\alpha = \overrightarrow{PQ}/\overrightarrow{PR}$, de modo que $Q = P + \alpha \overrightarrow{PR}$, y sea $Q'' = P' + \alpha \overrightarrow{P'R'}$. Basta probar que $Q'' = Q'$, lo cual equivale a que $Q = Q''$ o bien a que la recta QQ'' sea paralela a PP' . A su vez esto equivale a que $\overrightarrow{QQ''}$ sea múltiplo de $\overrightarrow{PP'}$. Ahora bien:

$$\begin{aligned} \overrightarrow{QQ''} &= \overrightarrow{QP} + \overrightarrow{PP'} + \overrightarrow{P'Q''} = -\alpha \overrightarrow{PR} + \overrightarrow{PP'} + \alpha \overrightarrow{P'R'} = \\ &= -\alpha(\overrightarrow{PP'} + \overrightarrow{P'R'} + \overrightarrow{R'R}) + \overrightarrow{PP'} + \alpha \overrightarrow{P'R'} = -\alpha \overrightarrow{PP'} + \alpha \lambda \overrightarrow{PP'} + \overrightarrow{PP'} \\ &= (\alpha \lambda - \alpha + 1) \overrightarrow{PP'}, \end{aligned}$$

como había que probar. ■

Aplicaciones afines Estudiamos ahora brevemente las aplicaciones que conservan la estructura afín:

Definición 3.16 Sean E y F dos espacios afines. Una *aplicación afín* o *afinidad* entre E y F es una aplicación $f : E \rightarrow F$ tal que para todo punto P de E se cumple

$$f(P) = f(O) + \vec{f}(\overrightarrow{OP}),$$

donde O es un punto prefijado en E , y $\vec{f} : \vec{E} \rightarrow \vec{F}$ es una aplicación lineal, llamada *aplicación lineal asociada* a f .

En primer lugar hemos de notar que si f admite una expresión como la indicada para un punto O y una cierta aplicación lineal \vec{f} , entonces f admite una expresión análoga para cualquier otro punto O' y con la misma aplicación \vec{f} . En efecto, puesto que $f(O') = f(O) + \vec{f}(\overrightarrow{OO'})$, resulta que

$$f(P) = f(O) + \vec{f}(\overrightarrow{OP}) = f(O') + \vec{f}(\overrightarrow{O'P}) - \vec{f}(\overrightarrow{OO'}) = f(O') + \vec{f}(\overrightarrow{O'P}).$$

En particular una afinidad determina su aplicación lineal asociada.

Notemos que la relación entre f y \vec{f} se puede escribir también en la forma $\vec{f}(\overrightarrow{PQ}) = \overrightarrow{f(P)f(Q)}$, para todo par de puntos P y Q . Teniendo esto en cuenta es fácil comprobar que la composición de aplicaciones afines es una aplicación afín, así como que la inversa de una biyección afín es una biyección afín. Además $\vec{f \circ g} = \vec{f} \circ \vec{g}$ y $\vec{f^{-1}} = (\vec{f})^{-1}$.

Si E tiene dimensión n , los puntos $P_0, \dots, P_n \in E$ son afinmente independientes en E y $Q_0, \dots, Q_n \in F$ son puntos cualesquiera, existe una única afinidad $f: E \rightarrow F$ que cumple $f(P_i) = Q_i$.

En efecto, tiene que ser la dada por $f(P_0) = Q_0$ y $\vec{f}(\overrightarrow{P_0P_i}) = \overrightarrow{Q_0Q_i}$.

Además f será una biyección afín si y sólo si F tiene dimensión n y Q_0, \dots, Q_n son afinmente independientes, pues esto equivale a que \vec{f} transforma una base de \vec{E} en una base de \vec{F} .

Es evidente que las biyecciones afines conservan todas las propiedades definibles a partir de la estructura de espacio afín: aplican variedades en variedades, paralelas en paralelas, semiplanos en semiplanos, ángulos en ángulos, triángulos en triángulos, conservan las relaciones de orden de las rectas si el cuerpo es un cuerpo ordenado, etc.

A su vez, esto nos muestra que la estructura de espacio afín no permite definir los conceptos de longitud de un segmento o de medida de un ángulo, pues es fácil ver que siempre existe una biyección afín que transforma cualquier terna de puntos no colineales en cualquier otra, por lo que ninguna definición de longitud de un segmento o de medida de un ángulo se conservaría por biyecciones afines, y éstas conservan todo lo definible en términos de la estructura de espacio afín.

Coordenadas cartesianas Con las consideraciones de las secciones precedentes ya podemos exponer el núcleo de la geometría analítica, en virtud de la cual los conceptos geométricos se caracterizan en términos de ecuaciones y desigualdades. Para ello necesitamos los conceptos de sistema de referencia y coordenadas de un punto.

Definición 3.17 Un *sistema de referencia* en un espacio afín E está formado por un punto O y una base $(\vec{e}_1, \dots, \vec{e}_n)$ del espacio \vec{E} .

Fijado un sistema de referencia, podemos identificar cada vector \vec{v} con sus coordenadas en la base del sistema. Así, $\vec{v} = (x_1, \dots, x_n)$ se interpreta como

que $\vec{v} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$. Llamaremos *vector de posición* de un punto P (siempre respecto al sistema de referencia fijado) al vector \overrightarrow{OP} . Las *coordenadas cartesianas* (lat. ‘de Descartes’) de un punto P serán las coordenadas de su vector de posición. Escribiremos $P(x_1, \dots, x_n)$ para indicar que las coordenadas de P en un sistema de referencia dado son (x_1, \dots, x_n) . Según lo dicho, esto equivale a que

$$P = O + x_1\vec{e}_1 + \cdots + x_n\vec{e}_n.$$

Lo habitual es considerar sistemas de referencia “ortonormales”, es decir, tales que los vectores \vec{e}_i tengan la misma longitud y sean perpendiculares dos a dos, pero ni “tener la misma longitud” ni “ser perpendicular” son conceptos que tengamos definidos en un espacio afín. Sin embargo, ninguno de ellos es necesario para las consideraciones siguientes:

Consideremos un hiperplano $H = P + \langle \vec{v}_1, \dots, \vec{v}_{n-1} \rangle$ en un espacio afín en el que hemos fijado un sistema de referencia de origen O y base $\vec{e}_1, \dots, \vec{e}_n$. Un punto $Q(x_1, \dots, x_n)$ está en H si y sólo si $\overrightarrow{PQ} \in \langle \vec{v}_1, \dots, \vec{v}_{n-1} \rangle$. Sea $\vec{w} \in \vec{E}$ un vector linealmente independiente de $\vec{v}_1, \dots, \vec{v}_{n-1}$ y consideremos la aplicación lineal $f : \vec{E} \rightarrow K$ determinada por que $f(\vec{v}_i) = 0$ y $f(\vec{w}) = 1$. Entonces $N(f) = \langle \vec{v}_1, \dots, \vec{v}_{n-1} \rangle$, luego $Q \in H$ si y sólo si $f(\overrightarrow{PQ}) = 0$. Si (p_1, \dots, p_n) son las coordenadas de P , entonces

$$\overrightarrow{PQ} = (x_1 - p_1)\vec{e}_1 + \cdots + (x_n - p_n)\vec{e}_n,$$

luego $Q \in H$ si y sólo si $f(\vec{e}_1)(x_1 - p_1) + \cdots + f(\vec{e}_n)(x_n - p_n) = 0$. Operando, esta ecuación tiene la forma

$$a_1x_1 + \cdots + a_nx_n = b,$$

donde necesariamente alguno de los coeficientes es no nulo.

Recíprocamente, el conjunto de puntos cuyas coordenadas satisfacen una ecuación de esta forma es un hiperplano. En efecto, el miembro izquierdo de la ecuación define una aplicación lineal no nula $g : K^n \rightarrow K$, que necesariamente será suprayectiva y su núcleo tendrá dimensión $n - 1$. Por otra parte, la aplicación $h : \vec{E} \rightarrow K^n$ que a cada vector le asigna sus coordenadas en la base $\vec{e}_1, \dots, \vec{e}_n$ es un isomorfismo, luego la composición $f : \vec{E} \rightarrow K$ es una aplicación lineal cuyo núcleo es un subespacio $\langle \vec{v}_1, \dots, \vec{v}_{n-1} \rangle$. Si P es un punto tal que $f(\overrightarrow{OP}) = b$, entonces $H = P + \langle \vec{v}_1, \dots, \vec{v}_{n-1} \rangle$ es un hiperplano cuyos puntos son exactamente aquellos cuyas coordenadas satisfacen la ecuación dada. En resumen:

Los puntos de un hiperplano en un espacio afín están caracterizados por que sus coordenadas en cualquier sistema de referencia satisfacen una determinada ecuación lineal no nula. Toda ecuación lineal no nula es la ecuación de un hiperplano.

Dos hiperplanos son paralelos si y sólo si no tienen puntos comunes, lo que fácilmente se traduce en que sus ecuaciones tienen los miembros izquierdos proporcionales pero el término derecho no respeta la proporción.

En particular, una recta en un espacio afín bidimensional está formada por los puntos cuyas coordenadas en un sistema de referencia dado satisfacen una ecuación de la forma $ax + by = c$.

Un plano en un espacio afín tridimensional está formado por los puntos cuyas coordenadas en un sistema de referencia dado satisfacen una ecuación de la forma $ax + by + cz = d$. Una recta se puede expresar a su vez como intersección de dos planos, luego sus puntos son aquellos cuyas coordenadas satisfacen un sistema de dos ecuaciones lineales independientes y compatibles

$$\left. \begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \end{aligned} \right\}$$

En general, los puntos de una variedad de dimensión m en un espacio afín de dimensión k están caracterizados por que sus coordenadas satisfacen un sistema de $n - m$ ecuaciones lineales independientes.

Veamos ahora el modo en que una afinidad transforma las coordenadas afines de los puntos. Sea $f : E \rightarrow F$ una afinidad, sea $(O; \vec{v}_1, \dots, \vec{v}_n)$ un sistema de referencia en E y $(O'; \vec{w}_1, \dots, \vec{w}_m)$ un sistema de referencia en F . Sean $A = (a_1, \dots, a_m)$ las coordenadas de $f(O)$ en O' , sea M la matriz asociada a la aplicación lineal \vec{f} en las bases de los sistemas de referencia. Sea P un punto en E con coordenadas $X = (x_1, \dots, x_n)$. Éstas son por definición las coordenadas de \overrightarrow{OP} en la base \vec{v}_i , luego las coordenadas de $\vec{f}(\overrightarrow{OP})$ en la base \vec{w}_i son XM . Puesto que

$$f(P) = f(O) + \vec{f}(\overrightarrow{OP}) = O' + \overrightarrow{O'f(O)} + \vec{f}(\overrightarrow{OP}),$$

vemos que las coordenadas de $\overrightarrow{O'f(P)}$ en la base \vec{w}_i son

$$Y = A + XM.$$

En resumen:

La relación entre las coordenadas X de un punto en un sistema de referencia O y las coordenadas Y de su imagen por una afinidad f es $Y = A + XM$, donde A son las coordenadas de $f(O)$ en el sistema O' y M es la matriz de \vec{f} .

Recíprocamente, es fácil ver que toda ecuación matricial $Y = A + XM$ está asociada a una única afinidad en unos sistemas de referencia prefijados.

El sumando A puede suprimirse si en el segundo espacio tomamos como origen de coordenadas el punto $O' = f(O)$. Sin embargo, cuando f es una afinidad de un espacio en sí mismo, resulta más conveniente considerar un único sistema de referencia, y entonces sólo podemos eliminar el sumando A si f tiene un punto fijo, es decir, un punto O tal que $f(O) = O$.

Espacios afines canónicos Observemos que hasta ahora no hemos puesto ningún ejemplo de espacio afín. Sin embargo, lo cierto es que existen espacios afines de cualquier dimensión sobre cualquier cuerpo K . En efecto, basta tomar como conjunto de puntos $E = K^n$ y como espacio vectorial $\vec{E} = K^n$, de modo que la suma de puntos y vectores sea la misma que la suma de vectores. Es inmediato comprobar que $(K^n, K^n, +)$ así definido cumple trivialmente la definición de espacio afín n -dimensional.

Las variedades lineales de K^n son de la forma $P + W$, donde W es un subespacio de K^n , es decir, son las clases de congruencia módulo los subespacios de K^n .

El *sistema de referencia canónico* en K^n es el formado por $(0, \dots, 0)$ como origen y la base canónica de K^n . Respecto a este sistema, cada punto coincide con sus propias coordenadas.

Más aún, es claro que si E es cualquier espacio afín n -dimensional sobre K , la aplicación que a cada punto le asigna sus coordenadas respecto a un sistema de referencia prefijado es una biyección afín (cuya expresión coordenada respecto a este sistema y al sistema canónico de K^n es simplemente $Y = X$).

Así pues, todo espacio afín n -dimensional sobre K es isomorfo a K^n (en el sentido de que existe una biyección afín entre ambos), luego todos los espacios afines sobre K tienen todas las mismas propiedades afines (las mismas propiedades que sean invariantes por biyecciones afines).

En estos términos, la exposición de la geometría analítica en el apéndice A de [IGE] o de [IC] se basa en sustituir la noción general de espacio afín por el caso concreto de \mathbb{R}^2 o \mathbb{R}^3 como espacio afín canónico de dimensión 2 o 3 sobre el cuerpo \mathbb{R} , en el que los puntos y los vectores “son lo mismo”. Ahora bien, el lector debería reflexionar sobre el hecho de que identificar los puntos de un espacio afín con los elementos de K^n es un tecnicismo similar al que adoptamos cuando identificamos los números naturales con conjuntos, como $2 = \{\{\emptyset\}\}$, o cuando identificamos los números enteros con conjuntos de pares de números naturales, o como cuando identificamos los números reales con conjuntos de números racionales o de sucesiones de Cauchy. En otras palabras, en el mismo sentido en que podemos decir que “en verdad” los números naturales, o los enteros, o los racionales, o los reales no son conjuntos, de modo que no tiene sentido preguntar cuáles son sus elementos, también podemos decir que “en verdad” ni los puntos ni los vectores de un espacio afín son n -tuplas de elementos de un cuerpo. Cada vector sólo puede ser identificado mediante una n -tupla desde el momento en que fijamos una base del espacio vectorial, y la n -tupla será distinta según la base que fijemos, aunque el vector sea siempre el mismo. Y a su vez, sólo podemos identificar un punto mediante una n -tupla desde el momento en que fijamos un sistema de referencia afín. En el espacio afín canónico existe un sistema de referencia canónico, pero “en verdad” no existe tal cosa. Un plano es lo que vemos cuando miramos una hoja de papel (haciendo abstracción de sus bordes) y en ella vemos puntos, pero no coordenadas. Sólo podemos “ver” coordenadas a partir del momento en que trazamos unos ejes y

fijamos una unidad de medida, pero si dos personas trazan ejes diferentes en el mismo plano, identificarán un mismo punto con pares de coordenadas distintas, y nadie podrá decir que las “coordenadas de verdad” del punto son unas u otras.

Coordenadas baricéntricas Si $(O; \vec{e}_1, \dots, \vec{e}_n)$ es un sistema de referencia de un espacio afín E , los puntos $A_0 = O$, $A_1 = O + \vec{e}_1, \dots, A_n = O + \vec{e}_n$ son afínmente independientes y, recíprocamente, un conjunto de $n + 1$ puntos afínmente independientes A_0, \dots, A_n determina el sistema de referencia $(A_0; \overrightarrow{A_0A_1}, \dots, \overrightarrow{A_0A_n})$. Por lo tanto todo punto P de E se expresa de forma única como

$$P = A_0 + x_1 \overrightarrow{A_0A_1} + \dots + x_n \overrightarrow{A_0A_n}.$$

En algunas ocasiones esta expresión no es satisfactoria, pues sitúa al punto A_0 en una situación asimétrica respecto de los otros, cuando en realidad la noción de puntos afínmente independientes es completamente simétrica. Por ello, a veces conviene considerar las coordenadas baricéntricas que definiremos a continuación.

Notemos en primer lugar, que para puntos cualesquiera A_1, \dots, A_n y escalares $\lambda_1 + \dots + \lambda_n = 0$, el vector

$$\sum_{i=1}^n \lambda_i \overrightarrow{OA_i}$$

es independiente de la elección del punto O , pues si tomamos otro punto O' vemos que

$$\sum_{i=1}^n \lambda_i \overrightarrow{O'A_i} - \sum_{i=1}^n \lambda_i \overrightarrow{OA_i} = \sum_{i=1}^n \lambda_i \overrightarrow{OO'} = \vec{0}.$$

De aquí se desprende que si $\lambda_1 + \dots + \lambda_n = \lambda \neq 0$ el punto

$$B = O + \frac{1}{\lambda} \sum_{i=1}^n \lambda_i \overrightarrow{OA_i}$$

es independiente de O . Al punto B se le llama *baricentro* de (A_i, λ_i) y lo representaremos por

$$\frac{1}{\lambda} (\lambda_1 A_1 + \dots + \lambda_n A_n) = O + \frac{1}{\lambda} \sum_{i=1}^n \lambda_i \overrightarrow{OA_i}.$$

Las sumas y productos del miembro izquierdo son meramente formales, sin un significado intrínseco. Tan sólo sugieren que las coordenadas del baricentro en el sistema de referencia $(O; \overrightarrow{OA_i})$ se obtienen multiplicando por λ_i las coordenadas de A_i , sumando y dividiendo entre λ . El baricentro recibe este nombre porque se corresponde con el centro de gravedad de un sistema de n partículas puntuales situadas en los puntos A_i con masas λ_i (véase la sección [An A.1]).

Retomando ahora el conjunto afínmente independiente A_0, \dots, A_n , ahora notamos que un punto P se puede expresar en la forma

$$P = A_0 + \sum_{i=1}^n \lambda_i \overrightarrow{A_0A_i} = \sum_{i=0}^n \lambda_i A_i,$$

donde

$$\lambda_0 = 1 - \sum_{i=1}^n \lambda_i.$$

Los números $(\lambda_0, \dots, \lambda_n)$ (sujetos a la condición $\lambda_0 + \dots + \lambda_n = 1$) están unívocamente determinados por P y se llaman *coordenadas baricéntricas* de P respecto al sistema (A_i) .

Teorema 3.18 *Los puntos A_0, \dots, A_n son afínmente independientes si y sólo si ninguno de ellos es un baricentro de los demás.*

DEMOSTRACIÓN: Si uno de los puntos, por ejemplo A_0 es un baricentro de los demás, entonces

$$A_0 = \sum_{i=1}^n \lambda_i A_i, \quad \text{con} \quad \sum_{i=1}^n \lambda_i = 1.$$

Alguno de los coeficientes ha de ser no nulo, digamos que $\lambda_n \neq 0$. Así

$$A_0 = A_n + \sum_{i=1}^{n-1} \lambda_i \overrightarrow{A_n A_i},$$

donde no todos los coeficientes son nulos, luego $\overrightarrow{A_n A_0}$ es combinación lineal de los restantes. Invertiendo el razonamiento tenemos la otra implicación. ■

3.3 Espacios euclídeos

Nos ocupamos ahora de extender la estructura de espacio afín para recoger el concepto de congruencia, el último concepto geométrico primitivo que nos queda por representar analíticamente. Para ello empezamos trabajando sintéticamente (en las mismas condiciones que en la sección 3.1) con el fin de encontrar una caracterización algebraica de las congruencias de segmentos y ángulos. Por simplicidad suponemos los axiomas de tridimensionalidad.

Sea R el cuerpo de los números reales que se corresponden con puntos de las rectas graduadas. Hemos demostrado que R tiene que ser un cuerpo ordenado pitagórico.

Por otra parte notamos que si \vec{v} y \vec{w} son dos vectores no nulos cualesquiera, O es un punto, $A = O + \vec{v}$ y $B = O + \vec{w}$, entonces el ángulo \widehat{AOB} (si está definido) es independiente de O , pues el triángulo \widehat{AOB} (fijada una unidad de medida) tiene lados $\|\vec{v}\|$, $\|\vec{w}\|$ y $\|\vec{v} - \vec{w}\|$, luego los triángulos construidos a partir de puntos O distintos son congruentes, y en particular el ángulo indicado es el mismo.

El ángulo \widehat{AOB} no está definido cuando los vectores \vec{v} y \vec{w} son linealmente dependientes (lo cual tampoco depende de O). Esto nos permite definir el *ángulo* entre dos vectores no nulos como el ángulo \widehat{AOB} si los puntos no están alineados, mientras que si $\vec{w} = \alpha \vec{v}$ con $\alpha < 0$ consideraremos que el ángulo es

llano y si $\alpha > 0$ diremos que el ángulo que forman es nulo y tiene medida 0. Cuando dos vectores forman un ángulo recto se dice que son *ortogonales* (gr. ‘en ángulo recto’).

Consideremos un sistema de referencia de origen O y con una base $\vec{e}_1, \vec{e}_2, \vec{e}_3$ formada por vectores ortogonales y de norma 1. Entonces, un punto arbitrario P cumplirá que

$$P = O + x\vec{e}_1 + y\vec{e}_2 + z\vec{e}_3,$$

para ciertos números reales (x, y, z) (sus coordenadas en el sistema de referencia indicado). Si $Q = O + x\vec{e}_1 + y\vec{e}_2$, vemos que la recta OQ está contenida en el plano XY , mientras que la recta QP es paralela al eje Z , que es perpendicular al plano XY (suponiendo que $Q \neq O$ y $Q \neq P$). Por lo tanto el triángulo \widehat{OQP} es rectángulo, y el teorema de Pitágoras nos da que

$$\|\vec{OP}\|^2 = \|\vec{OQ}\|^2 + \|\vec{QP}\|^2 = \|x\vec{e}_1 + y\vec{e}_2\|^2 + \|z\vec{e}_3\|^2.$$

Si $Q = O$ o $Q = P$ la conclusión es trivialmente cierta. Del mismo modo, si llamamos $R = O + x\vec{e}_1$ tenemos que el triángulo \widehat{OQR} es rectángulo (o bien $Q = O$ o $Q = R$), con lo que

$$\|\vec{OP}\|^2 = \|x\vec{e}_1\|^2 + \|y\vec{e}_2\|^2 + \|z\vec{e}_3\|^2 = x^2 + y^2 + z^2,$$

luego $\|\vec{OP}\| = \sqrt{x^2 + y^2 + z^2}$. Puesto que P es un punto arbitrario, tenemos que si un vector \vec{v} tiene coordenadas (x, y, z) respecto a una base formada por vectores ortogonales y unitarios, entonces

$$\|\vec{v}\| = \sqrt{x^2 + y^2 + z^2}.$$

Si \vec{v} y \vec{w} son dos vectores arbitrarios cuyas coordenadas en una base en las condiciones anteriores son respectivamente (x, y, z) y (x', y', z') , definimos su *producto escalar* como

$$\vec{v}\vec{w} = xx' + yy' + zz'.$$

En estos términos hemos probado que $\|\vec{v}\| = \sqrt{\vec{v}\vec{v}}$. Es claro además que se cumplen las propiedades siguientes:

1. $\vec{u}(\vec{v} + \vec{w}) = \vec{u}\vec{v} + \vec{u}\vec{w}$,
2. $(\vec{u} + \vec{v})\vec{w} = \vec{u}\vec{w} + \vec{v}\vec{w}$,
3. $(\alpha\vec{u})\vec{v} = \alpha(\vec{u}\vec{v}) = \vec{u}(\alpha\vec{v})$,
4. $\vec{u}\vec{v} = \vec{v}\vec{u}$.

Consideremos ahora dos vectores no nulos \vec{v} y \vec{w} . Supongamos en primer lugar que son linealmente independientes. Sean $A = O + \vec{v}$, $B = O + \vec{w}$. Entonces los puntos O, A, B no son colineales, luego forman un triángulo cuyos lados miden $\|\vec{OA}\| = \|\vec{v}\|$, $\|\vec{OB}\| = \|\vec{w}\|$, $\|\vec{AB}\| = \|\vec{w} - \vec{v}\|$. Por otra parte

$$\|\vec{AB}\|^2 = (\vec{w} - \vec{v})(\vec{w} - \vec{v}) = \vec{w}\vec{w} + \vec{v}\vec{v} - 2\vec{v}\vec{w} = \|\vec{OA}\|^2 + \|\vec{OB}\|^2 - 2\vec{v}\vec{w}.$$

Si comparamos con el teorema del coseno concluimos que

$$\vec{v}\vec{w} = \|\vec{v}\| \|\vec{w}\| \cos \widehat{v\vec{w}}.$$

Esta expresión sigue siendo válida cuando $\vec{w} = \alpha\vec{v}$ si convenimos en que el coseno de un ángulo nulo es 1 y el coseno de un ángulo llano es -1 . En efecto:

$$\vec{v}\vec{w} = \alpha\vec{v}\vec{v} = \pm|\alpha| \|\vec{v}\|^2 = \pm\|\vec{v}\| \|\vec{w}\|.$$

En particular esto prueba que el producto escalar de dos vectores no depende del sistema de referencia que elegimos para calcularlo, siempre y cuando los vectores de su base sean ortogonales y de norma 1 (en particular, sí que depende de la elección de una unidad de longitud).

Estas consideraciones condicionan ya la estructura algebraica que hemos de imponer a un espacio afín para definir en él unas relaciones de congruencia que satisfagan los axiomas de la geometría euclídea:

Definición 3.19 Un *espacio vectorial euclídeo* V es un espacio vectorial de dimensión finita sobre un cuerpo ordenado pitagórico R sobre el que hay definido un *producto escalar*, que es una aplicación $V \times V \rightarrow R$ que cumple las propiedades siguientes:

1. $\vec{v}\vec{v}$ es un cuadrado (en particular $\vec{v}\vec{v} \geq 0$) y $\vec{v}\vec{v} = 0$ si y sólo si $\vec{v} = 0$,
2. $\vec{u}(\vec{v} + \vec{w}) = \vec{u}\vec{v} + \vec{u}\vec{w}$,
3. $(\alpha\vec{u})\vec{v} = \alpha(\vec{u}\vec{v}) = \vec{u}(\alpha\vec{v})$,
4. $\vec{u}\vec{v} = \vec{v}\vec{u}$.

Un *espacio afín euclídeo* es un espacio afín E sobre un cuerpo pitagórico R tal que en \vec{E} hay definido un producto escalar que lo dota de estructura de espacio euclídeo. La *dimensión* de E se define como la dimensión de \vec{E} .

Definimos la *norma* de un vector en un espacio euclídeo como $\|\vec{v}\| = \sqrt{\vec{v}\vec{v}}$. Notemos que la raíz cuadrada existe en R por la propiedad 1. Hay que entender que tomamos la raíz cuadrada positiva.

La *distancia* entre dos puntos A y B de un espacio afín euclídeo será $\|\overrightarrow{AB}\|$. Diremos que dos vectores \vec{v} y \vec{w} son *ortogonales* si $\vec{v}\vec{w} = 0$. Lo representaremos por $\vec{v} \perp \vec{w}$. Un conjunto de vectores es *ortogonal* si no contiene al 0 y sus elementos son ortogonales dos a dos. Un conjunto de vectores es *ortonormal* si es ortogonal y todos sus vectores tienen norma 1.

Teorema 3.20 Si R es un cuerpo ordenado pitagórico, V es un R -espacio vectorial y $\vec{e}_1, \dots, \vec{e}_n$ es una base de V , entonces existe un único producto escalar en V respecto al cual la base dada es ortonormal.

DEMOSTRACIÓN: Supongamos que tenemos un producto escalar en V respecto al cual la base dada es ortonormal. Sean \vec{v} y \vec{w} dos vectores de V y sean $(x_1, \dots, x_n), (y_1, \dots, y_n)$ sus coordenadas en la base. Entonces

$$\vec{v}\vec{w} = \sum_{i=1}^n x_i \vec{e}_i \sum_{j=1}^n y_j \vec{e}_j = \sum_{i,j=1}^n x_i y_j (\vec{e}_i \vec{e}_j) = x_1 y_1 + \dots + x_n y_n.$$

Esto prueba que $\vec{v}\vec{w}$ está completamente determinado por la base, lo cual nos da la unicidad. Para probar la existencia basta observar que si definimos

$$\vec{v}\vec{w} = x_1 y_1 + \dots + x_n y_n,$$

donde $(x_1, \dots, x_n), (y_1, \dots, y_n)$ son las coordenadas de \vec{v} y \vec{w} en la base dada, entonces se cumplen todas las propiedades de la definición de producto escalar y la base dada es ortonormal. Notemos en particular que

$$\vec{v}\vec{v} = x_1^2 + \dots + x_n^2$$

es un cuadrado en R porque R es pitagórico (y es en este punto donde se requiere esta hipótesis). ■

En particular, si R es cualquier cuerpo ordenado pitagórico, podemos definir el *producto escalar canónico* en R^n como el producto escalar dado por

$$(x_1, \dots, x_n)(y_1, \dots, y_n) = x_1 y_1 + \dots + x_n y_n,$$

respecto del cual la base canónica de R^n es ortonormal.

Hemos probado que el espacio euclídeo en el sentido de la geometría sintética es un espacio afín euclídeo tridimensional en el sentido analítico. Probaremos que el recíproco también es cierto: todo espacio afín euclídeo tridimensional cumple los axiomas de la geometría euclídea.

En esta dirección, empezamos observando que la definición algebraica de espacio euclídeo no muestra directamente la existencia de bases ortonormales. Probar esto será nuestro primer objetivo. Para ello conviene notar algunos hechos elementales.

Se cumple que $\vec{0}\vec{v} = (\vec{0}\vec{v})\vec{v} = 0(\vec{0}\vec{v}) = 0$.

Si un conjunto de vectores $\{\vec{v}_1, \dots, \vec{v}_n\}$ es ortogonal, entonces es linealmente independiente.

En efecto, si $\alpha_1 \vec{v}_1 + \dots + \alpha_m \vec{v}_m = \vec{0}$ entonces

$$0 = \vec{v}_j \vec{0} = \alpha_1 \vec{v}_j \vec{v}_1 + \dots + \alpha_m \vec{v}_j \vec{v}_m = \alpha_j \vec{v}_j \vec{v}_j,$$

luego $\alpha_j = 0$.

Teorema 3.21 (Teorema de ortogonalización de Gram-Schmidt) *Si V es un espacio vectorial euclídeo, todo conjunto ortogonal de vectores de V se extiende hasta una base ortogonal de V .*

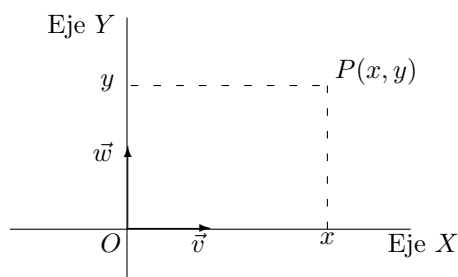
DEMOSTRACIÓN: Sea $A = \{\vec{v}_1, \dots, \vec{v}_r\}$ un conjunto de vectores ortogonales (y por lo tanto linealmente independientes). Si A es ya una base de V no hay nada que probar. En caso contrario tomamos un vector \vec{v} de V que no esté en $\langle A \rangle$. Consideremos un vector de la forma

$$\vec{w} = \vec{v} - \sum_{i=1}^r \alpha_i \vec{v}_i.$$

Claramente $\vec{v}_i \vec{w} = \vec{v}_i \vec{v} - \alpha_i \vec{v}_i \vec{v}_i$, luego si tomamos cada α_i como el único escalar que cumple $\vec{v}_i \vec{v} - \alpha_i \vec{v}_i \vec{v}_i = 0$ tenemos que $\vec{w} \perp A$. Así obtenemos un conjunto ortogonal con un vector más. Repitiendo el proceso llegamos hasta un conjunto ortogonal con tantos vectores como la dimensión de V , que será una base ortogonal. ■

La norma verifica la relación $\|\alpha \vec{v}\| = |\alpha| \|\vec{v}\|$ y si $\vec{v} \neq \vec{0}$ podemos calcular $\vec{u} = \frac{1}{\|\vec{v}\|} \vec{v}$, y resulta que $\|\vec{u}\| = 1$. Por lo tanto, si en una base ortogonal dividimos cada vector por su norma obtenemos una base ortonormal. Así pues, todo espacio vectorial euclídeo V tiene una base ortonormal.

Nota Las coordenadas de un punto respecto de un sistema de referencia ortonormal tienen una interpretación especialmente simple. En el caso bidimensional, si tenemos un sistema de referencia de origen O y una base ortonormal \vec{v} , \vec{w} , los puntos de coordenadas $(x, 0)$ forman una recta que se suele llamar eje X o ‘eje de abscisas’, mientras que los puntos con coordenadas $(0, y)$ forman una recta perpendicular que se suele llamar eje Y o ‘eje de ordenadas’, y entonces el punto P de coordenadas (x, y) es la intersección de la recta paralela al eje de ordenadas que pasa por $(x, 0)$ y la recta paralela al eje de abscisas que pasa por $(0, y)$. Además x e y son las distancias a O de los puntos $(x, 0)$ y $(0, y)$, respectivamente:



Igualmente podemos interpretar las coordenadas (x, y, z) de un punto del espacio respecto a un sistema de referencia ortonormal en el espacio tridimensional. ■

Los resultados del comienzo de la sección prueban que si en un espacio euclídeo arquimediano (en el sentido sintético) fijamos un sistema de referencia con base ortonormal y a cada punto le asignamos sus coordenadas respecto a dicho sistema, entonces el producto escalar se corresponde con el producto

escalar canónico que hemos definido en R^3 , luego la ortogonalidad y la distancia entre puntos se corresponden con los conceptos homónimos en R^3 definidos analíticamente.

El teorema siguiente es obvio desde un punto de vista sintético, pero es necesario para probar analíticamente las propiedades de la distancia entre puntos y la ortogonalidad de vectores.

Teorema 3.22 (Desigualdad de Cauchy-Schwarz) *Si \vec{v} y \vec{w} son dos vectores de un espacio vectorial euclídeo, entonces $|\vec{v}\vec{w}| \leq \|\vec{v}\| \|\vec{w}\|$ y se da la igualdad si y sólo si \vec{v} y \vec{w} son linealmente dependientes.*

DEMOSTRACIÓN: Si $\vec{w} = \vec{0}$ la igualdad se da trivialmente. Supongamos que $\vec{w} \neq \vec{0}$. Llamando $\vec{u} = \vec{w}/\|\vec{w}\|$, entonces $\|\vec{u}\| = 1$ y lo que hemos de probar es que $|\vec{v}\vec{u}| \leq \|\vec{v}\|$, y que se da la igualdad si y sólo si \vec{u} y \vec{v} son linealmente dependientes. Tomemos $\lambda \in R$. Entonces

$$0 \leq \|\vec{v} + \lambda\vec{u}\|^2 = (\vec{v} + \lambda\vec{u})(\vec{v} + \lambda\vec{u}) = \vec{v}\vec{v} + \lambda^2\vec{u}\vec{u} + 2\lambda\vec{v}\vec{u} = \|\vec{v}\|^2 + \lambda^2 + 2\lambda\vec{v}\vec{u},$$

En particular, si tomamos $\lambda = -\vec{v}\vec{u}$ queda que $\|\vec{v}\|^2 + (\vec{v}\vec{u})^2 - 2(\vec{v}\vec{u})^2 \geq 0$, o sea, $(\vec{v}\vec{u})^2 \leq \|\vec{v}\|^2$, luego $|\vec{v}\vec{u}| \leq \|\vec{v}\|$.

Además la igualdad se da si y sólo si $\|\vec{v} - (\vec{v}\vec{u})\vec{u}\| = 0$, o sea, si y sólo si $\vec{v} = (\vec{v}\vec{u})\vec{u}$. Por lo tanto, si se da la igualdad \vec{u} y \vec{v} son linealmente dependientes. Recíprocamente, si $\vec{v} = \lambda\vec{u}$, entonces $|\vec{v}\vec{u}| = |\lambda\vec{u}\vec{u}| = |\lambda| = \|\vec{v}\|$. ■

Ahora podemos probar las propiedades esenciales de la norma euclídea:

Teorema 3.23 *En todo espacio euclídeo se cumple:*

1. $\|\vec{v}\| \geq 0$ y $\|\vec{v}\| = 0$ si y sólo si $\vec{v} = \vec{0}$.
2. $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$.
3. $\|\alpha\vec{v}\| = |\alpha| \|\vec{v}\|$.

DEMOSTRACIÓN: La única propiedad que no es inmediata es la segunda, pero

$$\begin{aligned} \|\vec{v} + \vec{w}\|^2 &= (\vec{v} + \vec{w})(\vec{v} + \vec{w}) = \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\vec{v}\vec{w} \leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2|\vec{v}\vec{w}| \\ &\leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\|\vec{v}\| \|\vec{w}\| = (\|\vec{v}\| + \|\vec{w}\|)^2. \end{aligned}$$

Por lo tanto $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$. ■

La propiedad 2 del teorema anterior es la versión analítica de la desigualdad triangular. Es fácil ver que es una igualdad exactamente cuando uno de los vectores es un múltiplo positivo del otro.

Ahora nos ocupamos de las aplicaciones que conservan la estructura euclídea. La definición natural es la siguiente:

Definición 3.24 Una *inmersión isométrica lineal* $f : V \rightarrow W$ entre dos espacios vectoriales euclídeos sobre un mismo cuerpo R es una aplicación lineal que además cumple $f(\vec{v}_1)f(\vec{v}_2) = \vec{v}_1\vec{v}_2$, para todos los vectores $\vec{v}_1, \vec{v}_2 \in V$.

En particular, f cumple $\|f(\vec{v})\| = \|\vec{v}\|$ y esto a su vez implica que el núcleo de f es trivial, luego toda inmersión isométrica es inyectiva. Si además es biyectiva se dice que es una *isometría lineal* entre los espacios vectoriales dados.

Teorema 3.25 Sea $f : V \rightarrow W$ una aplicación lineal entre dos espacios vectoriales euclídeos sobre un mismo cuerpo R , sean $\vec{e}_1, \dots, \vec{e}_n$ y $\vec{f}_1, \dots, \vec{f}_n$ bases ortonormales de V y W respectivamente y sea M la matriz de f en dichas bases. Las afirmaciones siguientes son equivalentes:

1. f es una isometría lineal.
2. $f(\vec{e}_1), \dots, f(\vec{e}_n)$ es una base ortonormal de W .
3. $MM^t = I$ (la matriz identidad).

DEMOSTRACIÓN: $1 \Rightarrow 2$ se sigue inmediatamente de la definición de isometría lineal. Para probar $2 \Rightarrow 3$ basta tener en cuenta que las filas de M son las coordenadas en la base de W de los vectores $f(\vec{e}_1), \dots, f(\vec{e}_n)$, por lo que MM^t tiene en la posición i, j el producto $f(\vec{e}_i)f(\vec{e}_j)$.

Para probar $3 \Rightarrow 1$ observamos que si dos vectores \vec{v}_1 y $\vec{v}_2 \in V$ tienen coordenadas $x, y \in R^n$, entonces

$$f(\vec{v}_1)f(\vec{v}_2) = (xM)(yM) = xMM^ty^t = xy^t = \vec{v}_1\vec{v}_2,$$

luego f es una isometría lineal. ■

En particular vemos que dos espacios vectoriales euclídeos sobre un mismo cuerpo pitagórico R y de la misma dimensión son linealmente isométricos, pues cualquier isomorfismo entre ellos que transforme una base ortonormal en otra es una isometría lineal. En particular todo espacio vectorial euclídeo de dimensión n es linealmente isométrico³ al espacio canónico R^n .

Ahora cabría esperar que tomáramos como definición de isometría entre dos espacios afines euclídeos como una aplicación afín $f : E \rightarrow F$ cuya aplicación lineal asociada $\vec{f} : \vec{E} \rightarrow \vec{F}$ es una isometría lineal, pues éstas son las aplicaciones que conservan toda la estructura del espacio E . Sin embargo, vamos a probar que una definición más débil es en realidad equivalente:

Definición 3.26 Sean A y B dos subconjuntos de dos espacios afines euclídeos. Una *isometría* entre A y B es una aplicación biyectiva $f : A \rightarrow B$ tal que para todo par de puntos $P, Q \in A$ se cumple $d(P, Q) = d(f(P), f(Q))$.

³Quizá el lector se pregunte entonces para qué hemos definido el concepto de espacio afín euclídeo en lugar de trabajar directamente con los espacios R^n , si al final cualquier otro es isomorfo a uno de éstos. La razón es que, por ejemplo, si L es una variedad lineal en un espacio afín euclídeo E , entonces L es un espacio afín euclídeo con el espacio vectorial \vec{L} y la restricción del producto escalar. Así podemos tratar a cada variedad lineal como espacio afín euclídeo sin que los conceptos geométricos asociados a ella dependan de la elección de un isomorfismo en particular con un espacio R^n .

Tomemos un sistema de referencia del primer espacio con origen en un punto $O \in A$ y un sistema de referencia en el segundo espacio con origen en $O' = f(O)$.

Por simplificar la notación usaremos un apóstrofo para representar las imágenes por f , es decir, si $P \in A$, entonces P' representará a $f(P)$. Llamemos \tilde{A} al conjunto de los vectores \overrightarrow{OP} , con $P \in A$. Similarmente, sea \tilde{B} el conjunto de los vectores $\overrightarrow{O'P'}$ con $P' \in B$. Si $\vec{v} = \overrightarrow{OP} \in \tilde{A}$, llamaremos $\vec{v}' = \overrightarrow{O'P'} \in \tilde{B}$.

Así, $\vec{0} = \overrightarrow{OO} \in \tilde{A}$, $\vec{0}' = \overrightarrow{O'O'} \in \tilde{B}$ y si $\vec{v}, \vec{w} \in \tilde{A}$, se cumple $\|\vec{v} - \vec{w}\| = \|\vec{v}' - \vec{w}'\|$, pues si $\vec{v} = \overrightarrow{OA}$, $\vec{w} = \overrightarrow{OB}$, entonces

$$\|\vec{v} - \vec{w}\| = \|\overrightarrow{BA}\| = \|\overrightarrow{B'A'}\| = \|\vec{v}' - \vec{w}'\|.$$

En particular, tomando $\vec{w} = \vec{0}$ vemos que $\|\vec{v}\| = \|\vec{v}'\|$. En otras palabras, si f conserva las distancias, también conserva las normas. Más aún, la relación

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 - 2\vec{v}\vec{w}$$

implica que $\vec{v}\vec{w} = \vec{v}'\vec{w}'$, para todo par de vectores $\vec{v}, \vec{w} \in \tilde{A}$, es decir, que f conserva el producto escalar.

Veamos que si $\vec{v} \in \tilde{A}$ y $\alpha\vec{v} \in \tilde{A}$ entonces $(\alpha\vec{v})' = \alpha\vec{v}'$. Si $\vec{v} = \vec{0}$ es trivial. En caso contrario

$$|\vec{v}'(\alpha\vec{v})'| = |\vec{v}(\alpha\vec{v})| = |\alpha| \|\vec{v}\|^2 = \|\vec{v}'\| \|\alpha\vec{v}'\| = \|\vec{v}'\| \|(\alpha\vec{v})'\|.$$

Como la desigualdad de Cauchy-Schwarz es en este caso una igualdad, tenemos que $(\alpha\vec{v})' = \lambda\vec{v}'$, para cierto escalar λ . Multiplicando ambos miembros por \vec{v}' obtenemos $\alpha\|\vec{v}'\|^2 = \lambda\|\vec{v}'\|^2$, luego $\lambda = \alpha$, y se cumple lo afirmado.

Ahora probamos que si $\vec{u}, \vec{v}, \vec{u} + \vec{v} \in \tilde{A}$, entonces $(\vec{u} + \vec{v})' = \vec{u}' + \vec{v}'$. Tenemos

$$\|\vec{u}' + \vec{v}'\|^2 = \|\vec{u}'\|^2 + \|\vec{v}'\|^2 + 2\vec{u}'\vec{v}' = \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\vec{u}\vec{v} = \|\vec{u} + \vec{v}\|^2,$$

luego $\|\vec{u}' + \vec{v}'\| = \|\vec{u} + \vec{v}\|$. Por otra parte

$$(\vec{u} + \vec{v})'(\vec{u}' + \vec{v}') = (\vec{u} + \vec{v})'\vec{u}' + (\vec{u} + \vec{v})'\vec{v}' = (\vec{u} + \vec{v})\vec{u} + (\vec{u} + \vec{v})\vec{v} = \|\vec{u} + \vec{v}\|^2.$$

Reuniendo ambas igualdades $(\vec{u} + \vec{v})'(\vec{u}' + \vec{v}') = \|(\vec{u} + \vec{v})'\| \|\vec{u}' + \vec{v}'\|$. De nuevo la desigualdad de Cauchy-Schwarz es una igualdad, lo que implica que $\vec{u}' + \vec{v}' = \lambda(\vec{u} + \vec{v})'$. Consecuentemente,

$$\|\vec{u} + \vec{v}\|^2 = (\vec{u} + \vec{v})'(\vec{u}' + \vec{v}') = \lambda\|\vec{u}' + \vec{v}'\|^2 = \lambda\|\vec{u} + \vec{v}\|^2.$$

Esto implica $\lambda = 1$ (y por lo tanto la relación que queremos probar) salvo si $\vec{u} + \vec{v} = \vec{0}$, pero en este caso tenemos $\|\vec{u}' + \vec{v}'\| = 0$, luego $\vec{u}' + \vec{v}' = \vec{0} = (\vec{u} + \vec{v})'$.

Consideremos los subespacios vectoriales $W = \langle \tilde{A} \rangle$ y $W' = \langle \tilde{B} \rangle$. Tomamos una base de W contenida en \tilde{A} y definimos $\vec{f}: W \rightarrow W'$ como la aplicación lineal que sobre la base es $\vec{f}(\vec{v}) = \vec{v}'$. Los resultados que acabamos de probar

justifican que esta relación vale para todos los vectores de \tilde{A} . Ahora observamos que

$$\vec{f}(\vec{u})\vec{f}(\vec{v}) = \vec{u}\vec{v}, \quad \text{para todo } \vec{u}, \vec{v} \in W.$$

En efecto, basta tener en cuenta que esto es cierto para vectores de \tilde{A} y usar la linealidad de \vec{f} y del producto escalar. Esto significa que \vec{f} es una isometría lineal.

Consideramos ahora las variedades lineales $L = O + W$ y $L' = O' + W'$. Es claro que se trata de las menores variedades lineales que contienen a A y B respectivamente. Hemos probado que tienen la misma dimensión. Más aún, la afinidad $\tilde{f} : L \rightarrow L'$ dada por $\tilde{f}(P) = O' + \vec{f}(\overrightarrow{OP})$ es una isometría que extiende a f . Esta isometría es única, pues si g fuera otra, le aplicamos todo el razonamiento anterior, tomando ahora $A = L$, con lo que $\tilde{A} = W$. Con ello probamos que la aplicación $\vec{u} \mapsto \vec{u}'$ es lineal y coincide con \vec{f} en el conjunto \tilde{A} original (el generador de L), luego coincide con \vec{f} en L , luego g coincide con \tilde{f} .

El teorema siguiente recoge lo que hemos obtenido hasta ahora:

Teorema 3.27 *Toda isometría entre dos subconjuntos de dos espacios afines euclídeos se extiende a una única isometría entre las variedades lineales que generan. Además la extensión es una biyección afín cuya aplicación lineal asociada \vec{f} es una isometría lineal entre los espacios directores de las variedades.*

Ahora veamos que es posible extender la isometría a todo el espacio, aunque perdemos la unicidad. Para ello conviene introducir el concepto siguiente:

Definición 3.28 Dado un subespacio W de un espacio vectorial euclídeo V , se llama *complemento ortogonal* de W al espacio

$$W^\perp = \{\vec{v} \in V \mid \vec{v} \perp \vec{w} \text{ para todo } \vec{w} \in W\}.$$

Es claro que W^\perp es un subespacio vectorial, así como que $W \cap W^\perp = 0$. Tomemos una base ortogonal $\vec{v}_1, \dots, \vec{v}_r$ de W y extendámosla hasta una base ortogonal $\vec{v}_1, \dots, \vec{v}_n$ de V . Entonces es claro que $W^\perp = \langle \vec{v}_{r+1}, \dots, \vec{v}_n \rangle$, luego $V = W \oplus W^\perp$.

En general, si V_1, \dots, V_r son espacios vectoriales cuyos elementos son ortogonales entre sí, es fácil ver que su suma es directa, y en tal caso diremos que tienen *suma ortogonal* y la representaremos por $V_1 \perp \dots \perp V_r$. Hemos probado que $V = W \perp W^\perp$. Con todo esto ya podemos concluir:

Teorema 3.29 *Sean E y F dos espacios afines euclídeos de la misma dimensión n . Entonces*

1. *Las isometrías de E en F son las biyecciones afines f tales que \vec{f} es una isometría lineal.*
2. *Toda isometría entre un subconjunto A de E y un subconjunto B de F se extiende a una isometría de E en F .*
3. *Si A contiene un conjunto de $n + 1$ puntos afínmente independientes, entonces la extensión es única.*

DEMOSTRACIÓN: Sean $\langle A \rangle = O + \vec{W}$ y $\langle B \rangle = O' + \vec{W}'$, de modo que f se extienda a una isometría afín sobre $O + \vec{W}$. En particular W y W' tienen la misma dimensión, luego sus complementos ortogonales también. Extendamos \vec{f} a V (el espacio vectorial de E) asignando a una base ortonormal de W^\perp una base ortonormal de W'^\perp . Por el teorema 3.25 la extensión (que seguiremos llamando \vec{f}) es una isometría lineal, y la afinidad que induce en E es una isometría y extiende a f . ■

Así hemos demostrado un hecho no trivial, y es que cualquier biyección entre dos espacios afines euclídeos que conserve las distancias es, de hecho, una biyección afín que conserva el producto escalar.

A través de las isometrías podemos definir una noción general de congruencia de conjuntos arbitrarios que coincide con la que ya conocemos en el caso de segmentos, ángulos y triángulos.

Definición 3.30 Diremos que dos conjuntos A y B en un espacio afín euclídeo E son *congruentes* si existe una isometría $f : E \rightarrow E$ tal que $f[A] = B$. Los puntos de A y B correspondientes por (una isometría en particular) f se llaman puntos *homólogos*.

En el sentido de la geometría axiomática, dos segmentos son congruentes si y sólo si tienen la misma longitud, pero, si esto es así, una biyección entre los conjuntos de sus extremos es una isometría que se extiende a todo el espacio, y es fácil ver que transforma uno de los segmentos en el otro, luego ambos son congruentes en el sentido de la definición anterior. El recíproco es análogo.

Similarmente, si un ángulo \widehat{AOB} es congruente con otro de vértice O' , podemos tomar puntos en sus lados de modo que éste sea $\widehat{A'O'B'}$ y además $\overline{OA} \equiv \overline{O'A'}$, $\overline{OB} \equiv \overline{O'B'}$. Entonces se cumplirá también que $\overline{AB} \equiv \overline{A'B'}$, luego la correspondencia entre los conjuntos $\{A, O, B\}$ y $\{A', O', B'\}$ es una isometría, que se extiende a todo el espacio, y de nuevo es fácil ver que hace corresponder los ángulos, luego son semejantes en el sentido de la definición anterior. De igual modo se prueba el recíproco y la equivalencia correspondiente a triángulos.

Teorema 3.31 Si R es un cuerpo, todo espacio afín euclídeo E sobre R de dimensión ≥ 2 cumple los axiomas de incidencia de la geometría euclídea (incluyendo el axioma de las paralelas y los de tridimensionalidad si la dimensión es 3). Si R es un cuerpo ordenado, entonces E cumple también los axiomas de ordenación. Los axiomas de congruencia se cumplen si y sólo si R es pitagórico (y consideramos cualquier producto escalar en \vec{E}). La propiedad arquimediada se cumple si y sólo si $R \subset \mathbb{R}$ y el de continuidad se cumple si y sólo si $R = \mathbb{R}$.

DEMOSTRACIÓN: Sólo falta probar lo referente a los axiomas de congruencia. El axioma C1 se cumple trivialmente, pues es claro que la composición de isometrías es una isometría y que la inversa de una isometría es una isometría. Esto implica inmediatamente que las congruencias son relaciones de equivalencia.

También es fácil demostrar C3: si se cumple $A - B - C$ y $A' - B' - C'$ y $\overline{AB} \equiv \overline{A'B'}$, $\overline{BC} \equiv \overline{B'C'}$, esto significa que $\|\overline{AB}\| = \|\overline{A'B'}\|$, $\|\overline{BC}\| = \|\overline{B'C'}\|$ y que $\overline{AB} = \lambda\overline{AC}$, con $0 < \lambda < 1$ y $\overline{A'B'} = \lambda'\overline{A'C'}$, con $0 < \lambda' < 1$.

Ahora bien, $\overline{BC} = \overline{AC} - \overline{AB} = (1 - \lambda)\overline{AC}$, luego $\|\overline{BC}\| = (1 - \lambda)\|\overline{AC}\|$, luego $\lambda = 1 - \|\overline{BC}\|/\|\overline{AC}\|$, y razonando igual con los otros tres puntos llegamos a que $\lambda = \lambda'$.

Por consiguiente, $\|\overline{AC}\| = (1/\lambda)\|\overline{AB}\| = (1/\lambda)\|\overline{A'B'}\| = \|\overline{A'C'}\|$, de donde se sigue que $\overline{AC} \equiv \overline{A'C'}$.

Para probar C2 tomamos dos puntos $A \neq B$ y una semirrecta, que será de la forma $s = \{A' + \lambda\vec{v} \mid \lambda \in R, \lambda \geq 0\}$. Basta observar que el punto

$$B' = A' + \frac{\|\overline{AB}\|}{\|\vec{v}\|} \vec{v} \in s$$

cumple lo requerido por C2.

Para probar C4 consideremos en primer lugar dos semiplanos cualesquiera, que serán conjuntos de la forma

$$\pi = \{O' + \lambda\vec{v}_1 + \mu\vec{v}_2 \mid \mu \geq 0\} \quad \text{y} \quad \pi' = \{O' + \lambda\vec{w}_1 + \mu\vec{w}_2 \mid \mu \geq 0\},$$

donde podemos suponer que las bases (\vec{v}_1, \vec{v}_2) y (\vec{w}_1, \vec{w}_2) son ortonormales. Si llamamos \vec{f} a la aplicación lineal determinada por $\vec{f}(\vec{v}_i) = \vec{w}_i$, entonces \vec{f} es una isometría lineal, y es claro que la aplicación dada por $f(P) = O' + \vec{f}(\overline{OP})$ es una isometría entre los planos que contienen a π y a π' , tal que $f[\pi] = \pi'$ y que además transforma la semirrecta $O + \{\lambda\vec{v}_1 \mid \lambda \geq 0\}$, en la semirrecta $O' + \{\lambda\vec{w}_1 \mid \lambda \geq 0\}$.

De aquí se sigue que dados dos semiplanos y en sus fronteras sendas semirrectas, existe una isometría que transforma uno en el otro haciendo corresponder las semirrectas. Por lo tanto, si tenemos un ángulo, un semiplano y en su frontera una semirrecta (según el axioma C4) existe una isometría entre el semiplano que contiene al ángulo con un lado en su frontera y el semiplano dado, de modo que el lado del ángulo se corresponda con la semirrecta dada. La imagen del ángulo por la isometría indicada será un ángulo congruente con el dado, contenido en el semiplano dado y con un lado igual a la semirrecta dada, tal y como exige el axioma C4.

Para probar la unicidad suponemos que tenemos dos ángulos con un lado en común y contenidos en el mismo semiplano respecto a éste. Serán de la forma $A(O; \vec{v}_1, \vec{v}_2)$ y $A(O; \vec{v}_1, \vec{v}'_2)$. Además ambos están contenidos en el semiplano

$$\{O + \lambda\vec{v}_1 + \mu\vec{w} \mid \mu \geq 0\},$$

lo que se traduce en que $\vec{v}_2 = x\vec{v}_1 + y\vec{w}$ con $y > 0$ y $\vec{v}'_2 = x'\vec{v}_1 + y'\vec{w}$ con $y' > 0$. Podemos suponer que \vec{v}_1 , \vec{v}_2 y \vec{v}'_2 tienen todos norma 1.

Una isometría f entre los ángulos ha de cumplir⁴ $\vec{f}(\vec{v}_1) = \vec{v}_1$ y $\vec{f}(\vec{v}_2) = \vec{v}_2$ o bien $\vec{f}(\vec{v}_1) = \vec{v}_2$ y $\vec{f}(\vec{v}_2) = \vec{v}_1$. En cualquier caso tendremos $\vec{v}_1\vec{v}_2 = \vec{v}_1\vec{v}_2'$, luego $x^2 = x'^2$. Por otro lado,

$$x^2 + y^2 = \|\vec{v}_2\|^2 = \|\vec{v}_2'\|^2 = x'^2 + y'^2,$$

y como $y, y' > 0$, de hecho $y = y'$. Desarrollando igualmente la condición $\|\vec{v}_2 - \vec{v}_1\|^2 = \|\vec{v}_2' - \vec{v}_1\|^2$ se concluye $x = x'$, con lo que los ángulos son iguales.

La comprobación de C5 es similar. ■

También tenemos probado el recíproco:

Teorema 3.32 *Si \mathcal{E} es una geometría tridimensional euclídea arquimediana y R es el subcuerpo de \mathbb{R} formado por los números que se corresponden con los puntos de una recta graduada, entonces la aplicación $\mathcal{E} \rightarrow R^3$ que a cada punto le asigna sus coordenadas respecto de un sistema de referencia ortonormal hace corresponder todos los conceptos geométricos de \mathcal{E} con los análogos en R^3 definidos analíticamente.*

En particular la geometría que hemos definido analíticamente en \mathbb{R}^3 es salvo isomorfismo la única geometría tridimensional euclídea (en el sentido sintético) que satisface el axioma de continuidad. Todo lo dicho vale igualmente para la geometría bidimensional con los cambios obvios.

Es importante destacar que en el capítulo II hemos definido una geometría euclídea (tridimensional) como un conjunto de puntos con una serie de estructuras asociadas que cumple unas propiedades, pero hasta este momento no habíamos demostrado que realmente existieran ejemplos de tales conjuntos. Ahora ya sabemos que sí que existen, y que esencialmente sólo hay uno para cada cuerpo arquimediano pitagórico, y sólo uno (el espacio afín euclídeo canónico \mathbb{R}^3) si incluimos el axioma de continuidad.

3.4 Circunferencias

En un plano euclídeo no tiene por qué cumplirse el axioma de las circunferencias. Para ello el cuerpo de escalares R tiene que cumplir una condición más fuerte que ser pitagórico. Antes de mostrarlo vamos a calcular la expresión analítica de una circunferencia.

Dado un plano afín euclídeo y fijado en él un sistema de referencia ortonormal, la circunferencia de centro un punto (a, b) y radio $r > 0$ está formada por los puntos (x, y) cuya distancia a (a, b) es igual a r , es decir, $\|(x - a, y - b)\| = r$, o equivalentemente:

$$(x - a)^2 + (y - b)^2 = r^2.$$

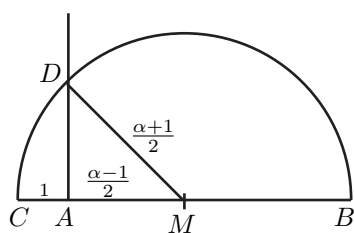
⁴El punto más delicado es comprobar que si $\widehat{AOB} = \widehat{A'O'B'}$ entonces $OA = O'A'$ y $OB = O'B'$ o bien $OA = O'B'$ y $OB = O'A'$. Por ejemplo, puede probarse que si una recta corta a un ángulo en un segmento, entonces los extremos del segmento están sobre los lados. De aquí se sigue fácilmente la propiedad indicada.

Desarrollando la ecuación, vemos que las circunferencias están formadas por los puntos cuyas coordenadas satisfacen una ecuación de la forma

$$x^2 + y^2 + Ax + By + C = 0, \quad (*)$$

donde $A = -2a$, $B = -2b$, $C = a^2 + b^2 - r^2$.

Supongamos ahora que en un plano euclídeo se cumple, no ya el axioma de las circunferencias, sino el teorema 1.90 sobre intersección de rectas y circunferencias, que es una consecuencia de dicho axioma.



Entonces, si $\alpha \in R$, consideramos un segmento AB de longitud α , tomamos un punto $C-A-B$ tal que $CA = 1$, calculamos el punto medio M de CB y consideramos la circunferencia de centro M y radio MB . Entonces el punto A es interior a la circunferencia, luego la perpendicular a CB por A la corta en un punto D , y un simple cálculo con el teorema de Pitágoras muestra que la longitud de AD es $\sqrt{\alpha}$.

Definición 3.33 Un cuerpo ordenado R es *euclídeo* si todo elemento positivo tiene raíz cuadrada.

Sabemos que \mathbb{R} es euclídeo, y es obvio que todo cuerpo euclídeo es pitagórico. Acabamos de probar parte del teorema siguiente:

Teorema 3.34 Sea E un plano afín euclídeo sobre un cuerpo pitagórico R . Las afirmaciones siguientes son equivalentes:

1. R es euclídeo.
2. E cumple el axioma de intersección de rectas y circunferencias, es decir, que si una recta pasa por un punto interior a una circunferencia entonces la corta al menos en un punto.
3. E cumple el axioma de las circunferencias, es decir, que si una circunferencia pasa por un punto interior y por un punto exterior de otra, entonces ambas tienen un punto en común.

DEMOSTRACIÓN: Sabemos que $3 \Rightarrow 2$ (es el teorema 1.90) y acabamos de probar que $2 \Rightarrow 1$. Sólo falta probar que $1 \Rightarrow 3$. Para ello consideramos dos circunferencias ω y ω' de radios $r, s \in R$. Podemos suponer que $0 < s < r$. Consideramos un sistema de referencia ortonormal $(O; \vec{v}, \vec{w})$ tal que O sea el centro de ω y $\vec{v} = \overrightarrow{OP} / \|\overrightarrow{OP}\|$, donde P es el centro de ω' . De este modo, las coordenadas de los centros son $(0, 0)$ y $(a, 0)$, para cierto $a \in R$, $a > 0$. Los puntos de ω son aquellos cuyas coordenadas cumplen la ecuación $x^2 + y^2 = r^2$, mientras que los puntos de ω' son los que cumplen la ecuación $(x - a)^2 + y^2 = s^2$.

Las circunferencias se cortarán si existe un par $(x, y) \in R^2$ que cumpla ambas ecuaciones, ya que entonces el punto con dichas coordenadas estará en la intersección.

La segunda ecuación es $x^2 + y^2 - 2ax + a^2 = s^2$, y sustituyendo la primera queda $r^2 - 2ax + a^2 = s^2$. Por lo tanto, tiene que ser

$$x = \frac{r^2 - s^2 + a^2}{2a} > 0.$$

La coordenada y debe cumplir

$$y^2 = r^2 - x^2 = \frac{4a^2r^2 - (r^2 - s^2 + a^2)^2}{4a^2}.$$

Es fácil ver que si y cumple esta ecuación, entonces (x, y) cumple ambas y las circunferencias se cortan. Por lo tanto, basta probar que el miembro derecho de la ecuación anterior es positivo, pues entonces tendrá raíz cuadrada y porque R es euclídeo. A su vez, basta probar que

$$4a^2r^2 \geq (r^2 - s^2 + a^2)^2,$$

lo cual equivale a que $2ar \geq r^2 - s^2 + a^2$, o también a que $(r - a)^2 \leq s^2$, o a que $|r - a| \leq s$, o a que $a - s \leq r \leq a + s$. Ahora bien, es fácil ver que si $r < a - s$ entonces los círculos limitados por las circunferencias son disjuntos, por lo que ninguna puede pasar por un punto interior a la otra, y si $a + s < r$ todos los puntos de ω' son interiores a ω , por lo que ω' no pasa por puntos exteriores de ω ni ω por puntos interiores de ω' . ■

3.5 Trigonometría

Terminamos este capítulo estudiando las razones trigonométricas desde el punto de vista de la geometría analítica. En primer lugar vamos a observar que las razones trigonométricas de un ángulo definidas en 2.57 tienen una caracterización analítica sencilla que es válida incluso en un plano afín euclídeo sobre un cuerpo pitagórico cualquiera R , no necesariamente arquimediano:

Definición 3.35 Sea V un espacio vectorial euclídeo sobre un cuerpo ordenado pitagórico R y sean $\vec{v}, \vec{w} \in V$ dos vectores no nulos. Definimos

$$\cos(\vec{v}, \vec{w}) = \frac{\vec{v}\vec{w}}{\|\vec{v}\|\|\vec{w}\|} \in R.$$

La desigualdad de Cauchy-Schwarz implica que $-1 \leq \cos(\vec{v}, \vec{w}) \leq 1$ y toma los valores ± 1 si y sólo si \vec{v} y \vec{w} son linealmente dependientes. Concretamente, toma el valor 1 si uno es un múltiplo positivo del otro y el valor -1 si es un múltiplo negativo. Además el coseno vale 0 si y sólo si $\vec{v} \perp \vec{w}$.

Otras propiedades obvias son que el coseno de dos vectores no depende del orden en que se consideren, ni se altera si los cambiamos por múltiplos positivos, así como que el coseno es invariante por isometrías lineales, es decir, que si $f : V \rightarrow V$ es una isometría lineal, entonces $\cos(\vec{v}, \vec{w}) = \cos(f(\vec{v}), f(\vec{w}))$.

Si E es un espacio afín euclídeo sobre un cuerpo ordenado pitagórico R , y P, O, Q son tres puntos no colineales en E , podemos definir el coseno del ángulo $\hat{A} = \widehat{POQ}$ como

$$\cos(\hat{A}) = \cos(\overrightarrow{OP}, \overrightarrow{OQ}),$$

que claramente no depende de la elección de los puntos P y Q en los lados del ángulo \hat{A} y es invariante por isometrías. Por lo tanto, dos ángulos congruentes tienen el mismo coseno y está definido $\cos \alpha$ cuando α es una clase de congruencia de ángulos, incluyendo los casos de la clase 0 de los ángulos nulos (las semirrectas) y la clase π de los ángulos llanos (los semiplanos), para los que $\cos 0 = 1$ y $\cos \pi = -1$. Es fácil probar la versión analítica del teorema del coseno:

Teorema 3.36 *Si $P \neq O \neq Q$ son tres puntos de un espacio afín euclídeo, se cumple que*

$$\|\overrightarrow{PQ}\|^2 = \|\overrightarrow{OP}\|^2 + \|\overrightarrow{OQ}\|^2 - 2\|\overrightarrow{OP}\| \|\overrightarrow{OQ}\| \cos \hat{A}(O; \overrightarrow{OP}, \overrightarrow{OQ}).$$

DEMOSTRACIÓN: Basta calcular

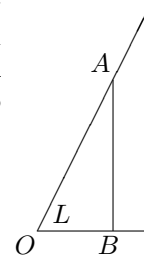
$$\|\overrightarrow{PQ}\|^2 = (\overrightarrow{OQ} - \overrightarrow{OP})(\overrightarrow{OQ} - \overrightarrow{OP}) = \|\overrightarrow{OP}\|^2 + \|\overrightarrow{OQ}\|^2 - 2\overrightarrow{OP} \cdot \overrightarrow{OQ}$$

y aplicar la definición del coseno. ■

El teorema de Pitágoras es un caso particular del teorema anterior.

Ahora es fácil justificar que el coseno que acabamos de definir es el mismo definido sintéticamente en 2.57. Como allí, dado un ángulo agudo L arbitrario de vértice O , tomamos un punto A en uno de sus lados y llamamos B al pie de la perpendicular al otro lado por A . Entonces, según la definición que hemos dado aquí,

$$\cos L = \frac{\overrightarrow{OA} \cdot \overrightarrow{OB}}{\|\overrightarrow{OA}\| \|\overrightarrow{OB}\|} = \frac{(\overrightarrow{OB} + \overrightarrow{BA}) \cdot \overrightarrow{OB}}{\|\overrightarrow{OA}\| \|\overrightarrow{OB}\|} = \frac{\|\overrightarrow{OB}\|^2}{\|\overrightarrow{OA}\| \|\overrightarrow{OB}\|} = \frac{\|\overrightarrow{OB}\|}{\|\overrightarrow{OA}\|},$$



luego se trata del mismo valor considerado en 2.57.

Si L es recto tenemos que $\cos L = 0$ tanto con la definición que hemos dado aquí como respecto a 2.57, y si L es obtuso entonces en 2.57 definimos $\cos L$ como el coseno del ángulo suplementario cambiado de signo, pero dicho ángulo complementario es el formado por \overrightarrow{OA} y $-\overrightarrow{OB}$, por lo que

$$\cos L = \cos(\overrightarrow{OA}, \overrightarrow{OB}) = -\cos(\overrightarrow{OA}, -\overrightarrow{OB})$$

coincide también con el definido en 2.57. Si definimos

$$\text{sen } L = \frac{\|\overrightarrow{AB}\|}{\|\overrightarrow{OA}\|} = \sqrt{1 - \cos^2 L} \in R$$

obtenemos el mismo seno definido en 2.57, y es invariante por isometrías por serlo el coseno, luego también podemos hablar de $\text{sen } \alpha$ cuando α es una clase de congruencia de ángulos. Notemos que $\text{sen } \alpha \in [0, 1]$.

Teorema 3.37 *Dos ángulos en un espacio afín euclídeo son congruentes si y sólo si tienen el mismo coseno.*

DEMOSTRACIÓN: Es claro que si dos ángulos son congruentes y uno es nulo o llano, lo mismo le sucede al otro, luego el teorema se cumple claramente en ambos casos. A partir de aquí consideramos un ángulo L que no sea nulo ni llano.

Fijemos un punto O y dos vectores ortogonales unitarios $\overrightarrow{OP_1}$, $\overrightarrow{OP_2}$, que determinan el semiplano S cuya frontera es la semirrecta l de origen O que contiene a P_1 y tal que $P_2 \in S$.

Hemos probado que los espacios afines euclídeos verifican el axioma C4, luego L es congruente a un único ángulo con un lado igual a l y el otro contenido en S . Este segundo lado contendrá un único punto P_L tal que $\|\overrightarrow{OP_L}\| = 1$, de modo que $L \equiv L'$ si y sólo si $P_L = P_{L'}$. Ahora bien, si expresamos

$$\overrightarrow{OP_L} = x\overrightarrow{OP_1} + y\overrightarrow{OP_2},$$

el hecho de que $P_L \in S$ implica que $y > 0$, y multiplicando por $\overrightarrow{OP_1}$ vemos que $x = \cos L$, luego $y = \sin L$. Por lo tanto, si $\cos L = \cos L'$, se cumple también que $\sin L = \sin L'$, luego $P_L = P_{L'}$, luego $L \equiv L'$. El recíproco es claro. ■

En términos de clases tenemos que $\cos \alpha = \cos \beta$ si y sólo si $\alpha = \beta$.

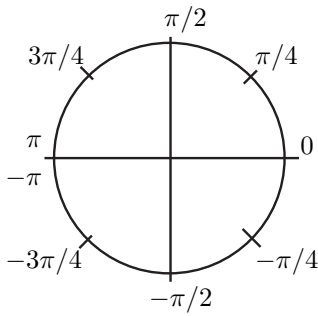
A partir de este punto nos restringimos al caso en que el cuerpo de escalares R es arquimediano, con lo que, fijado un ángulo unitario U , cada clase de congruencia de ángulos está determinada por la amplitud común $\alpha = L/U$ de cualquiera de los ángulos L de los que consta (definición 2.25), que es un número real en el intervalo $[0, \pi]$, donde π es ahora la amplitud de los ángulos llanos (respecto de la unidad U), aunque en principio no todo número en dicho intervalo tiene por qué ser la amplitud de un ángulo. Esto sucede en el caso en que $R = \mathbb{R}$.

Definición 3.38 Una *circunferencia graduada* es una circunferencia ω , de centro O y radio r , en la que se han señalado dos puntos P_0 y $P_{\pi/2}$, de modo que el ángulo $\widehat{P_0OP_{\pi/2}}$ sea recto. En tal caso los vectores

$$\vec{e}_1 = \frac{\overrightarrow{OP_0}}{\|\overrightarrow{OP_0}\|}, \quad \vec{e}_2 = \frac{\overrightarrow{OP_{\pi/2}}}{\|\overrightarrow{OP_{\pi/2}}\|}$$

son una base ortonormal y, recíprocamente, toda base ortonormal \vec{e}_1, \vec{e}_2 determina una graduación en cada circunferencia ω , a saber la dada por $P_0 = O + r\vec{e}_1$ y $P_{\pi/2} = O + r\vec{e}_2$.

Si P es un punto de ω con coordenadas (x, y) respecto del sistema de referencia ortonormal asociado a la graduación, definimos su *argumento* como la amplitud $\theta \in [0, \pi]$ del ángulo $\widehat{P_0OP}$ si $y \geq 0$, mientras que si $y \leq 0$ su argumento se define como $-\theta \in [-\pi, 0]$.



Con esto tenemos bien definido el argumento sobre todos los puntos de ω excepto $(-r, 0)$, el punto diametralmente opuesto a $P_0 = (r, 0)$, cuyo argumento debería ser por una parte π y por otra parte $-\pi$. Resolvemos esto asignando a cada punto, no un argumento, sino una clase de argumentos módulo 2π , es decir, si θ es el argumento de P según lo acabamos de definir, ahora definimos $\arg P = \{\theta + 2k\pi \mid k \in \mathbb{Z}\}$. De este modo, $\arg(-r, 0) = \{\pi + 2k\pi \mid k \in \mathbb{Z}\}$, que contiene tanto a π como a $-\pi$, y la definición resulta consistente. Tenemos así una aplicación inyectiva entre la circunferencia ω y el \mathbb{Z} -módulo cociente $\mathbb{R}/2\pi\mathbb{R}$. Si suponemos el axioma de continuidad, la aplicación $\arg : \omega \rightarrow \mathbb{R}/2\pi\mathbb{R}$ es biyectiva. Intuitivamente, el argumento de un punto indica el ángulo que hay que girar desde P_0 para llegar hasta él. Así, el hecho de que un mismo punto tenga argumentos $\pi/2$, $-3\pi/2$ y $5\pi/2$ significa que se llega hasta él girando un recto en un sentido, o tres rectos en sentido opuesto, o dando una vuelta entera de 4 rectos más otro recto. La figura muestra algunos argumentos.

De acuerdo con la definición de argumento, si un punto $P \in \omega$ de coordenadas (x, y) tiene un argumento $\theta \in [0, \pi]$, entonces

$$\cos \theta = \frac{x}{r}, \quad \text{sen } \theta = \frac{y}{r},$$

o, equivalentemente, $(x, y) = (r \cos \theta, r \text{sen } \theta)$, mientras que si el argumento está en $[-\pi, 0]$, entonces $(x, y) = (r \cos(-\theta), -r \text{sen}(-\theta))$. Por ello, si extendemos las funciones trigonométricas al intervalo $[-\pi, \pi]$ estableciendo que

$$\cos(-\theta) = \cos \theta, \quad \text{sen}(-\theta) = -\text{sen } \theta,$$

podemos afirmar que si un punto $P \in \omega$ tiene argumento $\theta \in [-\pi, \pi]$, entonces sus coordenadas son $(r \cos \theta, r \text{sen } \theta)$. A su vez, si extendemos las funciones trigonométricas hasta funciones $\text{sen}, \cos : \mathbb{R} \rightarrow [-1, 1]$ mediante el convenio

$$\text{sen}(\theta + 2k\pi) = \text{sen } \theta, \quad \cos(\theta + 2k\pi) = \cos \theta \quad \text{para todo } k \in \mathbb{Z},$$

podemos afirmar que si θ es cualquier argumento de $P \in \omega$, entonces sus coordenadas son $(r \cos \theta, r \text{sen } \theta)$.

En general, las *coordenadas polares* respecto de un sistema de referencia ortonormal $(O; \vec{e}_1, \vec{e}_2)$ de un punto $P \neq O$ se definen como su *módulo* $\rho = \overrightarrow{OP}$ y sus *argumentos* θ , que son los números reales que cumplen que las coordenadas de P son $(x, y) = (\rho \cos \theta, \rho \text{sen } \theta)$.

Ahora el resultado básico es el siguiente:

Teorema 3.39 Sean P y Q dos puntos en ω , sean $\arg P = [\alpha]$, $\arg Q = [\beta]$ de modo que $|\beta - \alpha| \leq \pi$. Entonces el ángulo \widehat{POQ} tiene amplitud $|\beta - \alpha|$.

DEMOSTRACIÓN: Notemos en general que cada clase módulo 2π tiene un único representante en el intervalo $[-\pi, \pi]$ (excepto π , que tiene dos). Esto implica que el valor de $|\beta - \alpha|$ menor o igual que π está unívocamente determinado por P y Q y no importa con qué representantes concretos lo calculamos. En principio tomamos α y β en $[-\pi, \pi]$.

Sean $I = (1, 0)$, $I' = (-1, 0)$. Si P y Q están ambos en el semiplano superior, podemos suponer que $0 \leq \alpha \leq \beta \leq \pi$. Descartando casos triviales tenemos que \widehat{IOP} está contenido en \widehat{IOQ} y sus amplitudes son α y β , luego es claro que la de \widehat{POQ} es $\beta - \alpha$. El mismo argumento vale si P y Q están ambos en el semiplano inferior.

Supongamos que P está en el semiplano superior y Q está en el inferior. Si Q es diametralmente opuesto a P , entonces \widehat{IOQ} es adyacente a \widehat{IOP} , luego sus amplitudes α y $-\beta$ suman π , como había que probar.

Sea P' el punto diametralmente opuesto a P . Podemos suponer que P es distinto de I, I' . Entonces la recta PP' deja a I y a I' en semiplanos distintos. Supongamos que Q e I están en el mismo semiplano. Entonces es claro que $\widehat{POQ} = \widehat{POI} + \widehat{IOQ}$ y de nuevo la suma de las amplitudes es $\alpha - \beta < \pi$.

Finalmente consideramos el caso en que Q e I' están en el mismo semiplano. Entonces $\widehat{POQ} = \widehat{POI'} + \widehat{QOI'}$. Las amplitudes de los sumandos son $\pi - \alpha$ y $\pi + \beta$, luego la amplitud de \widehat{POQ} es $2\pi + \beta - \alpha$. Si tomamos $2\pi + \beta$ como argumento de Q , se cumple lo pedido. ■

Isometrías lineales de un plano euclídeo Para estudiar las funciones trigonométricas conviene analizar primero las isometrías de un plano euclídeo E . En realidad aquí nos limitaremos a estudiar las isometrías $f : E \rightarrow E$ con un punto fijo O , es decir, un punto tal que $f(O) = O$. Vamos a considerar únicamente sistemas de referencia ortonormales de origen O , de modo que un sistema de referencia está determinado por una base ortonormal \vec{e}_1, \vec{e}_2 de \vec{E} . Si $P \in E$, entonces $f(P) = O + \vec{f}(\overrightarrow{OP})$, donde \vec{f} es la isometría lineal asociada a f . Si M_f es la matriz de \vec{f} en la base \vec{e}_1, \vec{e}_2 , entonces las coordenadas de $f(P)$ vienen dadas por $(x, y)M_f$, donde (x, y) son las coordenadas de P .

Los primeros resultados que vamos a obtener son muy generales, y no necesitamos suponer que el cuerpo R sea arquimediano, sino que valen para cualquier cuerpo ordenado pitagórico.

Definición 3.40 Una isometría $f : E \rightarrow E$ en un plano euclídeo es una *simetría axial* o un *giro* si existe un punto O tal que $f(O) = O$ y su matriz en un sistema de referencia ortonormal $(O; \vec{e}_1, \vec{e}_2)$ es de la forma

$$M_f = \begin{pmatrix} a & b \\ \pm b & \mp a \end{pmatrix}, \quad \text{con } a^2 + b^2 = 1,$$

donde el signo que precede a a es positivo en el caso de un giro y negativo en el caso de una simetría.

Veamos en primer lugar que toda isometría f con un punto fijo O es una simetría axial o un giro. Para ello observamos que si $\vec{f}(\vec{e}_1) = a\vec{e}_1 + b\vec{e}_2$, tiene que cumplirse $\|\vec{f}(\vec{e}_1)\| = 1$, lo cual equivale a $a^2 + b^2 = 1$. A su vez, $\vec{f}(\vec{e}_2)$ tiene que ser un vector ortogonal a $\vec{f}(\vec{e}_1)$ y de norma 1, pero sólo hay dos vectores en esas condiciones, los de coordenadas $(-b, a)$ y $(b, -a)$, luego la matriz de f en la base considerada tiene que ser la correspondiente a una simetría axial o bien a un giro.

Ahora falta probar que una misma isometría no puede cumplir la definición de simetría o la de giro según el sistema de referencia que tomemos. Para ello empezamos estudiando las simetrías axiales. Si la matriz de f en un sistema de referencia es

$$M_f = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \quad \text{con } a^2 + b^2 = 1,$$

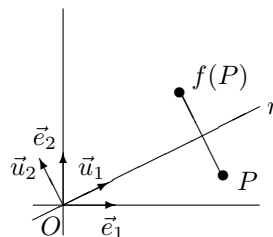
llamamos \vec{u}_1 y \vec{u}_2 a los vectores de coordenadas

$$\frac{1}{\sqrt{2-2a}}(b, 1-a), \quad \frac{1}{\sqrt{2-2a}}(a-1, b),$$

y es inmediato comprobar que forman una base ortonormal de \vec{E} y $f(\vec{u}_1) = \vec{u}_1$, $f(\vec{u}_2) = -\vec{u}_2$, por lo que la matriz de f en esta base es

$$M' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Esto implica que f deja invariantes a todos los puntos de la recta $r = O + \langle \vec{u}_1 \rangle$ y que, en general, si P no está en dicha recta, el segmento $\overline{Pf(P)}$ es perpendicular a r y r pasa por su punto medio. La recta r se llama *eje* de la simetría f . Es claro que cada recta del plano determina una única simetría axial que la tiene por eje.



Ahora supongamos que f es un giro, es decir, que admite una matriz de la forma

$$M_f = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \quad \text{con } a^2 + b^2 = 1.$$

Conviene observar que el producto de dos matrices de esta forma tiene esta forma, y que dos cualesquiera de ellas conmutan.

Un caso especial se da cuando $a = 1$ y, por consiguiente, $b = 0$, con lo que M_f es la matriz identidad y f es la aplicación identidad que fija a todos los puntos. Descartado este caso, consideremos un vector $\vec{v} \neq \vec{0}$ arbitrario, de coordenadas (x, y) . Entonces las coordenadas de $\vec{f}(\vec{v})$ son $(x, y)M$, luego

$$\vec{v} \vec{f}(\vec{v}) = (x, y)(ax - by, bx + ay) = ax^2 + ay^2 = a(x^2 + y^2),$$

luego $\cos(\vec{v}, \vec{f}(\vec{v})) = a$.

Así pues, si $P \neq O$, el punto $f(P)$ cumple que $\|\overrightarrow{Of(P)}\| = \|\overrightarrow{OP}\|$ y el ángulo $\widehat{POf(P)}$ es un ángulo constante α (independiente de P) tal que $\cos \alpha = a$. Por lo tanto, la matriz de f puede expresarse en la forma

$$M_{\pm\alpha} = \begin{pmatrix} \cos \alpha & \pm \operatorname{sen} \alpha \\ \mp \operatorname{sen} \alpha & \cos \alpha \end{pmatrix},$$

para cierto ángulo α .

En particular, si $a \neq 1$ y $P \neq O$, tenemos que los vectores \overrightarrow{OP} y $\overrightarrow{Of(P)}$ forman un ángulo α no nulo, luego tiene que ser $f(P) \neq P$. Así pues, un giro distinto de la identidad tiene a O como único punto fijo, mientras que para la identidad todos los puntos son fijos. Por el contrario, los puntos fijos de una simetría axial forman una recta, y estas propiedades son intrínsecas a la isometría, es decir, no dependen del sistema de referencia en el que consideramos su matriz. Por lo tanto, podemos concluir que toda isometría del plano con un punto fijo O es una simetría axial o un giro, pero no ambas cosas, que su matriz en cualquier sistema de referencia ortonormal de origen un punto fijo tendrá la forma correspondiente a la definición de giro o de simetría axial según sea el caso.

Ya hemos comprobado que la matriz de una simetría axial se simplifica considerablemente si tomamos un sistema de referencia en el que el eje de simetría sea uno de los ejes coordenados. En cambio, las matrices de los giros apenas se modifican al cambiar de sistema de referencia. En efecto, si un giro de centro O tiene matriz $M_{\pm\alpha}$ respecto de un sistema de referencia ortonormal de origen O , entonces tiene matriz $M_{\pm\alpha}$ (con el mismo α pero tal vez signo opuesto) respecto de cualquier otro sistema de referencia ortonormal de origen O , pues la matriz tiene que corresponder a un giro y el ángulo α está determinado por f . Notemos que si la matriz respecto de una base \vec{e}_1, \vec{e}_2 es M_α , la matriz respecto de la base $\vec{e}_1, -\vec{e}_2$ es $M_{-\alpha}$, luego el signo sí que depende de la elección de la base.

No obstante, fijado un sistema de referencia, podemos asociar a cada ángulo α el giro G_α de ángulo α (de matriz M_α) y el giro $G_{-\alpha}$ de ángulo $-\alpha$ (de matriz $M_{-\alpha}$). Son giros distintos salvo si $\alpha = 0$, en cuyo caso ambos son la aplicación identidad, o si $\alpha = \pi$, en cuyo caso ambos son la aplicación de matriz

$$M_{\pm\pi} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

de modo que cumple $G_{\pm\pi}(P) = O - \overrightarrow{OP}$, por lo que $G_{\pm\pi}$ se llama también simetría puntual respecto del origen O , pues O es el punto medio del segmento $\overline{PG_\pi(P)}$.

Observemos también que $M_{-\alpha} = M_\alpha^t = M_\alpha^{-1}$, por lo que G_α y $G_{-\alpha}$ son isometrías mutuamente inversas. Otro hecho elemental es el siguiente:

Teorema 3.41 *Si E es un plano euclídeo, $O \in E$ y \vec{u}, \vec{v} son dos vectores unitarios en \vec{E} , existe un único giro G de centro O tal que $\vec{G}(\vec{u}) = \vec{v}$.*

DEMOSTRACIÓN: Tomemos un sistema de referencia ortonormal $(O; \vec{e}_1, \vec{e}_2)$ con $\vec{e}_1 = \vec{u}$. Sea $\vec{v} = a\vec{e}_1 + b\vec{e}_2$. Como $\|\vec{v}\| = 1$ tiene que ser $a^2 + b^2 = 1$, luego un giro que cumpla lo requerido tiene que tener una matriz M_G cuya primera fila sea (a, b) , y esto fuerza a que su segunda fila sea $(-b, a)$, luego sólo hay un giro que cumpla lo pedido. ■

A partir de aquí volvemos al caso particular en que R es arquimediano para continuar estudiando las razones trigonométricas. Así cada clase de congruencia de ángulos está determinada por su amplitud $0 \leq \alpha \leq \pi$, y el convenio con el que hemos extendido las razones trigonométricas a amplitudes en el intervalo $[-\pi, \pi]$ nos permite expresar todo giro (respecto de un sistema de referencia ortonormal prefijado) en la forma G_α , con matriz

$$M_\alpha = \begin{pmatrix} \cos \alpha & \operatorname{sen} \alpha \\ -\operatorname{sen} \alpha & \cos \alpha \end{pmatrix},$$

donde ahora $\alpha \in [-\pi, \pi]$, por lo que ya no es necesario anteponer un posible signo negativo. Más en general, podemos admitir como α cualquier número real que sea la amplitud de un ángulo (no necesariamente en $[-\pi, \pi]$), de modo que $G_\alpha = G_{\alpha+2k\pi}$, para todo entero k . Ahora podemos probar lo siguiente:

Teorema 3.42 *Fijado un sistema de referencia ortonormal y llamando ρ_θ al punto P determinado por las coordenadas polares ρ y θ , el giro G_α cumple que $G_\alpha(\rho_\theta) = \rho_{\theta+\alpha}$.*

DEMOSTRACIÓN: Vamos a probar el teorema en primer lugar para planos euclídeos sobre el cuerpo \mathbb{R} . De este modo tenemos asegurado que todo número real es el argumento de un punto.

Sea $G : E \rightarrow E$ la aplicación definida por $G(O) = O$ y $G(\rho_\theta) = \rho_{\theta+\alpha}$ (aquí estamos usando que existe un punto de argumento $\theta + \alpha$). Veamos que es una isometría. En efecto, dados dos puntos P y Q , hemos de probar que $\|\overrightarrow{PQ}\| = \|\overrightarrow{G(Q)G(P)}\|$. Si $P = O$ el término izquierdo es la coordenada ρ de Q , mientras que el término derecho es la coordenada ρ de $G(Q)$, y ésta se conserva por definición. Si $P \neq O \neq Q$, entonces el teorema 3.39 prueba que $\widehat{POQ} = \widehat{G(P)OG(Q)}$ (incluyendo los valores posibles 0 y π si los puntos están alineados). Como también tenemos que $\|\overrightarrow{OP}\| = \|\overrightarrow{OG(P)}\|$ y $\|\overrightarrow{OQ}\| = \|\overrightarrow{OG(Q)}\|$, por el teorema del coseno tenemos que $\|\overrightarrow{PQ}\| = \|\overrightarrow{G(P)G(Q)}\|$.

En particular G es una biyección afín y, por definición, no tiene más punto fijo que O (salvo en el caso trivial $\alpha = 2k\pi$). Por consiguiente es un giro y, como $G(1_0) = 1_\alpha = (\cos \alpha, \operatorname{sen} \alpha)$, la primera fila de su matriz es precisamente $(\cos \alpha, \operatorname{sen} \alpha)$, luego ésta es necesariamente M_α y $G = G_\alpha$.

Para probar el caso general basta observar que, aunque R sea un subcuerpo de \mathbb{R} , las coordenadas de la imagen de ρ_θ por G_α vienen dadas por $(\rho \cos \theta, \rho \operatorname{sen} \theta)M_\alpha$, y el resultado no depende de si hacemos el cálculo en R o en \mathbb{R} . Por el caso ya probado, sabemos que dicho producto tiene que ser $(\rho \cos(\theta + \alpha), \rho \operatorname{sen}(\theta + \alpha))$, luego estas coordenadas tienen que estar en R^2 , luego en efecto existe el punto $\rho_{\theta+\alpha}$ y es $G_\alpha(\rho_\theta)$. ■

Nota La parte final de la prueba del teorema anterior implica que el conjunto $A_R \subset \mathbb{R}$ de los argumentos de los puntos de un plano euclídeo sobre R es un \mathbb{Z} -submódulo de \mathbb{R} , es decir, que la suma de dos argumentos es un argumento (que el opuesto también lo es era evidente desde un principio). En particular $A_{\mathbb{R}} = \mathbb{R}$. ■

Así pues, en términos de coordenadas polares, cada giro suma un valor constante α al argumento de cada punto. A partir de aquí es inmediato que $G_\alpha \circ G_\beta = G_{\alpha+\beta}$ para todos los $\alpha, \beta \in A_{\mathbb{R}}$. De aquí se sigue a su vez la relación $M_{\alpha+\beta} = M_\alpha M_\beta$, que a su vez implica las fórmulas para el seno y el coseno de una suma. El teorema siguiente recoge lo que tenemos probado (para el caso $R = \mathbb{R}$, que aquí es el más general):

Teorema 3.43 *Las funciones $\text{sen}, \text{cos} : \mathbb{R} \rightarrow [-1, 1]$ (respecto a una medida de ángulos prefijada en la que los ángulos llanos tengan amplitud π) verifican las propiedades siguientes:*

1. $\text{sen}(\alpha + 2k\pi) = \text{sen } \alpha, \quad \text{cos}(\alpha + 2k\pi) = \text{cos } \alpha, \quad \text{para todo } k \in \mathbb{Z}.$
2. $\text{sen } 0 = 0, \quad \text{cos } 0 = 1, \quad \text{sen}(\pi/2) = 1, \quad \text{cos}(\pi/2) = 0.$
3. $\text{sen}^2 \alpha + \text{cos}^2 \alpha = 1, \quad \text{sen}(-\alpha) = -\text{sen } \alpha, \quad \text{cos}(-\alpha) = \text{cos } \alpha.$
4. $\text{sen } \alpha \geq 0 \text{ si } \alpha \in [0, \pi], \quad \text{cos } \alpha \geq 0 \text{ si } \alpha \in [-\pi/2, \pi/2].$
5. *Para cada par $(x, y) \in \mathbb{R}^2$ tal que $x^2 + y^2 = 1$ existe un único $\alpha \in]-\pi, \pi]$ tal que $(x, y) = (\text{cos } \alpha, \text{sen } \alpha).$*
6. $\text{sen}(\alpha + \beta) = \text{sen } \alpha \text{cos } \beta + \text{cos } \alpha \text{sen } \beta,$
7. $\text{cos}(\alpha + \beta) = \text{cos } \alpha \text{cos } \beta - \text{sen } \alpha \text{sen } \beta.$

En particular, haciendo $\alpha = \beta$ obtenemos las fórmulas para los ángulos dobles:

$$\text{sen } 2\alpha = 2 \text{sen } \alpha \text{cos } \alpha, \quad \text{cos } 2\alpha = \text{cos}^2 \alpha - \text{sen}^2 \alpha.$$

De ellas se deducen sin esfuerzo las fórmulas para el ángulo mitad:

$$\text{sen } \frac{\alpha}{2} = \sqrt{\frac{1 - \text{cos } \alpha}{2}}, \quad \text{cos } \frac{\alpha}{2} = \sqrt{\frac{1 + \text{cos } \alpha}{2}}.$$

También es fácil deducir las fórmulas para las tangentes a partir de las correspondientes a los senos y los cosenos:

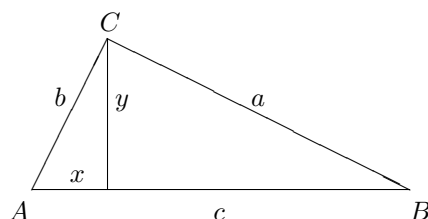
$$\tan(-\alpha) = -\tan \alpha, \quad \tan(\alpha + \beta) = \frac{\tan \alpha + \tan \beta}{1 - \tan \alpha \tan \beta}, \quad \tan 2\alpha = \frac{2 \tan \alpha}{1 - \tan^2 \alpha}.$$

Notemos que la tangente no está definida para $\alpha = \pm\pi/2$, pues el coseno se anula.

Veamos la demostración analítica de dos resultados básicos: que los ángulos de un triángulo suman dos rectos y el teorema de los senos:

Dado un triángulo \widehat{ABC} , tomando un sistema de referencia ortonormal con origen en A y un vector en la dirección de \overrightarrow{AB} podemos suponer que $A = (0, 0)$, $B = (c, 0)$, $C = (x, y)$, con $c, y \geq 0$. Se comprueba fácilmente que

$$\cos \hat{A} = \frac{x}{b}, \quad \text{sen } \hat{A} = \frac{y}{b}, \quad \cos \hat{B} = \frac{c-x}{a}, \quad \text{sen } \hat{B} = \frac{y}{a}.$$



Las fórmulas de la suma nos dan que

$$\begin{aligned} \text{sen}(\hat{A} + \hat{B}) &= \frac{yc}{ab} \geq 0, \\ \cos(\hat{A} + \hat{B}) &= \frac{cx - x^2 - y^2}{ab} = -\frac{\overrightarrow{AC} \cdot \overrightarrow{BC}}{ab} = -\cos \hat{C}. \end{aligned}$$

Estas relaciones implican que $0 \leq \hat{A} + \hat{B} \leq \pi$ y $\hat{A} + \hat{B} = \pi - \hat{C}$. Conservando la notación, tenemos

$$C = (b \cos \hat{A}, b \text{sen } \hat{A}), \quad B - C = (a \cos \hat{B}, -a \text{sen } \hat{B}),$$

y como $B = (c, 0)$, al sumar las segundas coordenadas queda

$$0 = b \text{sen } \hat{A} - a \text{sen } \hat{B},$$

que es una de las igualdades del teorema de los senos. La otra se prueba intercambiando los papeles los vértices (no entraremos en la relación con el radio de la circunferencia circunscrita). ■

Caracterización de las funciones trigonométricas Nuestra definición de las funciones seno y coseno se basa en la delicada construcción de la medida de ángulos, que no admite una formulación algebraica sencilla. Estas funciones pueden ser definidas mucho más fácilmente mediante técnicas analíticas (como en [ITAn 5.1]). Vamos a probar que cualquier par de funciones que cumplan las propiedades del teorema anterior son de hecho el seno y el coseno que hemos definido (respecto a una unidad de ángulos adecuada). Así tendremos justificado que nuestra definición geométrica equivale a cualquiera de las varias definiciones analíticas posibles.

Supongamos, pues, que tenemos dadas dos funciones $\text{sen}, \cos : \mathbb{R} \rightarrow [-1, 1]$ que satisfagan las propiedades del teorema anterior (respecto de cierto número real $\pi > 0$). Consideramos un plano euclídeo sobre \mathbb{R} .

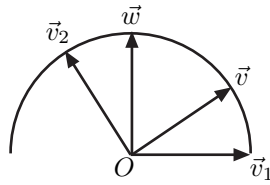
Dado $x \in [-1, 1]$, existe un único número $y \geq 0$ tal que $x^2 + y^2 = 1$, y así $(x, y) = (\cos \alpha, \sin \alpha)$ para un único $\alpha \in [0, \pi]$. En particular, dados dos vectores \vec{v}_1 y \vec{v}_2 , como $\cos(\vec{v}_1, \vec{v}_2) \in [-1, 1]$, tenemos que existe un único $\alpha \in [0, \pi]$ tal que $\cos \alpha = \cos(\vec{v}_1, \vec{v}_2)$. Diremos que α es la *amplitud* del ángulo entre \vec{v}_1 y \vec{v}_2 . Explícitamente, α está determinado por la relación

$$\cos \alpha = \frac{\vec{v}_1 \vec{v}_2}{\|\vec{v}_1\| \|\vec{v}_2\|}.$$

La amplitud de un ángulo \widehat{ABC} se define como la de los vectores \overrightarrow{BA} y \overrightarrow{BC} . Sabemos que esta definición no depende de la elección de A y C en los lados del ángulo, así como que se conserva por isometrías, es decir, que ángulos congruentes tienen la misma amplitud. Vamos a comprobar que esta definición de amplitud satisface las condiciones del teorema 2.29. Ya hemos probado la primera.

Para probar la segunda consideramos un ángulo $\hat{A} = A(O; \vec{v}_1, \vec{v}_2)$, donde podemos suponer que \vec{v}_1 y \vec{v}_2 tienen norma 1, y una semirrecta $O + \{\lambda \vec{v} \mid \lambda \geq 0\}$ contenida en el ángulo, lo cual equivale a que

$$\vec{v} = \lambda \vec{v}_1 + \mu \vec{v}_2, \quad \lambda, \mu \geq 0.$$



(También podemos suponer que \vec{v} tiene norma 1.) Sea \vec{w} un vector ortogonal a \vec{v}_1 en el plano del ángulo. Cambiando \vec{w} por $-\vec{w}$ si es preciso podemos suponer que $\vec{v}_2 = x\vec{v}_1 + y\vec{w}$, con $y > 0$. Sean α , β y γ las amplitudes de los ángulos $\widehat{v_1 v_2}$, $\widehat{v_1 v}$ y $\widehat{v v_2}$ respectivamente. Hemos de probar que $\alpha = \beta + \gamma$.

Tenemos $\cos \alpha = \vec{v}_1 \vec{v}_2 = x$, luego $y = \sin \alpha$ (pues $x^2 + y^2 = 1$ e $y > 0$). Así pues, $\vec{v}_2 = \cos \alpha \vec{v}_1 + \sin \alpha \vec{w}$. De aquí que

$$\vec{v} = (\lambda + \mu \cos \alpha) \vec{v}_1 + \mu \sin \alpha \vec{w}.$$

A su vez esto implica que $\cos \beta = \vec{v}_1 \vec{v} = \lambda + \mu \cos \alpha$, luego $\sin \beta = \mu \sin \alpha$. Usando las fórmulas 6 y 7 del teorema 3.43 vemos que

$$\begin{aligned} \cos(\alpha - \beta) &= \cos \alpha \cos \beta + \sin \alpha \sin \beta = \lambda \cos \alpha + \mu = \vec{v} \vec{v}_2 = \cos \gamma, \\ \sin(\alpha - \beta) &= \sin \alpha \cos \beta - \sin \beta \cos \alpha = \lambda \sin \alpha \geq 0. \end{aligned}$$

Puesto que $|\alpha - \beta| \leq \pi$, la última desigualdad implica que $0 \leq \alpha - \beta \leq \pi$, luego de $\cos(\alpha - \beta) = \cos \gamma$ podemos concluir $\alpha - \beta = \gamma$, como queríamos probar.

Llamemos $m(\hat{A})$ a la amplitud de un ángulo \hat{A} tal y como la hemos definido a partir de las funciones \sin y \cos . Sea $m'(\hat{A})$ la amplitud de \hat{A} en el sentido usual, digamos que tomando como unidad de ángulos el ángulo recto. Claramente los ángulos rectos tienen medida m igual a $\pi/2$. El teorema 2.29 nos da entonces que $m(\hat{A}) = (\pi/2)m'(\hat{A})$, para todo ángulo \hat{A} , dicho de otro modo, que la amplitud de un ángulo cualquiera en el sentido usual es $(2/\pi)m(\hat{A})$ ángulos rectos.

Si suponemos además que $\pi > 1$, entonces podemos tomar como unidad de ángulos $2/\pi$ veces un ángulo recto y, respecto a esta unidad, la amplitud de un ángulo cualquiera es precisamente $m(\hat{A})$.

Finalmente, dado $\alpha \in]0, \pi[$, fijamos un sistema de referencia ortonormal y consideramos los puntos $P(\cos \alpha, \sin \alpha)$ y $Q(1, 0)$. Entonces la perpendicular a OQ por P corta a OQ en el punto $(\cos \alpha, 0)$. Como $\|\overrightarrow{OP}\| = 1$, concluimos que $\sin \alpha$ y $\cos \alpha$ son el seno y el coseno de \hat{O} en el sentido geométrico, es decir, hemos probado:

Teorema 3.44 *Si unas funciones $\text{sen}, \text{cos} : \mathbb{R} \rightarrow [0, 1]$ satisfacen las propiedades del teorema 3.43 con $\pi > 1$, entonces existe una unidad de ángulos respecto a la cual $\text{sen } \alpha$ y $\text{cos } \alpha$ son el seno y el coseno de los ángulos de amplitud α (para todo $\alpha \in [0, \pi]$).*

Notemos que no hemos definido unas únicas funciones $\text{sen}, \text{cos} : \mathbb{R} \rightarrow [0, 1]$, sino un par de funciones $\text{sen}_\pi, \text{cos}_\pi$ para cada número real $\pi > 1$. No obstante, el teorema anterior implica que, si consideramos dos elecciones distintas,

$$\text{sen}_{\pi'}(\alpha) = \text{sen}_\pi\left(\frac{\pi}{\pi'} \alpha\right), \quad \text{cos}_{\pi'}(\alpha) = \text{cos}_\pi\left(\frac{\pi}{\pi'} \alpha\right),$$

pues los miembros derechos cumplen las propiedades del teorema 3.43 para π' . Esto significa que la diferencia entre ellas estriba simplemente en la elección del ángulo de amplitud unitaria.

Introducción analítica de la amplitud de ángulos Es importante insistir en que hemos demostrado la existencia de las funciones seno y coseno. No obstante, si se supone conocida su existencia —a través de una construcción mediante las técnicas del análisis matemático— junto con las propiedades del teorema 3.43, podemos introducir rápidamente la medida de ángulos en todo espacio afín euclídeo arquimediano. Para ello sólo tenemos que modificar la definición 3.35 de modo que, en lugar de definir el coseno $\text{cos}(\vec{v}, \vec{w})$, definimos la amplitud del ángulo formado por los vectores \vec{v} y \vec{w} como el número real $\alpha \in [0, \pi]$ que cumple

$$\cos \alpha = \frac{\vec{v}\vec{w}}{\|\vec{v}\|\|\vec{w}\|}.$$

A partir de ahí los argumentos posteriores a la definición se adaptan trivialmente para probar que el seno y el coseno de los ángulos agudos satisfacen la definición geométrica como cocientes de los catetos sobre la hipotenusa, y el teorema 3.37 se convierte en que dos ángulos son congruentes si y sólo si tienen la misma amplitud. Las propiedades de los giros, en particular la relación $G_{\alpha+\beta} = G_\alpha \circ G_\beta$ son ahora triviales, ya que contamos de antemano con las fórmulas para el seno y el coseno de una suma, y las pruebas que hemos dado del teorema de los senos, etc. valen sin cambio alguno. ■

Trigonometría no arquimediana Conviene observar que algunos de los resultados que hemos obtenido son válidos también en planos euclídeos sobre un cuerpo ordenado pitagórico cualquiera R , no necesariamente arquimediano. En este caso no tenemos definida la amplitud de un ángulo (pero sí la longitud de un segmento \overline{AB} , que es la norma $\|\overrightarrow{AB}\|$), aunque podemos hablar de ángulos α como clases de congruencia de ángulos, tal y como hemos hecho en los apartados precedentes cuando no suponíamos que el cuerpo fuera arquimediano. Como es habitual, llamamos 0 a la clase de los ángulos nulos, π a la clase de los ángulos llanos y $\pi/2$ a la clase de los ángulos rectos. También tenemos definidos los giros $G_{\pm\alpha}$, donde α es una clase de equivalencia de ángulos, y recordemos también que en el conjunto de todos los ángulos en un plano euclídeo tenemos parcialmente definida una suma (definición 1.66), de modo que $\alpha + \beta$ está definido si y sólo si β es menor o igual que es suplementario de α , y $\alpha - \beta$ está definido si y sólo si $\alpha \leq \beta$. Además, podemos probar lo siguiente:

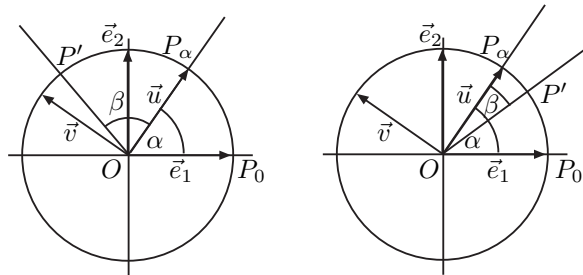
Teorema 3.45 *Si α y β son dos ángulos en un plano euclídeo E , entonces $G_{\alpha\pm\beta} = G_{\alpha} \circ G_{\pm\beta}$, donde hay que suponer que la suma o la resta de ángulos está definida.*

DEMOSTRACIÓN: Fijemos un sistema de referencia $(O, \vec{e}_1, \vec{e}_2)$ y llamemos $P_0 = O + \vec{e}_1$. Sea S el semiplano superior (es decir, el formado por los puntos de coordenadas (x, y) con $y \geq 0$).

Por el axioma C4 (que lo tenemos demostrado para los planos euclídeos) sabemos que $\alpha = \widehat{P_0OP_{\alpha}}$, para un punto P_{α} que es único si exigimos que $\|\overrightarrow{OP_{\alpha}}\| = 1$ y que $P_{\alpha} \in S$.

Similarmente, sea P' otro punto tal que $\|\overrightarrow{OP'}\| = 1$ y $\widehat{P_{\alpha}OP'} = \beta$. Aquí tenemos dos posibilidades: Si está definido $\alpha + \beta$, podemos tomar P' de modo que P_{α} esté contenido en $\widehat{P_0OP'}$, y si está definido $\alpha - \beta$, podemos tomar P' en $\widehat{P_0OP_{\alpha}}$. Se cumple entonces que $\widehat{P_0OP'} = \alpha \pm \beta$, según el caso.

Sea $\vec{u} = \overrightarrow{OP_{\alpha}}$. Como $P_{\alpha} \in S$, tenemos que la segunda coordenada de \vec{u} es positiva, luego concretamente $\vec{u} = \cos \alpha \vec{e}_1 + \sin \alpha \vec{e}_2$, luego el giro que transforma \vec{e}_1 en \vec{u} es G_{α} . Sea $\vec{v} = G_{\alpha}(\vec{e}_2)$. Así \vec{u}, \vec{v} es otra base ortonormal que determina otro sistema de referencia en E .



La matriz de cambio de base de \vec{u}, \vec{v} en \vec{e}_1, \vec{e}_2 es M_{α} , de modo que

$$(x, y) = (x', y')M_{\alpha},$$

donde (x, y) son las coordenadas de cualquier vector respecto de la primera base y (x', y') sus coordenadas en la segunda. Por lo tanto, las coordenadas de \vec{e}_1 en la base \vec{u}, \vec{v} son

$$(1, 0)M_\alpha^t = (\cos \alpha, -\operatorname{sen} \alpha).$$

Vemos así que $P_0 = O + \vec{e}_1$ está en el semiplano inferior respecto del sistema de referencia dado por \vec{u}, \vec{v} . Por el contrario, $O + \vec{v}$ está trivialmente en el semiplano superior.

En el caso correspondiente a $\alpha + \beta$ tenemos que P' y P_0 están en lados opuestos de la recta OP_α , mientras que en el caso correspondiente a $\alpha - \beta$ tenemos que P' y P_0 están en el mismo lado de dicha recta. Por lo tanto, en el primer caso P' está en el mismo lado que $O + \vec{v}$ y en el segundo están en lados opuestos. Equivalentemente, en el primer caso P' está en el semiplano superior respecto de la base \vec{u}, \vec{v} , y en el segundo caso está en el semiplano inferior.

Por consiguiente, la matriz respecto de la base \vec{u}, \vec{v} del giro que transforma \vec{u} en $\overrightarrow{OP'}$ es M_β en el primer caso y $M_{-\beta}$ en el segundo. Su matriz respecto de la base \vec{e}_1, \vec{e}_2 es $M_\alpha^t M_{\pm\beta} M_\alpha = M_{\pm\beta}$, pues las matrices de giros conmutan.

Concluimos que el giro que transforma \vec{u} en $\overrightarrow{OP'}$ (respecto de la base \vec{e}_1, \vec{e}_2) es $G_{\pm\beta}$ (donde el signo depende del caso) o, equivalentemente, que $G_{\pm\beta}(\vec{u}) = \overrightarrow{OP'}$. Por consiguiente, el giro $G_\alpha \circ G_{\pm\beta}$ es el giro que transforma \vec{e}_1 en $\overrightarrow{OP'}$, y es $G_{\alpha\pm\beta}$ porque P' está en el semiplano superior. ■

Para eliminar las restricciones a la suma o resta de ángulos podemos sustituir los números reales por los propios giros, en el sentido siguiente:

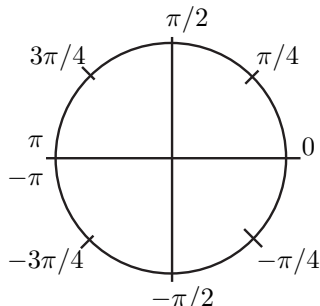
Definición 3.46 Llamaremos *ángulos generalizados* (respecto a un punto O en un plano euclídeo) a los giros de centro O . Si \mathcal{A}_O es el conjunto de los ángulos generalizados respecto de O , definimos en \mathcal{A}_O la suma dada por la composición: $G_1 + G_2 = G_1 \circ G_2$. Claramente es asociativa, conmutativa, tiene por elemento neutro a la identidad (que representaremos por 0) y todo ángulo generalizado α tiene un opuesto que representaremos por $-\alpha$.

Hemos probado que si \mathcal{A}_0 es el conjunto de todos los ángulos (entendidos como clases de congruencia), la aplicación $i : \mathcal{A}_0 \rightarrow \mathcal{A}_O$ dada por $i(\alpha) = G_\alpha$ (donde G_α se define respecto de un sistema de referencia ortonormal prefijado) conserva la suma y la resta, es decir, cuando $\alpha \pm \beta$ está definido, se cumple que $i(\alpha \pm \beta) = i(\alpha) \pm i(\beta)$. Esto nos permite identificar los ángulos con parte de los ángulos generalizados, con la diferencia de que cualquier par de ángulos generalizados se puede sumar o restar. Cada ángulo generalizado es de la forma $\pm\alpha$, donde α es un ángulo, y $\alpha \neq -\alpha$ salvo cuando $\alpha = 0$ o $\alpha = \pi$.

Vamos a usar los ángulos generalizados a modo de “números” que midan la amplitud de un ángulo, pero cuando queramos ver a un ángulo generalizado α como el giro que es, usaremos la notación G_α , es decir, α y G_α son lo mismo, pero escribiremos α cuando pensemos en α como “número” y G_α cuando pensemos en él como giro.

Observemos que, en términos de la aritmética de los ángulos generalizados, se cumple que $2\pi = \pi + \pi = \pi - \pi = 0$.

La definición 3.38 de circunferencia graduada vale igualmente aunque el cuerpo R no sea arquimediano, y en general podemos asignar, no un número real, pero sí un ángulo generalizado a cada punto de una circunferencia graduada:



En virtud del teorema 3.41, cada graduación de una circunferencia ω determina una biyección $\mathcal{A}_O \rightarrow \omega$ dada por $\alpha \mapsto P_\alpha = G_\alpha(P_0)$, donde G_α se calcula respecto del sistema de referencia asociado a la graduación. Notemos que $P_{\pi/2}$ en este sentido es el mismo $P_{\pi/2}$ que forma parte de la graduación. La figura muestra algunos puntos de la graduación de una circunferencia.

Podemos extender las funciones $\text{sen}, \text{cos} : \mathcal{A}_O \rightarrow [-1, 1] \subset R$ mediante

$$\cos(-\alpha) = \cos \alpha, \quad \text{sen}(-\alpha) = -\text{sen} \alpha.$$

De este modo, la matriz del giro G_α es

$$M_\alpha = \begin{pmatrix} \cos \alpha & \text{sen} \alpha \\ -\text{sen} \alpha & \cos \alpha \end{pmatrix},$$

tanto si α es un ángulo como si es el opuesto de un ángulo. Notemos que la determinación de cuál de dos ángulos opuestos $\pm\alpha$ tiene seno positivo y cuál tiene seno negativo depende de la elección del sistema de referencia.

En estos términos, si ω es una circunferencia graduada de radio r , como las coordenadas de P_0 son $(r, 0)$, las del punto P_α son $(r, 0)M_\alpha = (r \cos \alpha, r \text{sen} \alpha)$.

Definición 3.47 Dado un sistema de referencia $(O; \vec{e}_1, \vec{e}_2)$ en un plano euclídeo, llamaremos *coordenadas polares* de un punto $P \neq O$ respecto de dicho sistema de referencia a su *módulo* $\rho = \|\overrightarrow{OP}\| \in R$ y su *argumento* $\theta \in \mathcal{A}_O$, definido como el único ángulo generalizado para el que las coordenadas de P son

$$(x, y) = (\rho \cos \theta, \rho \text{sen} \theta).$$

La existencia del argumento se sigue de las observaciones previas a la definición, considerando la graduación de la circunferencia de centro O y radio ρ determinada por el sistema de referencia prefijado.

El hecho de que, por definición, $G_{\alpha+\beta} = \alpha + \beta = G_\alpha \circ G_\beta$ se traduce en que $M_{\alpha+\beta} = M_\alpha M_\beta$, y multiplicando las matrices se obtienen inmediatamente las relaciones

$$\begin{aligned} \text{sen}(\alpha + \beta) &= \text{sen} \alpha \cos \beta + \cos \alpha \text{sen} \beta, \\ \cos(\alpha + \beta) &= \cos \alpha \cos \beta - \text{sen} \alpha \text{sen} \beta. \end{aligned}$$

Usando las fórmulas para el seno y el coseno de una suma es inmediato que el teorema 3.42 es válido en este contexto general, y también valen sin cambio alguno todos los resultados que hemos demostrado tras el teorema 3.43.

Capítulo IV

Números complejos y cuaternios

En [A1 5.15] hemos introducido el cuerpo \mathbb{C} de los números complejos como la adjunción a \mathbb{R} de una “unidad imaginaria”, es decir, $i = \sqrt{-1}$. Esta definición “salvo isomorfismo” es la definición rigurosa de los números complejos formalmente más próxima a la forma no rigurosa en la que fueron introducidos históricamente (admitiendo la existencia de una raíz cuadrada de -1 y asumiendo las consecuencias), pues pone de manifiesto que los números complejos no son más que eso: los números reales a los que se les ha añadido una raíz cuadrada de -1 . Sin embargo, los números complejos no fueron realmente asimilados hasta que los matemáticos no se dieron cuenta de que podían ser representados como los puntos del plano. A mediados del siglo XIX, el matemático irlandés Sir William Rowan Hamilton estaba tratando en vano de encontrar una estructura algebraica similar que pudiera ser identificada con los puntos del espacio, hasta que el 16 de octubre de 1843 se dio cuenta de que podía definir una estructura semejante a los números complejos, no en \mathbb{R}^3 , sino en \mathbb{R}^4 . Así descubrió lo que desde entonces se conocen como los “cuaternios de Hamilton”, si bien habían sido descubiertos ya en 1840 por el matemático francés Olinde Rodrigues. En este capítulo vamos a estudiar ambas estructuras.

4.1 La geometría de los números complejos

Las propiedades algebraicas de los números complejos son válidas más en general si partimos de un cuerpo ordenado arbitrario R . El hecho de que sea un cuerpo ordenado implica en particular que $-1 < 0$ no puede ser un cuadrado, por lo que el polinomio $x^2 + 1$ es irreducible en R y al adjuntar una de sus raíces obtenemos una extensión de grado 2:

Definición 4.1 Si R es un cuerpo ordenado, definimos $C = R(i)$ como el cuerpo que resulta de adjuntarle a R una raíz i del polinomio $x^2 + 1$.

El teorema [Al 3.51] implica que siempre existe una extensión de R en la que el polinomio tiene tal raíz y el teorema [Al 5.14] implica que dos cualesquiera son R -isomorfas, por lo que es irrelevante cuál consideremos. Cuando $R = \mathbb{R}$ es el cuerpo de los números reales, entonces $C = \mathbb{C}$ es el cuerpo de los números complejos.

De acuerdo con [Al 5.5], todo elemento de C se expresa de forma única como $x + yi$, con $x, y \in R$ y, en términos de esta expresión, las operaciones son

$$(x+yi)+(x'+y'i) = (x+x')+(y+y')i, \quad (x+yi)(x'+y'i) = (xx'-yy')+(xy'+yx')i.$$

Tenemos que C es un R -espacio vectorial con base $\{1, i\}$. Para cada número $z = x + yi \in C$, los elementos x y y (que son las coordenadas de z en la base $\{1, i\}$) se llaman respectivamente *parte real* y *parte imaginaria* de z , y los representaremos por $\operatorname{Re} z$ e $\operatorname{Im} z$.

Tenemos así dos aplicaciones lineales $\operatorname{Re}, \operatorname{Im} : C \rightarrow R$ que determinan un isomorfismo de R -espacios vectoriales $C \cong R^2$.

Puesto que i y $-i$ son conjugados (pues son las raíces de $x^2 + 1$), el teorema [Al 5.14] implica que la conjugación $C \rightarrow C$ dada por $z = x + yi \mapsto \bar{z} = x - yi$ es un R -automorfismo de C , de modo que $z \in R$ si y sólo si $z = \bar{z}$.

Si R es un cuerpo pitagórico, entonces podemos considerar en C el producto escalar respecto al cual la base $\{1, i\}$ es ortonormal, de modo que la norma que induce viene dada por $\|z\| = \|x + yi\| = \sqrt{x^2 + y^2} \in R$. En este contexto es más habitual llamar a esta norma el *módulo* o *valor absoluto* de z y representarla por

$$|z| = \sqrt{x^2 + y^2} = \sqrt{z\bar{z}}.$$

Claramente cumple las propiedades siguientes:¹

1. $|z| \geq 0$ y $|z| = 0$ si y sólo si $z = 0$.
2. $|z_1 + z_2| \leq |z_1| + |z_2|$.
3. $|z_1 z_2| = |z_1| |z_2|$.

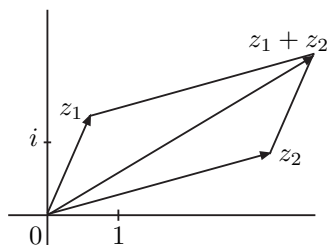
Las dos primeras propiedades son un caso particular del teorema 3.23, mientras que la tercera se sigue de que la conjugación conserva el producto.

El teorema [Al 5.49] implica que si R es un cuerpo ordenado euclídeo (es decir, si todo elemento positivo tiene raíz cuadrada), entonces todo elemento de C tiene raíz cuadrada. En particular toda ecuación de segundo grado con coeficientes en C tiene dos raíces en C (no necesariamente distintas).

Suponiendo únicamente que R sea un cuerpo pitagórico, acabamos de señalar que el cuerpo C es un espacio vectorial euclídeo de dimensión 2 (que a su vez puede considerarse como un espacio afín en el que los puntos son los propios vectores).

¹Si $R \subset \mathbb{R}$, estas propiedades expresan que C es un cuerpo métrico en el sentido de la definición [An 1.17].

En particular, los elementos de \mathbb{C} , los números complejos, pueden identificarse con los puntos del plano. Por ello es habitual referirse a \mathbb{C} como “el plano complejo”. El sistema de referencia canónico en C es el determinado por el punto 0 y la base $\{1, i\}$, de modo que la recta que pasa por 0 y 1 es “el eje real”, mientras que el que pasa por 0 e i es “el eje imaginario”.



En general, la suma de elementos de C tiene la interpretación usual según el cual $z_1 + z_2$ es el vector que traslada a un punto trasladándolo primero según z_1 y trasladando el resultado según z_2 . También es fácil dar una interpretación geométrica de la conjugación, pues no es sino la simetría respecto del eje real. Ahora vamos a dar una interpretación geométrica al producto.

Para ello observamos primeramente que, según 3.47, todo $z \in C$ admite una expresión en coordenadas polares²

$$z = \rho_\theta = \rho(\cos \theta + i \operatorname{sen} \theta),$$

donde $\rho = |z| \in R$.

En particular, si $u \in C$ tiene módulo 1, será de la forma $u = \cos \theta + i \operatorname{sen} \theta$. La aplicación $z \mapsto uz$ es un automorfismo de C como R -espacio vectorial. Como $u1 = u = \cos \theta + i \operatorname{sen} \theta$ y $ui = -\operatorname{sen} \theta + i \cos \theta$, tenemos que su matriz respecto de la base $1, i$ es

$$M_\theta = \begin{pmatrix} \cos \theta & \operatorname{sen} \theta \\ -\operatorname{sen} \theta & \cos \theta \end{pmatrix}.$$

Por lo tanto, la multiplicación por u es el giro G_θ de ángulo θ . Por consiguiente, si $z = \rho(\cos \theta + i \operatorname{sen} \theta)$ es un elemento no nulo de $R[i]$ y $z' = \rho'_{\theta'}$ es otro elemento no nulo, tenemos que

$$zz' = \rho(uz') = \rho G_\theta(\rho'_{\theta'}) = \rho \rho'_{\theta+\theta'} = \rho \rho'(\cos(\theta + \theta') + i \operatorname{sen}(\theta + \theta')) = (\rho \rho')_{\theta+\theta'}.$$

En definitiva:

El producto de dos elementos no nulos de C es el elemento determinado por que su módulo es el producto de los módulos y su argumento es la suma de los argumentos.

Por ejemplo, es claro que $i = 1_{\pi/2}$, luego la multiplicación por i se interpreta geoméricamente como un giro de 90° en el plano complejo. En particular $i^2 = 1_\pi = -1$ es la “explicación geométrica” de cómo puede tener -1 una raíz cuadrada en \mathbb{C} : al girar i un ángulo de 90° obtenemos -1 .

Así, el conjunto $S = \{z \in C \mid |z| = 1\}$, es un grupo multiplicativo que es isomorfo al grupo de los giros de centro 0 del plano C (el isomorfismo asigna a cada z la multiplicación por z).

²En general, θ es un ángulo generalizado, aunque si $R \subset \mathbb{R}$ podemos suponer que es un número real determinado salvo múltiplos de 2π . (Véase la página 123.)

Si $R \subset \mathbb{R}$ podemos tomar argumentos en \mathbb{R} y considerar el grupo aditivo $A = \{\theta \in \mathbb{R} \mid \cos \theta, \operatorname{sen} \theta \in R\}$, de modo que la aplicación $A \rightarrow S$ dada por $\theta \mapsto \cos \theta + i \operatorname{sen} \theta$ es un epimorfismo de grupos que induce un isomorfismo $A / \langle 2\pi \rangle \cong S$.

Ejemplo: Las raíces de la unidad Como $1^2 = (-1)^2 = 1$, podemos decir que 1 y -1 son las raíces cuadradas de la unidad, mientras que $1, -1, i, -i$ son las raíces cuartas de la unidad. Decimos “las” raíces porque no puede haber más, ya que estamos hablando de las raíces en el cuerpo C de los polinomios $z^2 - 1$ y $z^4 - 1$, y sabemos que el número de raíces de un polinomio no puede exceder a su grado.

Podemos plantearnos si en C existen igualmente tres raíces cúbicas de la unidad, y cinco raíces quintas, etc.

Fijemos un número natural $n \geq 2$ y supongamos que el polinomio $z^n - 1$ tiene n raíces distintas en C y sea C_n el conjunto de todas ellas. Es inmediato que si $u, v \in C$ cumplen $u^n = v^n = 1$, entonces $(uv^{-1})^n = 1$, lo que prueba que C_n es un subgrupo del grupo multiplicativo de C . Por el teorema [Al 4.50] tenemos que C_n es un grupo cíclico de orden n .

Las raíces n -simas de la unidad de orden n se llaman *raíces n -simas primitivas* de la unidad. Por ejemplo, de las cuatro raíces cuartas de la unidad $\pm 1, \pm i$, son primitivas (tienen orden 4) dos de ellas: $\pm i$.

En estos términos, lo que estamos suponiendo es que C contiene una raíz n -sima primitiva de la unidad (y, por consiguiente, tiene n raíces distintas). Vamos a describirlas geoméricamente.

Teorema 4.2 *Si en C existe una raíz n -sima primitiva de la unidad, entonces las raíces n -simas de la unidad dividen a la circunferencia unitaria en n arcos iguales.*

DEMOSTRACIÓN: Podemos suponer que $n \geq 3$, pues para $n = 2$ se cumple trivialmente el teorema. Sean $u, v \in C_n$ tales que el ángulo α que forman sea el mínimo posible. Como los giros son isometrías, tenemos que 1 y $w = vu^{-1}$ forman también un ángulo α .

Notemos que no puede ser $w = -1$, pues la minimalidad de α implicaría que no hay más elementos en C_n aparte de ± 1 , luego estaríamos en el caso $n = 2$ que hemos descartado. Por lo tanto $\operatorname{Im} w \neq 0$.

Si $\operatorname{Im} w < 0$ cambiamos w por \bar{w} , que también está en C_n y forma el mismo ángulo con 1 (porque la conjugación es una simetría, luego una isometría). De este modo $w = 1_\alpha$. Vamos a ver que w es una raíz n -sima primitiva de la unidad. Sus potencias son $w^m = 1_{m\alpha}$.

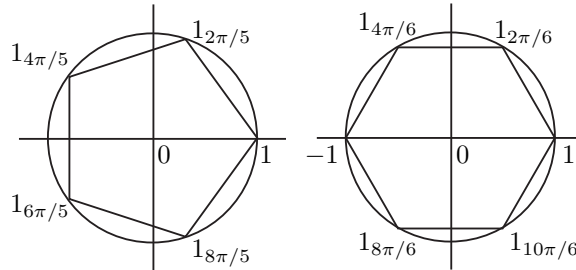
Sea k el máximo número natural tal que $k\alpha \leq \pi$, es decir, tal que el ángulo α se puede sumar consigo mismo k veces. Entonces las potencias $1, w, w^2, \dots, w^k$ son $k + 1$ puntos distintos de la semicircunferencia superior, cada uno de los cuales forma un ángulo α con el anterior y el posterior, y de modo que entre dos de ellos no hay ninguna otra raíz de la unidad (por la minimalidad de α).

Si $w^k \neq -1$, forma un ángulo $\beta < \alpha$ con -1 , mientras que w^{k+1} , que forma un ángulo α con w^k , forma un ángulo $\alpha - \beta$ con -1 . Pero tiene que ser $\beta = \alpha/2$, pues si $\beta < \alpha/2$ entonces \bar{w}^k sería una raíz n -sima de la unidad entre w^k y w^{k+1} , y si $\beta > \alpha/2$ entonces \bar{w}^{k+1} sería una raíz n -sima de la unidad entre w^k y w^{k+1} . Así pues, $\pi = k\alpha + \alpha/2$, o también $(2k + 1)\alpha = 2\pi$

Por consiguiente, $w^{k+1} = \bar{w}^k$, y de aquí que $w^{k+2} = \bar{w}^{k-1}$, con lo que w^{k+1}, \dots, w^{2k} son los conjugados de w^k, \dots, w^1 , luego tenemos $2k$ raíces distintas dos a dos que dividen la circunferencia en arcos de amplitud α , luego entre ellas no hay ninguna otra raíz. Por último $w^{2k+1} = w\bar{w} = 1$, de donde concluimos que las $2k + 1$ potencias de w son todas las raíces n -simas de la unidad (cualquier otra tendría que estar entre dos de dichas potencias, en contradicción con la minimalidad de α). En particular $n = 2k + 1$ y $\alpha = 2\pi/n$.

Si $w^k = -1$ se razona igualmente que $w^{k+1} = \bar{w}^{k-1}$ y llegamos a la misma conclusión con una mínima variante del argumento. Por lo tanto, w es una raíz primitiva y sus potencias dividen la circunferencia en arcos iguales de amplitud $\alpha = 2\pi/n$. ■

Nota Si $R = \mathbb{R}$ tenemos un argumento mucho más simple que, de hecho, no requiere suponer que \mathbb{C} contiene una raíz n -sima primitiva de la unidad, sino que demuestra que así es: basta considerar los números complejos $1_{2k\pi/n}$, para $k = 0, \dots, n - 1$, y observar que son n raíces n -simas distintas de la unidad.



La figura muestra las raíces quintas y las raíces sextas de la unidad en el plano complejo. ■

El recíproco del teorema anterior se demuestra fácilmente: si una circunferencia puede dividirse en n arcos iguales (necesariamente de amplitud $2\pi/n$) entonces el polinomio $z^n - 1$ tiene n raíces distintas en C . En efecto, es fácil ver que si una circunferencia puede dividirse en n arcos iguales, lo mismo le sucede a una circunferencia de centro 0. Dividiendo los puntos entre el radio de la circunferencia, podemos suponer que es la circunferencia de centro 0 y radio 1, y multiplicando todos los puntos por el inverso de uno de ellos, podemos suponer que uno de los puntos de la división es 1, pero entonces los n puntos tienen argumentos $2k\pi/n$, para $k = 0, \dots, n - 1$, luego son $1_{2k\pi/n}$, que claramente son n raíces n -simas de la unidad. Por consiguiente:

Teorema 4.3 Una circunferencia de C puede dividirse en n arcos de amplitud $2\pi/n$ si y sólo si el polinomio $z^n - 1$ tiene n raíces distintas en C .

Ejemplo Si $\alpha = 2\pi/7$, hemos visto que $\zeta = 1_\alpha = \cos \alpha + i \operatorname{sen} \alpha$ es una raíz séptima de la unidad, luego es raíz del polinomio

$$x^7 - 1 = (x - 1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1),$$

pero $\zeta \neq 1$, luego

$$\zeta^6 + \zeta^5 + \zeta^4 + \zeta^3 + \zeta^2 + \zeta + 1 = 0.$$

Ahora bien, $\zeta^k = 1_{k\alpha} = \cos k\alpha + i \operatorname{sen} k\alpha$, luego, al tomar partes reales en la igualdad anterior, queda que

$$\cos 6\alpha + \cos 5\alpha + \cos 4\alpha + \cos 3\alpha + \cos 2\alpha + \cos \alpha + 1 = 0.$$

Pero

$$\begin{aligned} \cos 6\alpha &= \cos 12\pi/7 = \cos(-2\pi/7) = \cos \alpha, \\ \cos 5\alpha &= \cos 10\pi/7 = \cos(-4\pi/7) = \cos 2\alpha, \\ \cos 4\alpha &= \cos 8\pi/7 = \cos(-6\pi/7) = \cos 3\alpha, \end{aligned}$$

luego

$$2 \cos 3\alpha + 2 \cos 2\alpha + 2 \cos \alpha + 1 = 0.$$

Si ahora usamos las fórmulas

$$\cos 2\alpha = \cos^2 \alpha - 1, \quad \cos 3\alpha = 4 \cos^3 \alpha - 3 \cos \alpha,$$

concluimos que $\cos \alpha$ es raíz del polinomio

$$8x^3 + 4x^2 - 4x - 1,$$

que es claramente el polinomio mínimo de $\cos 2\pi/7$.

Por ejemplo, es fácil ver (véase la sección [IGE 3.2]) que el área de un heptágono regular de lado l es

$$S = \frac{7l^2}{4} \sqrt{\frac{1 + \cos \alpha}{1 - \cos \alpha}}.$$

Ahora que conocemos el polinomio mínimo de $\cos \alpha$, podemos simplificar esta fórmula. Para ello observamos que, si $x = \cos \alpha$, entonces

$$(8x^3 + 4x^2 - 4x - 8) + 7 = 0,$$

donde hemos hecho un ajuste para que el polinomio entre paréntesis tenga raíz 1. Así:

$$(x - 1)(8x^2 + 12x + 8) + 7 = 0,$$

luego

$$\frac{7}{1 - x} = 8x^2 + 12x + 8,$$

y, por consiguiente,

$$S^2 = \frac{7l^4}{4^2}(1+x)(8x^2+12x+8) = \frac{7l^4}{4^2}(8x^3+20x^2+20x+8).$$

Usando ahora que $8x^3 = -4x^2 + 4x + 1$, queda

$$S^2 = \frac{7l^4}{4^2}(16x^2+24x+9) = \frac{7l^4}{4^2}(4x+3)^2,$$

luego

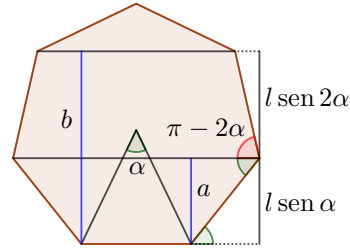
$$S = \sqrt{7}(\cos \alpha + 3/4)l^2.$$

De aquí se deduce una expresión no trivial para el área de un heptágono regular. Si llamamos a y b a las distancias a un lado de las dos diagonales paralelas, entonces

$$S = \sqrt{7}ab.$$

En efecto, en la figura vemos que

$$a = l \sin \alpha, \quad b = l \sin \alpha + l \sin 2\alpha,$$



luego, llamando como antes $x = \cos \alpha$,

$$\begin{aligned} \sqrt{7}ab &= l^2(\sin^2 \alpha + 2 \sin^2 \alpha \cos \alpha) = \sqrt{7}l^2(1 - \cos^2 \alpha)(1 + 2 \cos \alpha) \\ &= \sqrt{7}l^2(2x^3 - x^2 + 2x + 1) = \frac{\sqrt{7}}{4}l^2(-8x^3 - 4x^2 + 8x + 4). \end{aligned}$$

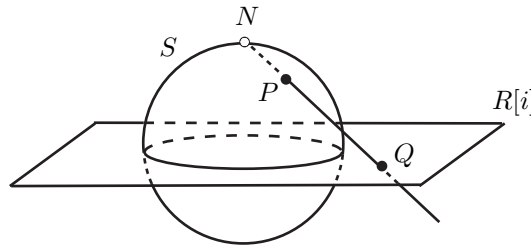
Usando de nuevo que $-8x^3 = 4x^2 - 4x - 1$ queda

$$\sqrt{7}ab = \frac{\sqrt{7}}{4}l^2(4x+3) = \frac{\sqrt{7}}{4}l^2(4 \cos \alpha + 3) = S.$$

■

La esfera de Riemann Veamos ahora otra interpretación geométrica útil de los números complejos. Seguimos trabajando en general sobre un cuerpo ordenado pitagórico R , y consideramos la esfera

$$S^2 = \{(x, y, z) \in R^3 \mid x^2 + y^2 + z^2 = 1\}.$$



Llamamos $N = (0, 0, 1)$ al “polo Norte” de S^2 y consideramos la proyección estereográfica $f : S^2 \setminus \{N\} \rightarrow C$ que a cada punto P de la esfera le asigna el punto Q donde la recta NP corta al plano $z = 0$ (al que le quitamos la tercera coordenada nula para tener un elemento de $C = \mathbb{R}^2$).

Al final de la sección [An 2.5] están hechos los cálculos que muestran que f viene dada por

$$f(x, y, z) = \frac{x}{1-z} + \frac{y}{1-z}i,$$

Así como que su inversa es

$$g(u + vi) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1} \right). \quad (4.1)$$

Aunque los cálculos de [An] están hechos para $R = \mathbb{R}$, es inmediato que valen literalmente para cualquier cuerpo ordenado R .

Definición 4.4 Llamaremos $C^\infty = C \cup \{\infty\}$, donde ∞ es cualquier elemento que no pertenezca a C , de modo que la proyección estereográfica puede extenderse a una biyección $f : S^2 \rightarrow C^\infty$ haciendo $f(N) = \infty$. En particular, C^∞ recibe el nombre de *esfera de Riemann*.

En [An] demostramos que en la esfera de Riemann podemos definir una topología de forma natural que la hace homeomorfa a la esfera S^2 .

En el apéndice A se prueba, entre otras cosas, que la proyección estereográfica hace corresponder las circunferencias de la esfera con las rectas y las circunferencias de C^∞ (entendiendo que las rectas de C^∞ pasan todas por ∞).

4.2 Cuaternios

El hecho de que \mathbb{R}^2 admita una estructura de cuerpo (o, más precisamente, que exista un cuerpo \mathbb{C} con \mathbb{R} como subcuerpo y que tenga dimensión 2 como espacio vectorial sobre \mathbb{R}) tiene numerosas aplicaciones, tanto a la geometría como al análisis matemático.

Por ejemplo, en el capítulo siguiente aprovecharemos de forma esencial este hecho para “capturar” el concepto de figura constructible con regla y compás. Resulta natural plantearse si es posible definir una estructura análoga de cuerpo en \mathbb{R}^n para otros valores de $n > 2$ (o, más precisamente, si existen cuerpos que contengan a \mathbb{R} como subcuerpo y que tengan dimensión n como espacios vectoriales sobre \mathbb{R}), pero sucede que la respuesta es negativa (y esencialmente se debe al hecho de que \mathbb{C} es algebraicamente cerrado). En otros términos, no existe ningún cuerpo cuyos elementos tengan la forma $a+bi+cj$, donde $a, b, c \in \mathbb{R}$ son tres números arbitrarios, de modo que los elementos con $b = c = 0$ formen un subcuerpo identificable con \mathbb{R} .

Sin embargo, vamos a probar aquí que si añadimos un término más, sí que es posible definir una estructura en \mathbb{R}^4 que nos permita expresar sus elementos en la forma $a + bi + cj + dk$ que “casi” cumpla los axiomas de cuerpo y en la que

los elementos con $b = c = d = 0$ formen un subcuerpo isomorfo a \mathbb{R} y, más aún, los elementos $a + bi$ (es decir, con $c = d = 0$) formen un subcuerpo isomorfo a \mathbb{C} .

La idea es repetir “casi” la construcción con la que hemos pasado de \mathbb{R} a \mathbb{C} , es decir, igual que hemos construido \mathbb{C} de modo que sus elementos son de la forma $a + bi$, con $a, b \in \mathbb{R}$ y hemos determinado su estructura algebraica estableciendo que $i^2 = -1$, ahora vamos a considerar elementos de la forma

$$w + zj = (a + bi) + (c + di)j = a + bi + cj + dij, \quad \text{con } w, z \in \mathbb{C}, \quad j^2 = -1.$$

No es posible obtener una estructura de cuerpo de esta forma, pues si $i^2 = j^2$, entonces $i^2 - j^2 = (i + j)(i - j) = 0$, luego $i = \pm j$, pero entonces $w + zj = w \pm zi$ es un número complejo y no hemos obtenido realmente un cuerpo de dimensión 4 sobre \mathbb{R} , sino el propio \mathbb{C} .

Ahora bien, el “casi” al que hemos hecho referencia en varias ocasiones consiste en renunciar a la propiedad conmutativa y buscar un anillo de división en lugar de un cuerpo. En tal caso ya no tiene por qué ser cierto que $i^2 - j^2 = (i + j)(i - j)$, y el “argumento de imposibilidad” deja de ser válido.

Concretamente, vamos a violar la conmutatividad de la forma más simple posible, que consiste en establecer que las dos unidades imaginarias i, j no conmuten, sino que cumplan la relación $ij = -ji$.

Si llamamos $k = ij$, entonces existe una única estructura de anillo de división sobre $\mathbb{R}^4 = \mathbb{C}^2$ de modo que si llamamos $1, i, j, k$ a la base canónica, con lo que todo elemento de \mathbb{R}^4 se expresa de forma única como $a + bi + cj + dk$, se cumplan las condiciones indicadas, a saber, que $i^2 = j^2 = -1$, $ij = k = -ji$. Por ejemplo, podemos calcular $k^2 = ijij = i(-ij)j = -i^2j^2 = -1$. Por lo tanto, la situación es más simétrica incluso de lo previsto:

$$i^2 = j^2 = k^2 = -1, \quad ij = k = -ji.$$

Si queremos calcular, por ejemplo, ik , vemos que es $ijj = -j$, y así podemos calcular cualquier producto de los vectores básicos.

No obstante, comprobar en estos términos que realmente estas especificaciones determinan un producto que satisface los axiomas de anillo de división sería una labor muy tediosa. En su lugar, vamos a dar una definición alternativa que nos conducirá al mismo resultado pero que volverá inmediatas todas las comprobaciones. Se trata de conseguir un producto no conmutativo usando una de las pocas operaciones no conmutativas que conocemos: el producto de matrices.

Teorema 4.5 *Sea R un cuerpo ordenado pitagórico, sea $C = R^2$ con la estructura de cuerpo que hemos definido en este capítulo y sea M el conjunto de todas las matrices de $\text{Mat}_2(C)$ de la forma³*

$$q = \begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix} = \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix}.$$

Entonces M es un subanillo unitario de $\text{Mat}_2(C)$.

³En el fondo, los conjugados y el signo negativo en la fila inferior se disponen pensando en que el determinante de la matriz sea $|w|^2 + |z|^2 = a^2 + b^2 + c^2 + d^2$.

DEMOSTRACIÓN: Es inmediato que la suma (o resta) de matrices de la forma indicada tiene la forma indicada, e igualmente se comprueba con el producto:

$$\begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix} \begin{pmatrix} w' & z' \\ -\bar{z}' & \bar{w}' \end{pmatrix} = \begin{pmatrix} ww' - z\bar{z}' & wz' + zw' \\ -\bar{z}w' - \bar{w}\bar{z}' & -\bar{z}z' + \bar{w}\bar{w}' \end{pmatrix} \quad (4.2)$$

es claro que la entrada (2, 2) del producto es el conjugado de la entrada (1, 1) y que la entrada (2, 1) es el opuesto del conjugado de la entrada (1, 2), luego el producto está en Q . También es inmediato que la matriz identidad está en M , luego M es un anillo unitario. ■

Podríamos definir el anillo de los cuaternios sobre R como el anillo M considerado en el teorema anterior, pero vamos a tomar la definición más simple posible desde un punto de vista conjuntista:

Definición 4.6 Sea R un cuerpo ordenado pitagórico, sea M el anillo dado por el teorema anterior y consideremos la biyección $f : R^4 \rightarrow M$ dada por

$$f(a, b, c, d) = \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix}.$$

Definimos el anillo H de los *cuaternios* sobre R como $H = R^4$ con la suma y el producto que convierten a f en un isomorfismo de anillos, es decir,

$$q_1 + q_2 = f^{-1}(f(q_1) + f(q_2)), \quad q_1 q_2 = f^{-1}(f(q_1)f(q_2)).$$

En particular, cuando $R = \mathbb{R}$ escribiremos \mathbb{H} y nos referiremos a él como el *anillo de los cuaternios* (sin necesidad de especificar que el cuerpo de partida es \mathbb{R}).

Vamos a probar que H es un anillo de división. En primer lugar observamos que la suma de cuaternios se puede definir directamente con facilidad:

$$(a, b, c, d) + (a', b', c', d') = (a + a', b + b', c + c', d + d').$$

Sólo hay que comprobar que

$$f(a, b, c, d) + f(a', b', c', d') = f(a + a', b + b', c + c', d + d'),$$

lo cual es inmediato. La utilidad de definir las operaciones de H a través de la biyección f es no tener que describir explícitamente el producto de H en términos de las ocho coordenadas de los factores.

En segundo lugar observamos que

$$(a, 0, 0, 0)(a', b', c', d') = (aa', ab', ac', ad'),$$

para lo cual basta ver que $f(a, 0, 0, 0)f(a', b', c', d') = f(aa', ab', ac', ad')$, lo cual también es inmediato, dado que $f(a, 0, 0, 0) = aI$, donde I es la matriz identidad.

De aquí se sigue que la aplicación $R \rightarrow H$ dada por $a \mapsto (a, 0, 0, 0)$ es un monomorfismo de anillos, de modo que podemos identificar cada $a \in R$ con su imagen, y entonces R pasa a ser un subcuerpo de H . Más aún, si definimos

$$i = (0, 1, 0, 0), \quad j = (0, 0, 1, 0), \quad k = (0, 0, 0, 1),$$

entonces

$$(a, b, c, d) = (a, 0, 0, 0) + (0, b, 0, 0) + (0, 0, c, 0) + (0, 0, 0, d) = a + bi + cj + dk.$$

Así pues, concluimos:

Todo cuaternio $q \in H$ se expresa de forma única como

$$q = a + bi + cj + dk, \quad \text{con } a, b, c, d \in R.$$

Observemos que en la representación matricial de los cuaternios (es decir, a través de f):

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad k \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

Entonces se comprueba inmediatamente que $i^2 = j^2 = k^2 = -1$. A su vez, esto implica inmediatamente que los subespacios vectoriales $\langle 1, i \rangle$, $\langle 1, j \rangle$, $\langle 1, k \rangle$, es decir, los subconjuntos formados por los cuaternios de la forma $a + bi$, $a + cj$ y $a + dk$, respectivamente, son subcuerpos de H isomorfos a C . Concretamente, consideraremos $C \subset H$ identificándolo con el subcuerpo formado por los cuaternios de la forma $a + bi$.

También es inmediato comprobar que $ij = k$, por lo que todo cuaternio puede escribirse de forma única como

$$q = a + bi + cj + dij = a + bi + (c + di)j = w + zj, \quad \text{con } w, z \in C.$$

En estos términos, es inmediato que

$$f(w + zj) = \begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix}.$$

Teorema 4.7 *Si R es un cuerpo ordenado pitagórico, la aplicación $H \rightarrow H$ dada por*

$$\bar{q} = \overline{a + bi + cj + dk} = a - bi - cj - dk$$

cumple $\overline{q_1 + q_2} = \bar{q}_1 + \bar{q}_2$, $\overline{q_1 q_2} = \bar{q}_2 \bar{q}_1$. Además $q\bar{q} = a^2 + b^2 + c^2 + d^2$.

DEMOSTRACIÓN: Claramente es biyectiva y conserva las sumas. Para probar que invierte los productos observamos que a través del isomorfismo f se corresponde con la aplicación

$$\overline{\begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix}} = \begin{pmatrix} \bar{w} & -z \\ \bar{z} & w \end{pmatrix}.$$

Teniendo esto en cuenta ya es fácil comprobar mediante la ecuación (4.2) que $f(\bar{q}_2\bar{q}_1) = f(\overline{q_1q_2})$, lo que a su vez implica que $\bar{q}_1\bar{q}_2 = \bar{q}_2\bar{q}_1$.

Por otra parte,

$$\begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix} \overline{\begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix}} = \begin{pmatrix} |w|^2 + |z|^2 & 0 \\ 0 & |w|^2 + |z|^2 \end{pmatrix},$$

de donde se sigue inmediatamente la segunda parte del teorema. ■

Definición 4.8 Si R es un cuerpo no arquimediano y H es su anillo de cuaternios, llamaremos *conjugado* de un $q \in H$ al cuaternio

$$\bar{q} = \overline{a + bi + cj + dk} = a - bi - cj - dk.$$

Hemos probado que la conjugación es un *antiautomorfismo* de H (lo que significa que invierte los productos) que extiende a la conjugación compleja

Tenemos que H es un espacio vectorial sobre R de base $1, i, j, k$. Nos referiremos a esta base como la *base canónica* de H . El producto escalar canónico de H es el que hace que la base canónica sea ortonormal. Definimos el *valor absoluto* de un cuaternio q como la norma respecto del valor absoluto canónico, es decir:

$$|q| = \sqrt{q\bar{q}} = \sqrt{a^2 + b^2 + c^2 + d^2} \in R.$$

El hecho de que la conjugación sea un automorfismo implica inmediatamente la tercera de las propiedades siguientes. Las otras dos son válidas para cualquier norma derivada de un producto escalar:

1. $|q| \geq 0$ y $|q| = 0$ si y sólo si $q = 0$.
2. $|q_1 + q_2| \leq |q_1| + |q_2|$.
3. $|q_1q_2| = |q_1||q_2|$.

Ahora es inmediato que H es un anillo de división, es decir, que todo $q \in H$ no nulo tiene un inverso para el producto. Basta observar que dicho inverso es

$$q^{-1} = \frac{\bar{q}}{|q|^2}.$$

Sin embargo, no es un cuerpo, porque $ij = k = -ji$ (como se comprueba inmediatamente a través de la representación matricial).

Más precisamente, es fácil ver que si restringimos el producto de H al conjunto $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ obtenemos una estructura de grupo no abeliano que por [Al 6.42] es isomorfo al grupo cuaternio Q_8 , ya que la única alternativa sería que fuera isomorfo al grupo diédrico D_8 , pero no es el caso, porque no tiene más que un elemento de orden 2 (el -1).

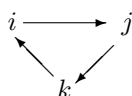
El teorema siguiente es ahora inmediato:

Teorema 4.9 Si R es un cuerpo ordenado pitagórico, entonces el anillo de división H de los cuaternios sobre R está determinado salvo isomorfismo por que sus elementos se expresan de forma única como $q = a + bi + cz + dk$ con $a, b, c, d \in R$ y que el producto en el conjunto $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ determina una estructura de grupo de tipo Q_8 .

Explícitamente:

$$i^2 = j^2 = k^2 = -1,$$

y al multiplicar dos de los tres elementos i, j, k se obtiene el tercero si el producto se hace en el sentido que indican las flechas:



y el tercero cambiado de signo si se hace en sentido contrario, es decir:

$$ij = k, \quad jk = i, \quad ki = j, \quad ji = -k, \quad kj = -i, \quad ik = -j.$$

Esto basta para calcular explícitamente cualquier producto de cuaternios, aunque en muchas ocasiones puede ser más conveniente la representación matricial. En particular tenemos una nueva representación matricial del grupo Q_8 distinta de la que hemos dado para introducirlo en [TG], a saber, Q_8 es isomorfo al subgrupo generado en $\text{Mat}_2(\mathbb{C})$ por las matrices

$$i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Polinomios con infinitas raíces Los anillos de cuaternios muestran que el hecho de que todo polinomio en un cuerpo tiene a lo sumo tantas raíces como indica su grado depende esencialmente de la conmutatividad del cuerpo. En efecto, todo cuaternio de la forma $q = bi + cj + dk$ con $|q| = 1$ cumple $\bar{q} = -q$, luego $q^2 = -q\bar{q} = -1$, luego es raíz del polinomio $x^2 + 1$. Es claro entonces que este polinomio tiene infinitas soluciones en H , por ejemplo, todas las de la forma $bi + \sqrt{1-b^2}j$, con $-1 \leq b \leq 1$. ■

Sumas de cuatro cuadrados En este apartado consideramos el anillo de cuaternios sobre el cuerpo \mathbb{Q} de los números racionales, y en él consideramos

$$H_{\mathbb{Z}} = \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{Z}\},$$

que claramente es un subanillo. Entonces, es obvio que un número natural n es suma de cuatro cuadrados si y sólo si $n = |q|^2$, para un cierto $q \in H_{\mathbb{Z}}$. Ahora es inmediato el hecho siguiente:

Si dos números naturales son suma de cuatro cuadrados, su producto también lo es.

En efecto, si $m = |q|^2$ y $n = |q'|^2$, para ciertos $q, q' \in H_{\mathbb{Z}}$, entonces $qq' \in H_{\mathbb{Z}}$ y $|qq'|^2 = |q|^2|q'|^2 = mn$.

Este hecho se utiliza en la demostración del teorema de Lagrange, según el cual, de hecho, todo número natural puede expresarse como suma de cuatro cuadrados (véase [ITA1 3.5]). ■

4.3 La geometría de los cuaternios

En toda esta sección R será un cuerpo ordenado pitagórico y H el anillo de división de los cuaternios sobre R . Aunque H es un espacio vectorial de dimensión 4 sobre R , en realidad “está pensado” para estudiar la geometría tridimensional de R , pues de los cuatro elementos de la base canónica, $1, i, j, k$, el primero tiene un comportamiento distinto al de los otros tres. Por ello conviene introducir las definiciones siguientes:

Definición 4.10 Llamaremos \vec{E} al subespacio vectorial de H generado por los vectores i, j, k , de modo que todo cuaternio $q \in H$ se expresa de forma única como

$$q = r + \vec{v} = r + xi + yj + zk, \quad \text{con } r \in R, \vec{v} \in \vec{E}.$$

Diremos que r es la *parte escalar* de q , mientras que \vec{v} es su *parte vectorial*. A los elementos de \vec{E} lo llamaremos *vectores* o *cuaternios puros*. A los elementos de R los llamaremos *escalares*.

Consideraremos a \vec{E} como espacio euclídeo con el producto escalar respecto al que la base canónica i, j, k es ortonormal (que no es sino la restricción del producto escalar de H). Usaremos la notación $q \cdot q'$ para referirnos al producto escalar de dos cuaternios, mientras que qq' será el producto de la estructura de anillo de división. Por ejemplo, $i \cdot j = 0$, mientras que $ij = k$.

El producto vectorial Observemos la expresión explícita para el producto de dos vectores:

$$\begin{aligned} \vec{v}\vec{v}' &= (xi + yj + zk)(x'i + y'j + z'k) = -xx' - yy' - zz' \\ &\quad + (yz' - zy')i - (x'z - xz')j + (xy' - x'y)k. \end{aligned}$$

Vemos que la parte escalar del producto es $-\vec{v} \cdot \vec{v}'$. Conviene dar nombre a la parte vectorial:

El *producto vectorial* de dos vectores $\vec{v} = xi + yj + zk$ y $\vec{v}' = x'i + y'j + z'k$ se define como⁴

$$\vec{v} \times \vec{v}' = (yz' - zy')i - (x'z - xz')j + (xy' - x'y)k.$$

⁴En términos de determinantes, esta definición puede expresarse así:

$$\vec{v} \times \vec{v}' = \begin{vmatrix} i & j & k \\ x & y & z \\ x' & y' & z' \end{vmatrix}.$$

De este modo, concluimos que

$$\vec{v} \vec{v}' = -\vec{v} \cdot \vec{v}' + \vec{v} \times \vec{v}'. \quad (4.3)$$

Así pues, el producto escalar de dos vectores nos da la parte escalar del producto (cambiada de signo) y el producto vectorial nos da la parte vectorial. El producto escalar ya lo conocemos, así que vamos a estudiar el producto vectorial. De la propia definición se sigue inmediatamente que es anticonmutativo:

$$\vec{v} \times \vec{v}' = -\vec{v}' \times \vec{v}.$$

Despejando en (4.3) obtenemos que

$$\vec{v} \times \vec{v}' = \vec{v} \vec{v}' + \vec{v} \cdot \vec{v}'.$$

De aquí se sigue trivialmente que el producto vectorial es bilineal:

$$\begin{aligned} (\vec{u} + \vec{v}) \times \vec{w} &= \vec{u} \times \vec{w} + \vec{v} \times \vec{w}, & \vec{u} \times (\vec{v} + \vec{w}) &= \vec{u} \times \vec{v} + \vec{u} \times \vec{w}, \\ (r\vec{u}) \times \vec{v} &= r(\vec{u} \times \vec{v}) = \vec{u} \times (r\vec{v}). \end{aligned}$$

Además, teniendo en cuenta que $\vec{v} \vec{v} = -\vec{v}$, vemos también que

$$\vec{v} \times \vec{v} = -\vec{v}(-\vec{v}) + \vec{v} \cdot \vec{v} = -|\vec{v}|^2 + |\vec{v}|^2 = 0.$$

Más en general, dos vectores linealmente dependientes tiene producto vectorial nulo. Enseguida veremos que el recíproco también es cierto.

De (4.3) se sigue que, para vectores ortogonales, el producto vectorial coincide con el producto. En particular $i \times j = k$, etc.

Es importante tener presente que el producto escalar no es asociativo. Por ejemplo,

$$(i \times j) \times j = k \times j = -i, \quad i \times (j \times j) = i \times 0 = 0.$$

No obstante, la asociatividad del producto de H nos da una relación de interés:

$$\begin{aligned} \vec{u}(\vec{v} \vec{w}) &= \vec{u}(-\vec{v} \cdot \vec{w} + \vec{v} \times \vec{w}) = -(\vec{v} \cdot \vec{w})\vec{u} + \vec{u}(\vec{v} \times \vec{w}) \\ &= -\vec{u} \cdot (\vec{v} \times \vec{w}) + \vec{u} \times (\vec{v} \times \vec{w}) - (\vec{v} \cdot \vec{w})\vec{u}. \\ (\vec{u} \vec{v})\vec{w} &= -(\vec{u} \times \vec{v}) \cdot \vec{w} + (\vec{u} \times \vec{v}) \times \vec{w} - (\vec{u} \cdot \vec{v})\vec{w}. \end{aligned}$$

Al igualar las partes reales queda $\vec{u} \cdot (\vec{v} \times \vec{w}) = \vec{w} \cdot (\vec{u} \times \vec{v})$, e intercambiando los papeles de los vectores podemos añadir una segunda identidad:

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = \vec{v} \cdot (\vec{w} \times \vec{u}) = \vec{w} \cdot (\vec{u} \times \vec{v}).$$

En otros términos, en un producto mixto de la forma $\vec{u} \cdot (\vec{v} \times \vec{w})$ podemos permutar cíclicamente los vectores sin alterar el resultado (si la permutación no es cíclica se cambia el signo).

Con esto ya estamos en condiciones de interpretar geoméricamente el producto vectorial. En primer lugar observamos que

$$\vec{v} \cdot (\vec{v} \times \vec{w}) = \vec{w} \cdot (\vec{v} \times \vec{v}) = 0, \quad \vec{w} \cdot (\vec{v} \times \vec{w}) = \vec{v} \cdot (\vec{w} \times \vec{w}) = 0,$$

luego si \vec{v} y \vec{w} son vectores no nulos, el vector $\vec{v} \times \vec{w}$ es ortogonal a ambos. Esto determina su dirección. Por otra parte, de (4.3) se sigue que

$$|\vec{v} \times \vec{v}'|^2 = (\vec{v} \cdot \vec{v}')^2 + |\vec{v} \times \vec{v}'|^2,$$

luego

$$|\vec{v} \times \vec{v}'|^2 = |\vec{v}|^2 |\vec{v}'|^2 - |\vec{v}|^2 |\vec{v}'|^2 \cos^2(\widehat{\vec{v}, \vec{v}'}) = |\vec{v}|^2 |\vec{v}'|^2 \sin^2(\widehat{\vec{v}, \vec{v}'}),$$

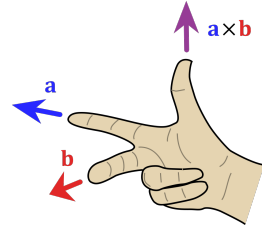
y en definitiva:

Teorema 4.11 *Si \vec{v} y \vec{w} son dos vectores no nulos, entonces $\vec{v} \times \vec{w}$ es un vector ortogonal a ambos de módulo*

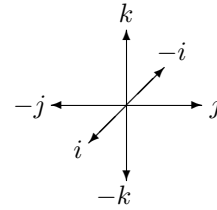
$$|\vec{v} \times \vec{w}| = |\vec{v}| |\vec{w}| \sin(\widehat{\vec{v}, \vec{w}}).$$

Nota Observemos que en realidad, si \vec{v} y \vec{w} son linealmente independientes, existen dos vectores ortogonales a \vec{v} y \vec{w} con el módulo indicado, es decir, el teorema anterior determina la dirección y el módulo de $\vec{v} \times \vec{w}$, pero no su sentido.

Cuando discutamos el concepto de “orientación”, en la sección 5.6, veremos que el sentido del producto vectorial está determinado por la “regla de la mano derecha”, Esto significa que si representamos los vectores i, j, k con el criterio usual, según el cual cuando el dedo índice de la mano derecha apunta hacia i y el dedo medio apunta hacia j , el pulgar apunta hacia k , entonces se cumple en general que si ponemos la mano derecha con el índice apuntando hacia un vector \vec{u} y el dedo medio apuntando hacia otro vector \vec{v} , entonces el pulgar señala el sentido de $\vec{u} \times \vec{v}$.



De momento el lector puede comprobar que la regla de la mano derecha se cumple para los productos vectoriales de los vectores $\pm i, \pm j, \pm k$. ■



Veamos una última propiedad del producto vectorial:

Teorema 4.12 (Fórmula de Lagrange)

$$\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \cdot \vec{w}) \vec{v} - (\vec{u} \cdot \vec{v}) \vec{w}.$$

DEMOSTRACIÓN: Los dos miembros son lineales en $\vec{u}, \vec{v}, \vec{w}$, por lo que expresando cada vector como combinación lineal de los vectores i, j, k y desarrollando ambos miembros, basta probar que la igualdad se cumple cuando las tres variables $\vec{u}, \vec{v}, \vec{w}$ toman valores i, j, k (con lo que en particular son una base ortonormal). Esto en principio nos deja 27 casos que comprobar. Ahora bien,

si $\vec{v} = \vec{w}$ o los tres vectores son distintos dos a dos, ambos miembros valen 0, luego podemos suponer que $\vec{v} \neq \vec{w}$ y que \vec{u} coincide con uno de los dos. Por último, también es claro que si intercambiamos \vec{v} y \vec{w} ambos miembros cambian de signo, luego basta probar el caso

$$\vec{w} \times (\vec{v} \times \vec{w}) = \vec{v},$$

donde $\vec{v}, \vec{w} \in \{i, j, k\}$ son distintos dos a dos. Ahora bien, el producto vectorial de vectores ortogonales coincide con el producto de H , luego

$$\vec{w} \times (\vec{v} \times \vec{w}) = \vec{w} \vec{v} \vec{w} = -\vec{w} \vec{w} \vec{v} = \vec{v}. \quad \blacksquare$$

Giros tridimensionales Veamos que los cuaternios pueden usarse para describir y estudiar los giros tridimensionales. Para ello conviene observar que existe un análogo para cuaternios de la forma polar de un número complejo:

Si $q = r + \vec{v} \in H$ tiene módulo 1, entonces $|q|^2 = r^2 + |\vec{v}|^2 = 1$, luego existe un ángulo $0 \leq \alpha \leq \pi$ tal que $r = \cos \alpha$, $|\vec{v}| = \sin \alpha$ (notemos que $|\vec{v}| \geq 0$, luego tiene que ser $\alpha \leq \pi$). Así pues:

$$q = \cos \alpha + \sin \alpha \vec{u},$$

donde $|\vec{u}| = 1$. En general, todo cuaternio no nulo puede expresarse en la forma

$$q = \rho(\cos \alpha + \sin \alpha \vec{u}),$$

donde $\rho = |q|$ y $|\vec{u}| = 1$ (para cualquier ángulo α). El vector \vec{u} está determinado por q salvo un signo, porque q admite una representación alternativa cambiando \vec{u} por $-\vec{u}$ y α por $-\alpha$.

Sea H_1 el grupo multiplicativo de los cuaternios de módulo 1 y veamos que H_1 actúa sobre \vec{E} por conjugación, es decir, que si $q \in H_1$ y $\vec{v} \in \vec{E}$, entonces

$$\vec{v}^q = q^{-1} \vec{v} q \in \vec{E}.$$

Para ello observamos que el inverso de un cuaternio de módulo 1 es su conjugado, luego en forma polar tenemos:

$$q = \cos \alpha + \sin \alpha \vec{u}, \quad q^{-1} = \cos \alpha - \sin \alpha \vec{u}.$$

Como \vec{v} puede descomponerse en suma de un múltiplo de \vec{u} y de un vector ortogonal a \vec{u} , basta probar que $\vec{v}^q \in \vec{E}$ en estos dos casos. En el primero, teniendo en cuenta que $\vec{u} \vec{u} = -|\vec{u}|^2 = -1$, tenemos que

$$\begin{aligned} \vec{u}^q &= (\cos \alpha \vec{u} + \sin \alpha)(\cos \alpha + \sin \alpha \vec{u}) = \\ &= \cos^2 \alpha \vec{u} + \sin \alpha \cos \alpha - \sin \alpha \cos \alpha + \sin^2 \alpha \vec{u} = \vec{u} \in \vec{E}. \end{aligned}$$

Ahora suponemos que $\vec{u} \cdot \vec{v} = 0$. Entonces

$$\begin{aligned} \vec{v}^q &= (\cos \alpha \vec{v} - \sin \alpha \vec{u} \times \vec{v})(\cos \alpha + \sin \alpha \vec{u}) = \\ &= \cos^2 \alpha \vec{v} - \sin \alpha \cos \alpha \vec{u} \times \vec{v} + \sin \alpha \cos \alpha \vec{v} \times \vec{u} - \sin^2 \alpha (\vec{u} \times \vec{v}) \times \vec{u}. \end{aligned}$$

Por la fórmula de Legendre:

$$(\vec{u} \times \vec{v}) \times \vec{u} = -\vec{u} \times (\vec{u} \times \vec{v}) = \vec{v},$$

luego queda

$$\begin{aligned} \vec{v}^q &= (\cos^2 \alpha - \operatorname{sen}^2 \alpha) \vec{v} + 2 \operatorname{sen} \alpha \cos \alpha (\vec{v} \times \vec{u}) \\ &= \cos 2\alpha \vec{v} + \operatorname{sen} 2\alpha (\vec{v} \times \vec{u}) \in \vec{E}. \end{aligned}$$

Nota En realidad $\vec{v}^q \in \vec{E}$ para todo q , sin suponer que sea unitario, pues si $\rho \neq 0$ tenemos que $\vec{v}^{\rho q} = \vec{v}^q$, pero precisamente por esto no perdemos generalidad si consideramos únicamente conjugaciones por cuaternios unitarios. ■

Si llamamos $G_q : \vec{E} \rightarrow \vec{E}$ a la conjugación por q , tenemos que es una isometría lineal en \vec{E} , pues obviamente es lineal, y biyectiva (su inversa es $G_{q^{-1}}$), y

$$|G_q(\vec{v})| = |q^{-1}\vec{v}q| = |q|^{-1}|\vec{v}||q| = |\vec{v}|.$$

Esto implica que, como biyección afín, es una isometría, luego es una isometría lineal. Más precisamente, podemos tomar un vector unitario \vec{v} ortogonal a \vec{u} , y entonces $\vec{v}, \vec{u} \times \vec{v}, \vec{u}$ forman una base ortonormal de \vec{E} , respecto a la cual, la matriz de G_q es de la forma

$$M = \begin{pmatrix} \cos 2\alpha & -\operatorname{sen} 2\alpha & 0 \\ \operatorname{sen} 2\alpha & \cos 2\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

En efecto, nos falta calcular la segunda fila, pero, usando la fórmula de Lagrange:

$$G_q(\vec{u} \times \vec{v}) = \cos 2\alpha (\vec{u} \times \vec{v}) + \operatorname{sen} 2\alpha (\vec{u} \times \vec{v}) \times \vec{u} = \cos 2\alpha (\vec{u} \times \vec{v}) + \operatorname{sen} 2\alpha \vec{u}.$$

Definición 4.13 Si E es un espacio euclídeo de dimensión 3, un *giro* en E es una isometría $G : E \rightarrow E$ que tenga un punto fijo O y que, respecto a un cierto sistema de referencia ortonormal $(O; \vec{v}_1, \vec{v}_2, \vec{v}_3)$ su matriz sea de la forma

$$M_\alpha = \begin{pmatrix} \cos \alpha & \operatorname{sen} \alpha & 0 \\ -\operatorname{sen} \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

para un cierto ángulo α .

Es claro que M_0 es la identidad. Para ángulos $\alpha \neq 0$ (módulo 2π) tenemos que los puntos fijos de G son exactamente los de la recta $r = O + \langle \vec{v}_3 \rangle$, que recibe el nombre de *eje* del giro. La restricción de G a cualquier plano ortogonal al eje de giro es claramente un giro de ángulo α (y centro la intersección del eje con el plano), en el sentido de 3.40. Como en el caso bidimensional, el ángulo de giro está determinado salvo el signo, en el sentido de que un giro de ángulo α respecto de una base tiene ángulo $-\alpha$ respecto de la base que resulta de cambiar el signo a uno de sus vectores.

Es claro que la matriz de un giro tiene la forma indicada en la definición exactamente respecto de aquellas bases ortonormales cuyo tercer vector tiene la dirección del eje de giro.

En estos términos hemos probado que si $q = \cos \alpha + \operatorname{sen} \alpha \vec{u} \in H_1$, entonces $G_q : \vec{E} \rightarrow \vec{E}$ es un giro, que es trivial (la identidad) cuando $q = \pm 1$ y tiene por eje la recta $\langle \vec{u} \rangle$ en caso contrario. Concretamente, si fijamos un vector unitario \vec{v} ortogonal a \vec{u} , tenemos que G_q es un giro de ángulo -2α respecto de la base $\vec{v}, \vec{u} \times \vec{v}, \vec{u}$, y como siempre podemos tomar un $q \in H_1$ con parte real $\cos(\alpha/2)$ y parte vectorial en la dirección de cualquier recta prefijada que pase por el origen, concluimos que todo giro de \vec{E} de centro el origen es de la forma G_q , para cierto $q \in H_1$.

De aquí deducimos un hecho que no es inmediato:

Teorema 4.14 *Si E es un espacio euclídeo tridimensional, el conjunto $\operatorname{Is}_O^+(E)$ de todos los giros en E cuyo eje pasa por un mismo punto O es un grupo con la composición de aplicaciones.*

En efecto, no perdemos generalidad si suponemos que el espacio es el espacio \vec{E} de los cuaternios vectoriales y que O es el vector nulo. Entonces todo giro es de la forma G_q , para $q \in H_1$, y $G_q \circ G_{q'} = G_{qq'}$, mientras que $G_q^{-1} = G_{q^{-1}}$.

Más aún, lo que tenemos es que si \mathcal{G} es el grupo de todos los giros de \vec{E} cuyo eje pasa por O , entonces $G : H_1 \rightarrow \mathcal{G}$ es un epimorfismo de grupos, cuyo núcleo es $\{\pm 1\}$, luego $\mathcal{G} \cong H_1/\{\pm 1\}$.

Explícitamente, si $q = q_0 + q_1i + q_2j + q_3k \in H_1$, entonces

$$\begin{aligned} G_q(i) = q^{-1}iq &= (q_0 - q_1i - q_2j - q_3k)i(q_0 + q_1i + q_2j + q_3k) \\ &= (q_1 + q_0i - q_3j + q_2k)(q_0 + q_1i + q_2j + q_3k) \\ &= (q_0^2 + q_1^2 - q_2^2 - q_3^2)i + (2q_1q_2 - 2q_0q_3)j + (2q_1q_3 + 2q_0q_2)k, \end{aligned}$$

e igualmente se calculan $G_q(j)$, $G_q(k)$, con lo que deducimos que la matriz de G_q en la base canónica (i, j, k) es

$$M_q = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2q_1q_2 - 2q_0q_3 & 2q_1q_3 + 2q_0q_2 \\ 2q_1q_2 + 2q_0q_3 & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2q_1q_3 - 2q_0q_1 \\ 2q_1q_3 - 2q_0q_2 & 2q_2q_3 + 2q_0q_1 & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}.$$

Recíprocamente, si conocemos la matriz $M_q = (m_{ij})$ de un giro, podemos calcular q observando en primer lugar que

$$m_{11} + m_{22} + m_{33} = 3q_0^2 - q_1^2 - q_2^2 - q_3^2 = 4q_0^2 - |q|^2 = 4q_0 - 1,$$

de donde obtenemos la primera de las cuatro fórmulas siguientes (y las otras tres son inmediatas):

$$\begin{aligned} q_0 &= \frac{\sqrt{1 + m_{11} + m_{22} + m_{33}}}{2}, \\ q_1 &= \frac{m_{32} - m_{23}}{4q_0}, \quad q_2 = \frac{m_{13} - m_{31}}{4q_0}, \quad q_3 = \frac{m_{21} - m_{12}}{4q_0}. \end{aligned}$$

Notemos que en el cálculo de q_0 es irrelevante que tomemos la raíz cuadrada positiva o la negativa, porque q está determinado salvo signo.

Ejemplo Consideremos el giro de 120° respecto a la recta que pasa por el origen con vector director $(1, 1, 1)$. Corresponde al cuaternio unitario

$$q = \cos 60^\circ - \operatorname{sen} 60^\circ \frac{1}{\sqrt{3}}(i + j + k) = \frac{1}{2}(1 - i - j - k).$$

Entonces, como cabía esperar,

$$G_q(i) = \frac{1}{2}(1 + i + j + k)i\frac{1}{2}(1 - i - j - k) = \frac{1}{4}(-1 + i + j - k)(1 - i - j - k) = j,$$

e igualmente $G_q(j) = k$, $G_q(k) = i$. La matriz del giro es

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

■

Automorfismos de Q_8 Más en general, podemos considerar el subgrupo G de \mathcal{G} formado por los giros que se restringen a permutaciones del conjunto $\Omega = \{\pm i, \pm j, \pm k\}$. En el ejemplo anterior hemos calculado un elemento de G .

Observemos que si $g \in G$ está representado por el cuaternio q , es decir, si $g = G_q$, entonces, si $x, y \in \Omega$ cumplen $x \neq \pm y$, se cumple que $xy \in \Omega$ y

$$g(xy) = (xy)^q = x^q y^q = g(x)g(y).$$

Por otra parte, $g(-x) = -g(x)$ (un giro transforma vectores opuestos en vectores opuestos, luego la fórmula anterior vale para todo par $x, y \in \Omega$ sin convenimos en que $g(1) = 1$ y $g(-1) = -1$). Más aún, vale para todo par de elementos $x, y \in Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$, por lo que tenemos una aplicación $\phi : G \rightarrow \operatorname{Aut} Q_8$ que a cada giro de G le hace corresponder el automorfismo de Q_8 que sobre Ω actúa como G y fija a ± 1 .

Claramente, ϕ es un homomorfismo de grupos (la composición de dos giros se corresponde con la composición de los automorfismos correspondientes). De hecho, es un monomorfismo, pues si un giro deja invariantes a i, j, k , es que es la identidad.

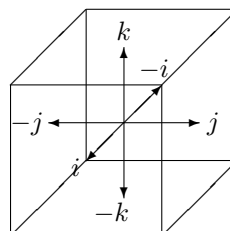
Observemos ahora que G tiene orden 24, porque siempre podemos tomar un giro $g \in G$ que transforme i en cualquiera de los 6 elementos de Ω y, una vez elegido $g(i)$, entonces $g(j)$ puede ser cualquier elemento de Ω distinto de $\pm g(i)$, lo que nos da 4 posibilidades, y una vez elegidos $g(i), g(j)$, necesariamente $g(k) = g(i)g(j)$, luego tenemos exactamente 24 posibilidades para los giros de G .

En particular, Q_8 tiene al menos 24 automorfismos (las imágenes de los giros de G), pero no puede tener más por el mismo motivo anterior: un automorfismo f está completamente determinado por las imágenes que toma sobre i, j (porque

éstos generan Q_8 , y $f(i)$ tiene que estar en Ω (luego, como máximo, tenemos 6 posibilidades) y, una vez elegido $f(i)$, necesariamente $f(j)$ tiene que ser un elemento de Ω distinto de $\pm f(i)$, luego como máximo tenemos 4 posibilidades y, en total, $|\text{Aut } Q_8| \leq 24$, luego $\text{Aut } Q_8 \cong G$ son grupos de orden 24.

Notemos que también podemos pensar en el grupo G como el grupo de los giros que dejan fijo un cubo con centro en el origen de coordenadas. Los seis elementos de Ω se corresponden con las seis caras del cubo. Al final de la sección 2.6 de [TG] hemos visto que el grupo de las simetrías de un cubo es isomorfo a Σ_4 , así que podemos concluir que

$$G \cong \text{Aut } Q_8 \cong \Sigma_4.$$



En el ejemplo anterior hemos construido un giro de G de orden 3. Cambiando el eje por $(\pm 1, \pm 1, \pm 1)$ obtenemos 8 giros de orden 3 (dos vectores opuestos determinan el mismo eje de giro, pero los giros de 120° respecto de ambos son giros opuestos, luego ciertamente tenemos 8 giros), que se corresponden con los 8 ciclos de longitud 3 en Σ_4 .

Es fácil ver que los seis giros de 90° respecto de los ejes (correspondientes a los ciclos de longitud 4) se obtienen mediante los cuaternios

$$\frac{1}{\sqrt{2}}(1 \pm i), \quad \frac{1}{\sqrt{2}}(1 \pm j), \quad \frac{1}{\sqrt{2}}(1 \pm k),$$

y los giros de 180° son sus cuadrados: i, j, k (que se corresponden con los tres pares de trasposiciones). Las seis trasposiciones se corresponden con los giros de 180° respecto de los ejes de la forma $(\pm 1, \pm 1, 0)$ (los que pasan por los puntos medios de dos pares de aristas opuestas), y se corresponden con cuaternios de la forma

$$\frac{1}{\sqrt{2}}(i \pm j).$$

■

Ejemplo Consideremos ahora las matrices

$$\sigma = \frac{1}{7} \begin{pmatrix} 6 & 2 & -3 \\ 2 & 3 & 6 \\ 3 & -6 & 2 \end{pmatrix}, \quad \tau = \frac{1}{7} \begin{pmatrix} 2 & 6 & -3 \\ -6 & 3 & 2 \\ 3 & 2 & 6 \end{pmatrix}.$$

No es evidente que correspondan a giros, pero admitiendo que es así, vamos a calcular los cuaternios asociados. Para la primera tenemos que

$$q_0 = \frac{\sqrt{11/7 + 1}}{2} = \frac{3}{\sqrt{14}},$$

$$q_1 = -\frac{12/7}{12/\sqrt{14}} = -\frac{2}{\sqrt{14}}, \quad q_2 = -\frac{6/7}{12/\sqrt{14}} = -\frac{1}{\sqrt{14}}, \quad q_3 = 0,$$

luego

$$q_\sigma = \frac{3}{\sqrt{14}} - \frac{2}{\sqrt{14}}i - \frac{1}{\sqrt{14}}j,$$

luego se trata de un giro de $\alpha = \frac{1}{2} \arccos \frac{3}{\sqrt{14}} \approx 18.35^\circ$ respecto del eje de vector director $(-2, -1, 0)$. Una forma de comprobar que la matriz dada corresponde realmente a un giro es calcular la matriz asociada a q_σ y comprobar que, en efecto, es la matriz dada. Similarmente se calcula

$$q_\tau = \frac{3}{\sqrt{14}} - \frac{1}{\sqrt{14}}j + \frac{2}{\sqrt{14}}k.$$

En muchos aspectos, representar un giro mediante un cuaternio unitario es más conveniente que hacerlo mediante una matriz. Para empezar, requiere sólo cuatro números en lugar de 9, muestra claramente el ángulo y el eje de giro, y permite calcular fácilmente el eje de giro de una composición. Por ejemplo, el producto $\sigma\tau$ corresponde a al giro asociado al cuaternio

$$q_\sigma q_\tau = \frac{1}{14}(3 - 2i - j)(3 - j + 2k) = \frac{1}{14}(8 - 8i - 2j + 8k).$$

Por lo tanto, la composición de los dos giros corresponde a un giro de ángulo $\frac{1}{2} \arccos \frac{8}{14} \approx 27.57^\circ$ alrededor de la recta de vector director $(-4, -1, 4)$. ■

Capítulo V

Bijecciones afines

La geometría sintética suele ser la más indicada para tratar problemas “locales”, en torno a figuras que involucran pocos puntos y rectas, mientras que la geometría analítica resulta más apta, por no decir imprescindible, para cuestiones “globales”, como son las relacionadas con las transformaciones del espacio en sí mismo. Dedicamos este capítulo a profundizar en las propiedades de estas transformaciones. Conviene distinguir siempre entre los resultados afines, válidos en todo espacio afín, de los genuinamente euclídeos, pues la geometría afín es una herramienta algebraica susceptible de ser aplicada al estudio de cuerpos arbitrarios, mientras que la estructura euclídea es más particular.

5.1 El grupo afín y el grupo lineal

Definición 5.1 Dado un espacio vectorial V sobre un cuerpo K , el *grupo lineal* de V , representado por $\text{LG}(V)$, es el conjunto de todos los automorfismos de V , es decir, el conjunto de todas las aplicaciones lineales biyectivas de V en sí mismo, que es un grupo con la composición de aplicaciones. Según observamos en [Al 6.20], si V tiene dimensión n , entonces $\text{LG}(V)$ es isomorfo al grupo $\text{LG}(n, K)$ de las matrices inversibles $n \times n$ con coeficientes en K .

Dado un espacio afín E , llamaremos *grupo afín* de E al grupo $\text{GA}(E)$ formado por las biyecciones afines de E en sí mismo, también con la composición de aplicaciones.

Recordemos (definición 3.16) que una afinidad $f \in \text{GA}(E)$ es una aplicación que actúa sobre cada punto P en la forma

$$f(P) = f(O) + \vec{f}(\overrightarrow{OP}),$$

donde $\vec{f} \in \text{LG}(\vec{E})$ y O es un punto de E . Tras 3.16 probamos que una afinidad admite una expresión de esta forma para cualquier punto O prefijado y siempre con la misma aplicación \vec{f} .

Es inmediato comprobar que la aplicación

$$\begin{array}{ccc} \text{GA}(E) & \longrightarrow & \text{LG}(\vec{E}) \\ f & \mapsto & \vec{f} \end{array}$$

es un epimorfismo de grupos. Su núcleo está formado por las aplicaciones de la forma

$$f(P) = f(O) + \overrightarrow{OP} = P + \overrightarrow{Pf(O)} + \overrightarrow{OP} = P + \overrightarrow{Of(O)},$$

donde $\overrightarrow{Of(O)}$ es un vector arbitrario de V . Así pues, el núcleo del epimorfismo anterior es el grupo $T(E)$ de las *traslaciones* de E , esto es, el grupo formado por las afinidades $T_{\vec{v}}$ dadas por $T_{\vec{v}}(P) = P + \vec{v}$.

Así pues, tenemos que $T(E) \trianglelefteq \text{GA}(E)$ y $\text{GA}(E)/T(E) \cong \text{LG}(V)$. Por otro lado, es obvio que la aplicación $\vec{v} \mapsto T_{\vec{v}}$ es un isomorfismo $\vec{E} \cong T(E)$ (donde consideramos a \vec{E} como grupo con la suma). En particular vemos que $T(E)$ es un grupo abeliano. Diremos que dos conjuntos son *trasladados* si uno es la imagen del otro por una traslación.

También podemos sumergir $\text{LG}(\vec{E})$ en $\text{GA}(E)$. Dado un punto $O \in E$, el conjunto $\text{GA}_O(E)$ formado por los elementos $f \in \text{GA}(E)$ tales que $f(O) = O$ es claramente un subgrupo¹ de $\text{GA}(E)$ isomorfo a $\text{LG}(\vec{E})$. Concretamente, cada $\vec{f} \in \text{LG}(\vec{E})$ se corresponde con la afinidad \vec{f}_O dada por

$$\vec{f}_O(P) = O + \vec{f}(\overrightarrow{OP}).$$

El isomorfismo inverso es simplemente $f \mapsto \vec{f}$.

Fijado O , toda biyección afín f se expresa como

$$f(P) = f(O) + \vec{f}(\overrightarrow{OP}) = O + \vec{f}(\overrightarrow{OP}) + \overrightarrow{Of(O)} = \vec{f}_O(P) + \overrightarrow{Of(O)},$$

o sea, como composición de $\vec{f}_O \in \text{GA}_O(E)$ con la traslación de dirección $\overrightarrow{Of(O)}$. Esto significa que $\text{GA}(E) = \text{GA}_O(E) T(E)$. Por otro lado es claro que las traslaciones no tienen puntos fijos, luego $\text{GA}_O(E) \cap T(E) = 1$.

Se comprueba fácilmente que $T_{\vec{v}}^{\vec{f}_O} = T_{\vec{f}(\vec{v})}$. Esto significa que la acción de $\text{GA}_O(E)$ sobre $T(E)$ por conjugación se corresponde a través de los isomorfismos que hemos indicado con la acción natural de $\text{LG}(\vec{E})$ sobre \vec{E} .

Si una afinidad tiene un punto fijo O , entonces su expresión respecto a un sistema de referencia con origen O es particularmente simple, por ello es interesante el teorema siguiente, que describe el conjunto de puntos fijos de una afinidad y en un caso particular garantiza incluso su existencia:

¹De hecho, lo que tenemos es que $\text{GA}(E)$ actúa sobre E y $\text{GA}_O(E)$ es precisamente el estabilizador de O .

Teorema 5.2 Sea $f : E \rightarrow E$ una afinidad en un espacio afín E . Entonces

1. El conjunto $\vec{I}(f)$ de puntos fijos (o invariantes) de f es un subespacio vectorial de \vec{E} .
2. Si f tiene un punto fijo O , entonces el conjunto $I(f)$ de puntos fijos de f es la variedad lineal $O + \vec{I}(f)$.
3. Si $\vec{I}(f) = 0$, entonces f tiene un único punto fijo.

DEMOSTRACIÓN: 1 es evidente. En 2 es claro que todos los puntos de $O + \vec{I}(f)$ son puntos fijos de f . Recíprocamente, si $f(P) = P$ esto equivale a que $P = O + \vec{f}(\vec{OP})$, luego $\vec{f}(\vec{OP}) = \vec{OP}$ y por lo tanto $\vec{OP} \in \vec{I}(f)$.

Veamos 3. Sea O un origen arbitrario. Estamos buscando un punto P tal que $P = f(O) + \vec{f}(\vec{OP}) = O + \vec{Of}(O) + \vec{f}(\vec{OP})$, es decir, tal que

$$\vec{OP} - \vec{f}(\vec{OP}) = \vec{Of}(O).$$

La aplicación $1 - \vec{f}$ (donde 1 representa a la identidad) es lineal y por hipótesis su núcleo es 0, luego es inyectiva, y por consiguiente suprayectiva, luego ciertamente existe un vector \vec{OP} tal que $(1 - \vec{f})(\vec{OP}) = \vec{Of}(O)$, como queríamos probar. La unicidad del punto fijo se sigue del apartado anterior. ■

5.2 Homotecias

Las homotecias son una familia importante de aplicaciones afines, pues contienen esencialmente el concepto de semejanza de figuras, que hasta ahora hemos estudiado únicamente para el caso de los triángulos. Comenzamos definiéndolas algebraicamente y después las caracterizaremos geométricamente.

Definición 5.3 Sea E un espacio afín sobre un cuerpo K . La *homotecia* de centro $O \in E$ y razón $k \in K^* = K \setminus \{0\}$ es la aplicación $H(O, k) \in \text{GA}(E)$ dada por

$$H(O, k)(P) = O + k\vec{OP}.$$

Llamaremos $H(O, E)$ al grupo de todas las homotecias de centro O en E . Dos conjuntos son *homotéticos* (gr. ‘con el mismo aspecto’) si uno es la imagen del otro por una homotecia.²

Es inmediato comprobar que la aplicación $k \mapsto H(O, k)$ es un isomorfismo $K^* \cong H(O, E)$.

Las homotecias se pueden caracterizar por sus aplicaciones lineales asociadas. Llamaremos *homotecia lineal* de razón $k \in K^*$ en un espacio vectorial V sobre un cuerpo K a la aplicación dada por $f(\vec{v}) = k\vec{v}$. Claramente las homotecias lineales forman un grupo $H(V)$ isomorfo a K^* .

²Notar que esta relación es reflexiva y simétrica, pero enseguida veremos que no es transitiva.

Teorema 5.4 Una afinidad f de un espacio afín E en sí mismo es una homotecia de razón $k \neq 1$ si y sólo si \vec{f} es una homotecia lineal de razón k .

DEMOSTRACIÓN: Si $f = H(O, k)$, entonces $f(P) = O + \vec{f}(\overrightarrow{OP}) = O + k\overrightarrow{OP}$, luego $\vec{f}(\overrightarrow{OP}) = k\overrightarrow{OP}$ para todo vector \overrightarrow{OP} de \vec{E} , es decir, \vec{f} es la homotecia lineal de razón k .

Recíprocamente, si \vec{f} es la homotecia de razón $k \neq 1$, entonces es claro que $\vec{f}(\vec{0}) = \vec{0}$, luego por 5.2 sabemos que f tiene un punto fijo O , de donde $f = H(O, k)$. ■

Notemos que la única homotecia de razón 1 es la identidad. Ahora es claro que la composición de dos homotecias de centros distintos no es necesariamente una homotecia. Concretamente, la composición de una homotecia de razón k con otra de razón $1/k$ da lugar a una afinidad con $\vec{f} = 1$, pero si los centros no son el mismo es fácil ver que $f \neq 1$, y entonces f es una traslación. Pronto veremos que la composición de dos homotecias es siempre una homotecia o una traslación. Antes necesitamos algunos conceptos auxiliares:

Definición 5.5 Diremos que dos variedades lineales L y M de un espacio afín E son *suplementarias* si $\vec{E} = \vec{L} \oplus \vec{M}$.

Notemos que si L y M son suplementarias entonces existe un punto O tal que $L \cap M = \{O\}$. En efecto, dado $P \in L$ y $Q \in M$, entonces $\overrightarrow{PQ} = \vec{v} + \vec{w}$, donde $\vec{v} \in \vec{L}$ y $\vec{w} \in \vec{M}$, luego $O = P + \vec{v} = Q - \vec{w}$ cumple $O \in L \cap M$. Si $O' \in L \cap M$, entonces $\overrightarrow{OO'} \in \vec{L} \cap \vec{M} = 0$, luego $O = O'$.

Dadas dos variedades lineales suplementarias L y M en un espacio afín E , todo punto de E se expresa de forma única como $O + \vec{v} + \vec{w}$, donde $L \cap M = \{O\}$, $\vec{v} \in \vec{L}$ y $\vec{w} \in \vec{M}$. Llamaremos *proyección* de E en L paralela a M a la aplicación $p : E \rightarrow L$ dada por $p(O + \vec{v} + \vec{w}) = O + \vec{v}$.

Alternativamente, $p(P)$ es el único punto de intersección entre las variedades suplementarias L y $P + \vec{M}$. También es claro que p es una afinidad, pues

$$p(P) = O + \vec{p}(\overrightarrow{OP}),$$

donde \vec{p} es la proyección de $\vec{L} \oplus \vec{M}$ sobre \vec{L} .

Ahora estamos en condiciones de probar una versión general del teorema de Tales. Podría pensarse que el teorema de Tales es genuinamente euclídeo, por lo que no puede enunciarse en espacios afines, pero no es así, y ello se debe a la observación siguiente: si P, Q, R y S son cuatro puntos colineales en un espacio afín sobre un cuerpo K y $P \neq Q$, entonces $\overrightarrow{RS} = \alpha \overrightarrow{PQ}$, para un único $\alpha \in K$. Por lo tanto podemos definir

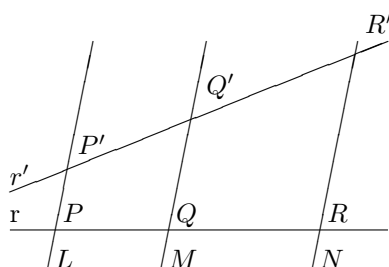
$$\frac{\overrightarrow{RS}}{\overrightarrow{PQ}} = \alpha.$$

Notemos que esta definición generaliza a la que dimos al principio de la sección [IGE 3.6]. En un espacio afín no podemos comparar vectores que no correspondan a puntos colineales,³ pero para formular el teorema de Tales no necesitamos más:

Teorema 5.6 (Teorema de Tales) Sean L, M, N tres hiperplanos paralelos en un espacio afín E . Si r es cualquier recta no paralela a estos hiperplanos, entonces la razón

$$\frac{\overrightarrow{PQ}}{\overrightarrow{PR}}, \quad \text{donde } r \cap L = \{P\}, \quad r \cap M = \{Q\}, \quad r \cap N = \{R\},$$

es independiente de r .

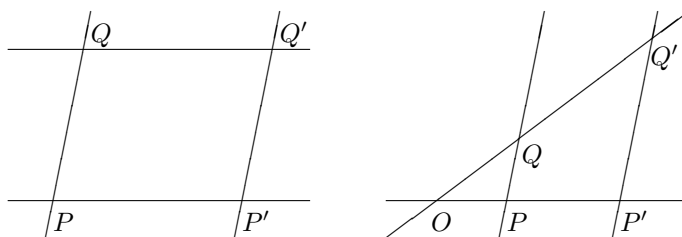


DEMOSTRACIÓN: Sea W el espacio director de los tres hiperplanos. Sea r' otra recta no paralela a L, M, N que los corte en los puntos P', Q', R' . Sea $p : E \rightarrow r'$ la proyección paralela a W . Entonces es claro que $p(P) = P', p(Q) = Q', p(R) = R'$. Por lo tanto $\overrightarrow{P'R'} = \vec{p}(\overrightarrow{PR})$ y $\overrightarrow{P'Q'} = \vec{p}(\overrightarrow{PQ})$. Si se cumple $\overrightarrow{PQ} = \alpha \overrightarrow{PR}$, aplicando \vec{p} obtenemos $\overrightarrow{P'Q'} = \alpha \overrightarrow{P'R'}$. ■

Como consecuencia obtenemos:

Teorema 5.7 Sean P, Q, P', Q' puntos distintos de un espacio afín tales que las rectas PQ y $P'Q'$ sean paralelas.

1. Si las rectas PP' y QQ' son paralelas, entonces $Q' = T_{\vec{v}}(Q)$, con $\vec{v} = \overrightarrow{PP'}$.
2. Si las rectas PP' y QQ' se cortan en O , entonces $Q' = H(O, k)(Q)$, donde $k = \overrightarrow{OP'} / \overrightarrow{OP}$.



³Podríamos haber definido más en general la razón entre dos vectores con la misma dirección, aunque los puntos que los definen no sean colineales, pero esto ya no sería un invariante afín (las afinidades no lo conservan) mientras que sobre puntos colineales sí lo es, como es fácil comprobar.

DEMOSTRACIÓN: 1) Hemos de probar que $Q' = Q + \overrightarrow{PP'}$. Para ello notamos que $Q + \overrightarrow{PP'}$ está en la recta QQ' y por otro lado

$$Q + \overrightarrow{PP'} = P + \overrightarrow{PQ} + \overrightarrow{PP'} = P' + \overrightarrow{PQ}$$

está en la recta $P'Q'$, luego se trata del punto de corte entre ambas, Q' , como queríamos probar.

2) Hemos de probar que $Q' = O + k\overrightarrow{OQ}$, o sea, $\overrightarrow{OQ'}/\overrightarrow{OQ} = \overrightarrow{OP'}/\overrightarrow{OP}$, pero esto es consecuencia inmediata del teorema de Tales (aplicado al plano que contiene las rectas consideradas y teniendo en cuenta además la recta paralela a PQ que pasa por O). ■

Esto nos lleva a la caracterización geométrica de las homotecias que habíamos anunciado:

Teorema 5.8 *Sea E un espacio afín de dimensión mayor o igual que 2. El conjunto de las biyecciones de E en sí mismo que transforman cada recta en una recta paralela es el subgrupo de $\text{GA}(E)$ formado por las homotecias y las traslaciones de E . Lo representaremos por $\text{HT}(E)$.*

DEMOSTRACIÓN: Las biyecciones indicadas forman obviamente un grupo. Es fácil ver que contiene a las homotecias y a las traslaciones. Por ejemplo, si $f = H(O, k)$ y $r = P + \langle \vec{v} \rangle$ es una recta, la imagen de un punto $Q = P + \lambda \vec{v}$ es $O + k\overrightarrow{OQ}$, pero $k\overrightarrow{OQ} = k\overrightarrow{OP} + k\overrightarrow{PQ} = k\overrightarrow{OP} + \lambda k, \vec{v}$, luego concluimos que $f[r] = (O + k\overrightarrow{OP}) + \langle \vec{v} \rangle$, que es una recta paralela a r .

Veamos que cualquier elemento f del grupo es una homotecia o una traslación. Distinguimos varios casos.

Si f tiene dos puntos fijos, Q y Q' , entonces, dado cualquier punto P que no esté en la recta QQ' , tenemos que $f[PQ]$ es una recta paralela a PQ que contiene a P , luego es PQ . Igualmente $f[PQ'] = PQ'$. Como P es el único punto en común entre PQ y PQ' , necesariamente $f(P) = P$. Así pues, f fija a todos los puntos fuera de la recta QQ' . Si R está en la recta QQ' y P es un punto exterior a esta recta, entonces todos los puntos de PR salvo quizá R son fijados, luego R también. Por consiguiente f es la identidad.

Si f tiene un único punto fijo O , tomemos un punto cualquiera $P \neq O$. Sea $P' = f(P)$, que es un punto distinto de O y P y colineal con ellos. Si Q es cualquier punto exterior a la recta OP , entonces $Q' = f(Q)$ cumple que las rectas QQ' y PP' se cortan en O y las rectas PQ y $P'Q'$ son paralelas. Por el teorema anterior $f(Q) = H(O, k)(Q)$, donde $k = \overrightarrow{OP'}/\overrightarrow{OP}$. Esto vale en principio para todo Q fuera de la recta OP , pero razonando con otro punto P fuera de esta recta, también vale para ella.

Si f no tiene puntos fijos, consideremos un punto cualquiera P y su imagen $P' = f(P)$. Para cualquier punto Q exterior a la recta PP' el punto $Q' = f(Q)$ cumple que PQ es paralela a $P'Q'$. Además PP' es paralela a QQ' , pues si ambas rectas tuvieran un punto en común, éste sería un punto fijo de f . Por

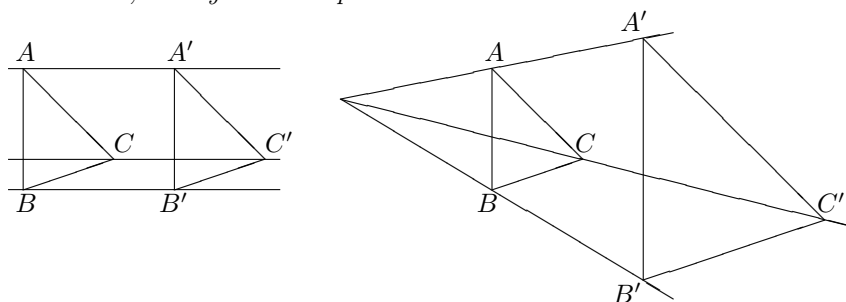
el teorema anterior $f(Q) = T_{\vec{v}}(Q)$, donde $\vec{v} = \overrightarrow{PP'}$. En principio esto vale para todo punto Q fuera de la recta PP' , pero partiendo de otra recta, también vale para ella. ■

Ejercicio: Probar que toda traslación puede obtenerse como composición de dos homotecias.

Ejercicio: Probar que los grupos $H(O, E)$ y $H(O', E)$ son conjugados en $\text{HT}(E)$ mediante la traslación de vector $\overrightarrow{OO'}$.

Veamos ahora una consecuencia muy importante del teorema de Tales (comárese con [IGE 4.59] y las notas posteriores):

Teorema 5.9 (Teorema de Desargues) *En un espacio afín, sean \widehat{ABC} y $\widehat{A'B'C'}$ dos triángulos con $A \neq A'$, $B \neq B'$, $C \neq C'$ y tales que las rectas AB , AC , BC sean paralelas respectivamente a $A'B'$, $A'C'$ y $B'C'$. Entonces las rectas AA' , BB' y CC' son paralelas o concurrentes.*



DEMOSTRACIÓN: Notemos que no se exige que los triángulos estén contenidos en el mismo plano. Basta probar que si dos de las tres rectas AA' , BB' , CC' se cortan en un punto, la tercera pasa por dicho punto (si dos de las rectas coinciden el teorema se cumple trivialmente, las tres rectas no pueden coincidir). Supongamos que AA' y CC' concurren en O . Si $A = O$ las rectas AC y $A'C'$ no serían paralelas. En general, O ha de ser distinto de los seis vértices. Por el teorema de Tales $\overrightarrow{OA}/\overrightarrow{OA'} = \overrightarrow{OC}/\overrightarrow{OC'} = k$. Como $A \neq A'$ ha de ser $k \neq 1$. Sea $f = H(O, k)$ y $B'' = f(B)$. Por el teorema 5.7 tenemos que B'' está en la paralela a AB que pasa por A' , es decir, en $A'B'$, y por el mismo motivo está en $B'C'$, luego ha de ser $B'' = B'$. Esto prueba que O , B y B' están alineados (de hecho los triángulos son homotéticos). ■

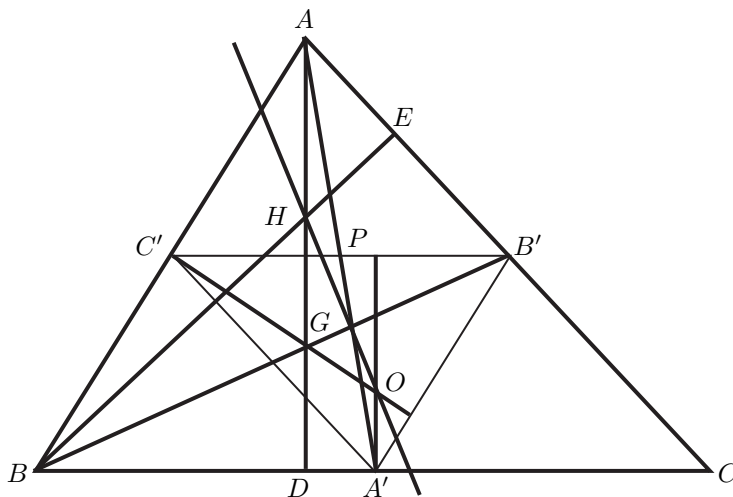
Ejercicio: Refinar el teorema anterior (analizando los casos que no ha hecho falta analizar) para concluir que bajo las hipótesis del teorema de Desargues los triángulos son trasladados u homotéticos.

El teorema de Desargues es equivalente a su recíproco:

Teorema 5.10 (Teorema de Desargues) *En un espacio afín, sean \widehat{ABC} y $\widehat{A'B'C'}$ dos triángulos con $A \neq A'$, $B \neq B'$, $C \neq C'$ y tales que las rectas AA' , BB' y CC' sean paralelas o concurrentes en un punto O distinto de todos los vértices. Si los triángulos tienen dos pares de lados paralelos, también el tercer par lo es.*

DEMOSTRACIÓN: Supongamos que AB es paralela a $A'B'$ y que AC es paralela a $A'C'$. No puede ser que ambos pares de rectas sean coincidentes (o sería $A = A'$) supongamos que las rectas AB y $A'B'$ son distintas y sea B'' el punto de corte entre $A'B'$ y la recta paralela a BC por C' . Entonces $B \neq B''$ y podemos aplicar el teorema anterior para concluir que las rectas AA' , BB'' , CC' son paralelas o concurrentes, lo que conjuntamente con la hipótesis implica que $BB'' = BB'$, luego $B' = B''$, pues ambos son el punto de corte de BB'' con $A'B'$. ■

Aplicaciones Concluimos esta sección con algunas aplicaciones de los resultados precedentes. En primer lugar probaremos una interesante propiedad de los triángulos descubierta por Euler [IGE 4.8]: El circuncentro, el ortocentro y el baricentro son siempre colineales. Llegaremos a ello estudiando el llamado *triángulo medial*, que es el triángulo que une los pies de las medianas.



En primer lugar observamos que el teorema de Tales implica que los lados del triángulo medial $\widehat{A'B'C'}$ son paralelos a los de \widehat{ABC} , así como que los del primero miden la mitad que los del segundo. El teorema de Desargues implica que ambos son homotéticos, y la razón de la homotecia es -2 . El teorema de Tales prueba también que la mediana AA' pasa por el punto medio P de $B'C'$, luego PA' es una mediana del triángulo medial. En general, las medianas del triángulo medial están contenidas en las del triángulo de partida, luego ambos triángulos tienen el mismo baricentro G .

Es fácil ver que una homotecia conserva el punto medio de un segmento, luego el punto medio P del lado $B'C'$ es el homólogo del punto medio A' del lado BC . Por lo tanto, la mediana PA' del triángulo medial se transforma en la mediana $A'A$ del triángulo dado, es decir, la homotecia deja fijas las medianas (comunes) de ambos triángulos. En particular fija al baricentro G , luego éste ha de ser el centro de la homotecia.

Por otra parte, es obvio que las alturas del triángulo medial son las bisectrices de \widehat{ABC} , luego el circuncentro O de \widehat{ABC} es el ortocentro de su triángulo medial.

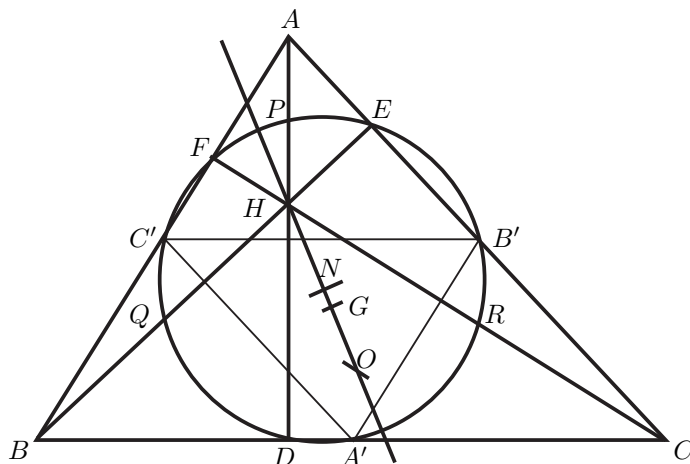
La recta homotética de la altura PA' del triángulo medial es su paralela por el punto A , es decir, la altura AD del triángulo dado. En general, cada altura se corresponde con una altura y por lo tanto el circuncentro O (ortocentro de $\widehat{A'B'C'}$) es el homólogo del ortocentro H de \widehat{ABC} .

Esto prueba que O , H y G se encuentran sobre la misma recta. Más aún, G está entre O y H y $HG = 2GO$.

Hay un caso especial, en el que $G = O$ (y necesariamente entonces $G = H$). Esto equivale a que las mediatrices coincidan con las alturas, lo que a su vez equivale a que el triángulo sea equilátero.

Definición 5.11 Se llama *recta de Euler* de un triángulo no equilátero a la recta que contiene al su circuncentro, su baricentro y su ortocentro.

Todavía podemos llevar más lejos nuestro análisis. Sea N el punto medio del segmento \overline{OH} . Es claro que su homólogo a través de la homotecia anterior es el circuncentro O de \widehat{ABC} , luego N ha de ser el circuncentro del triángulo medial.



Consideramos ahora la homotecia de centro H y razón $1/2$. Ésta transforma el circuncentro O de \widehat{ABC} en el punto N , luego transforma la circunferencia circunscrita de \widehat{ABC} en una circunferencia ω de centro N que contiene a los puntos medios de los segmentos \overline{HA} , \overline{HB} y \overline{HC} , llamémoslos P , Q , R . El triángulo $\widehat{A'B'C'}$ se transforma en \widehat{PQR} mediante la homotecia de centro G y razón -2 seguida de la homotecia de centro H y razón $1/2$ (pasando por \widehat{ABC}), pero estas transforman a N en O y O en N de nuevo, luego la composición fija a N , pero ha de ser una homotecia o una traslación y, como fija a N , ha de ser una homotecia de centro N y razón ± 1 . Es fácil ver que P no puede coincidir con A' , luego la homotecia no puede ser la identidad, luego en definitiva se trata de la homotecia de centro N y razón -1 .

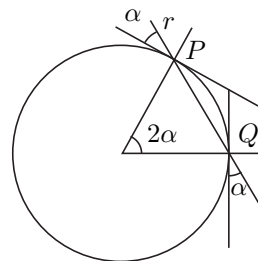
Esto implica que el círculo ω contiene también a los pies de las medianas A' , B' y C' . Más aún, los segmentos $\overline{PA'}$, $\overline{QB'}$, $\overline{RC'}$ son diámetros de ω .

Finalmente, si el pie de la altura D es distinto de A' el triángulo $\widehat{PDA'}$ es rectángulo, luego su circuncentro es N , luego D está también en ω , y lo mismo es válido para E y F . Con esto hemos probado [IGE 4.9]:

Teorema 5.12 *En todo triángulo $T = \widehat{ABC}$ con ortocentro H , los puntos medios de los segmentos \overline{HA} , \overline{HB} y \overline{HC} , los pies de las alturas y los pies de las medianas están sobre una misma circunferencia ω cuyo centro es el punto medio entre el ortocentro y el circuncentro de T .*

Poncelet llamó a la circunferencia del teorema anterior el *circunferencia de los nueve puntos* del triángulo T . Observemos que es la circunferencia circunscrita tanto a su triángulo medial como a su triángulo órtico (el que tiene por vértices los pies de las alturas). Veamos otra propiedad interesante. Para enunciarla necesitamos definir el ángulo entre una circunferencia y una recta secante:

Definición 5.13 Dada una circunferencia ω y una recta r que la corte en dos puntos P y Q , llamaremos *ángulo* entre r y ω en P al ángulo (menor) determinado por r y la tangente a ω por P . Si r es tangente a ω en P convendremos en que el ángulo entre r y ω es nulo.



Es fácil ver que r determina el mismo ángulo α tanto en P como en Q , así como que éste es la mitad del arco menor determinado por P y Q en ω , y por lo tanto igual a cualquier ángulo inscrito en ω que abarque el arco \widehat{PQ} .

Teorema 5.14 *Los ángulos con que un lado BC de un triángulo corta a su circunferencia de los nueve puntos tienen amplitud $|\hat{B} - \hat{C}|$.*

DEMOSTRACIÓN: Con la notación anterior, el lado corta a la circunferencia en los puntos A' y D . Es fácil ver que $D = A'$ si y sólo si $B = C$ (en cuyo caso el lado es tangente a la circunferencia), por lo que podemos suponer que $D \neq A'$. Hemos de calcular cualquier ángulo inscrito en la circunferencia que abarque a $\widehat{DA'}$, por ejemplo

$$\widehat{DPA'} = \widehat{DPN} = \widehat{DAO},$$

pues las rectas PN y AO son paralelas (son homotéticas desde H).

Claramente $\widehat{DAC} = \pi/2 - \hat{C}$, mientras que

$$\widehat{OAC} = \widehat{OCA} = \frac{\pi}{2} - \frac{\widehat{AOC}}{2} = \frac{\pi}{2} - \hat{B},$$

pues \hat{B} está inscrito en la circunferencia circunscrita a \widehat{ABC} .

Esto implica que $\widehat{DAO} = |\hat{B} - \hat{C}|$. ■

5.3 El teorema fundamental de la geometría afín

Vamos a dar una caracterización puramente geométrica (no algebraica) de las biyecciones afines entre dos espacios afines sobre un mismo cuerpo euclídeo arquimediano.

Definición 5.15 Si E y E' son dos espacios afines de la misma dimensión $n \geq 2$ sobre un mismo cuerpo K , una *colineación* entre ellos es una aplicación biyectiva $f : E \rightarrow E'$ que transforma puntos colineales en puntos colineales.

Vamos a probar que, si el cuerpo K es euclídeo y arquimediano, las colineaciones coinciden con las biyecciones afines. No obstante, vamos a probar algunos resultados generales sobre colineaciones $f : E \rightarrow E'$ que vale para cuerpos arbitrarios.

1. Si $A = \{P_0, \dots, P_m\}$ es un conjunto de puntos de E afínmente independientes, entonces $f[\langle A \rangle] \subset \langle f[A] \rangle$.

Lo probamos por inducción sobre m . Para $m = 1$ se cumple por la definición de colineación. Supongámoslo cierto para conjuntos con m puntos y tomemos $P \in \langle A \rangle$. Usando coordenadas baricéntricas tenemos

$$P = \lambda_0 P_0 + \dots + \lambda_m P_m, \quad \text{con } \lambda_0 + \dots + \lambda_m = 1.$$

Puesto que $m \geq 1$ hay al menos un $\lambda_i \neq 1$. Supongamos por ejemplo que se trata de λ_m . Entonces

$$P = (1 - \lambda_m)P' + \lambda_m P_m,$$

donde

$$P' = \frac{1}{1 - \lambda_m} (\lambda_0 P_0 + \dots + \lambda_{m-1} P_{m-1}) \in \langle P_0, \dots, P_{m-1} \rangle.$$

Entonces $f(P)$ está en la recta $f(P')f(P_m)$ y $f(P')$ está en la variedad $\langle f(P_0), \dots, f(P_{m-1}) \rangle$ por hipótesis de inducción. Así pues, $f(P) \in \langle f[A] \rangle$.

2. Las imágenes por f de $m + 1$ puntos afínmente independientes son afínmente independientes

Podemos completar el conjunto hasta un sistema de referencia afín y suponer, por lo tanto, que $m = n$. Si A es un conjunto de $n + 1$ puntos afínmente independientes, por el apartado anterior $E' = f[\langle A \rangle] \subset \langle f[A] \rangle$, luego los puntos de $f[A]$ han de ser independientes.

3. f transforma cada recta PQ en la recta $f(P)f(Q)$.

La hipótesis del teorema nos da una inclusión. Sea R' un punto de $f(P)f(Q)$. Como f es biyectiva $R' = f(R)$ para un cierto R . Si R no estuviera en PQ entonces los puntos P, Q, R serían afínmente independientes, luego por el punto anterior lo mismo ocurriría con $f(P), f(Q), f(R)$, lo cual es una contradicción.

4. f transforma rectas paralelas en rectas paralelas.

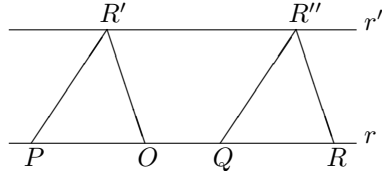
Sean PQ y $P'Q'$ rectas paralelas distintas en E . Entonces los puntos P, Q, P' son afinmente independientes y generan el plano que contiene a las dos rectas. Por el paso 1 tenemos que $f(P), f(Q), f(P')$ son afinmente independientes, luego generan un plano que contiene a $f(Q')$ por 2, luego de hecho contiene a las rectas $f[PQ]$ y $f[P'Q']$. Como f es biyectiva estas rectas no tienen puntos comunes, luego son paralelas.

5. Para todos los puntos $O, P, Q \in E$, se cumple

$$f(O + \overrightarrow{PQ}) = f(O) + \overrightarrow{f(P)f(Q)}.$$

Sea $R = O + \overrightarrow{PQ}$. Si O, P, Q no están alineados, lo mismo sucede con $f(O), f(P), f(Q)$. Por el teorema 5.7 tenemos que R es la intersección de la paralela a OP por Q y la paralela a OQ por P . Por consiguiente $f(R)$ es la intersección de la paralela a $f(O)f(P)$ por $f(Q)$ y la paralela a $f(O)f(Q)$ por $f(P)$. De nuevo por el teorema 5.7 podemos concluir que $f(R) = f(O) + \overrightarrow{f(O)f(Q)}$.

Si O, P, Q están contenidos en una recta r , tomamos una paralela cualquiera r' y en ella un punto R' . La paralela a PR' por Q corta a r' en un punto $R'' = R' + \overrightarrow{PQ}$ y la paralela a OR' por R'' corta a r en $O + \overrightarrow{PQ} = R$. Al aplicar f a todos estos puntos y rectas se conservan las relaciones de incidencia y paralelismo, luego tenemos que $f(R'') = f(R') + \overrightarrow{f(P)f(Q)}$ y $f(R) = f(O) + \overrightarrow{f(P)f(Q)}$.

6. La aplicación $\vec{f}: \vec{E} \rightarrow \vec{E}'$ dada por $\vec{f}(\overrightarrow{PQ}) = \overrightarrow{f(P)f(Q)}$ está bien definida, de modo que $f(O + \vec{v}) = f(O) + \vec{f}(\vec{v})$.

Se trata de probar que si $\overrightarrow{PQ} = \overrightarrow{RS}$, entonces $\overrightarrow{f(P)f(Q)} = \overrightarrow{f(R)f(S)}$. En efecto, tenemos que $S = R + \overrightarrow{RS} = R + \overrightarrow{PQ}$, luego $f(S) = f(R) + \overrightarrow{f(P)f(Q)}$, lo que nos da la igualdad requerida.

7. Si $\vec{v}, \vec{w} \in \vec{E}$, se cumple $\vec{f}(\vec{v} + \vec{w}) = \vec{f}(\vec{v}) + \vec{f}(\vec{w})$.

Fijado cualquier punto $O \in E$, tenemos que

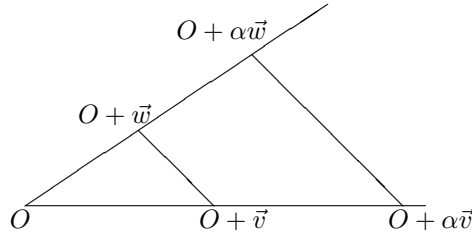
$$f(O) + \vec{f}(\vec{v} + \vec{w}) = f(O + \vec{v} + \vec{w}) = f(O + \vec{v}) + \vec{f}(\vec{w}) = f(O) + \vec{f}(\vec{v}) + \vec{f}(\vec{w}),$$

de donde se tiene la igualdad requerida.

El hecho de que f sea biyectiva se traduce en que \vec{f} también lo es, luego es un isomorfismo entre los grupos aditivos de \vec{E} y \vec{E}' . En general, no es una aplicación lineal. Ahora bien, que f transforme rectas paralelas en rectas paralelas se traduce en que si $\vec{v} \in \vec{E}$ es no nulo y $\alpha \in K$, entonces $\vec{f}(\alpha\vec{v}) = \sigma(\alpha, \vec{v})\vec{f}(\vec{v})$, para cierto $\sigma(\alpha, \vec{v}) \in K$.

8. En las condiciones anteriores, $\sigma(\alpha, \vec{v})$ no depende de \vec{v} .

Sean \vec{v} y \vec{w} dos vectores linealmente independientes y fijemos un punto $O \in E$ cualquiera. Entonces tenemos la situación que muestra la figura siguiente:



Como f conserva el paralelismo, tenemos una figura análoga con

$$f(O) + \sigma(\alpha, \vec{v})\vec{f}(\vec{v}), \quad f(O) + \sigma(\alpha, \vec{w})\vec{f}(\vec{w})$$

y por el teorema de Tales,

$$\sigma(\alpha, \vec{v}) = \frac{\vec{f}(\alpha\vec{v})}{\vec{f}(\vec{v})} = \frac{\vec{f}(\alpha\vec{w})}{\vec{f}(\vec{w})} = \sigma(\alpha, \vec{w}).$$

Si \vec{v} y \vec{w} son linealmente dependientes, tomamos un tercer vector \vec{u} independiente de ambos y aplicamos el caso ya probado para concluir que $\sigma(\alpha, \vec{v}) = \sigma(\alpha, \vec{u}) = \sigma(\alpha, \vec{w})$.

Por lo tanto, a partir de aquí escribiremos $\sigma(\alpha)$ en lugar de $\sigma(\alpha, \vec{v})$, con lo que tenemos una aplicación $\sigma : K \rightarrow K$ tal que, para todo $\vec{v} \in \vec{E}$ y todo $\alpha \in K$, se cumple que $\vec{f}(\alpha\vec{v}) = \sigma(\alpha)\vec{f}(\vec{v})$.

9. La aplicación $\sigma : K \rightarrow K$ que acabamos de definir es un automorfismo de cuerpos.

Si $\vec{v} \in \vec{E}$ es cualquier vector no nulo, tenemos que

$$\begin{aligned} \sigma(\alpha + \beta)\vec{f}(\vec{v}) &= \vec{f}((\alpha + \beta)\vec{v}) = \vec{f}(\alpha\vec{v} + \beta\vec{v}) = \vec{f}(\alpha\vec{v}) + \vec{f}(\beta\vec{v}) = \\ &= \sigma(\alpha)\vec{f}(\vec{v}) + \sigma(\beta)\vec{f}(\vec{v}) = (\sigma(\alpha) + \sigma(\beta))\vec{f}(\vec{v}), \end{aligned}$$

de donde se sigue que $\sigma(\alpha + \beta) = \sigma(\alpha) + \sigma(\beta)$ y análogamente se concluye que $\sigma(\alpha\beta) = \sigma(\alpha)\sigma(\beta)$. La biyectividad de f implica inmediatamente la de σ , luego es un automorfismo de cuerpos.

Y esto es todo lo que podemos demostrar en general sobre la forma algebraica de las colineaciones. Para enunciar lo que hemos obtenido necesitamos una definición:

Definición 5.16 Si V y W son espacios vectoriales sobre un mismo cuerpo K , una aplicación $f : V \rightarrow W$ es *semilineal* si existe un automorfismo $\sigma : K \rightarrow K$ tal que, para todos los vectores $v_1, v_2 \in V$ y todo $\alpha \in K$, se cumple

$$f(v_1 + v_2) = f(v_1) + f(v_2), \quad f(\alpha v_1) = \sigma(\alpha)f(v_1).$$

Observemos que la composición de aplicaciones semilineales es semilineal, tomando como automorfismo la composición de los automorfismos correspondientes.

Si E y E' son dos espacios afines sobre un mismo cuerpo K , una aplicación $f : E \rightarrow E'$ es *semiafín* si para todo punto P de E se cumple

$$f(P) = f(O) + \vec{f}(\overrightarrow{OP}),$$

donde O es un punto prefijado en E , y $\vec{f} : \vec{E} \rightarrow \vec{E}'$ es una aplicación semilineal.

El mismo argumento elemental empleado tras la definición 3.16 prueba que esta definición no depende de la elección del punto O , de modo que si f es semiafín, la aplicación \vec{f} está determinada por f . La composición de aplicaciones semiafines es claramente semiafín.

En estos términos, hemos demostrado el teorema siguiente:

Teorema 5.17 (Teorema fundamental de la geometría afín) *Si E y E' son dos espacios afines de la misma dimensión $n \geq 2$ sobre un cuerpo K , las colineaciones $f : E \rightarrow E'$ coinciden con las biyecciones semiafines.*

En realidad falta demostrar que toda biyección semiafín es una colineación, pero eso es inmediato. En el caso de los espacios afines reales, la situación es más simple en virtud del teorema siguiente:

Teorema 5.18 *Si $K \subset \mathbb{R}$ es un cuerpo euclídeo, su único automorfismo es la identidad.*

DEMOSTRACIÓN: Sea $\sigma : K \rightarrow K$ un automorfismo. Como K es euclídeo, todo $\alpha \in K$ que cumpla $\alpha \geq 0$ tiene raíz cuadrada, de modo que $\alpha = \beta^2$, luego $\sigma(\alpha) = \sigma(\beta)^2 \geq 0$. Así pues, σ transforma elementos positivos en elementos positivos. Más aún, se cumple $\alpha \leq \beta$ si y sólo si $\beta - \alpha \geq 0$, si y sólo si $\sigma(\beta) - \sigma(\alpha) \geq 0$, si y sólo si $\sigma(\alpha) \leq \sigma(\beta)$. Por lo tanto, σ conserva el orden.

Por otra parte, como $\sigma(1) = 1$, es claro que σ fija a todos los números naturales, luego a todos los números enteros, luego a todos los números racionales.

Finalmente, si existiera un $\alpha \in K$ tal que $\alpha \neq \sigma(\alpha)$, supongamos por ejemplo que $\alpha < \sigma(\alpha)$. Entonces existe $q \in \mathbb{Q}$ tal que $\alpha < q < \sigma(\alpha)$, pero esto es imposible, porque entonces $\sigma(\alpha) < \sigma(q) = q$, contradicción. ■

Por lo tanto, el teorema fundamental de la geometría afín para espacios afines sobre cuerpos euclídeos arquimedianos se reduce a que las colineaciones entre tales espacios coinciden con las biyecciones afines.

5.4 Isometrías y semejanzas

Nos ocupamos ahora de los espacios euclídeos (siempre sobre cuerpos euclídeos). Recordemos que una isometría de un espacio afín euclídeo E es una afinidad biyectiva cuya aplicación lineal asociada es una isometría lineal. A su vez, las isometrías lineales son los automorfismos \vec{f} de \vec{E} tales que

$$\vec{f}(\vec{v})\vec{f}(\vec{w}) = \vec{v}\vec{w}, \quad \text{para todo } \vec{v}, \vec{w} \in \vec{E},$$

lo cual a su vez equivale a que $\|\vec{f}(\vec{v})\| = \|\vec{v}\|$ para todo $\vec{v} \in \vec{E}$, debido a la relación

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 - 2\vec{v}\vec{w}.$$

Otra caracterización importante es que una aplicación lineal $\vec{f}: \vec{E} \rightarrow \vec{E}$ es una isometría si y sólo si transforma una base ortonormal en una base ortonormal, pues si conserva el producto escalar sobre una base, por linealidad lo conserva sobre vectores arbitrarios.

Definición 5.19 Si E es un espacio euclídeo, en el grupo afín $\text{GA}(E)$ podemos considerar el subgrupo $\text{Is}(E)$ de las isometrías de E , y en el grupo $\text{LG}(\vec{E})$ podemos considerar el subgrupo $\text{O}(\vec{E})$ de las isometrías lineales, de modo que el epimorfismo natural $\text{GA}(E) \rightarrow \text{LG}(\vec{E})$ se restringe a un epimorfismo $\text{Is}(E) \rightarrow \text{O}(\vec{E})$ cuyo núcleo sigue siendo el grupo $\text{T}(E) \cong \vec{E}$ de las traslaciones en E , pues éste está contenido en $\text{Is}(E)$.

Diremos que una matriz cuadrada A en un cuerpo K es *ortogonal* si cumple $AA^t = I_n$, donde A^t es la matriz traspuesta de A y I_n es la matriz identidad. Así, A es ortogonal si su inversa es su traspuesta. El *grupo ortogonal* de K es el grupo $\text{O}(n, K)$ de las matrices ortogonales $n \times n$ con coeficientes en K .

Notemos que, al tomar determinantes, resulta que $|A|^2 = 1$, luego una matriz ortogonal cumple necesariamente $|A| = \pm 1$.

El teorema 3.25 afirma que si \vec{E} es un espacio vectorial euclídeo y $f: \vec{E} \rightarrow \vec{E}$ una aplicación lineal, entonces f es una isometría lineal si y sólo si su matriz respecto de una base ortonormal es ortogonal (y en tal caso la matriz de f en cualquier base ortonormal es ortogonal).

Esto significa que, si \vec{E} tiene dimensión n , el isomorfismo $\text{LG}(\vec{E}) \cong \text{LG}(n, K)$ entre el grupo lineal de \vec{E} y el grupo de las matrices inversibles $n \times n$ que asigna a cada automorfismo su matriz en una base prefijada se restringe a un isomorfismo $\text{O}(\vec{E}) \cong \text{O}(n, K)$ (si tomamos la base de modo que sea ortonormal).

Exactamente igual que en la sección 5.1, podemos identificar $\text{O}(\vec{E})$ con un subgrupo de $\text{Is}(E)$ fijando un punto O y considerando el grupo $\text{Is}_O(E)$ de todas las isometrías que fijan a O .

Una biyección $f: A \rightarrow B$ entre dos subconjuntos de E es una *semejanza* de razón $k > 0$ si para todo par de puntos $P, Q \in A$ se cumple

$$\|\overrightarrow{f(P)f(Q)}\| = k\|\overrightarrow{PQ}\|.$$

Si existe una semejanza entre A y B se dice que son *semejantes*, y cada par de puntos P , $f(P)$ correspondientes por una semejanza f se llaman puntos *homólogos* (respecto a f).

Es claro que dos triángulos son semejantes en este sentido si y sólo si lo son en el sentido que ya teníamos definido, es decir, si sus lados son proporcionales.

También es obvio que la inversa de una semejanza de razón k es una semejanza de razón $1/k$ y que la composición de semejanzas es una semejanza de razón igual al producto de las razones. Las semejanzas de razón 1 son las isometrías.

Una homotecia de razón k es una semejanza de razón $|k|$. Por consiguiente, si f es una semejanza de razón k entre dos subconjuntos A y B de un espacio afín euclídeo E , al componerla con una homotecia g de razón $1/k$ obtenemos una semejanza de razón 1, es decir, una isometría. Por 3.29 sabemos que $f \circ g$ se extiende a una isometría de E en sí mismo, llamémosla h . Entonces $h \circ g^{-1}$ es una semejanza de E en sí mismo que extiende a f . Esto prueba el teorema siguiente:

Teorema 5.20 *Toda semejanza de razón k entre dos subconjuntos de un espacio afín euclídeo E se extiende a una semejanza de razón k de E en sí mismo. Toda semejanza de razón k de E en sí mismo es una afinidad y se puede expresar como composición de una isometría y una homotecia de razón k .*

Definición 5.21 Llamaremos $\text{Sem}(E)$ al grupo de las semejanzas de un espacio afín euclídeo E en sí mismo.

Hemos probado que $\text{Sem}(E) \leq \text{GA}(E)$. La aplicación que a cada semejanza le asigna su razón es un homomorfismo de $\text{Sem}(E)$ en $]0, +\infty[$ cuyo núcleo es claramente $\text{Is}(E)$.

Vamos a caracterizar las semejanzas en términos de sus aplicaciones lineales asociadas. Si V es un espacio vectorial euclídeo, diremos que $f \in \text{LG}(V)$ es una *semejanza lineal* de razón $k > 0$ si $\|f(\vec{v})\| = k \|\vec{v}\|$ para todo $\vec{v} \in V$. Llamaremos $\text{Sem}(V)$ al grupo de las semejanzas lineales de V .

Es claro que una afinidad f es una semejanza si y sólo si \vec{f} es una semejanza lineal. Tenemos, pues, un epimorfismo $\text{Sem}(E) \rightarrow \text{Sem}(\vec{E})$ cuyo núcleo es el grupo $T(E)$ de las traslaciones de E . Por otra parte, el hecho de que toda semejanza de E se descomponga en producto de una isometría y una homotecia se traduce, a través de este epimorfismo, en que toda semejanza lineal es producto de una isometría lineal y una homotecia lineal. Más aún, es fácil ver que si llamamos $\text{H}_+(\vec{E})$ al grupo de las homotecias lineales de \vec{E} de razón positiva, se cumple

$$\text{Sem}(\vec{E}) = \text{O}(\vec{E}) \times \text{H}_+(\vec{E}).$$

Para ello hay que probar que los dos subgrupos de la derecha son normales, pero ello se sigue inmediatamente del hecho de que las homotecias lineales conmutan con todos los elementos de $\text{LG}(\vec{E})$.

Ejercicio: Probar que $\text{HT}(E)$ es un subgrupo normal de $\text{Sem}(E)$ y

$$\text{Sem}(E)/\text{HT}(E) \cong \text{Sem}(\vec{E})/H(\vec{E}).$$

Una propiedad interesante de las semejanzas propiamente dichas, es decir, las que no son isometrías, es que siempre tienen un único punto fijo.

Teorema 5.22 *Toda semejanza de razón $k \neq 1$ en un espacio afín euclídeo tiene un único punto fijo llamado centro de la semejanza.*

DEMOSTRACIÓN: Es consecuencia inmediata del teorema 5.2, pues si f cumple las hipótesis entonces \vec{f} es una semejanza lineal de razón $k \neq 1$, y es claro que su único punto fijo es 0. ■

En general, las semejanzas lineales no conservan las normas, luego tampoco el producto escalar. Sin embargo, el hecho de que toda semejanza lineal f de razón k se exprese como composición de una isometría lineal por una homotecia lineal de razón k implica la relación $f(\vec{v})f(\vec{w}) = k^2\vec{v}\vec{w}$. Esto implica que las semejanzas lineales conservan ángulos, pues

$$\cos \widehat{f(\vec{v})f(\vec{w})} = \frac{f(\vec{v})f(\vec{w})}{\|f(\vec{v})\| \|f(\vec{w})\|} = \frac{k^2\vec{v}\vec{w}}{k^2\|\vec{v}\| \|\vec{w}\|} = \cos \widehat{\vec{v}\vec{w}}.$$

En particular las semejanzas lineales conservan la ortogonalidad. Vamos a probar el recíproco, es decir, que las únicas aplicaciones lineales que conservan la ortogonalidad (o a fortiori los ángulos) son las semejanzas lineales. De aquí deduciremos el análogo para las semejanzas afines.

Teorema 5.23 *Sea V un espacio vectorial euclídeo y $f : V \rightarrow V$ una aplicación lineal no nula. Entonces son equivalentes:*

1. f es una semejanza lineal.
2. f conserva la ortogonalidad de vectores.
3. f conserva los ángulos entre vectores.

DEMOSTRACIÓN: Es suficiente probar que 2 implica 1. Suponemos, pues, que si $\vec{v}\vec{w} = 0$ entonces $f(\vec{v})f(\vec{w}) = \vec{0}$ (nótese que no suponemos la doble implicación).

Fijado un vector $\vec{v} \in V$, la aplicación $u(\vec{w}) = f(\vec{v})f(\vec{w})$ es lineal. Supongamos que existe un \vec{w} tal que $u(\vec{w}) \neq 0$. Entonces el núcleo de u tiene dimensión $n - 1$ (donde n es la dimensión de V), y por hipótesis contiene a $\langle \vec{v} \rangle^\perp$, cuya dimensión es la misma. Por consiguiente u se anula exactamente sobre los vectores ortogonales a \vec{v} .

Definimos $k(\vec{v})$ como el único escalar que cumple

$$u(\vec{w}) = f(\vec{v})f(\vec{w}) = k(\vec{v})\vec{v}\vec{w}.$$

Todo vector de V se expresa como $\vec{t} + \alpha\vec{w}$, donde $\vec{t} \perp \vec{w}$, con lo que la relación anterior vale en realidad no sólo para el \vec{w} que hemos tomado sino de hecho para todo $\vec{w} \in V$. Si u es nula la relación sigue siendo válida sin más que tomar $k(\vec{v}) = 0$.

Por simetría, si $\vec{v}\vec{w} \neq 0$ se ha de cumplir $k(\vec{v}) = k(\vec{w})$. Si $\vec{v}\vec{w} = 0$ con ambos factores no nulos, entonces $\vec{y} = \vec{v} + \vec{w}$ cumple que $\vec{v}\vec{y} \neq 0 \neq \vec{w}\vec{y}$, luego $k(\vec{v}) = k(\vec{y}) = k(\vec{w})$. Esto prueba que $k = k(\vec{v})$ es independiente de $\vec{v} \neq \vec{0}$. Además

$$\|f(\vec{v})\|^2 = k\|\vec{v}\|^2, \quad \text{para todo } \vec{v} \in V.$$

Como existen vectores con imagen no nula, ha de ser $k > 0$, de donde f es una semejanza de razón \sqrt{k} . ■

Teorema 5.24 *Sea E un espacio afín euclídeo y $f : E \rightarrow E$ una aplicación biyectiva. Las afirmaciones siguientes son equivalentes:*

1. f es una semejanza.
2. Para toda terna de puntos distintos se cumple $\widehat{ABC} = f(A)\widehat{f(B)f(C)}$.
3. Para todos los puntos $A \neq B, C \neq D$, si $AB \perp CD$, entonces

$$f(A)f(B) \perp f(C)f(D).$$

DEMOSTRACIÓN: Usando el teorema anterior es fácil ver que $1 \Rightarrow 2 \Rightarrow 3$. Para probar $3 \Rightarrow 1$ basta demostrar que f es afín, y por el teorema fundamental esto equivale a probar que f transforma puntos alineados en puntos alineados.

Sean A, B, C tres puntos alineados. Podemos formar una base ortogonal de \vec{E} de la forma $\overrightarrow{AA_i}$, para $i = 1, \dots, n$ con $A_1 = B$. Por hipótesis los vectores $\overrightarrow{f(A)f(A_i)}$ son ortogonales dos a dos, luego son una base ortogonal de \vec{E} . Podemos expresar

$$\overrightarrow{f(A)f(C)} = \lambda_1 \overrightarrow{f(A)f(A_1)} + \dots + \lambda_n \overrightarrow{f(A)f(A_n)}. \quad (5.1)$$

Ahora bien, si $i \neq 1$ entonces $AA_i \perp AC$, luego también $f(A)f(A_i) \perp f(A)f(C)$. Multiplicando (5.1) por $\overrightarrow{f(A)f(A_i)}$ obtenemos $\lambda_i = 0$, luego (5.1) se reduce a $\overrightarrow{f(A)f(C)} = \lambda_1 \overrightarrow{f(A)f(A_1)}$, lo que significa que $f(A), f(A_1) = f(B)$ y $f(C)$ están alineados. ■

Existen tres clases básicas de isometrías: las traslaciones, los giros y las simetrías. Ya conocemos las primeras y parcialmente los giros. Terminamos esta sección analizando con más detalle los giros y las simetrías.

Giros La definición siguiente generaliza las definiciones de giro 3.40 y 4.13 a espacios euclídeos de dimensión arbitraria:

Definición 5.25 Una isometría f de un espacio afín euclídeo de dimensión ≥ 2 es un *giro* de ángulo θ si tiene un punto fijo y la matriz de \vec{f} respecto de una cierta base ortonormal es de la forma

$$\begin{pmatrix} \cos \theta & \operatorname{sen} \theta & & & \\ -\operatorname{sen} \theta & \cos \theta & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix},$$

donde las entradas no especificadas fuera de la diagonal son todas nulas.

Observamos que la matriz es la identidad si y sólo si θ es múltiplo de 2π , en cuyo caso f es la identidad. Exceptuado este caso, si E tiene dimensión n , la aplicación f tiene una variedad de puntos fijos de dimensión $n - 2$.

Si E es un plano, entonces f tiene un único punto fijo O y f es un giro de ángulo θ alrededor de O en el sentido de la definición 3.40. El punto O se llama *centro* del giro. Recordemos que hay dos giros de ángulo θ con un mismo centro O , que sobre un mismo sistema de referencia de origen O se corresponden con los ángulos $\pm\theta$, de modo que al cambiar el signo de un vector de la base el giro de ángulo θ pasa a tener ángulo $-\theta$ y viceversa).

Si la dimensión de E es 3 entonces la variedad de puntos fijos de un giro f distinto de la identidad es una recta r llamada *eje* del giro. Sea $\vec{v}_1, \vec{v}_2, \vec{v}_3$ la base ortonormal en la que \vec{f} tiene una matriz según la definición. Si O es cualquier punto del eje de giro r , entonces $r = O + \langle \vec{v}_3 \rangle$ y el plano ortogonal a r dado por $O + \langle \vec{v}_1, \vec{v}_2 \rangle$ es fijado por f (en el sentido de que $f[\Pi] = \Pi$). En el sistema de referencia $(O; \vec{v}_1, \vec{v}_2)$ la matriz de la aplicación lineal asociada a la restricción es la de un giro de ángulo θ en dos dimensiones. Así pues, f gira los puntos de cada plano perpendicular a r respecto al centro O en el cual r corta a tal plano.

En general, si G es una variedad de dimensión $n - 2$ en E , un giro de ángulo θ respecto a G es una isometría que deja invariante a cada punto de G y que se comporta como un giro (plano) de ángulo θ sobre cada plano perpendicular⁴ a G (con centro en el punto de intersección). Hay exactamente dos giros en estas condiciones.

Simetrías Introducimos ahora el concepto de simetría. La idea es que la imagen de un punto P por una simetría respecto a una variedad L se obtiene trazando la recta perpendicular a L por P y tomando en ésta el punto P' que se encuentra a la misma distancia que P del punto de corte. De este modo, el punto de corte es el punto medio del segmento $\overline{PP'}$. Vamos a formular analíticamente esta descripción de las simetrías.

En primer lugar, dado un espacio afín euclídeo E y una variedad L distinta de todo E , por cada punto P exterior a L pasa una única recta r perpendicular a L .

⁴Aquí entendemos que dos variedades L y L' son perpendiculares si tienen un punto en común, necesariamente único, y $\vec{L} \perp \vec{L}'$. Nótese que en este sentido fuerte dos planos del espacio tridimensional nunca pueden ser perpendiculares.

En efecto, dado $Q \in L$ descomponemos $\overrightarrow{PQ} = \vec{v} + \vec{w}$, con $\vec{v} \in \vec{L}$, $\vec{w} \in \vec{L}^\perp$. Sea $R = P + \vec{w} = Q - \vec{v}$. Claramente $R \in L$ y $r = PR$ es perpendicular a L y pasa por P . Además es única, pues si una recta PR' cumple lo mismo, es decir, $\overrightarrow{PR'} \perp \vec{L}$, entonces $\overrightarrow{RR'}$ sería ortogonal a \overrightarrow{PR} y $\overrightarrow{PR'}$, lo cual es contradictorio salvo si $R = R'$.

Por otra parte, dados dos puntos P y P' , tenemos que $P' = P + \overrightarrow{PP'}$, luego el punto medio entre ambos será $M = P + \frac{1}{2}\overrightarrow{PP'}$, que es el único punto que cumple $\overrightarrow{PM} = \overrightarrow{MP'}$. Esto sirve como definición de punto medio en cualquier espacio afín,⁵ no necesariamente euclídeo. Una definición equivalente pero simétrica es

$$M = \frac{1}{2}P + \frac{1}{2}P'$$

(ver pág 106). Esta definición tiene sentido incluso si $P = P'$, en cuyo caso el punto medio es $M = P = P'$.

Notemos que dados dos puntos P y M siempre existe un único punto P' tal que M sea el punto medio de P y P' . Necesariamente P' ha de ser el baricentro $P' = 2M - P$.

Con todas estas observaciones ya podemos justificar la existencia de simetrías respecto a variedades cualesquiera:

Definición 5.26 En un espacio afín euclídeo E , llamaremos *simetría* respecto a una variedad L de E a la única aplicación $s_L : E \rightarrow E$ que fija a todos los puntos de L y que a cada punto P exterior a L le asigna el único punto $s_L(P)$ tal que la perpendicular a L por P corta a L en el punto medio entre P y $s_L(P)$.

En particular la simetría respecto a E es la identidad. No es necesario distinguir la forma en que una simetría respecto a una variedad L actúa sobre los puntos de L y sobre los puntos exteriores a L . En cualquier caso, el punto $P' = s_L(P)$ está determinado por las relaciones

$$\overrightarrow{PP'} \in \vec{L}^\perp \quad \text{y} \quad \frac{1}{2}P + \frac{1}{2}P' \in L.$$

Las simetrías respecto a puntos, rectas y planos se llaman respectivamente simetrías *puntuales*, *axiales* (lat. 'respecto a un eje') y *especulares* (lat. 'respecto a un espejo'). Las simetrías respecto a hiperplanos se llaman *reflexiones*.

Teorema 5.27 Las simetrías de un espacio afín euclídeo son isometrías. Además una isometría f es una simetría si y sólo si $f \circ f = 1$.

DEMOSTRACIÓN: Sea E un espacio afín euclídeo y L una variedad lineal de E , que podemos suponer distinta de todo el espacio. Tomemos una base ortonormal $\vec{v}_1, \dots, \vec{v}_r$ de \vec{L} y completémosla hasta una base ortonormal $\vec{v}_1, \dots, \vec{v}_n$ de \vec{E} . Tomemos un punto O en L y definamos $s_L(P) = O + \vec{s}_L(\overrightarrow{OP})$, donde \vec{s}_L

⁵En realidad hemos de suponer que el cuerpo tiene característica distinta de 2.

es la isometría lineal que en la base anterior tiene matriz diagonal con diagonal igual a $\underbrace{[1, \dots, 1]}_r, \underbrace{[-1, \dots, -1]}_{n-r}$. (Es una isometría porque la matriz es ortogonal.)

Dado un punto $P = O + \alpha_1 \vec{v}_1 + \dots + \alpha_n \vec{v}_n$, es claro que

$$\overrightarrow{Ps_L(P)} = -2\alpha_{r+1}\vec{v}_{r+1} - \dots - 2\alpha_n\vec{v}_n \in \vec{L}^\perp,$$

$$\frac{1}{2}P + \frac{1}{2}s_L(P) = O + \frac{1}{2}\overrightarrow{OP} + \frac{1}{2}\overrightarrow{Os_L(P)} = O + \alpha_1\vec{v}_1 + \dots + \alpha_r\vec{v}_r \in L.$$

Por consiguiente la isometría s_L es la simetría respecto a L .

Sea ahora f una isometría tal que $f \circ f = 1$. Podemos suponer que no es la identidad. Si $P \in E$ cumple que $f(P) \neq P$, como $f(f(P)) = P$, tenemos que f intercambia P y $f(P)$. Como f es una isometría, es claro que la imagen del punto medio M de P y $f(P)$ tiene que ser el punto medio de $f(P)$ y P , es decir, $f(M) = M$.

Así pues, f tiene puntos fijos. Sea L la variedad formada por todos ellos. Si L es un punto, es claro que f es la simetría respecto de L . En caso contrario, para todo $P \in E \setminus L$, sea $M \in L$ el punto medio de P y $P' = f(P)$ y sea $Q \in L$ cualquier otro punto, entonces los triángulos \widehat{PMQ} y $\widehat{P'MQ}$ tienen sus lados iguales, luego también sus ángulos, de donde se sigue que \widehat{PMQ} es recto, y como esto vale para todo $Q \in L$, la recta $\overline{PP'}$ es perpendicular a L , luego f es la simetría s_L . ■

Toda traslación puede expresarse como composición de dos homotecias, y el teorema siguiente implica que también puede expresarse como composición de dos simetrías respecto a dos variedades paralelas de dimensión arbitraria:

Teorema 5.28 *Si L es una variedad lineal en un espacio afín euclídeo, \vec{v} un vector ortogonal a \vec{L} y $L' = \vec{v} + L$, entonces $s_L \circ s_{L'} = T_{2\vec{v}}$.*

DEMOSTRACIÓN: Tomemos un punto arbitrario P . Sea $P' = s_L(P)$ y sea $P'' = s_{L'}(P')$. Por definición de simetría

$$M = \frac{1}{2}P + \frac{1}{2}P' \in L \quad \text{y} \quad \overrightarrow{PP'} \in \vec{L}^\perp.$$

Llamemos $M' = M + \vec{v} \in L'$. Entonces $\overrightarrow{P'M'} = \overrightarrow{P'M} + \overrightarrow{MM'} = \frac{1}{2}\overrightarrow{P'P} + \vec{v} \in \vec{L}^\perp$, y de nuevo por la definición de simetría $\overrightarrow{P'P''} = 2\overrightarrow{P'M'}$. Así pues,

$$\overrightarrow{PP''} = \overrightarrow{PP'} + \overrightarrow{P'P''} = 2\overrightarrow{MP'} + 2\overrightarrow{P'M'} = 2\overrightarrow{MM'} = 2\vec{v}. \quad \blacksquare$$

5.5 Clasificación de isometrías

Vamos a clasificar las isometrías de un espacio euclídeo sobre un cuerpo realmente cerrado R . El teorema siguiente reduce el problema al caso de las isometrías lineales:

Teorema 5.29 Sea E un espacio afín euclídeo y $f \in \text{Is}(E)$. Entonces f se descompone de forma única como $f = g \circ T_{\vec{v}}$, donde g es una isometría con un conjunto no vacío G de puntos fijos y $\vec{v} \in \vec{G}$. Además en tal caso $\vec{G} = S(\vec{f}, 1)$ y $f = T_{\vec{v}} \circ g$.

DEMOSTRACIÓN: Probemos primero la unicidad. Dada una descomposición $f = g \circ T_{\vec{v}}$ en las condiciones del enunciado, es claro que $\vec{f} = \vec{g}$, luego tenemos que $\vec{G} = S(\vec{g}, 1) = S(\vec{f}, 1)$. Si P es un punto fijo de g , se cumple que $f(P) = P + \vec{v}$, luego $\overrightarrow{Pf(P)} = \vec{v} \in S(\vec{f}, 1)$.

Si $f = g' \circ T_{\vec{v}'}$ es otra descomposición en las condiciones del enunciado y P' es un punto fijo de g' , tendremos igualmente que $\overrightarrow{P'f(P')} = \vec{v}' \in S(\vec{f}, 1)$. Un simple cálculo nos da

$$\vec{v} - \vec{v}' = \overrightarrow{PP'} - \vec{f}(\overrightarrow{PP'}) \in S(\vec{f}, 1) \cap \text{Im}(\vec{f} - 1).$$

Si probamos que $E = S(\vec{f}, 1) \perp \text{Im}(\vec{f} - 1)$ tendremos que $\vec{v} = \vec{v}'$, luego también $g = g'$.

Notemos que $S(\vec{f}, 1)$ es el núcleo de $\vec{f} - 1$, luego basta probar que los dos espacios son ortogonales y por las dimensiones su suma será todo E . Ahora bien, dados $\vec{x} \in S(\vec{f}, 1)$ e $\vec{y} = \vec{f}(\vec{z}) - \vec{z} \in \text{Im}(\vec{f} - 1)$, se cumple

$$\vec{x}\vec{y} = \vec{x}\vec{f}(\vec{z}) - \vec{x}\vec{z} = \vec{f}(\vec{x})\vec{f}(\vec{z}) - \vec{x}\vec{z} = \vec{x}\vec{z} - \vec{x}\vec{z} = 0,$$

pues \vec{f} conserva el producto escalar.

Para probar la existencia fijamos un punto O arbitrario y descomponemos

$$\overrightarrow{Of(O)} = \vec{v} + \vec{f}(\vec{w}) - \vec{w},$$

con $\vec{v} \in S(\vec{f}, 1)$ y $\vec{f}(\vec{w}) - \vec{w} \in \text{Im}(\vec{f} - 1)$ (usando la descomposición de E que acabamos de obtener). Entonces, para todo punto P se cumple

$$\overrightarrow{Pf(P)} = \overrightarrow{PO} + \overrightarrow{Of(O)} + \vec{f}(\overrightarrow{OP}) = \vec{v} + (\vec{f} - 1)(\vec{w} + \overrightarrow{OP}).$$

En particular, si tomamos A de modo que $\overrightarrow{OA} = -\vec{w}$ la fórmula anterior se reduce a $\overrightarrow{Af(A)} = \vec{v} \in S(\vec{f}, 1)$. Equivalentemente, $f(A) = A + \vec{v}$.

Definimos $g = f \circ T_{-\vec{v}}$. Así g es ciertamente una isometría, se cumple $f = g \circ T_{\vec{v}}$ y además $g(A) = f(A) - \vec{v} = A$, luego el conjunto G de sus puntos fijos es no vacío. Más aún, es claro que $\vec{G} = S(\vec{g}, 1) = S(\vec{f}, 1)$ y por consiguiente $\vec{v} \in \vec{G}$.

Por último, se cumple que $f = T_{\vec{v}} \circ g$ porque para todo punto P tenemos

$$g(P + \vec{v}) = g(P) + \vec{g}(\vec{v}) = g(P) + \vec{f}(\vec{v}) = g(P) + \vec{v} = f(P). \quad \blacksquare$$

Un par de observaciones complementarias: en el teorema anterior (por la unicidad) se cumplirá $\vec{v} = \vec{0}$ si y sólo si f tiene un punto fijo, y si f no tiene puntos fijos, entonces $\dim G \geq 1$ (pues entonces $\vec{0} \neq \vec{v} \in \vec{G}$).

En otras palabras: si clasificamos las isometrías con al menos un punto fijo, las que no tienen puntos fijos se obtienen a partir de las que tienen más de un punto fijo componiéndolas con traslaciones respecto de vectores que lleven puntos fijos a puntos fijos.

Si nos restringimos, pues, a isometrías con al menos un punto fijo O , tomando un sistema de referencia centrado en O , todo se reduce a estudiar las isometrías lineales de un espacio vectorial euclídeo V . Para ello consideraremos sus matrices posibles respecto a bases ortonormales y nos formulamos entonces las preguntas habituales en este contexto: ¿qué relación hay entre dos matrices de una misma isometría lineal en dos bases ortonormales distintas? y ¿cómo saber si dos matrices corresponden a la misma isometría en bases distintas? Como siempre, la primera pregunta es fácil y concreta el planteamiento de la segunda:

Definición 5.30 Diremos que dos matrices $S, T \in \text{Mat}_n(R)$ son *congruentes* si existe una matriz $P \in O(n, R)$ tal que $T = P^{-1}SP$.

La interpretación de esta relación es sencilla. Supongamos que S y T son dos matrices de un mismo endomorfismo h en dos bases ortonormales B y C . Entonces sabemos que $T = P^{-1}SP$, donde P es la matriz de cambio de base entre B y C o, en otros términos, la matriz de la identidad en \mathbb{R}^n respecto de las bases B y C . Ahora bien, como la identidad es una isometría y las bases son ortonormales, sabemos que P ha de ser una matriz ortogonal, luego S y T son congruentes. Recíprocamente, si S es la matriz de h en una base ortonormal y T es una matriz congruente con S , entonces existe una base ortonormal de V respecto a la cual la matriz de h es T .

En la sección [Al 6.3] encontramos formas canónicas para las matrices de los endomorfismos de un espacio vectorial de dimensión finita, pero la teoría no es aplicable directamente a nuestro contexto, porque necesitamos formas canónicas para la relación de congruencia, no la de semejanza. En principio, hay dos inconvenientes:

- 1) El paso de una matriz a otra semejante puede exigir el paso a una base no ortonormal, de modo que la matriz resultante no sea congruente con la dada.
- 2) La forma canónica (en el sentido de la sección [Al 6.3] de una matriz ortogonal no tiene por qué ser ortogonal, y entonces no sirve como forma canónica respecto de la congruencia.

Vamos a ver que 1) no es realmente un problema, sino que dos matrices ortogonales son congruentes si y sólo si son semejantes. Una vez probado esto, resolveremos el problema planteado por 2) eligiendo en la clase de semejanza de cada matriz ortogonal otra forma canónica que sea también ortogonal. La clave en lo tocante a 1) la proporciona el teorema siguiente, donde usamos por primera vez la hipótesis de que el cuerpo R es realmente cerrado:

Teorema 5.31 *Sea A una matriz ortogonal. Entonces las raíces complejas del polinomio mínimo de A son simples y de módulo 1.*

DEMOSTRACIÓN: Cuando hablamos de raíces complejas nos referimos a raíces en el cuerpo algebraicamente cerrado $C = R[i]$. Vamos a considerar $V = C^n$ como C -espacio vectorial y $h : V \rightarrow V$ el isomorfismo definido por A en la base canónica. Así, $h(v) = vA$.

Veamos ante todo que el polinomio mínimo de A es el mismo tanto si la consideramos como matriz real o como matriz compleja. Sea $p(x)$ el polinomio mínimo real y $q(x)$ el polinomio mínimo complejo. Como $p(A) = 0$ es claro que $q(x) \mid p(x)$ (el recíproco no vale porque $p(x)$ divide a todos los polinomios reales que anulan a A). Descomponemos $q(x) = q_1(x) + q_2(x)i$, con $q_1(x), q_2(x) \in R[x]$.

Como $q(A) = q_1(A) + q_2(A)i = 0$, ha de ser $q_1(A) = q_2(A) = 0$, luego $p(x)$ divide a ambos, luego $p(x) \mid q(x)$.

Sea α una raíz del polinomio mínimo de A . Entonces es un valor propio, luego existe un $v \in V$ no nulo tal que $vA = \alpha v$. Trasponiendo y conjugando esta ecuación obtenemos $A^t v^t = \alpha v^t$ y $\bar{v}A = \bar{\alpha}\bar{v}$, donde \bar{v} es el vector que resulta de conjugar cada coordenada de v . La matriz A no necesita ser conjugada porque sus coeficientes son reales. La linealidad de la conjugación hace que la igualdad siga siendo cierta.

Ahora multiplicamos las ecuaciones: $\bar{v}AA^t v^t = \alpha\bar{\alpha}\bar{v}v^t$, y como A es ortogonal la expresión se reduce a $\bar{v}v^t = |\alpha|^2 \bar{v}v^t$.

Pero $\bar{v}v^t$ es la suma de los cuadrados de los módulos de las coordenadas de v . Como v no es nulo, $\bar{v}v^t \neq 0$, luego $|\alpha|^2 = 1$ y así, $|\alpha| = 1$.

Supongamos que el polinomio mínimo tiene una raíz múltiple α . Entonces un divisor elemental de A es de la forma $(x - \alpha)^r$ con $r \geq 2$. Sea $v \in V$ tal que su polinomio mínimo sea $(x - \alpha)^r$ y sea $u = (x - \alpha)^{r-2}v$. Entonces $(x - \alpha)u \neq 0$, pero $(x - \alpha)^2 u = 0$.

En otros términos, si $w = (x - \alpha)u$, tenemos que $w \neq 0$ pero $(x - \alpha)w = 0$, o también:

$$w = uA - \alpha u \neq 0 \quad \text{y} \quad wA = \alpha w.$$

Despejamos en la primera ecuación y trasponemos: $A^t u^t = w^t + \alpha u^t$, conjugamos la segunda: $\bar{w}A = \bar{\alpha}\bar{w}$, las multiplicamos: $\bar{w}AA^t u^t = \bar{\alpha}\bar{w}w^t + \alpha\bar{\alpha}\bar{w}u^t$, aplicamos que A es ortogonal y que α tiene módulo 1, con lo que llegamos a $\bar{w}u^t = \bar{\alpha}\bar{w}w^t + \bar{w}u^t$. O sea, $\bar{\alpha}\bar{w}w^t = 0$, de donde $\bar{w}w^t = 0$, pero $\bar{w}w^t$ es la suma de los cuadrados de los módulos de las coordenadas de $w \neq 0$, luego no puede ser cero. ■

Así pues, los factores irreducibles (en $R[x]$) del polinomio mínimo de una matriz ortogonal han de ser o bien $x - 1$, o bien $x + 1$, o bien de la forma

$$(x - (\cos \theta + i \operatorname{sen} \theta))(x - (\cos \theta - i \operatorname{sen} \theta)) = x^2 - 2 \cos \theta x + 1,$$

para un cierto $\theta \in]0, \pi[$ (aquí es esencial la clausura algebraica de C). Además han de ser distintos dos a dos. Los divisores elementales dividen al polinomio mínimo, luego han de tener todos exponente 1, es decir, han de ser irreducibles.

El polinomio característico es el producto de los divisores elementales, luego éstos son precisamente los factores irreducibles del polinomio característico, luego el polinomio característico de una matriz ortogonal determina los divisores elementales y, consecuentemente, también los factores invariantes.

Ejemplo Consideremos la matriz

$$B = \begin{pmatrix} 1/9 & -4/9 & 8/9 \\ -4/9 & 7/9 & 4/9 \\ 8/9 & 4/9 & 1/9 \end{pmatrix}$$

Es fácil comprobar que se trata de una matriz ortogonal. Podríamos calcular sus factores invariantes con el algoritmo dado en la sección [Al 6.3], pero basta calcular el polinomio característico, que resulta ser $(x+1)(x-1)^2$. Como sabemos que los divisores elementales son irreducibles, han de ser $x+1$, $x-1$ y $x-1$, de donde los factores invariantes son $x-1$ y $(x+1)(x-1)$. ■

Esto significa que dos matrices ortogonales son semejantes si y sólo si tienen el mismo polinomio característico, pero vamos a probar que, según anunciábamos, de hecho en tal caso son congruentes (con lo que la semejanza equivale a la congruencia).

Teorema 5.32 *Sea V un espacio vectorial euclídeo y $h : V \rightarrow V$ una isometría con divisores elementales p_1, \dots, p_r . Entonces V admite una descomposición de la forma*

$$V = \langle v_1 \rangle_{k[x]} \perp \cdots \perp \langle v_r \rangle_{k[x]},$$

donde $\text{pol m}^\text{ín } v_i = p_i$.

DEMOSTRACIÓN: Vamos a probar por inducción sobre la dimensión de V que existe una descomposición como la indicada de modo que los polinomios mínimos de los generadores son irreducibles (sin afirmar que sean los divisores elementales de h). Esto implica inmediatamente el teorema por la unicidad de los divisores elementales. Si la dimensión de V es 1 es evidente.

Sea p un divisor elemental de h y sea $v \in V$ tal que $\text{pol m}^\text{ín } v = p$. Sea $W = \langle v \rangle_{k[x]}$. Entonces $V = W \perp W^\perp$. Tenemos que W es invariante y como h es biyectiva de hecho $h[W] = W$. Puesto que h conserva el producto escalar, es claro que $h[W^\perp] = h[W]^\perp = W^\perp$, luego podemos restringir h a una isometría de W^\perp y aplicar la hipótesis de inducción. ■

Si tomamos una base ortonormal de cada uno de los subespacios de la descomposición del teorema anterior, su unión nos da una base ortonormal de V en la cual la matriz de h está formada por cajas $[M_1, \dots, M_r]$, donde M_i es la matriz de la restricción de h al subespacio i -ésimo.

Como los divisores elementales tienen grado 1 o 2, éstas son las dimensiones de las cajas. Si $p_i = x-1$ podemos tomar como base el vector v_i y, puesto que $h(v_i) = v_i$, la caja correspondiente es un 1.

Si $p_i = x+1$ tomamos igualmente la base v_i y ahora $h(v_i) = -v_i$, luego la caja es un -1 .

Si $p_i = x^2 - 2 \cos \theta x + 1$, entonces la caja M_i es una matriz de $O(2)$ cuyo único divisor elemental (luego también su único factor invariante, luego su polinomio característico) es p_i , luego su determinante es el término independiente 1 y su

traza es $2\cos\theta$. Estas condiciones determinan completamente la matriz. En efecto, dada una matriz cualquiera

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

la ecuación $AA^t = I$ implica $a^2 + b^2 = 1$, $c^2 + d^2 = 1$, $ac + bd = 0$. La ecuación $A^tA = I$ nos da también $a^2 + c^2 = 1$, $b^2 + d^2 = 1$, de donde se sigue inmediatamente $c = \pm b$ y $d = \pm a$. La ecuación $ac + bd = 0$ implica que los signos han de ser distintos, es decir, $d = \pm a$ y $c = \mp b$. Si exigimos que el determinante sea 1, es decir, $ad - bc = 1$, llegamos a que $d = a$ y $c = -b$. La matriz es, por lo tanto,

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Al considerar la traza vemos que $a = \cos\theta$ y la ecuación $a^2 + b^2 = 1$ implica que $b = \pm\sin\theta$. Cambiando el signo al segundo vector de la base si es preciso, podemos exigir que $b = \sin\theta$. En definitiva obtenemos la matriz

$$M_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

que corresponde al giro de ángulo θ . Vamos a resumir lo que hemos obtenido.

Definición 5.33 Diremos que una matriz ortogonal está en *forma canónica* si es de la forma $[M_1, \dots, M_r]$, donde cada caja M_i es ± 1 o bien M_θ , para un ángulo $\theta \in]0, \pi[$.

Hemos probado que toda isometría admite en una cierta base ortonormal una matriz en forma canónica con tantas cajas iguales a 1 como divisores elementales iguales a $x - 1$, tantas cajas iguales a -1 como divisores elementales iguales a $x + 1$ y tantas cajas de la forma M_θ como divisores elementales de la forma $x^2 - 2\cos\theta x + 1$ (para $\theta \in]0, \pi[$).

El hecho de que las cajas estén determinadas por (y de hecho determinen) los divisores elementales implica que dos formas canónicas no son congruentes salvo que se diferencien en una permutación de las cajas.

Si a esto añadimos que los divisores elementales están determinados por el polinomio característico, concluimos:

Teorema 5.34 *Dos matrices ortogonales son congruentes si y sólo si son semejantes, si y sólo si tienen el mismo polinomio característico.*

El teorema 5.29 junto con la determinación de las formas canónicas de las isometrías lineales constituyen una clasificación de las isometrías de un espacio euclídeo, que puede hacerse más explícita para dimensiones pequeñas. En general, cuando hablemos del determinante, la traza, etc. de una isometría nos referimos, naturalmente, a los de su aplicación lineal asociada.

Isometrías de una recta Si E es una recta euclídea, cada isometría lineal de $O(\vec{E})$ admite, respecto de cierta base ortonormal, una matriz de la forma $[\pm 1]$. Si el signo es positivo se trata de la identidad y si es negativo se trata de la aplicación $\vec{f}(\vec{v}) = -\vec{v}$, por lo que $O(\vec{E}) \cong \{\pm 1\}$ es un grupo de orden 2.

Equivalentemente, el grupo $O_O(\vec{E})$ formado por las isometrías que fijan al punto O tiene orden 2, y está formado por la identidad y la simetría respecto de O . Según el teorema 5.29, el grupo de isometrías $O(E)$ está formado por las traslaciones y las simetrías puntuales. Se distinguen porque las traslaciones tienen determinante 1 y las simetrías tienen determinante -1 .

Isometrías del plano Si E es un plano euclídeo, una isometría de $O(\vec{E})$ (o una isometría de $O_O(E)$, para cualquier punto O prefijado), admitirá una forma canónica de la forma M_θ (incluyendo aquí las posibilidades $[\pm 1, \pm 1]$, correspondientes a $\theta = 0$ o $\theta = \pi$) si su determinante es 1, mientras que si el determinante es -1 la única posibilidad es $[1, -1]$.

El segundo caso corresponde a una simetría axial que tiene por eje la recta determinada por el primer vector de la base ortonormal en la que la matriz de la isometría es canónica. En el primer caso tenemos giros, y en la discusión tras la definición 3.40 vimos que un giro tiene una matriz de la forma M_θ respecto de cualquier base ortonormal (pues un cambio de base a lo sumo cambia θ por $-\theta$).

Esto se traduce en que el grupo $O^+(2, R)$ de las matrices ortogonales 2×2 de determinante 1 está formado por las matrices de la forma M_θ . En particular, para $R = \mathbb{R}$ tenemos un epimorfismo de grupos

$$\mathbb{R} \longrightarrow O^+(2, \mathbb{R})$$

dado por $\theta \mapsto M_\theta$ que induce un isomorfismo

$$O^+(2, \mathbb{R}) \cong \mathbb{R}/\langle 2\pi \rangle.$$

A su vez, una isometría $f \in \text{Is}(E)$ de determinante 1 con un punto fijo será un giro (o la identidad), mientras que si no tiene puntos fijos entonces \vec{f} ha de ser la identidad (por el teorema 5.2), luego f es una traslación. Si el determinante es -1 y f tiene un punto fijo, será una simetría axial, mientras que si no tiene puntos fijos el teorema 5.29 implica que f es la composición de una simetría axial con una traslación respecto a un vector en la dirección del eje de simetría.

Isometrías del espacio Consideremos ahora el caso en que E es un espacio tridimensional euclídeo. La forma canónica de una isometría de $O(\vec{E})$ con determinante 1 tiene que ser de la forma $[M_\theta, 1]$. Esto significa que el grupo $O^+(\vec{E})$ formado por las isometrías lineales de determinante 1 es precisamente el grupo de giros descrito tras la definición 4.13, y si llamamos $O^+(3, R)$ al grupo de las matrices ortogonales 3×3 de determinante 1, tenemos los isomorfismos

$$O^+(3, R) \cong O^+(\vec{E}) \cong H_1/\langle \pm 1 \rangle$$

que representan estos grupos en términos del grupo de los cuaternios unitarios.

Por otra parte, la forma canónica de una isometría de determinante -1 tiene que ser de la forma $[M_\theta, -1]$, que corresponde a un giro (tal vez la identidad) compuesto con una simetría especular respecto del plano perpendicular al eje de giro.

Por consiguiente, una isometría de $\text{Is}(E)$ de determinante 1 con un punto fijo será un giro (tal vez la identidad) y, si no tiene puntos fijos, por el teorema 5.29 será la composición de un giro con una traslación respecto a un vector en la dirección del eje de giro. Si el giro es la identidad tenemos simplemente una traslación. En caso contrario la isometría es lo que se llama un *giro helicoidal*, pues describe el movimiento de una hélice, que gira y avanza al mismo tiempo.

Si el determinante es -1 y \vec{f} es un giro (no trivial) compuesto con una simetría, entonces \vec{f} no tiene puntos fijos, luego f ha de tener exactamente uno, luego f es la composición de un giro con una simetría especular respecto a un plano perpendicular al eje de giro (el punto fijo de f es la intersección entre el eje y el plano). Si el giro es la identidad, entonces, o bien f tiene puntos fijos y se trata de una simetría especular, o bien no los tiene, y se trata de una simetría especular compuesta con una traslación en dirección perpendicular al plano de simetría.

La tabla siguiente resume la clasificación:

Dim.	Det.	Con punto fijo	Sin puntos fijos
1	+1	Identidad	Traslación
	-1	Simetría puntual	—
2	+1	Giro	Traslación
	-1	Simetría axial	Simetría con traslación
3	+1	Giro	Traslación o Giro helicoidal
	-1	Simetría especular	Simetría con traslación o Giro y simetría

Ejercicio: ¿Por qué no aparecen en la tabla las simetrías puntuales en dimensión 2 y 3 ni las simetrías axiales en dimensión 3?

Ejercicio: Clasificar las isometrías en dimensión 4.

Definimos ahora en general los grupos $O^+(n, R)$ que acabamos de introducir en casos particulares:

Definición 5.35 La aplicación determinante induce un epimorfismo de grupos $O(n, R) \rightarrow \{\pm 1\}$, cuyo núcleo, el subgrupo formado por las matrices ortogonales de determinante 1, se representa por $O^+(n, R)$.

Similarmente, la aplicación que a cada isometría le asigna el determinante de su isometría lineal asociada es también un epimorfismo $\text{Is}(E) \rightarrow \{\pm 1\}$, cuyo núcleo se representa por $\text{Is}^+(E)$.

Las isometrías de determinante 1 se llaman *isometrías directas* o *movimientos*. Las isometrías de determinante -1 se denominan *isometrías inversas*.

En la sección siguiente explicaremos estas denominaciones. De momento terminamos demostrando lo siguiente:

Teorema 5.36 *Si E es un espacio euclídeo de dimensión $n \geq 2$ sobre un cuerpo realmente cerrado, el grupo $\text{Is}(E)$ está generado por las reflexiones, mientras que $\text{Is}^+(E)$ está generado por los giros.*

DEMOSTRACIÓN: Consideremos en primer lugar una isometría f (que no sea la identidad) con un punto fijo O . Consideremos un sistema de referencia de origen O respecto al cual \vec{f} tenga matriz $[M_1, \dots, M_r]$, es claro que podemos expresarla como composición de isometrías con punto fijo O que en el mismo sistema de referencia tengan matrices formadas por una sola de las cajas, que puede ser un -1 , en cuyo caso se trata de una reflexión, o bien una caja M_θ , en cuyo caso se trata de un giro. Veamos ahora que todo giro es composición de dos reflexiones.

En efecto, pongamos que el giro g tiene matriz $[M_\theta, 1, \dots, 1]$ y consideremos su composición con la reflexión h de matriz $[-1, 1, \dots, 1]$. El resultado es una isometría $u = g \circ h$ con punto fijo O cuya matriz es de la forma $[M, 1, \dots, 1]$, donde M es una matriz ortogonal 2×2 con determinante -1 . Por lo tanto, su forma canónica tiene que ser $[-1, 1]$, luego, cambiando dos vectores de la base, obtenemos otro sistema de referencia en que u tiene matriz $[-1, 1, \dots, 1]$, luego es una reflexión. Por lo tanto $g = u \circ h$ es composición de dos reflexiones.

Concluimos que toda isometría con un punto fijo es composición de reflexiones. Si una isometría no tiene puntos fijos, es composición de una de este tipo con una traslación, pero según 5.28 toda traslación es composición de dos reflexiones, luego tenemos probada la primera parte del teorema.

La segunda se obtiene con una ligera variante. Si \vec{f} tiene determinante 1, el número de cajas -1 en su forma canónica tiene que ser par. En lugar de descomponer las cajas individualmente, podemos considerar isometrías con forma canónica $[-1, -1, 1, \dots, 1]$, que son giros de ángulo π . Eso prueba que \vec{f} es composición de giros. Por último, el teorema 5.28 permite descomponer también una traslación como producto de dos simetrías respecto de un espacio de codimensión 2, pero sus formas canónicas son entonces $[-1, -1, 1, \dots, 1]$, luego también son giros de ángulo π . ■

5.6 Orientación

El determinante de una biyección afín tiene una interpretación geométrica. De momento vamos a interpretar su signo. Para ello empezamos dando unas definiciones generales:

Definición 5.37 Sea V un espacio vectorial de dimensión finita sobre un cuerpo ordenado R . En el conjunto de todas las bases ordenadas de V establecemos la relación de equivalencia según la cual dos bases están relacionadas si y sólo si la matriz de cambio de base tiene determinante positivo.

Esta relación determina claramente dos clases de equivalencia, a las que llamaremos *orientaciones* de V . Así, diremos que dos bases ordenadas de V *tienen la misma orientación* si pertenecen a la misma clase de equivalencia, es decir, si el determinante de la matriz de cambio de base es positivo.

Teniendo en cuenta las propiedades de los determinantes, es inmediato que si permutamos dos vectores de una base ordenada o si cambiamos uno de sus vectores por un múltiplo negativo, invertimos la orientación de la base.

Un *espacio vectorial orientado* es un par (V, O) formado por un espacio vectorial V y una de sus dos orientaciones, de modo que a las bases de O las llamamos *bases positivamente orientadas*, mientras que las restantes son *bases negativamente orientadas*.

En lo sucesivo consideraremos los espacios R^n como espacios orientados con el convenio de que la base canónica está positivamente orientada.

Un automorfismo $f : V \rightarrow V$ es *positivo* o *negativo* (o también *directo* o *inverso*) según que su determinante tenga signo positivo o negativo.

Observemos que la matriz de f en una base (v_1, \dots, v_n) es también la matriz de cambio de base de ésta a $(f(v_1), \dots, f(v_n))$, por lo que es inmediato que un automorfismo es positivo si y sólo si conserva la orientación de las bases, y es negativo si y sólo si invierte la orientación de las bases.

Recíprocamente, podemos decir que dos bases tienen la misma orientación si y sólo si es posible transformar una en otra mediante un automorfismo positivo (el determinado por la matriz de cambio de base).

Esta equivalencia es importante porque vamos a dar una interpretación geométrica de la orientación interpretando geoméricamente los automorfismos positivos. Pero antes conviene trasladar estos conceptos al caso de espacios afines:

Un *espacio afín orientado* es un par (E, O) , donde O es una orientación del espacio vectorial \vec{E} . Un sistema de referencia de E se dice *positivo* o *negativamente orientado* según que lo sea su base asociada.

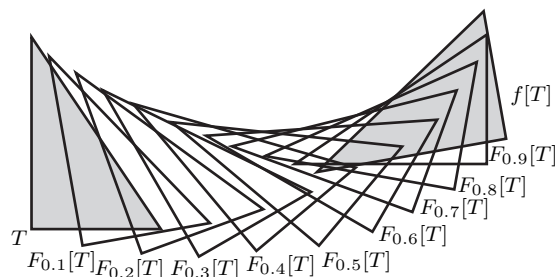
Similarmente, una biyección afín es *positiva* o *negativa* (o *directa* o *inversa*) según que lo sea su aplicación lineal asociada.

Notemos que esta definición generaliza a la que hemos dado en la sección anterior para isometrías.

Claramente, una biyección afín es positiva o negativa según si conserva o invierte la orientación de los sistemas de referencia afines, y dos sistemas de referencia afines tienen la misma orientación si y sólo si uno puede transformarse en el otro mediante una biyección afín directa. Si los sistemas son ortonormales, la biyección afín será de hecho una isometría directa.

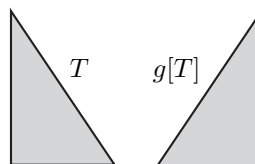
Vamos a dar en primer lugar una interpretación de las isometrías directas/inversas y luego la generalizaremos a biyecciones afines arbitrarias. La interpretación es válida para espacios vectoriales sobre \mathbb{R} .

Teorema 5.38 Sea E un espacio euclídeo sobre el cuerpo \mathbb{R} de los números reales. Una isometría $f : E \rightarrow E$ es directa si y sólo si existe una aplicación continua $F : [0, 1] \times E \rightarrow E$ tal que las aplicaciones $F_t : E \rightarrow E$ dadas por $F_t(x) = F(t, x)$ son isometrías, F_0 es la identidad y $F_1 = f$.



Por ejemplo, la figura muestra un triángulo T y su imagen $f[T]$ por una isometría directa. Que sea directa se traduce en que es posible realizar la transformación de forma gradual. Podemos pensar que la variable t representa el tiempo, de modo que el triángulo T “se mueve” desde la posición T que ocupa en el instante $t = 0$ hasta la posición $f[T]$, a la que llega en $t = 1$, y entonces $F_t[T]$ es la posición del triángulo en el instante t . En el enunciado del teorema es esencial la continuidad de F , que se traduce en que cuanto más parecidos sean los instantes t y t' , más parecidos serán las figuras $F_t[T]$ y $F_{t'}[T]$, de modo que no hay ninguna alteración “brusca” en ningún momento. Es por esto que las isometrías directas se llaman también movimientos.

En cambio, los dos triángulos de la figura son también congruentes, por lo que existe una (única) isometría g que transforma uno en el otro, una simetría axial, que es una isometría inversa. Tenía que ser inversa, porque es imposible mover T hasta que pase a ocupar la posición que ocupa $g[T]$. Para ello en algún momento hay que “darle la vuelta” al triángulo, y eso se traduce en una discontinuidad de cualquier función F que pretenda estar en las condiciones del teorema.



DEMOSTRACIÓN: Consideramos en primer lugar en el caso en que f es un giro. Entonces, en un sistema de referencia adecuado, la expresión de f en coordenadas es

$$f(x) = (x_1 \cos \theta - x_2 \operatorname{sen} \theta, x_1 \operatorname{sen} \theta + x_2 \cos \theta, x_3, \dots, x_n),$$

y basta definir F de modo que su expresión en coordenadas sea

$$F(t, x) = (x_1 \cos t\theta - x_2 \operatorname{sen} t\theta, x_1 \operatorname{sen} t\theta + x_2 \cos t\theta, x_3, \dots, x_n).$$

Claramente es continua⁶ y cumple lo requerido.

⁶Aquí suponemos conocida la continuidad de las funciones $\operatorname{sen} x$, $\cos x$, tal y como se prueba en [ITAn 5.1].

En el caso general, el teorema 5.36 nos da que $f = f_1 \circ \dots \circ f_m$, donde cada f_i es un giro. Sea $F_i : [0, 1] \times E \rightarrow E$ una aplicación en las condiciones del teorema válida para f_i . Entonces $F_t = F_{1t} \circ \dots \circ F_{mt}$ cumple claramente lo requerido.

Recíprocamente, supongamos que existe la aplicación F y veamos que la isometría f es directa. En realidad, no necesitamos suponer que todas las aplicaciones F_t sean isometrías. Basta suponer que son biyecciones afines y que, por lo tanto, tienen determinante no nulo.

Componiendo con la aplicación que a cada punto le asigna sus coordenadas en un sistema de referencia dado, podemos suponer que $F : [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. Entonces $\vec{F}_t(x) = F_t(x) - F_t(0)$ y la componente $a_{ij}(t)$ de la matriz de \vec{F}_t es la componente j -ésima de $F(t, e_i) - F(t, 0)$, donde e_i es el i -ésimo vector de la base canónica. De aquí se sigue que las funciones $a_{ij} : [0, 1] \rightarrow \mathbb{R}$ son continuas, y el determinante de la matriz es una suma de productos de estos coeficientes, luego la función $d : [0, 1] \rightarrow \mathbb{R}$ dada por $d(t) = \det F_t$ es continua.

Como $d(0) = \det F_0 = \det I = 1 > 0$ y por hipótesis d no toma nunca el valor 0, la continuidad exige que d tome únicamente valores positivos, luego $d(1) = \det f > 0$, como había que probar. ■

Ahora generalizamos el teorema al caso de biyecciones afines arbitrarias:

Teorema 5.39 *Sea E un espacio afín sobre el cuerpo \mathbb{R} de los números reales. Una biyección afín $f : E \rightarrow E$ es directa si y sólo si existe una aplicación continua $F : [0, 1] \times E \rightarrow E$ tal que las aplicaciones $F_t : E \rightarrow E$ dadas por $F_t(x) = F(t, x)$ son biyecciones afines, F_0 es la identidad y $F_1 = f$.*

DEMOSTRACIÓN: Vamos a probar primero el teorema cambiando E por un espacio vectorial euclídeo V y f por un automorfismo. Vamos a usar el teorema [Al 6.22] según el cual (véase también la prueba) todo $f \in \text{LG}(V)$ se expresa como composición de transvecciones y una dilatación. Como las transvecciones tienen determinante 1, el determinante de la dilatación tiene que ser el de f , luego tiene que ser positivo. Por consiguiente, por el mismo argumento empleado en el teorema anterior, basta probar el teorema para transvecciones y dilataciones de determinante positivo.

Si f es una transvección, entonces su expresión en coordenadas respecto a un sistema de referencia adecuado es

$$f(x) = (x_1 + \alpha x_2, x_2, \dots, x_n)$$

y basta definir F de modo que su expresión en coordenadas sea

$$F(t, x) = (x_1 + t \alpha x_2, x_2, \dots, x_n).$$

Si f es una dilatación de determinante $r > 0$, en un sistema de referencia adecuado su expresión en coordenadas es

$$f(x) = (rx_1, x_2, \dots, x_n),$$

y basta definir F de modo que, en coordenadas,

$$F(t, x) = ((1 - t + tr)x_1, x_2, \dots, x_n).$$

Esto termina la prueba para automorfismos. En el caso general, podemos suponer que $E = \mathbb{R}^n$. Si f es una biyección afín, $f(P) = f(0) + \vec{f}(P)$. Sea $G : [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ la aplicación que cumple el teorema para \vec{f} , y definimos

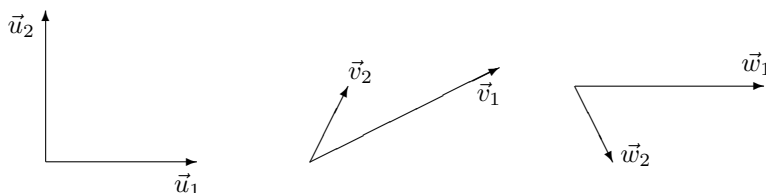
$$F(t, P) = tf(0) + G(t, P).$$

Claramente F es continua y cumple lo requerido.

Recíprocamente, si existe F , entonces f tiene determinante positivo, porque el argumento dado en el teorema anterior vale literalmente en este caso sin cambio alguno. ■

Como indicábamos, para espacios sobre cuerpos ordenados distintos de \mathbb{R} no es válido el argumento de continuidad que hemos dado, sino que falla especialmente en el punto donde usamos la propiedad de los valores intermedios de las funciones continuas, que esencialmente equivale a que los subespacios conexos de \mathbb{R} son los intervalos. No obstante, el concepto de orientación lo tenemos definido en cualquier espacio euclídeo sobre un cuerpo ordenado (pitagórico) y su comportamiento es análogo en todos los aspectos al que tiene en los espacios reales.

Ahora podemos analizar a su vez la interpretación del concepto de orientación de una base. Por ejemplo, consideremos las tres bases orientadas de la figura:



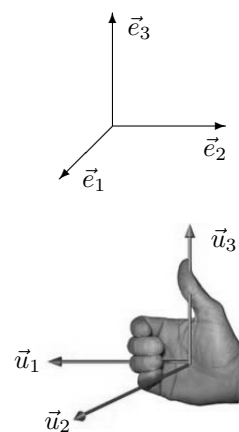
Podemos concluir intuitivamente, es decir, sin necesidad de hacer ningún cálculo, que las dos primeras tienen la misma orientación, mientras que la tercera tiene la orientación opuesta. En efecto, para transformar la primera en la segunda sólo hay que “estirar” un poco \vec{u}_1 , “acortar” \vec{u}_2 y reducir el ángulo entre ellos hasta que ocupen las posiciones de \vec{v}_1 y \vec{v}_2 , y todo ello puede hacerse de forma continua sin que en ningún momento de la transformación los vectores dejen de formar una base.

En cambio, si intentamos transformar la primera base en la tercera, no hay problema en “estirar” continuamente \vec{u}_1 para que pase a ser \vec{w}_1 , pero si intentamos girar \vec{u}_2 hasta que ocupe el lugar de \vec{w}_2 , necesariamente, en un momento del giro los dos vectores tienen que alinearse, con lo que dejan de ser una base. Por el contrario, sí que podemos transformar continuamente \vec{u}_1 en \vec{w}_2 y \vec{u}_2 en \vec{w}_1 sin perder nunca la independencia lineal. Esto se debe a que la orientación de (\vec{w}_2, \vec{w}_1) es la opuesta de la de (\vec{w}_1, \vec{w}_2) .

El convenio habitual es considerar como orientación positiva de un plano la determinada por la primera de las tres bases anteriores. Puede describirse especificando que el giro que transforma \vec{u}_1 en \vec{u}_2 siguiendo el arco más corto se produce en sentido antihorario, es decir, en el sentido contrario al de las agujas de un reloj. Pero conviene ser consciente de que todas las palabras que podemos usar para distinguir una orientación de otra (como “sentido horario” y “sentido antihorario”, o “izquierda” y “derecha”, etc.) sólo pueden definirse intuitivamente, es decir, mediante una imagen. Si pudiéramos hablar por radio, pero sin transmitir imágenes, con un habitante de otro planeta y no supiéramos si allí son diestros o zurdos, o en qué sentido giran las agujas de los relojes que construyen, sería imposible usando meramente el lenguaje determinar si usan las palabras “izquierda” y “derecha”, “sentido horario” y “sentido antihorario” en el mismo sentido que nosotros o justo en el opuesto. Matemáticamente, podemos justificar —y lo hemos hecho— que en un espacio afín hay dos orientaciones, pero no hay ningún criterio que nos permita seleccionar una en concreto como positiva. (En el espacio R^n hemos tomado como orientación positiva la de la base canónica, pero en un espacio afín arbitrario no hay bases canónicas, cosa que se pone de manifiesto cuando pretendemos representar gráficamente \mathbb{R}^2 , pues no hay nada que determine si “lo correcto” es representar la base canónica como la base (\vec{u}_1, \vec{u}_2) anterior, que es el convenio usual, o como (\vec{u}_2, \vec{u}_1) .)

Similarmente, el convenio usual para las representaciones tridimensionales es dibujar la base canónica de \mathbb{R}^3 con la orientación de la figura. Ésta puede describirse mediante la “regla de la mano derecha”: si ponemos la mano derecha con el índice curvado y el pulgar perpendicular al arco determinado por aquél, si el giro que lleva de la base a la punta del índice es el que transforma el primer vector de la base en el segundo (por el arco más corto), entonces el tercero debe apuntar en el sentido que señala el pulgar.

Este criterio permite reconocer la orientación de cualquier base representada gráficamente, aunque no sea ortonormal. El criterio es que el tercer vector apunte hacia el mismo semiespacio que el pulgar.



La orientación del producto vectorial Para terminar, ahora podemos completar la caracterización geométrica del producto vectorial definido en la sección 4.3. Si R es un cuerpo pitagórico, podemos identificar R^3 con el espacio de vectores del anillo de división $H = R^4$ de los cuaternios sobre R , con lo que tenemos definido el producto vectorial de dos vectores $x, y \in R^3$ mediante

$$x \wedge y = \begin{vmatrix} i & j & k \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix},$$

y sabemos que, si x e y son linealmente independientes, entonces $z = x \wedge y$ es perpendicular a ambos. Ahora podemos añadir que su sentido es el que hace

que la base x, y, z sea positiva (respecto de la base canónica o, en términos de los cuaternios, respecto de la base (i, j, k)). Basta observar que el determinante de la matriz de cambio de base es

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} = z_1^2 + z_2^2 + z_3^2 > 0,$$

donde hemos desarrollado el determinante por la tercera fila. ■

Capítulo VI

Construcciones con regla y compás

En la sección [IGE 6.4] dimos una caracterización algebraica de los puntos del plano constructibles con regla y compás y demostramos que, identificados con números complejos, los puntos constructibles son algebraicos. Más aún, enunciamos sin demostración que otra condición necesaria es que el grado de su polinomio mínimo sea potencia de 2, lo que nos permitió demostrar que determinadas construcciones geométricas clásicas (como la cuadratura del círculo, la trisección del ángulo, etc.) no pueden realizarse exclusivamente con regla y compás. En este capítulo profundizaremos en el estudio de la constructibilidad.

6.1 Constructibilidad

Empezamos recordando la definición [IGE 6.11] de constructibilidad con regla y compás. Allí la enunciamos identificando directamente los puntos del plano con los números complejos. En términos puramente geométricos ligeramente más generales (y añadiendo la posibilidad de partir de un conjunto de puntos dados) puede formularse así:

Definición 6.1 Sea E un plano euclídeo (sobre un cuerpo euclídeo R) y Γ un conjunto de puntos de E . Diremos que un punto $P \in E$ es *constructible (con regla y compás)* a partir de Γ si existe una sucesión finita de puntos P_1, \dots, P_n tal que $P = P_n$ y cada P_i cumpla una de las condiciones siguientes:

1. $P_i \in \Gamma$.
2. P_i es la intersección de dos rectas que pasan cada una por dos puntos previos.
3. P_i es está en la intersección de una recta que pasa por dos puntos previos y una circunferencia cuyo centro es un punto previo y cuyo radio es la distancia entre dos puntos previos.

4. P_i está en la intersección de dos circunferencias con centro en dos puntos previos y radios que sean las distancias entre dos pares de puntos previos.

Diremos que una recta es *constructible* a partir de Γ si pasa por dos puntos constructibles a partir de Γ , y que una circunferencia es *constructible* a partir de Γ si su centro es constructible a partir de Γ y su radio es la distancia entre dos puntos constructibles a partir de Γ .

Una *construcción* con regla y compás es cualquier sucesión P_1, \dots, P_n que cumpla las condiciones de la definición anterior. Observemos que si P y Q son constructibles a partir de Γ y P_1, \dots, P_n y Q_1, \dots, Q_m son construcciones de P y Q respectivamente, entonces la sucesión $P_1, \dots, P_n, Q_1, \dots, Q_m$ es también una construcción. Más en general, cualquier conjunto finito de puntos constructibles puede incluirse en una misma construcción, de donde se sigue que los puntos de intersección entre dos rectas constructibles, dos circunferencias constructibles o una recta y una circunferencia constructibles son también constructibles.

A partir de aquí el lector puede deducir las definiciones obvias de cuándo se consideran construidas otras figuras geométricas. Por ejemplo, una semirrecta es constructible si lo es su extremo y la recta que la contiene, un segmento es constructible si lo son sus extremos, un ángulo es constructible si lo son su vértice y sus lados, etc.

Es inmediato que a partir del conjunto vacío no es posible construir nada, y a partir de un conjunto con un solo punto sólo es constructible dicho punto. Por lo tanto, se necesita un mínimo de dos puntos para iniciar cualquier construcción no trivial. Si marcamos arbitrariamente dos puntos P_0 y P_1 en un plano euclídeo, podemos fijar un sistema de referencia ortonormal respecto al que sus coordenadas sean $(0, 0)$ y $(1, 0)$, y es claro entonces que un punto $P(x, y)$ será constructible respecto de P_0, P_1 si y sólo si (x, y) es constructible en R^2 respecto de $(0, 0)$ y $(1, 0)$. Así pues, considerar a estos dos puntos como constructibles no supone realmente ninguna pérdida de generalidad.

Si R es un cuerpo ordenado euclídeo, llamaremos \mathcal{C}_R al conjunto de puntos de R^2 constructibles con regla y compás a partir de los puntos $(0, 0)$ y $(1, 0)$.

Consideraremos a $C = R^2$ con la estructura de cuerpo usual, de modo que R se identifica con los puntos de la forma $(r, 0)$, con $r \in R$ y de modo que $C = R[i]$, con $i^2 = -1$.

En particular, así $R \subset C$ y podemos considerar el conjunto $R_{rc} = \mathcal{C}_R \cap R$ formado por los puntos de R (del eje real de C) constructibles con regla y compás.

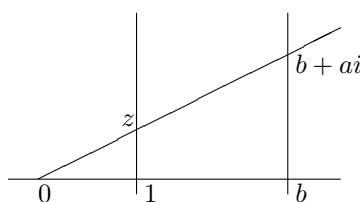
Remitimos a la sección [IGE 6.4] para la justificación de que cualquier construcción geométrica clásica con regla y compás justifica la constructibilidad de una figura en el sentido algebraico. Así, en [IGE] hemos dado construcciones elementales que justifican que es posible construir con regla y compás la mediatriz de un segmento dado, o la paralela y la perpendicular a una recta dada por un punto dado, o la bisectriz de un ángulo dado, etc.

El teorema [IGE 6.12] afirma que un elemento de C es constructible si y sólo si lo son su parte real y su parte imaginaria o, equivalentemente, que $\mathcal{C}_R = R_{rc}^2$. Por conveniencia repetimos aquí la demostración del teorema [IGE 6.14]:

Teorema 6.2 *Si R es un cuerpo euclídeo y $C = R[i]$, el conjunto \mathcal{C}_R de elementos de C constructibles con regla y compás es un subcuerpo de C y $R_{rc} = \mathcal{C}_R \cap R$ es un cuerpo ordenado euclídeo. Además $\mathcal{C}_R = R_{rc}[i]$.*

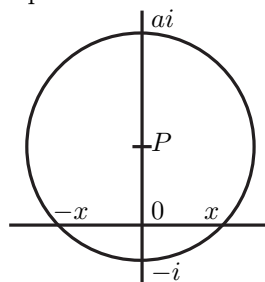
DEMOSTRACIÓN: Veamos en primer lugar que R_{rc} es un subcuerpo de R . Sean $a, b \in R_{rc}$. Si $b = 0$, entonces $a - b = a \in R_{rc}$. En otro caso, la circunferencia de centro a y radio $|b|$ es constructible y corta al eje real en $a + b$ y $a - b$, luego $a - b \in R_{rc}$. Vemos así que R_{rc} es un subgrupo aditivo de R .

Ahora supongamos que $b \neq 0$ y veamos que $ab^{-1} \in R_{rc}$, lo que probará que es un subcuerpo. Podemos suponer que $a, b > 0$. Usando las construcciones que hemos visto antes es claro que todos los puntos marcados en la figura son constructibles. Además, la distancia de z a 1 es $\tan \alpha = ab^{-1}$.



Dados z_1 y z_2 en \mathcal{C}_R , tenemos que sus partes reales e imaginarias son constructibles, de donde se sigue fácilmente, aplicando la parte ya probada, que las partes reales e imaginarias de $z_1 - z_2$ y $z_1 z_2^{-1}$ (si $z_2 \neq 0$) son constructibles, luego también lo son estos números, luego \mathcal{C}_R es un cuerpo.

Falta probar que si $a \in R_{rc}$ y $a > 0$, entonces $\sqrt{a} \in R_{rc}$. Sea P el punto medio entre ai y $-i$ (que se obtiene de la bisectriz del segmento). La circunferencia de centro P que pasa por $-i$ corta al eje real en dos puntos $\pm x$. La potencia de 0 respecto a esta circunferencia (en módulo) es por una parte x^2 y por otra parte a , luego $x = \sqrt{a}$. ■



En particular todos los números racionales son constructibles, pues \mathbb{Q} es el menor cuerpo contenido en C , pero también es fácil dar una prueba directa de la constructibilidad de los números racionales. Como $\mathcal{C}_R = R_{rc}[i]$, el teorema [Al 5.49] nos da que todo elemento de \mathcal{C}_R tiene raíz cuadrada en \mathcal{C}_R .

Ahora damos una primera caracterización algebraica de la constructibilidad que precisa el teorema [IGE 6.16], según el cual todos los números constructibles son algebraicos. En primer lugar consideramos números reales:

Teorema 6.3 *Sea R un cuerpo ordenado euclídeo. Un $z \in R$ está en R_{rc} si y sólo si existe una cadena de cuerpos*

$$\mathbb{Q} = K_0 \subset K_1 \subset \dots \subset K_n \subset R$$

de modo que $|K_{j+1} : K_j| = 2$ y $z \in K_n$.

DEMOSTRACIÓN: Sea $z_1, \dots, z_n = z$ una construcción de z y vamos a construir una sucesión de cuerpos en las condiciones del enunciado de modo que cada cuerpo K_j contenga la parte real y la parte imaginaria de z_j . Razonando inductivamente, basta probar que si un cuerpo K contiene las partes reales e imaginarias de los puntos anteriores a z_j entonces las partes real e imaginaria de z_j están en K o en una extensión cuadrática de K .

Pongamos que $z_j = x + yi$. Si $z_j = 0, 1$ entonces trivialmente $z_j \in K$. Supongamos que z_j es la intersección de dos rectas que pasan por dos puntos de entre los anteriores, en particular dos puntos con coordenadas en K . Dados dos puntos (a, b) y (c, d) en K^2 , sabemos que la ecuación lineal

$$(d - b)(x - a) = (c - a)(y - b)$$

corresponde a una recta en K^2 , y claramente pasa por (a, b) y (c, d) , luego es la ecuación de la recta que pasa por los dos puntos dados. Al operarla se convierte en una ecuación $Ax + By = C$ con coeficientes en K . Tenemos, pues, que $z_j = (x, y)$ es la solución única de dos ecuaciones de este tipo, y al resolver explícitamente el sistema obtenemos una expresión de x e y en función de los coeficientes, que prueba que ambos están en K .

Supongamos ahora que z_j está en la intersección de una recta que pasa por dos puntos de K^2 y una circunferencia con centro K^2 y radio r igual a la distancia entre dos puntos de K^2 . Entonces $r^2 \in K$, luego tenemos que (x, y) satisface las ecuaciones

$$\left. \begin{array}{l} Ax + By = C \\ x^2 + y^2 + ax + by + c = 0 \end{array} \right\}$$

donde todos los coeficientes están en K . Uno de los coeficientes A o B ha de ser no nulo. Al despejar la variable correspondiente en la primera ecuación y sustituirla en la segunda obtenemos una ecuación de segundo grado con coeficientes en K que ha de tener entre sus soluciones una de las coordenadas de z_j , luego ésta se encuentra en una extensión de $K(\sqrt{\alpha})$ de K , donde α es el discriminante de la ecuación, que está en K . La otra coordenada se despeja de la primera ecuación, luego está también en $K(\sqrt{\alpha})$. Claramente $|K(\sqrt{\alpha}) : K| \leq 2$.

Por último suponemos que z_j está en la intersección de dos circunferencias constructibles a partir de los puntos anteriores. Por el mismo razonamiento de antes, sus coordenadas cumplen dos ecuaciones de la forma

$$\left. \begin{array}{l} x^2 + y^2 + ax + by + c = 0 \\ x^2 + y^2 + a'x + b'y + c' = 0 \end{array} \right\}$$

con los coeficientes en K . Restando ambas ecuaciones obtenemos la ecuación

$$(a - a')x + (b - b')y + c - c' = 0,$$

que también es satisfecha por las coordenadas de z_j y cuyos coeficientes no pueden ser todos nulos, o si no las dos circunferencias serían la misma. En definitiva tenemos que z_j está también en una recta cuya ecuación tiene sus coeficientes en K y este caso se reduce al anterior.

Recíprocamente, si z cumple las condiciones del enunciado, entonces es claro que $K_{j+1} = K_j(\sqrt{\alpha_j})$ para un cierto $\alpha_j \in K_j$, $\alpha_j > 0$, y si $K_j \subset R_{rc}$, también $\sqrt{\alpha_j} \in R_{rc}$, luego $K_{j+1} \subset R_{rc}$. Como obviamente $K_0 \subset R_{rc}$, concluimos que $z \in K_n \subset R_{rc}$. ■

A partir del teorema anterior es fácil probar una variante para números complejos:

Teorema 6.4 *Sea R un cuerpo ordenado euclídeo. Un $z \in C$ está en \mathcal{C}_R si y sólo si existe una cadena de cuerpos*

$$\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_n \subset C$$

de modo que $|F_{j+1} : F_j| = 2$ y $z \in F_n$.

DEMOSTRACIÓN: Si $z \in \mathcal{C}_R$, entonces su parte real y su parte imaginaria están en R_{rc} . La prueba del teorema anterior muestra que podemos obtener una cadena de cuerpos $\mathbb{Q} = K_0 \subset \cdots \subset K_n \subset R$ de modo que K_n contenga la parte real y la parte imaginaria de z , y entonces basta tomar la cadena de cuerpos

$$\mathbb{Q} \subset K_0(i) \subset \cdots \subset K_n(i) \subset C.$$

Recíprocamente, si existe una cadena en las condiciones del enunciado, es claro que $F_{j+1} = F_j(\sqrt{\alpha_j})$ para un cierto $\alpha_j \in F_j$, y si se cumple $F_j \subset \mathcal{C}_R$, también $\sqrt{\alpha_j} \in \mathcal{C}_R$, luego $F_{j+1} \subset \mathcal{C}_R$. Como obviamente $F_0 \subset \mathcal{C}_R$, concluimos que $z \in F_n \subset \mathcal{C}_R$. ■

Ahora tenemos una condición necesaria sencilla para que un número sea constructible. Notemos que es muy restrictiva:

Teorema 6.5 *Si R es un cuerpo ordenado euclídeo y $z \in C_R$, entonces z es algebraico sobre \mathbb{Q} y $|\mathbb{Q}(z) : \mathbb{Q}|$ es potencia de 2.*

En particular el teorema anterior afirma que la extensión \mathcal{C}_R/\mathbb{Q} es algebraica. Sucede que la condición $|\mathbb{Q}(z) : \mathbb{Q}| = 2^n$ no es suficiente para garantizar que z sea constructible.

El teorema siguiente muestra que, esencialmente, sólo hay un cuerpo de números constructibles:

Teorema 6.6 *Si R es un cuerpo ordenado euclídeo, entonces:*

1. R_{rc} es el menor subcuerpo euclídeo de R . En particular, si $R \subset R'$ son cuerpos euclídeos, entonces $R_{rc} = R'_{rc}$ y $\mathcal{C}_R = \mathcal{C}_{R'}$.
2. La extensión \mathcal{C}_R/\mathbb{Q} es de Galois.
3. Si R y R' son dos cuerpos euclídeos cualesquiera, entonces $R_{rc} \cong R'_{rc}$ y $\mathcal{C}_R \cong \mathcal{C}_{R'}$.

DEMOSTRACIÓN: 1) Sea $R' \subset R$ un cuerpo euclídeo y sea $z \in R_{rc}$. Consideramos una sucesión de cuerpos

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_n \subset R$$

en las condiciones del teorema 6.3 con $z \in K_n$. Si $K_j \subset R'$, como claramente $K_{j+1} = K_j(\sqrt{\alpha_j})$, para cierto $\alpha_j \in K_j \subset R'$, necesariamente positivo y R' es euclídeo, resulta que $K_{j+1} \subset R'$. Así $z \in K_n \subset R'$.

En particular, si $R \subset R'$, el menor subcuerpo euclídeo de R es también el menor subcuerpo euclídeo de R' , luego $R_{rc} = R'_{rc}$, y por consiguiente también $\mathbb{C}_R = R_{rc}[i] = R'_{rc}[i] = \mathbb{C}_{R'}$.

Vamos a probar 2) para $R = \mathbb{R}$. El caso general se deducirá de 3). Sabemos que la extensión $\mathbb{C}_{\mathbb{R}}/\mathbb{Q}$ es algebraica, y obviamente separable, porque los cuerpos son de característica 0. Falta probar que si $z \in \mathbb{C}_{\mathbb{R}}$ y $p(x) = \text{pol m\u00edn}(z, \mathbb{Q})$, entonces $p(x)$ se escinde en $\mathbb{C}_{\mathbb{R}}[x]$. Sabemos que se escinde en $\mathbb{C}[x]$, luego basta probar que toda raíz z' de $p(x)$ en \mathbb{C} est\u00e1 de hecho en $\mathbb{C}_{\mathbb{R}}$.

Sea $\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_n \subset \mathbb{C}$ una sucesi\u00f3n en las condiciones del teorema 6.4 con $z \in F_n$ y sea K la clausura normal de F_n sobre \mathbb{Q} . Entonces $z' \in K$ y existe $\sigma \in G(K/\mathbb{Q})$ tal que $\sigma(z) = z'$. Claramente, la sucesi\u00f3n de cuerpos $\mathbb{Q} = \sigma[F_0] \subset \cdots \subset \sigma[F_n] \subset K \subset \mathbb{C}$ cumple las condiciones de 6.4 y prueba que $z' \in \mathbb{C}_{\mathbb{R}}$.

3) Basta probar el resultado para $R' = \mathbb{R}$. Como $\mathbb{R}_{rc}/\mathbb{Q}$ es algebraica, el cuerpo \mathbb{R}_{rc} es numerable, y a partir de una enumeraci\u00f3n de sus elementos es f\u00e1cil construir una sucesi\u00f3n $\{z_j\}_{j=1}^{\infty}$ que contenga a todos los elementos de \mathbb{R}_{rc} y de modo que cada sucesi\u00f3n finita $\{z_j\}_{j=1}^n$ cumpla las condiciones de la definici\u00f3n 6.1. El argumento del teorema 6.3 nos da entonces una sucesi\u00f3n $\{K_j\}_{j=0}^{\infty}$ de subcuerpos de \mathbb{R} tal que

$$K_0 = \mathbb{Q}, \quad K_j \subset K_{j+1}, \quad |K_{j+1} : K_j| = 2, \quad \mathbb{R}_{rc} = \bigcup_{j=0}^{\infty} K_j.$$

Veamos que existe una sucesi\u00f3n $\{\sigma_j\}_{j=0}^{\infty}$ de monomorfismos de cuerpos ordenados $\sigma_j : K_j \rightarrow R_{rc}$. Tomamos como σ_0 la identidad en \mathbb{Q} , y supuesto definido σ_j , como $K_{j+1} = K_j(\sqrt{\alpha_j})$, con $\alpha_j \in K_j$, necesariamente positivo, y $\sigma_j(\alpha_j) > 0$ tiene ra\u00edz cuadrada en R_{rc} , el teorema [Al 5.12] nos permite extender σ_j hasta K_{j+1} , y es claro que la extensi\u00f3n conserva el orden. La uni\u00f3n de todos los σ_j nos da un monomorfismo $\sigma : \mathbb{R}_{rc} \rightarrow R_{rc}$.

Ahora bien, si $z' \in R_c$, sea $\mathbb{Q} = K'_0 \subset \cdots \subset K'_n \subset R$ una sucesi\u00f3n de cuerpos en las condiciones del teorema 6.3 con $z' \in K'_n$. Entonces se cumple que cada $K'_j \subset \sigma[\mathbb{R}_{rc}]$. En efecto, para $j = 0$ es obvio, y si vale para j , tenemos que $K'_{j+1} = K'_j(\pm\sqrt{\alpha_j})$, con $\alpha_j \in K'_j$ positivo, luego $\sigma^{-1}(\alpha_j) \in \mathbb{R}_{rc}$ positivo, luego sus ra\u00edces cuadradas tambi\u00e9n est\u00e1n en \mathbb{R}_{rc} , y la imagen por σ de una de las dos tiene que ser $\sqrt{\alpha_j}$, luego $\sqrt{\alpha_j} \in \sigma[\mathbb{R}_{rc}]$, luego $K'_{j+1} \subset \sigma[\mathbb{R}_{rc}]$, luego $z' \in K'_n \subset \sigma[\mathbb{R}_{rc}]$.

Esto prueba que $\sigma : \mathbb{R}_{rc} \rightarrow R_{rc}$ es un isomorfismo, que se extiende a su vez a un isomorfismo $\mathbb{C}_{\mathbb{R}} \cong \mathbb{C}_R$. ■

Definición 6.7 Llamaremos $\mathcal{C} = \mathcal{C}_{\mathbb{R}}$ al cuerpo de los números complejos constructibles y $\mathcal{C}_r = \mathbb{R}_{rc} = \mathcal{C} \cap \mathbb{R}$ al cuerpo ordenado euclídeo de los números reales constructibles, de modo que $\mathcal{C} = \mathcal{C}_r[i]$.

Hemos probado que \mathcal{C}/\mathbb{Q} es una extensión de Galois y que \mathcal{C}_r es el menor cuerpo ordenado euclídeo, en el sentido de que cualquier otro contiene una copia isomorfa de \mathcal{C}_r . Más aún, hemos visto que los puntos constructibles de un plano euclídeo no aumentan si extendemos el cuerpo de escalares, sino que son siempre los mismos. El plano constructible $\mathcal{C} = \mathcal{C}_r^2$ es el menor plano que cumple los axiomas de la geometría euclídea plana (incluyendo el de las circunferencias, pero no el de continuidad), y una figura es constructible con regla y compás si y sólo si existe en dicho plano. Todo esto hace que no perdamos generalidad si nos limitamos a estudiar los cuerpos \mathcal{C} y \mathcal{C}_r (lo que nos sitúa en la geometría arquimediana).

Ahora podemos demostrar un recíproco parcial del teorema 6.5, que constituye una aplicación no trivial de la teoría de Galois y la teoría de grupos:

Teorema 6.8 Si K es una extensión finita de Galois de \mathbb{Q} , entonces $K \subset \mathcal{C}$ si y sólo si $|K : \mathbb{Q}|$ es potencia de 2.

DEMOSTRACIÓN: El grupo de Galois $G(K/\mathbb{Q})$ es un 2-grupo, luego por la observación tras el teorema [TG 3.6] contiene una cadena de subgrupos normales:

$$1 = G_n \trianglelefteq G_{n-1} \trianglelefteq \cdots \trianglelefteq G_0 = G(K/\mathbb{Q})$$

de modo que cada uno tiene índice 2 en el siguiente. Por el teorema de Galois los cuerpos fijados forman una cadena de cuerpos

$$\mathbb{Q} = K_0 \subset \cdots \subset K_n = K$$

cada uno de los cuales tiene índice 2 en el siguiente, luego $K \subset \mathcal{C}$ por 6.4.

La implicación opuesta se sigue de que la extensión tiene un elemento primitivo, necesariamente constructible, de modo que $K = \mathbb{Q}(z)$, y basta aplicar el teorema 6.5. ■

Insistimos, no obstante, en que si un número $z \in \mathbb{C}$ cumple, por ejemplo, $|\mathbb{Q}(z) : \mathbb{Q}| = 4$, eso no garantiza que z sea constructible, pues la extensión no tiene por qué ser de Galois. De hecho, si K es la clausura normal de $\mathbb{Q}(z)$ sobre \mathbb{Q} , como la extensión \mathcal{C}/\mathbb{Q} es de Galois, para que z sea constructible es necesario y suficiente que $K \subset \mathcal{C}$, pero el grado $|K : \mathbb{Q}|$ podría ser 12 o 24.

Ejemplo: La trisección del ángulo Sabemos que es posible trazar con regla y compás la bisectriz de cualquier ángulo dado. Sin embargo, no es posible en general dividir cualquier ángulo dado en tres partes iguales, ni siquiera suponiendo que el ángulo dado sea constructible.

Observemos en primer lugar que si con regla y compás podemos trazar un ángulo, \overline{AOB} , podemos usar el compás para marcar un punto A' en el lado \overline{OA}

de modo que $\overline{OA'} = 1$, y entonces la perpendicular por A' al lado opuesto corta a éste en un punto B' tal que $\overline{OB'}$ es el coseno del ángulo. Recíprocamente, a partir del coseno es fácil construir el ángulo, luego concluimos que un ángulo es constructible con regla y compás si y sólo si lo es su coseno.

La fórmula para el coseno de una suma nos da que

$$\begin{aligned}\cos 3\alpha &= \cos(2\alpha + \alpha) = \cos 2\alpha \cos \alpha - \operatorname{sen} 2\alpha \operatorname{sen} \alpha = \\ &(\cos^2 \alpha - \operatorname{sen}^2 \alpha) \cos \alpha - 2 \operatorname{sen}^2 \alpha \cos \alpha = \cos^3 \alpha - 3(1 - \cos^2 \alpha) \cos \alpha.\end{aligned}$$

En definitiva:

$$\cos 3\alpha = 4 \cos^3 \alpha - 3 \cos \alpha.$$

Equivalentemente:

$$4 \cos^3 \frac{\alpha}{3} - 3 \cos \frac{\alpha}{3} = \cos \alpha.$$

Por ejemplo, $\cos 60^\circ = 1/2$, luego el ángulo de 60° es constructible, pero su trisección cumple

$$4 \cos^3 20^\circ - 3 \cos 20^\circ = \frac{1}{2}.$$

El polinomio $p(x) = x^3 - \frac{3}{4}x - \frac{1}{8}$ es irreducible en $\mathbb{Q}[x]$. Para probarlo basta ver que no tiene raíces en \mathbb{Q} . Si q fuera una raíz, tendríamos que

$$8q^3 - 6q - 1 = 0,$$

luego $2q$ sería una raíz racional de $x^3 - 3x - 1$. Ahora bien, por [Al 3.35] las raíces racionales de este polinomio tienen que ser enteras. Si n es una raíz, todo primo que divida a n debe dividir a 1, luego tiene que ser $n = \pm 1$, pero se comprueba que ± 1 no son raíces.

Concluimos que $p(x)$ es el polinomio mínimo de $\cos 20^\circ$, y como tiene grado 3, podemos concluir que el coseno no es constructible, luego el ángulo de 20° tampoco lo es, luego el ángulo de 60° es un ángulo constructible que no puede trisecarse con regla y compás. ■

Remitimos a [IGE] para más detalles sobre el problema clásico de la trisección del ángulo o el de la duplicación del cubo, o la cuadratura del círculo, etc., ninguno de los cuales tiene solución con regla y compás, como se desprende fácilmente de los resultados precedentes.

6.2 Construcción de polígonos regulares

Es especialmente interesante la teoría en torno a la construcción con regla y compás de polígonos regulares. El axioma de continuidad implica que existen polígonos regulares de cualquier centro y radio para todo $n \geq 3$, pues de él se sigue que el cuerpo de escalares es \mathbb{R} , luego los puntos del plano se identifican con los de \mathbb{C} , que es un cuerpo algebraicamente cerrado y, en general, si el

cuerpo $C = R[i]$ es algebraicamente cerrado, entonces contiene una raíz n -sima primitiva de la unidad

$$\zeta_n = \cos \frac{2\pi}{n} + i \operatorname{sen} \frac{2\pi}{n},$$

cuyas potencias son los vértices de un polígono regular de centro 0 y radio 1 (véase la discusión tras [ITAn 5.10]). A partir de él se puede construir fácilmente un polígono regular de cualquier centro y radio.

Ahora bien, si no suponemos el axioma de continuidad, o al menos que el cuerpo $C = R[i]$ sea algebraicamente cerrado, ya no es necesario que existan polígonos regulares de todos los lados posibles. En particular, los polígonos regulares no son necesariamente constructibles con regla y compás.

Ejemplo Los eneágonos regulares (polígonos de 9 lados) no son constructibles con regla y compás, pues ello equivaldría a que lo fuera el ángulo de 40° , y si éste lo fuera, también lo sería el de 20° (porque todo ángulo se puede bisecar), y al final de la sección anterior hemos visto que no lo es. ■

En general, es claro que los polígonos regulares de n lados son constructibles si y sólo si lo son los ángulos de amplitud $2\pi/n$, lo cual equivale a su vez a que $\cos \frac{2\pi}{n} \in \mathcal{C}$, o a que $\zeta_n \in \mathcal{C}$, o a que $\mathbb{Q}(\zeta_n) \subset \mathcal{C}$. Teniendo en cuenta que la extensión ciclotómica $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ es finita de Galois de grado $\phi(n)$ (por el teorema [Al 5.84]), el teorema 6.8 nos da inmediatamente esta caracterización:

Teorema 6.9 *Los polígonos regulares de n lados son constructibles con regla y compás si y sólo si $\phi(n)$ es potencia de 2.*

Por ejemplo, $\phi(9) = 6$, lo que nos da una prueba alternativa de que los eneágonos regulares no son constructibles con regla y compás, y a su vez de que los ángulos de 40° y 20° no son constructibles.

Como $\phi(7) = 6$, el heptágono regular tampoco es constructible con regla y compás, y es, de hecho, el menor polígono regular no constructible.

Vamos a dar una caracterización aritmética de cuándo $\phi(n)$ es potencia de 2. Factoricemos $n = 2^{r_0} p_1^{r_1} \cdots p_k^{r_k}$, donde p_1, \dots, p_k son primos impares distintos. Entonces $\phi(n) = (p_1 - 1)p_1^{r_1 - 1} \cdots (p_k - 1)p_k^{r_k - 1}$ y la única posibilidad para que este número sea una potencia de dos es que todos los exponentes sean 1 y que $p_i - 1$ sea potencia de 2 para todo i . Recordemos que los primos de la forma $2^n + 1$ se llaman primos de Fermat, y son, de hecho, de la forma $2^{2^n} + 1$ (véase la sección [ITAI 5.3]). Así pues, hemos probado lo siguiente:

Teorema 6.10 *El polígono regular de n lados es constructible con regla y compás si y sólo si n es producto de una potencia de 2 y de primos de Fermat con exponente 1.*

Los únicos primos de Fermat que se conocen son

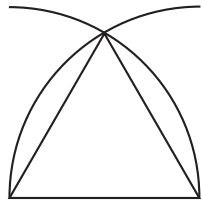
$$3, \quad 5, \quad 17, \quad 257, \quad 65537.$$

A partir de un polígono de n lados es muy fácil construir un polígono de $2n$ lados (basta bisecar sus ángulos o sus lados), luego sólo tiene interés la construcción de polígonos con un número impar de lados. A partir de los primos de Fermat que conocemos sólo puede construirse un número finito de tales polígonos, concretamente 31. Los primeros tienen los siguientes números de lados:

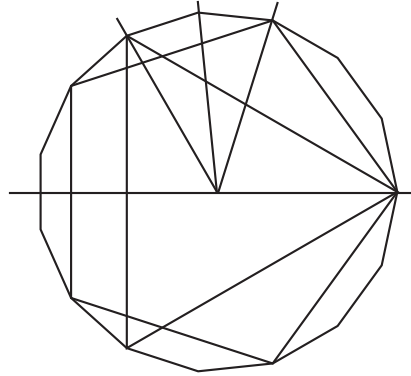
$$3, 5, 15, 17, 51, 85, 255, 257, \dots$$

La construcción de los tres primeros era conocida por los griegos, mientras que la del heptadecágono regular fue descubierta por Gauss.

Vamos a mostrar construcciones explícitas de estos cuatro polígonos. En primer lugar notemos que es muy fácil construir un triángulo equilátero y que, si sabemos construir un pentágono regular, también sabemos construir un pentadecágono regular: basta restar de un ángulo de amplitud $2\pi/3$ otro de amplitud $2\pi/5$ para obtener uno de amplitud $4\pi/15$, y después bisecar el resultado:



Construcción de un triángulo equilátero



Construcción de un pentadecágono regular

Sólo necesitamos, pues, encontrar una construcción para el pentágono y el heptadecágono regular. La construcción del pentágono es muy simple, y nos servirá de guía para obtener la del heptadecágono.

6.2.1 Construcción del pentágono regular

Sea $K = \mathbb{Q}(\zeta)$, donde $\zeta = 1_{2\pi/5}$, el cuerpo ciclotómico quinto. Su grado es 4 y su grupo de Galois G es cíclico. Un generador es el automorfismo σ que cumple $\sigma(\zeta) = \zeta^2$. Hay un único cuerpo intermedio L , el cuerpo fijado por σ^2 , al cual pertenecen $\eta_1 = \zeta + \zeta^4$ y $\eta_2 = \zeta^2 + \zeta^3$, que constituyen una \mathbb{Q} -base de L .

La conjugación compleja induce un automorfismo en K de orden 2, luego ha de ser σ^2 . Por lo tanto $\zeta^4 = \bar{\zeta}$, de donde $\eta_1 = 2 \operatorname{Re} \zeta = 2 \cos 72^\circ > 0$.

Teniendo en cuenta que ζ es una raíz de $x^4 + x^3 + x^2 + x + 1$, es inmediato que $\eta_1 + \eta_2 = \eta_1 \eta_2 = -1$, luego η_1 y η_2 son las raíces del polinomio $x^2 + x - 1$. Estas raíces son

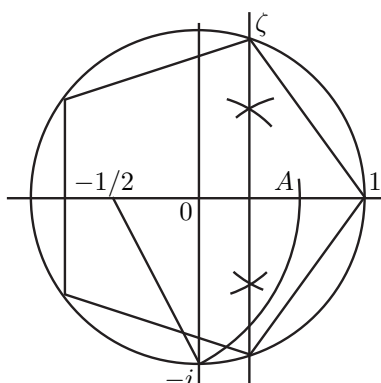
$$\eta_1 = \frac{-1 + \sqrt{5}}{2} \quad \text{y} \quad \eta_2 = \frac{-1 - \sqrt{5}}{2}$$

(sabemos que η_1 es la raíz positiva). Así pues,¹

$$\cos 72^\circ = \frac{-1 + \sqrt{5}}{4}.$$

Aunque no lo vamos a necesitar, hemos obtenido que $\sqrt{5} \in L$, luego $L = \mathbb{Q}(\sqrt{5})$.

Una forma rápida de construir el pentágono es trazar un arco con el compás apoyado en $-1/2$ y el otro extremo en $-i$, hasta cortar el eje real en el punto A . La distancia de $-1/2$ a $-i$ es $\sqrt{5}/2$, luego $A = \eta_1$. Después trazamos la mediatriz del segmento $\overline{0A}$, que pasa por $\cos 72^\circ$ y cortará al círculo de centro 0 y radio 1 en el punto ζ :



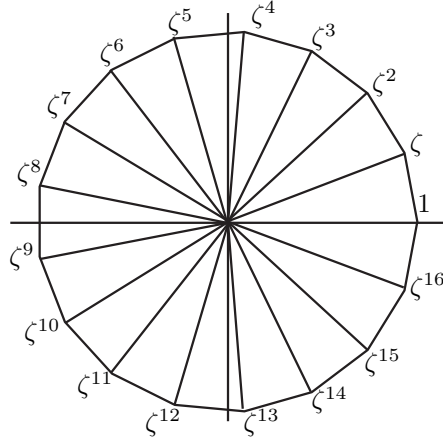
6.2.2 Construcción del heptadecágono regular

Sea $\zeta = 1_{2\pi/17}$ y $K = \mathbb{Q}(\zeta)$. Ahora el grupo de Galois G tiene orden 16. Una raíz primitiva módulo 17 es 3, luego el automorfismo σ determinado por $\sigma(\zeta) = \zeta^3$ tiene orden 16 y por lo tanto es un generador de G . El grupo G tiene tres subgrupos, a saber, $\langle \sigma^8 \rangle$, $\langle \sigma^4 \rangle$ y $\langle \sigma^2 \rangle$, que por el teorema de Galois se corresponden con tres cuerpos intermedios: $\mathbb{Q} \subset L_1 \subset L_2 \subset L_3 \subset K$, de grados 2, 4 y 8.

Para llegar desde los números racionales hasta ζ (y obtener una expresión radical de $\cos 2\pi/17$) hemos de dar cuatro pasos, lo que significa que la expresión de ζ será demasiado complicada para manejarla con comodidad. En lugar de eso nos limitaremos a describir ζ en función de su polinomio mínimo respecto a L_3 , y los coeficientes de este polinomio en función de sus polinomios mínimos en L_2 y así sucesivamente.

En el caso del pentágono los números η_1 y η_2 estaban determinados por que sabíamos que eran las raíces de $x^2 + x - 1$ y porque además sabíamos que η_1 era la raíz positiva. Aquí también tendremos que hacer estimaciones de signos con el mismo fin, para lo cual deberemos tener presente la distribución de las raíces de la unidad:

¹Alternativamente, el valor de $\cos 72^\circ$ se obtiene inmediatamente aplicando el teorema de Pitágoras aun triángulo áureo.



Como en el caso del pentágono, el automorfismo de orden 2, o sea, σ^8 es la conjugación compleja y así, ζ y ζ^{16} son conjugados, al igual que ζ^2 y ζ^{15} , etc. (como muestra la figura). El cuerpo L_3 (fijado por σ^8) contiene a los pares de raíces conjugadas:

$$\begin{aligned} \lambda_1 &= \zeta + \zeta^{16}, & \lambda_2 &= \zeta^2 + \zeta^{15}, & \lambda_3 &= \zeta^3 + \zeta^{14}, & \lambda_4 &= \zeta^4 + \zeta^{13}, \\ \lambda_5 &= \zeta^5 + \zeta^{12}, & \lambda_6 &= \zeta^6 + \zeta^{11}, & \lambda_7 &= \zeta^7 + \zeta^{10}, & \lambda_8 &= \zeta^8 + \zeta^9, \end{aligned}$$

y como $\zeta\zeta^{16} = 1$, es claro que ζ y ζ^{16} son las raíces del polinomio $x^2 - \lambda_1x + 1$. En general tenemos que

$$\begin{array}{lll} \zeta & \text{y} & \zeta^{16} \text{ son las raíces de } x^2 - \lambda_1x + 1 \\ \zeta^2 & \text{y} & \zeta^{15} \text{ son las raíces de } x^2 - \lambda_2x + 1 \\ \zeta^3 & \text{y} & \zeta^{14} \text{ son las raíces de } x^2 - \lambda_3x + 1 \\ \zeta^4 & \text{y} & \zeta^{13} \text{ son las raíces de } x^2 - \lambda_4x + 1 \\ \zeta^5 & \text{y} & \zeta^{12} \text{ son las raíces de } x^2 - \lambda_5x + 1 \\ \zeta^6 & \text{y} & \zeta^{11} \text{ son las raíces de } x^2 - \lambda_6x + 1 \\ \zeta^7 & \text{y} & \zeta^{10} \text{ son las raíces de } x^2 - \lambda_7x + 1 \\ \zeta^8 & \text{y} & \zeta^9 \text{ son las raíces de } x^2 - \lambda_8x + 1 \end{array}$$

En realidad para la construcción del heptadecágono nos interesan más los λ_k que las raíces de la unidad, pues claramente $\lambda_k = 2 \operatorname{Re} \zeta^k = 2 \cos(2k\pi/17)$. De aquí se sigue claramente (ver la figura anterior) que

$$\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > 0 > \lambda_5 > \lambda_6 > \lambda_7 > \lambda_8.$$

Una vez construido un λ_k , basta dividirlo entre 2 y levantar la perpendicular para obtener ζ^k y su conjugado como las intersecciones de dicha recta y la circunferencia unidad. Para construir los λ_k vamos a expresarlos en función del cuerpo L_2 , que es el cuerpo fijado por σ^4 y que, por tanto, tiene por base a las sumas:

$$\begin{aligned} \xi_1 &= \zeta + \zeta^{16} + \zeta^4 + \zeta^{13}, & \xi_2 &= \zeta^2 + \zeta^{15} + \zeta^8 + \zeta^9, \\ \xi_3 &= \zeta^3 + \zeta^{14} + \zeta^5 + \zeta^{12}, & \xi_4 &= \zeta^6 + \zeta^{11} + \zeta^7 + \zeta^{10}. \end{aligned}$$

(Si aplicamos a ζ las potencias de σ^4 obtenemos ζ , ζ^{16} , ζ^4 y ζ^{13} , por lo que σ^4 permuta los sumandos de ξ_1 , que es, por lo tanto, invariante). Es obvio que

$$\xi_1 = \lambda_1 + \lambda_4, \quad \xi_2 = \lambda_2 + \lambda_8, \quad \xi_3 = \lambda_3 + \lambda_5, \quad \xi_4 = \lambda_6 + \lambda_7.$$

Por otro lado

$$\begin{aligned} \lambda_1 \cdot \lambda_4 &= (\zeta + \zeta^{16})(\zeta^4 + \zeta^{13}) = \zeta^5 + \zeta^{14} + \zeta^3 + \zeta^{12} = \xi_3 \\ \lambda_2 \cdot \lambda_8 &= (\zeta^2 + \zeta^{15})(\zeta^8 + \zeta^9) = \zeta^{10} + \zeta^{11} + \zeta^6 + \zeta^7 = \xi_4 \\ \lambda_3 \cdot \lambda_5 &= (\zeta^3 + \zeta^{14})(\zeta^5 + \zeta^{12}) = \zeta^8 + \zeta^{15} + \zeta^2 + \zeta^9 = \xi_2 \\ \lambda_6 \cdot \lambda_7 &= (\zeta^6 + \zeta^{11})(\zeta^7 + \zeta^{10}) = \zeta^{13} + \zeta^{16} + \zeta + \zeta^4 = \xi_1 \end{aligned}$$

de donde se desprende que

$$\begin{array}{ll} \lambda_1 \text{ y } \lambda_4 & \text{son las raíces de } x^2 - \xi_1 x + \xi_3, \\ \lambda_2 \text{ y } \lambda_8 & \text{son las raíces de } x^2 - \xi_2 x + \xi_4, \\ \lambda_3 \text{ y } \lambda_5 & \text{son las raíces de } x^2 - \xi_3 x + \xi_2, \\ \lambda_6 \text{ y } \lambda_7 & \text{son las raíces de } x^2 - \xi_4 x + \xi_1. \end{array}$$

Finalmente, L_1 es el cuerpo fijado por σ^2 , y su base son las sumas:

$$\begin{aligned} \eta_1 &= \zeta + \zeta^9 + \zeta^{13} + \zeta^{15} + \zeta^{16} + \zeta^8 + \zeta^4 + \zeta^2 \\ \eta_2 &= \zeta^3 + \zeta^{10} + \zeta^5 + \zeta^{11} + \zeta^{14} + \zeta^7 + \zeta^{12} + \zeta^6. \end{aligned}$$

Un simple cálculo nos da $\eta_1 + \eta_2 = -1$, $\eta_1 \eta_2 = -4$, luego son las raíces del polinomio $x^2 + x - 4$.

Por otro lado se comprueba que ξ_1 y ξ_2 son las raíces de $x^2 - \eta_1 x - 1$, mientras que ξ_3 y ξ_4 son las raíces de $x^2 - \eta_2 x - 1$.

Estos datos nos permiten ascender gradualmente hasta construir ζ . No obstante, para agilizar la construcción nos valdremos de un truco:

Consideremos el ángulo θ que cumple $0 < 4\theta < \pi/2$ y $\tan 4\theta = 4$. Las raíces del polinomio $x^2 + x - 4$ son $2 \tan 2\theta$ y $-2/\tan 2\theta$, pues su producto es -4 y su suma vale

$$2 \tan 2\theta - \frac{2}{\tan 2\theta} = 2 \left(\frac{\tan^2 2\theta - 1}{\tan 2\theta} \right) = -\frac{4}{\tan 4\theta} = -1$$

(por la fórmula de la tangente del ángulo doble).

Por otra parte sabemos que las raíces son η_1 y η_2 , luego

$$\eta_1 = 2 \tan 2\theta \quad \text{y} \quad \eta_2 = -2 \tan 2\theta,$$

(hay que comprobar que $\eta_1 > 0$, pero eso se sigue de que

$$\eta_1 = \lambda_1 + \lambda_2 + \lambda_4 + \lambda_8 = 2 \left(\cos \frac{2\pi}{17} + \cos \left(2 \frac{2\pi}{17} \right) + \cos \left(4 \frac{2\pi}{17} \right) + \cos \left(8 \frac{2\pi}{17} \right) \right)$$

y los tres primeros sumandos son positivos, más aún, los dos primeros suman más de 1, luego compensan al último.)

Las raíces de $x^2 - \eta_1 x - 1$ son $\tan(\theta + \pi/4)$ y $\tan(\theta - \pi/4)$, pues su producto es -1 (los ángulos se diferencian en $\pi/2$, luego la tangente de uno es la inversa de la tangente del otro, y además tienen signos opuestos) y su suma es

$$\begin{aligned} \tan\left(\theta + \frac{\pi}{4}\right) + \tan\left(\theta - \frac{\pi}{4}\right) &= \frac{\tan\theta + 1}{1 - \tan\theta} + \frac{\tan\theta - 1}{1 + \tan\theta} \\ &= \frac{4 \tan\theta}{1 - \tan^2\theta} = 2 \tan 2\theta = \eta_1. \end{aligned}$$

Por lo tanto $\xi_1 = \tan(\theta + \pi/4)$ y $\xi_2 = \tan(\theta - \pi/4)$ (pues $\xi_1 = \lambda_1 + \lambda_4 > 0$). Con el mismo razonamiento se concluye que $\xi_3 = \tan\theta$ y $\xi_4 = -1/\tan\theta$. Ya tenemos suficiente para realizar la construcción:

Consideramos los puntos $A = 1$ y $B = i$. Dividimos el segmento \overline{OB} en cuatro partes, con lo que obtenemos el punto $I = (1/4)i$. El ángulo \widehat{OIA} , tiene tangente 4, luego es 4θ . Lo bisecamos dos veces y obtenemos θ igual al ángulo \widehat{OIE} (figura 6.1).

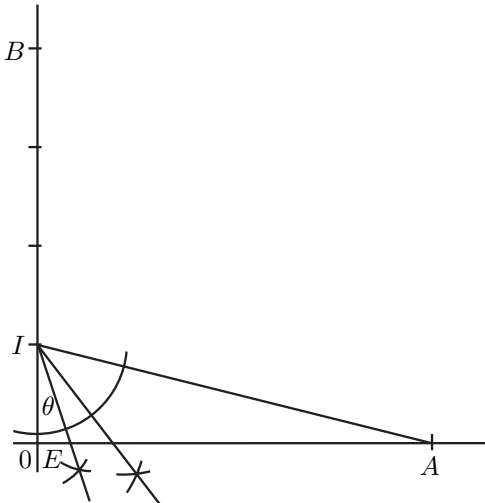


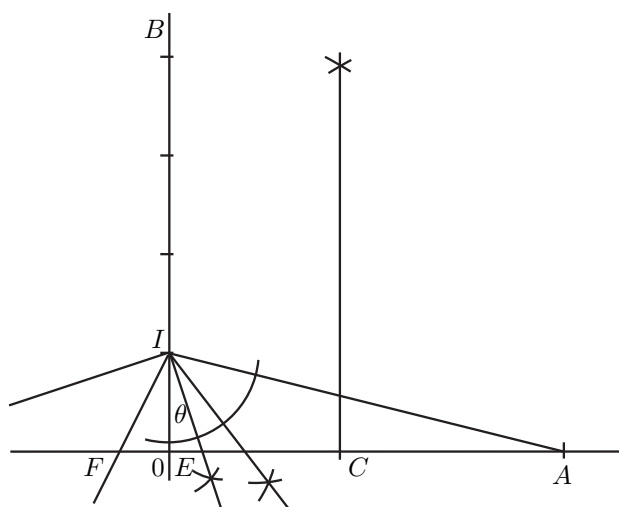
Figura 6.1: Construcción de θ .

Ahora construimos un punto F de modo que el ángulo \widehat{FIE} sea igual a $\pi/4$, con lo que el ángulo \widehat{OIF} es igual a $\pi/2 - \theta$. Llamamos C al punto medio de \overline{AF} (figura 6.2).

Según nuestros cálculos $E = \xi_3/4$, $F = -\xi_2/4$ y la distancia de A a F es $\overline{AF} = (4 - \xi_2)/4$. Por lo tanto la distancia de F a C (que es igual a la distancia de C a A) es $(4 - \xi_2)/8$. Consecuentemente

$$C = \overline{OC} = \frac{4 - \xi_2}{8} + \frac{\xi_2}{4} = \frac{\xi_2 + 4}{8}.$$

Trazamos el círculo de centro C y radio \overline{AC} , que cortará al eje imaginario en

Figura 6.2: Construcción de C .

un punto K . Por el teorema de Pitágoras

$$K = \overline{OK} = \sqrt{\overline{OK}^2 - \overline{OC}^2} = \sqrt{\overline{CA}^2 - \overline{OC}^2} = \frac{\sqrt{-\xi_2}}{2},$$

$$\overline{EK} = \sqrt{\overline{OK}^2 + \overline{OE}^2} = \frac{\sqrt{\xi_3 - 4\xi_2}}{4}.$$

A continuación trazamos el círculo de centro E y radio \overline{EK} , que cortan al eje horizontal en los puntos N_5 y N_3 . Entonces

$$N_3 = E + \overline{EK} = \frac{\xi_3 + \sqrt{\xi_3^2 - 4\xi_2}}{2} = \frac{\lambda_3}{2},$$

pues λ_3 es la raíz positiva del polinomio $x^2 - \xi_3 x + \xi_2$. Así pues, $N_3 = \cos 6\pi/17$. Por otra parte,

$$N_5 = \overline{EN_5} - E = \overline{EK} - E = \frac{\xi_3^2 - 2\xi_2}{4} - \frac{\xi_3}{4} = \frac{-\xi_3 + \sqrt{\xi_3^2 - 4\xi_2}}{2} = -\frac{\lambda_5}{2},$$

luego $N_5 = \cos 10\pi/17$.

Levantando perpendiculares por N_3 y N_5 y trazando el círculo de centro 0 y radio 1 obtenemos ζ_3 y ζ_5 (figura 6.3).

Para acabar, bisecamos el ángulo $\widehat{\zeta_3 0 \zeta_5}$ con lo que obtenemos ζ_4 y, uniendo ζ_3 con ζ_4 obtenemos el lado del heptadecágono regular (figura 6.4).

6.3 Construcciones con regla

En las secciones siguientes vamos a estudiar algunas variantes de la constructibilidad con regla y compás que tienen interés por motivos diferentes, pero

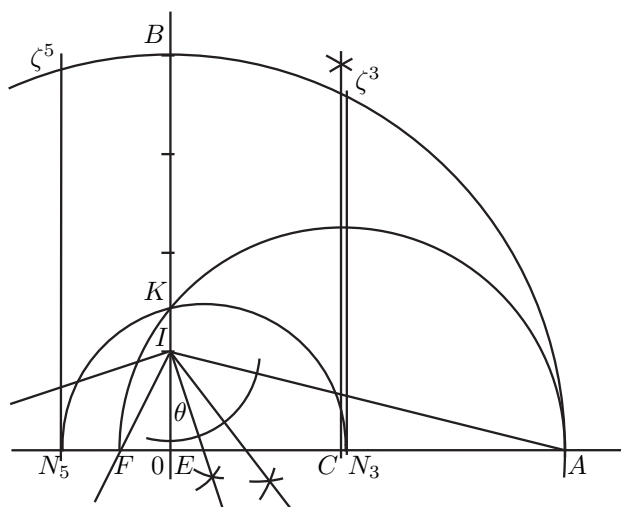
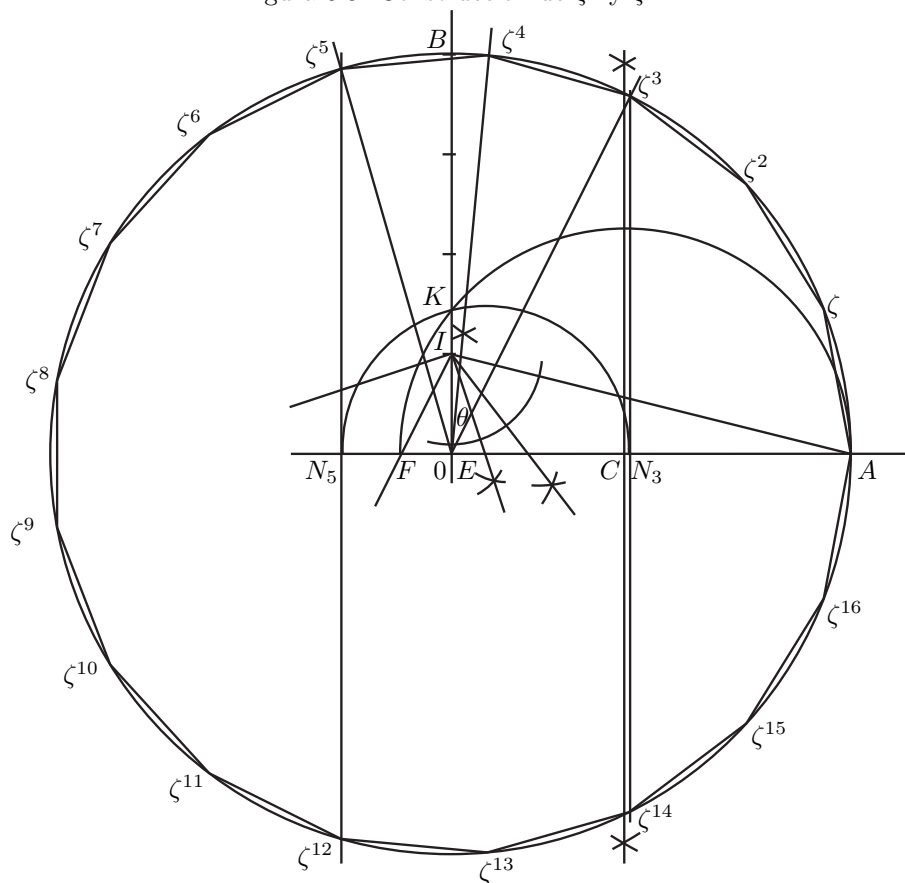
Figura 6.3: Construcción de ζ^3 y ζ^5 .

Figura 6.4: Construcción del heptadecágono regular

antes de entrar en ello resulta útil determinar cuál es exactamente el papel que representa la regla en las construcciones con regla y compás, para lo cual vamos a estudiar los puntos constructibles sólo con regla. Como no vamos a considerar circunferencias, en esta sección podemos trabajar con un plano euclídeo sobre un cuerpo pitagórico R cualesquiera, no necesariamente euclídeo.

Definición 6.11 Sea E un plano euclídeo y Γ un conjunto de puntos de E . Diremos que un punto $P \in E$ es *constructible con regla* a partir de Γ si cumple la definición 6.1 de constructibilidad con regla y compás, pero suprimiendo las condiciones 2) y 3).

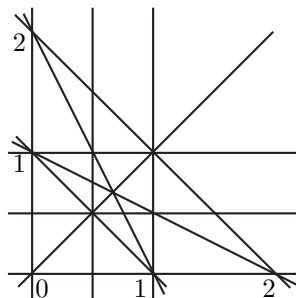
Es fácil ver que si partimos de dos puntos dados en el plano, con la mera ayuda de una regla sólo podemos construir la recta que pasa por ellos, así que necesitamos considerar un conjunto mayor de puntos dados. Dejamos para más adelante la discusión sobre qué puntos se necesitan realmente y de momento vamos a tomar un conjunto que, como vamos a ver, resulta suficiente:

Si R es un cuerpo ordenado pitagórico y $\Gamma \subset C = R^2$, llamaremos $\mathcal{C}_R^r(\Gamma)$ al conjunto de los puntos de C constructibles con regla a partir del conjunto $\Gamma \cup \{(1, 0), (2, 0), (0, 1), (0, 2)\}$, es decir, vamos a considerar que partimos siempre de un plano en el que se han marcado los cuatro puntos indicados. En esta sección “constructible” significará siempre “constructible con regla”.

El teorema siguiente muestra el juego que dan los cuatro puntos que hemos tomado como inicio de nuestras construcciones:

Teorema 6.12 *Los puntos $(0, 0), (1, 0), (0, 1), (1, 1)$ son constructibles, así como los puntos medios de los segmentos que determinan. Toda recta constructible a partir de un conjunto Γ contiene tres puntos constructibles tales que uno es el punto medio de los otros dos.*

DEMOSTRACIÓN: A partir de los cuatro puntos que son constructibles por definición podemos formar seis rectas. Dos de ellas se cortan en $(0, 0)$ y otras dos en $(2/3, 2/3)$. Uniendo estos dos puntos obtenemos la recta $y = x$, que corta a otras dos en $(1/2, 1/2)$ y en $(1, 1)$. Esto nos permite trazar las rectas $x = 1, y = 1$, que nos dan a su vez los puntos $(1/2, 1), (1, 1/2)$, que a su vez nos dan las rectas $x = 1/2$ e $y = 1/2$, que a su vez nos dan los puntos $(1/2, 0)$ y $(0, 1/2)$.



En particular tenemos construidas tres paralelas horizontales y tres verticales. Toda recta constructible a partir de Γ tiene que cortar a uno de los dos grupos de paralelas, las cuales determinan en ella tres puntos constructibles, uno de los cuales es el punto medio de los otros dos. ■

Ahora necesitamos un resultado técnico que resulta fundamental para las construcciones con regla:

Teorema 6.13 Sean A, B, C, D cuatro puntos no colineales tres a tres. Supongamos que las rectas AD y BC se cortan en un punto E y que AC y BD se cortan en un punto F . Entonces AB es paralela a CD si y sólo si EF biseca a uno de los segmentos \overline{AB} y \overline{CD} (y en tal caso los biseca a ambos).

DEMOSTRACIÓN: Podemos suponer que $A - D - E$, pues si $D - A - E$ basta intercambiar los nombres de A y D por una parte y B y C por otra y tenemos las mismas hipótesis con el mismo E y el mismo F pero con $A - D - E$, y si $A - E - D$, basta observar que $B - D - F$, pues entonces intercambiando A con B tenemos las mismas hipótesis con E y F intercambiados, luego pasa a cumplirse $A - D - E$.

En efecto, si $A - E - D$, aplicamos el axioma B4 al triángulo \widehat{ADF} y a la recta BC , lo cual nos da que $A - C - F$, y considerando ahora \widehat{ABF} y la recta CD , llegamos a que $B - D - F$. (Véase la figura siguiente cambiando A con B y E con F .)

Suponemos, pues que $A - D - E$. Entonces el axioma B4 nos da que $B - C - E$. Por lo tanto, C y D están en los lados del ángulo \widehat{AEB} , luego la semirrecta \overrightarrow{EF} está contenida en dicho ángulo, luego tiene que cortar a los segmentos \overline{CD} y \overline{AB} en puntos L y M .

Si AB es paralela a CD , Los triángulos \widehat{AME} y \widehat{DLE} están en posición de Tales, luego son semejantes, al igual que \widehat{BME} y \widehat{CLE} . Por lo tanto

$$\frac{\overline{AM}}{\overline{DL}} = \frac{\overline{EM}}{\overline{EL}} = \frac{\overline{BM}}{\overline{CL}} \Rightarrow \frac{\overline{AM}}{\overline{BM}} = \frac{\overline{DL}}{\overline{CL}}.$$

También es claro que \widehat{AMF} es semejante a \widehat{CLF} , pues ambos tienen los ángulos iguales, e igualmente \widehat{BMF} y \widehat{DLF} . Por lo tanto

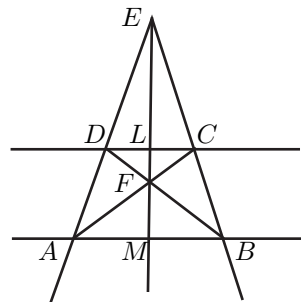
$$\frac{\overline{AM}}{\overline{CL}} = \frac{\overline{FM}}{\overline{FL}} = \frac{\overline{BM}}{\overline{DL}} \Rightarrow \frac{\overline{AM}}{\overline{BM}} = \frac{\overline{CL}}{\overline{DL}}.$$

Esto sólo puede ser si

$$\frac{\overline{AM}}{\overline{BM}} = \frac{\overline{CL}}{\overline{DL}} = 1,$$

luego M y L son los puntos medios de los segmentos respectivos. Recíprocamente, supongamos, por ejemplo, que M es el punto medio de \overline{AB} , sea C' el punto donde la paralela a AB por D corta a BE y sea F' el punto de corte entre BD y AC' (las rectas se cortan, porque BD no separa a E y C' , pero sí a E y A , luego separa a A y C'). Por la parte ya probada, la recta EF' pasa por M , luego es EF , luego $F = F'$, pues ambos son el punto de corte con BD , luego $C = C'$, porque ambos son el punto de corte de AF con EB , luego concluimos que DC es paralela a AB . ■

De aquí deducimos un método para trazar paralelas con regla:



Teorema 6.14 *Sea r una recta constructible a partir de un conjunto de puntos Γ y sea C un punto constructible a partir de Γ exterior a r . Entonces la paralela a r por C es constructible a partir de Γ .*

DEMOSTRACIÓN: Sean $A - M - B$ tres puntos constructibles en r de modo que M sea el punto medio de los otros dos. La recta BC tiene al menos un tercer punto constructible E . Podemos suponer que EM no es paralela a AC . Si lo fuera, tomamos un tercer punto constructible X en EM (que no estará en AC) y entonces AX tiene que cortar a BC en otro punto constructible E' distinto de B , C o E , luego basta tomar E' en lugar de E .

Sea F' el punto de corte de EM y AC (la figura del teorema anterior ilustra también la situación actual). La paralela a r por C corta a AE en un punto D . Tenemos que probar que D es constructible. Si AC fuera paralela a DB , entonces $ABCD$ sería un paralelogramo de centro E , de donde se sigue fácilmente que EM sería paralela a AC , en contra de lo supuesto, luego podemos considerar el punto F donde se cortan DB y AC . Por el teorema anterior la recta EF pasa por M , luego $F = F'$ es constructible, luego también lo es FB y, por consiguiente su punto de corte con AE , que es D . ■

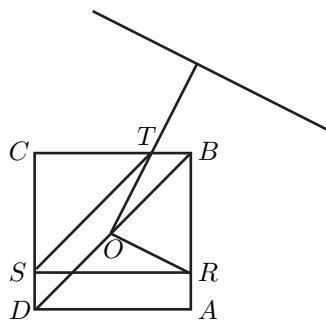
Teorema 6.15 *El punto medio de dos puntos constructibles a partir de un conjunto Γ es constructible a partir de Γ .*

DEMOSTRACIÓN: Sean A y B dos puntos constructibles a partir de Γ . Sabemos que existen puntos constructibles exteriores a la recta AB , luego existe una recta paralela a AB constructible a partir de Γ , en la cual podemos tomar tres puntos constructibles $C - L - D$, donde L es el punto medio. Si AD es paralela a BC o DB es paralela a AC , entonces tenemos un paralelogramo y la paralela a sus lados por L es constructible y corta a \overline{AB} en su punto medio M .

En caso contrario sea E el punto de corte de AD y BC y sea F el punto de corte de AC y BD . Por el teorema 6.13 tenemos que EF corta a \overline{AB} en su punto medio M . ■

Teorema 6.16 *La perpendicular a una recta constructible a partir de un conjunto Γ por un punto constructible a partir de Γ es constructible a partir de Γ .*

DEMOSTRACIÓN: Por el teorema 6.12 sabemos que existe un cuadrado de vértices y centro constructibles. Si la recta es paralela a uno de los lados del cuadrado la conclusión es trivial. En caso contrario, sea R el punto donde la paralela a la recta dada por el centro del cuadrado corta a uno de sus lados, por ejemplo, \overline{AB} . Sea S el punto donde la paralela a AD por R corta a CD . Sea T el punto donde la paralela a la diagonal DB por S corta a CT .



Es claro entonces que $\overline{AR} \equiv \overline{SD} \equiv \overline{BT}$, de donde se sigue que los triángulos \widehat{ARO} y \widehat{BTO} son iguales (pues tienen dos lados iguales que forman un ángulo de 45°). De aquí se sigue fácilmente que el ángulo \widehat{ROT} es recto como \widehat{AOB} , luego la recta OT es perpendicular a la recta dada. La paralela a OT por el punto dado de partida es la recta requerida. ■

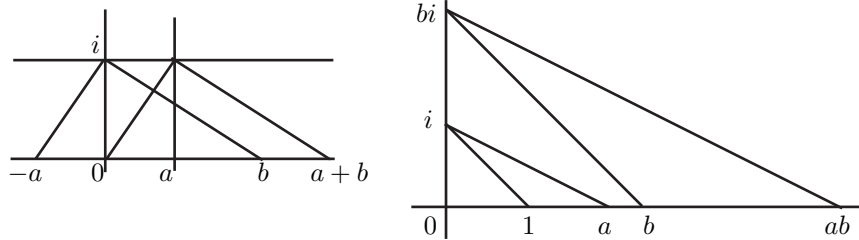
Podemos identificar $C = R^2 = R[i]$ del modo habitual, con lo que $R \subset C$ se identifica con los puntos del eje real. Definimos entonces $R_r(\Gamma) = \mathcal{C}_R^r(\Gamma) \cap R$ como el conjunto de los elementos de R constructibles con regla a partir de Γ . El teorema siguiente prueba que $\mathcal{C}_R^r(\Gamma) = R_r(\Gamma)^2$.

Teorema 6.17 *Un punto $z \in C$ es constructible con regla a partir de un conjunto de puntos Γ si y sólo si lo son su parte real y su parte imaginaria.*

DEMOSTRACIÓN: Si $z = a + bi$ es constructible, trazando paralelas a los ejes por z obtenemos que a y bi son constructibles. Por otra parte, la recta $x + y = 1$ es constructible, luego podemos trazar su paralela por bi , la cual corta al eje real en b , que es, por consiguiente, constructible. El recíproco se prueba análogamente. ■

Teorema 6.18 *Si $\Gamma \subset C$, el conjunto $\mathcal{C}_R^r(\Gamma)$ es un subcuerpo de C y $R_r(\Gamma)$ es un subcuerpo de R , de modo que $\mathcal{C}_R^r(\Gamma) = R_r(\Gamma)[i]$.*

DEMOSTRACIÓN: Basta probar que $R_r(\Gamma)$ es un cuerpo de R , pues entonces es claro que $\mathcal{C}_R^r(\Gamma) = R_r(\Gamma)[i]$. La figura de la izquierda muestra cómo construir el opuesto de un número $a \in R_r(\Gamma)$ y la suma de dos números $a, b \in R_r(\Gamma)$ sin más que trazar las paralelas oportunas:



La figura de la derecha muestra cómo construir el producto de dos números $a, b \in R_r(\Gamma)$. No perdemos generalidad si los suponemos positivos: Unimos a con i , trazamos la paralela a dicha recta que pasa por b , que corta al eje imaginario en bi . Trazamos la recta que une i con a y la paralela a esta recta que pasa por bi . El punto X donde ésta corta al eje real es ab . En efecto, por el teorema de Tales, $x/a = b/1$.

Para construir $1/a$ trazamos la paralela a la recta que une a con i que pasa por 1 , lo que nos da i/a , y luego la paralela a la recta que une i con 1 que pasa por dicho punto, lo que nos da $1/a$. ■

Finalmente podemos determinar cuáles son los puntos constructibles con regla (escribimos R_r y \mathcal{C}_R^r en lugar de $R_r(\Gamma)$ y $\mathcal{C}_R^r(\Gamma)$ cuando $\Gamma = \emptyset$, con lo que los únicos puntos de partida admitidos son $1, 2, i, 2i$):

Teorema 6.19 *Si R es un cuerpo ordenado pitagórico, entonces $R_r = \mathbb{Q}$ y, por consiguiente, $\mathcal{C}_R^r = \mathbb{Q}(i)$.*

DEMOSTRACIÓN: Por el teorema anterior R_r es un cuerpo, que obviamente contiene a \mathbb{Q} . Para probar la inclusión contraria hay que ver que todos los puntos constructibles con regla tienen su parte real y su parte imaginaria en \mathbb{Q} , pero esto es inmediato por el mismo argumento empleado en la primera parte del teorema 6.3, ya que si en una construcción con regla todos los puntos previos a uno dado tienen sus coordenadas en \mathbb{Q} , las rectas que pasan por dos de dichos puntos anteriores admiten una ecuación $Ax + By = C$ con coeficientes en \mathbb{Q} , y al resolver el sistema formado por dos de dichas ecuaciones, vemos que el resultado está en \mathbb{Q} . ■

Así pues, la geometría de la regla es la geometría de \mathbb{Q} .

Construcciones con puntos arbitrarios Vamos a discutir ahora con más detalle la cuestión de la elección de los puntos de partida para las construcciones con regla. Hemos visto que la elección de los puntos $1, 2, i, 2i$ ha sido suficiente para obtener todos los puntos con coordenadas racionales, pero parece una elección excesivamente *ad hoc*. Imaginemos que tenemos un papel en blanco en el que están marcados únicamente los vértices de un cuadrado. Eligiendo adecuadamente el sistema de referencia podemos suponer que son los puntos $0, 1, i, 1+i$. ¿Qué podemos construir a partir de ellos sin más ayuda que una regla?

Si el lector se pone a construir, se convencerá enseguida de que sólo puede construir seis rectas (las prolongaciones de los lados y las diagonales del cuadrado) y un quinto punto (su centro). Aparentemente estamos en un callejón sin salida, pero no es exactamente así. Tenemos a nuestra disposición una técnica que no hemos contemplado en las definiciones de constructibilidad que hemos dado, y es la de tomar puntos arbitrarios.

Se puede objetar que eso “es trampa”, pues un punto arbitrario no puede considerarse construido, pero no es exactamente “una trampa” si con su ayuda obtenemos puntos que son independientes de las elecciones realizadas. Explicamos esta idea con un teorema:

Teorema 6.20 *Dados cuatro puntos A, B, C, M tales que M sea el punto medio de \overline{AB} y C no sea colineal con ellos, es posible construir con regla la paralela a AB que pasa por C con la ayuda de puntos auxiliares arbitrarios.*

DEMOSTRACIÓN: La prueba es una simplificación de la del teorema 6.14. Marcamos un punto arbitrario E en la recta BC , por ejemplo, tal que $B-C-E$, lo cual garantiza que la recta EM corta a AC en un punto F y llamamos D al punto donde se cortan AE y BF . El argumento del teorema 6.14 prueba que la recta CD es paralela a AB . ■

De acuerdo con la definición que hemos dado de puntos y rectas constructibles, ni el punto E empleado en la construcción anterior ni los puntos y rectas subsiguientes (en particular, la paralela buscada) son constructibles a partir de los cuatro puntos dados, pero lo cierto es que el proceso descrito es un algoritmo que termina con la paralela requerida en el papel sin hacer uso de nada más que la regla.

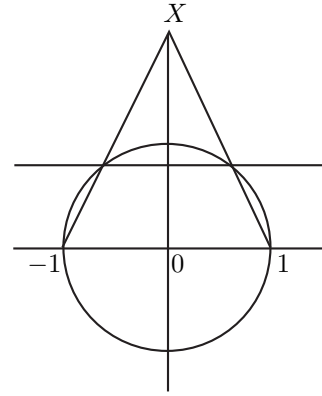
Por lo tanto, si partimos de un papel en el que están marcados los vértices $0, 1, i, 1 + i$ del cuadrado unitario, podemos construir el centro M del cuadrado, y el teorema anterior aplicado a $A = i, B = 1, C = 1 + i$ nos permite trazar la paralela a AB por C , la cual corta a los ejes en 2 y $2i$, y a partir de ahí ya podemos construir todos los puntos con coordenadas racionales según hemos visto.

En suma, en la práctica sólo necesitamos los vértices del cuadrado unitario para construir con regla cualquier punto de coordenadas racionales.

Quizá el lector se plantee si no convendría incorporar en la definición de constructibilidad con regla, o incluso con regla y compás, la técnica de tomar puntos arbitrarios en las construcciones a condición de que la construcción final sea independiente de las elecciones. Sin embargo, no merece la pena, pues una vez hemos construido los puntos 2 y $2i$, las elecciones arbitrarias ya no aportan nada, pues toda recta tiene infinitos puntos constructibles distribuidos densamente en ella, de modo que, si en un momento dado ayuda tomar un punto arbitrario, nada nos impide tomarlo constructible.

Construcciones a partir de una circunferencia Imaginemos ahora que nos dan un papel en el que están marcados únicamente los puntos $0, 1$ y la circunferencia de centro 0 y radio 1 . ¿Qué podemos construir a partir de ahí sin más ayuda que una regla?

Si no admitimos elecciones arbitrarias, es fácil ver que poca cosa: el eje real y, por consiguiente, el punto -1 . Nada más. Ahora bien, eligiendo arbitrariamente un punto interior a la circunferencia y a su vez otro punto arbitrario para realizar la construcción del teorema anterior, obtenemos una recta paralela al eje real que corta a la circunferencia en dos puntos. Al unir estos puntos con 1 y -1 respectivamente obtenemos dos rectas que se cortan en un punto X en el eje imaginario, y al unir X con 0 obtenemos dicho eje, y con él los puntos $\pm i$. A su vez, recurriendo nuevamente a puntos arbitrarios, podemos trazar las rectas $x = 1$ e $y = 1$, que nos dan el punto $1 + i$, a partir del cual sabemos cómo construir los puntos 2 y $2i$.



A partir de aquí ya no son necesarias más elecciones arbitrarias, así que en lugar de incluirlas en la definición siguiente, incluimos más fácilmente los puntos 2 y $2i$ como puntos de partida admitidos:

Definición 6.21 Un punto P de un plano euclídeo (sobre un cuerpo ordenado euclídeo) es *constructible con regla a partir de una circunferencia* respecto de un sistema de referencia ortonormal prefijado si existe una sucesión de puntos P_1, \dots, P_n tal que $P_n = P$ y cada P_i cumple una de las condiciones siguientes:

1. P_i es uno de los puntos $(0, 0)$, $(2, 0)$, $(0, 2)$.
2. P_i es la intersección de dos rectas que pasan por dos pares de puntos anteriores en la sucesión.
3. P_i está en la intersección de una recta que pasa por un par de puntos anteriores en la sucesión y la circunferencia de centro $(0, 0)$ y radio 1.

Según hemos visto, los puntos constructibles según esta definición pueden obtenerse sólo con regla a partir del $(0, 0)$ y el $(0, 1)$ (y usando la circunferencia unitaria) admitiendo elecciones arbitrarias de puntos.

A partir de ahora “constructible” significará “constructible con regla a partir de una circunferencia”.

Es fácil ver, aunque no es todo lo obvio que podría parecer, que todos los resultados que hemos obtenido para construcciones con regla (excepto 6.19) son válidos para construcciones con regla a partir de una circunferencia, pues lo que se puede hacer sólo con regla, se puede también si además tenemos la circunferencia disponible.

Para entender por qué decimos que este hecho no es todo lo obvio que podría parecer consideremos, por ejemplo, la prueba de que la paralela a una recta constructible r por un punto constructible C es constructible. La dificultad es que ahora r y P son constructibles con regla a partir de una circunferencia, lo cual no significa necesariamente que sean constructible con regla. Ahora bien, podemos aplicar el teorema 6.14 tomando como Γ el conjunto formado por dos puntos constructibles de la recta r y el punto C . Entonces r y C son trivialmente constructibles con regla a partir de Γ , luego concluimos que la recta paralela es constructible con regla a partir de Γ , pero como los elementos de Γ son constructibles con regla a partir de una circunferencia, es claro que la recta paralela también lo es.

En suma, sólo hay que tener presente que el hecho de que hayamos probado versiones relativas (a un conjunto de puntos Γ) de los resultados sobre construcciones con regla es fundamental para poder aplicarlos ahora a las construcciones con regla a partir de una circunferencia.

Otro ejemplo, si $a, b \in R$ son constructibles, podemos asegurar que $a + b$ es constructible. En efecto, sabemos que a, b son trivialmente constructibles con regla a partir de $\Gamma = \{a, b\}$, luego 6.18 nos da que $a + b$ es constructible con regla a partir de Γ , de donde se sigue que $a + b$ es constructible con regla a partir de una circunferencia.

En general, ahora sabemos que si R es un cuerpo ordenado euclídeo, el conjunto \mathcal{C}_R^* de los puntos constructibles de $C = R^2$ es un subcuerpo del cuerpo \mathcal{C} de todos los puntos constructibles con regla y compás, y que $R_r^* = \mathcal{C}_R^* \cap R$ es un subcuerpo de \mathcal{C}_r , de modo que $\mathcal{C}_R^* = \mathcal{C}_r[i]$. Sin embargo, el teorema 6.19 no es cierto para estas construcciones. Por el contrario:

Teorema 6.22 (Poncelet-Steiner) *Los puntos constructibles con regla a partir de una circunferencia son los puntos constructibles con regla y compás.*

DEMOSTRACIÓN: Basta probar que $R_r^* \subset \mathcal{C}_r$ es un cuerpo euclídeo, pues entonces tiene que darse la igualdad $R_r^* = \mathcal{C}_r$, ya que \mathcal{C}_r es el menor cuerpo ordenado euclídeo. Para ello observamos que si $a \in R_r^*$ es positivo, entonces

$$\sqrt{a} = \sqrt{\left(\frac{a+1}{2}\right)^2 - \left(\frac{a-1}{2}\right)^2} = \frac{a+1}{2} \sqrt{1 - \left(\frac{a-1}{a+1}\right)^2}.$$

Sabemos que $(a-1)/(a+1) \in R_r^*$ (porque éste es un cuerpo) y podemos trazar la recta perpendicular al eje real que pasa por dicho punto, la cual corta a la circunferencia unidad en

$$\sqrt{1 - \left(\frac{a-1}{a+1}\right)^2} \in R_r^*,$$

y el producto por $(a+1)/2$ sigue en el cuerpo, luego $\sqrt{a} \in R_r^*$. ■

Esto significa que en toda construcción con regla y compás (a partir de los puntos 0 y 1) no es necesario usar el compás más que una vez, para trazar la circunferencia de centro 0 y radio 1, y a partir de ahí basta usar la regla.

En particular, podríamos definir lo que podríamos llamar “construcciones con regla y compás oxidado”, consistentes en que el compás tiene una abertura fija de una unidad, de modo que sólo pueden trazarse circunferencias de radio 1, pero el teorema anterior prueba que los puntos constructibles en esas condiciones son, de hecho, todos los puntos constructibles con regla y compás.

6.4 Construcciones con compás

A la vista de los resultados de la sección anterior, resulta natural plantearse qué puntos pueden construirse sólo con el compás. La respuesta es la siguiente:

Teorema 6.23 (Mohr-Mascheroni) *Los puntos constructibles con compás son los mismos que los constructibles con regla y compás.*

Para probarlo basta demostrar que con el compás es posible construir la intersección de dos rectas que pasan por sendos pares de puntos dados y la intersección de una recta que pasa por dos puntos dados con una circunferencia de centro dado y radio igual a la distancia de dos puntos dados.

Empezamos observando lo siguiente:

El punto simétrico de un punto dado respecto de una recta que pasa por dos puntos dados es constructible

Basta trazar las circunferencias con centros en los puntos de la recta y que pasan por el punto dado. El otro punto de corte será el simétrico que buscamos.

Esto basta para resolver la mitad del problema: Dados dos puntos de una recta y el centro y el radio de una circunferencia, para calcular los puntos de corte calculamos el simétrico del centro de la circunferencia respecto de la recta y trazamos la circunferencia del mismo radio y centro dicho punto simétrico. Los puntos de corte de las dos circunferencias son los puntos de intersección buscados.

La parte complicada es calcular la intersección de dos rectas dadas. Para ello probamos dos resultados previos:

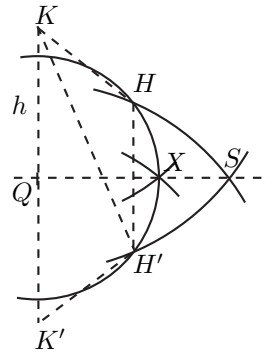
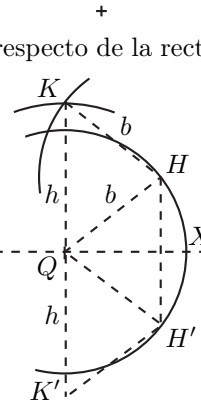
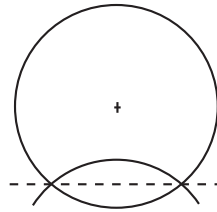
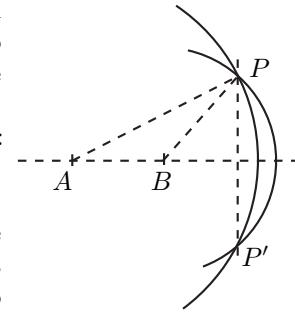
La suma de dos segmentos de extremos dados es constructible.

Pongamos que uno de los segmentos es \overline{PQ} y que el otro tiene longitud b . Empezamos trazando la circunferencia de centro Q y radio b . Sobre ella, seleccionamos un punto arbitrario H y calculamos su simétrico H' respecto de la recta PQ . Llamamos h a la longitud del segmento $\overline{HH'}$.

La circunferencia de centro Q y radio h corta a la circunferencia de centro H y radio b en un punto K . Sea K' su simétrico respecto de PQ . Se forma así un trapezoide que se descompone en tres triángulos isósceles iguales. Si llamamos d a la diagonal $\overline{KH'}$, podemos formar un triángulo rectángulo de hipotenusa d , un cateto $3h/2$ y otro cateto igual a la altura del triángulo isósceles, cuyo cuadrado es $b^2 - (h/2)^2$. El teorema de Pitágoras nos da que

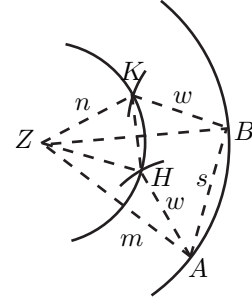
$$d^2 = 9h^2/4 + b^2 - h^2/4 = 2h^2 + b^2.$$

La figura muestra el punto X que queremos calcular, es decir, el punto que está sobre la recta PQ a una distancia b de Q . Si llamamos $x = \overline{KX}$, el triángulo rectángulo \overline{KXQ} nos da que $x^2 = b^2 + h^2$, luego $d^2 = x^2 + h^2$. Por lo tanto, los arcos de centros K y K' y radio d se cortan en un punto S que forma un triángulo rectángulo \overline{KQS} de hipotenusa d y un cateto igual a h , luego el otro cateto \overline{QS} mide x . Esto nos permite trazar los arcos de centros K y K' y radio x , que determinan un punto X que forma el triángulo rectángulo \overline{KQX} de hipotenusa x y un cateto igual a h , luego el otro cateto \overline{QX} mide b , lo que significa que X es el punto que buscábamos.



Dados tres segmentos de longitudes m, n, s , es posible construir un segmento de longitud $(n/m)s$.

Trazamos dos circunferencias concéntricas de centro Z y radios m y n . En la circunferencia de radio m trazamos un punto arbitrario A y luego marcamos otro B a distancia S . Esto presupone que s es menor que el diámetro de la circunferencia, es decir, que $2m$. Si no fuera así, podemos sustituir m y n por múltiplos suficientemente grandes, que son constructibles porque hemos probado que la suma de segmentos constructibles es constructible.



A continuación trazamos un arco desde A con abertura arbitraria w para que corte a la otra circunferencia en un punto H , y con la misma abertura trazamos un arco desde B que corte a la otra circunferencia en un punto K . Ahora observamos que \widehat{ZKB} y \widehat{ZHA} son congruentes, luego $\widehat{AZH} \equiv \widehat{BZK}$, luego $\widehat{AZB} \equiv \widehat{HZK}$, porque el segundo ángulo resulta de restarle al primero \widehat{AZH} y luego sumarle el ángulo igual \widehat{BZH} . Por consiguiente, los triángulos \widehat{AZB} y \widehat{HZK} son similares (son isósceles con el mismo ángulo entre los lados iguales). Concluimos que

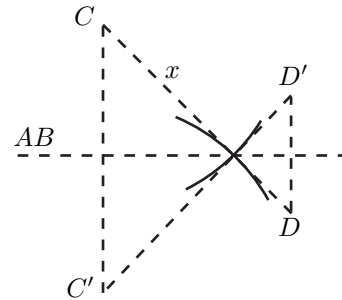
$$\frac{\overline{HK}}{n} = \frac{s}{m},$$

luego \overline{HK} es el segmento buscado.

Finalmente consideramos dos rectas AB y CD y vamos a construir su punto de corte. En primer lugar trazamos los simétricos de C y D respecto de AB . Se forman así dos triángulos semejantes, de modo que

$$\frac{x}{\overline{CC'}} = \frac{\overline{CD} - x}{\overline{DD'}},$$

de donde $x = \frac{\overline{CC'}}{\overline{CC'} + \overline{DD'}} \overline{CD}$. Por los resultados previos sabemos que x es constructible, y las circunferencias de centros C, C' y radio x se cortan en el punto de intersección de las rectas. ■



6.5 Construcciones con regla marcada

En la sección [IGE 6.5] vimos que el uso de una regla marcada permite llevar a cabo construcciones que son imposibles con una regla “clásica” sin marcas. Recordemos que una regla marcada es simplemente una regla en la que se han señalado dos marcas, la distancia entre las cuales podemos tomar como unidad de medida.

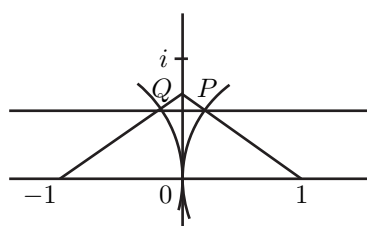
A una regla marcada se le pueden dar muchos usos. Para presentar el más elemental veamos qué podemos construir sin más ayuda que una regla marcada partiendo únicamente de los puntos 0 y 1.

Para empezar, usando la regla “al modo clásico”, podemos trazar el eje real uniéndolo los dos puntos dados. El primer uso específico de una regla marcada es el siguiente:

Uso 1 *Dada una recta y un punto en ella, podemos construir los dos puntos situados en la recta que distan una unidad del punto dado.*

Basta poner una marca sobre el punto dado y señalar el punto donde cae la otra marca.

Al aplicar esto al eje real y el punto 0, la regla marcada nos permite construir el 1 (que ya lo teníamos) y el -1 . De hecho, siguiendo así podemos construir cualquier número entero, pero de momento nos basta con considerar los puntos -1 , 0 y 1. Como 0 es el punto medio de los otros dos, usando puntos arbitrarios podemos trazar una paralela al eje real por un punto arbitrario, que debemos elegir “a ojo” cuidando de que diste del eje real menos de una unidad.



Ahora podemos hacer un segundo uso de nuestra regla marcada: situamos una marca sobre el 1 y giramos la regla hasta que la otra marca se encuentre sobre la recta que hemos trazado, lo que determina un punto P (en realidad, así podemos determinar dos puntos distintos, pero llamamos P al más cercano al eje imaginario). Repitiendo el proceso con -1 obtenemos un punto Q , y prolongando las dos rectas que pasan por 1 y P y por -1 y Q , obtenemos un punto del eje imaginario. Uniéndolo con 0 obtenemos el eje, y usando de nuevo las marcas obtenemos el punto i y a su vez $2i$.

Antes de seguir nuestro análisis destacamos en qué ha consistido el segundo uso de la regla marcada:

Uso 2 *Dada una recta y un punto exterior a ella, podemos construir (si existen) los puntos de la recta que distan una unidad del punto dado.*

Notemos que tanto el primer como el segundo uso que hemos dado a la regla marcada pueden conseguirse igualmente con un compás. De hecho, basta con un “compás oxidado”. Por lo tanto, hasta ahora no hemos hecho nada que no pueda hacerse también con una regla no marcada y un compás.

Ahora que ya contamos con los puntos 1, 2, i , $2i$, el mismo argumento empleado con la constructibilidad con regla respecto de una circunferencia muestra que todos los resultados sobre constructibilidad con regla relativa a un conjunto de puntos arbitrario Γ se traducen en los mismos resultados para la constructibilidad con regla marcada, así, podemos construir (ya sin elecciones arbitrarias) rectas paralelas, perpendiculares, etc. Y las coordenadas de puntos constructibles con regla marcada forman un subcuerpo de $C = R^2$.

Ahora observamos que la prueba del teorema de Poncelet-Steiner 6.22, es decir, la prueba de que todo número positivo constructible tiene raíz cuadrada constructible, vale igualmente para la constructibilidad con regla marcada, pues lo único que requiere es, dado un número $0 < x < 1$, levantar la perpendicular por x al eje real y determinar en ella un punto que diste de 0 una unidad, lo cual es el segundo uso que hemos descrito de la regla marcada.

La conclusión es que los puntos constructibles con los dos usos que hemos descrito de la regla marcada son exactamente los constructibles con regla y compás. Sin embargo, la regla marcada admite otros usos que superan la capacidad de la regla y el compás. Veamos uno de ellos (la *neusis* recta-recta):

Uso 3 *Dadas dos rectas y un punto cualquiera P , podemos construir (si existen) los puntos A y B situados uno en cada recta, colineales con P y que distan entre sí una unidad.*

Para ello situamos una marca sobre una de las rectas y movemos la regla de modo que esté siempre sobre P y la marca no deje de estar sobre la recta, hasta que la otra marca se encuentre sobre la otra recta. Los pares de puntos posibles A y B pueden ser dos, uno o ninguno.

No hemos dicho nada hasta ahora sobre el cuerpo de escalares R que estamos considerando. Para los dos primeros usos de la regla marcada tenemos que exigir que R sea euclídeo, para no encontrarnos con que un punto que “debería existir” en realidad no existe en R^2 . (Por ejemplo, no podemos probar que todo número positivo constructible con regla marcada tiene raíz cuadrada constructible si resulta que no tiene raíz cuadrada en R .) Sin embargo, esto no es suficiente para que el tercer uso pueda aplicarse sin inconvenientes por las posibles deficiencias algebraicas de R . Para evitar problemas de este tipo consideramos cuerpos realmente cerrados [Al 5.71], como \mathbb{R} o como el cuerpo \mathbb{R}_a de los números reales algebraicos (sobre \mathbb{Q}).

En lugar de dar una definición precisa de los puntos constructibles con estos tres usos de las reglas marcadas, vamos a sustituir los dos primeros usos por el compás, pues ya sabemos que es equivalente y la definición se simplifica:

Definición 6.24 Un punto P de un plano euclídeo E sobre un cuerpo ordenado realmente cerrado es *constructible con regla marcada* a partir de un conjunto de puntos $\Gamma \subset E$ si existe una sucesión de puntos P_1, \dots, P_n tal que $P_n = P$ y cada P_i cumple una de las condiciones siguientes:

1. $P_i \in \Gamma$.
2. P_i es la intersección de dos rectas que pasan cada una por dos puntos previos.
3. P_i es está en la intersección de una recta que pasa por dos puntos previos y una circunferencia cuyo centro es un punto previo y cuyo radio es la distancia entre dos puntos previos.

4. P_i está en la intersección de dos circunferencias con centro en dos puntos previos y radios que sean las distancias entre dos pares de puntos previos.
5. Existen dos rectas que pasan cada una por un par de puntos previos de la sucesión y un punto previo P de la sucesión de modo que P_i es uno de los dos puntos A y B situados uno en cada recta, colineales con P y distan entre sí una unidad.

Si R es un cuerpo realmente cerrado y $\Gamma \subset C = R^2$, llamaremos $\mathcal{C}_R^{rm}(\Gamma)$ al cuerpo formado por los puntos de C constructibles con regla marcada a partir del conjunto $\Gamma \cup \{(0, 0), (1, 0)\}$. A su vez, $R_{rm}(\Gamma) = \mathcal{C}_R^{rm}(\Gamma) \cap R$.

El punto 5) describe la *neusis*, y hemos visto que los puntos 3) y 4) (que describen los usos del compás) pueden sustituirse por descripciones de los usos 1) y 2) de la regla marcada, aunque entonces hay que tomar como puntos de partida los puntos $(1, 0), (2, 0), (0, 1), (0, 2)$ para evitar la necesidad de elecciones arbitrarias.

Hemos exigido que el cuerpo R sea realmente cerrado con la intención de que no “nos falten puntos” en R a la hora de usar la regla marcada, pero en principio no tenemos garantías de que esto resuelva el posible problema. Tenemos que justificar que al aplicar la *neusis*, o lo que ahora es la regla 5 de la definición anterior, al partir de puntos con coordenadas en R obtenemos necesariamente puntos algebraicos sobre C , por lo que existirán o no en función de la situación geométrica, pero no pasarán de no existir a existir por sustituir R por un cuerpo mayor, ya que C es algebraicamente cerrado.

Para ello analizamos una situación ligeramente más general que la descrita en la regla 5. Sea E un plano afín sobre el cuerpo realmente cerrado R en el que hemos fijado un sistema de referencia ortonormal. Sea $K \subset R$ un cuerpo cualquiera, consideremos dos rectas r y s en E que pasen cada una por dos pares de puntos con coordenadas en K , sea P otro punto con coordenadas en K y sean A y B dos puntos de C situados cada uno en una de las rectas dadas y cuya distancia sea un $d \in K$ (en nuestro contexto sólo nos interesa el caso $d = 1$, pero ahora conviene razonar en general). Vamos a probar que las coordenadas de A y B son necesariamente algebraicas sobre K .

En primer lugar observamos que, como O tiene sus coordenadas $(u, v) \in K^2$, si cambiamos el sistema de referencia para que su origen sea O , las coordenadas de un punto cualquiera se transforman de (x, y) a $(x - u, y - v)$, por lo que las nuevas estarán en una extensión algebraica de K si y sólo si las antiguas están en esa misma extensión. Además, los puntos que antes tenían coordenadas en K siguen teniéndolas. Por consiguiente, no perdemos generalidad si suponemos que O tiene coordenadas $(0, 0)$.

Si r pasa por puntos con coordenadas $(a, b), (c, d) \in K$, un vector director es $(u, v) = (c - a, d - b) \in K^2$, y un vector ortogonal es $(-v, u) \in K^2$. La recta perpendicular a r y que pasa por O tiene ecuación $-vx + uy = 0$.

Como r también admite una ecuación con coeficientes en K , la intersección de ambas, es decir, el pie X de la perpendicular a r por O tiene sus coordenadas

$(x, y) \in K^2$. Ahora cambiamos la base del sistema de referencia, pasando a

$$\left(-\frac{y}{\sqrt{x^2 + y^2}}, \frac{x}{\sqrt{x^2 + y^2}} \right), \quad \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right).$$

Los puntos que antes tenían sus coordenadas en K^2 ahora las tienen en K'^2 , donde $K' = K(\sqrt{x^2 + y^2})$ (que puede ser el propio K o una extensión cuadrática de K). Si las coordenadas de los puntos A y B respecto del nuevo sistema de referencia están en una extensión algebraica K'' de K (que es lo que vamos a probar), las coordenadas originales también lo estarán (pues son combinaciones lineales de las nuevas con coeficientes en K').

Respecto del nuevo sistema de referencia, las coordenadas de X son de la forma $(0, b)$, donde $b = \sqrt{x^2 + y^2} \in K'$, luego r tiene ecuación $y = b$, para cierto $b \in K'$.

Ahora vamos a determinar el lugar geométrico de los puntos A tales que la recta OA corta a la recta $y = b$ en un punto B cuya distancia a A es igual a d . Si uno de tales puntos A tiene coordenadas (x, y) y la recta OA tiene ecuación $y = ax$ (con $a \in K'$), el punto B es $(b/a, b)$, y para que la distancia de A a B sea d se tiene que cumplir que

$$(x - b/a)^2 + (y - b)^2 = d^2.$$

Pero (x, y) también cumple $y = ax$, luego

$$(x - bx/y)^2 + (y - b)^2 = d^2 \quad \Rightarrow \quad x^2 \frac{(y - b)^2}{y^2} + (y - b)^2 = d^2,$$

luego

$$(x^2 + y^2)(y - b)^2 = d^2 y^2.$$

Los puntos (x, y) que cumplen esta ecuación forman una curva llamada *concoide de Nicomedes*, y es el primer ejemplo no trivial que vemos —si consideramos ejemplos triviales las rectas y las circunferencias— de *curva algebraica*, es decir, de curva definida por una ecuación polinómica.

A veces es más práctica la ecuación de la concoide en coordenadas polares. Si sustituimos $x = \rho \cos \theta$, $y = \rho \sen \theta$ queda

$$\rho^2(\rho \sen \theta - b)^2 = d^2 \rho^2 \sen^2 \theta \quad \Rightarrow \quad \left(\rho - \frac{b}{\sen \theta} \right)^2 = d^2$$

de donde²

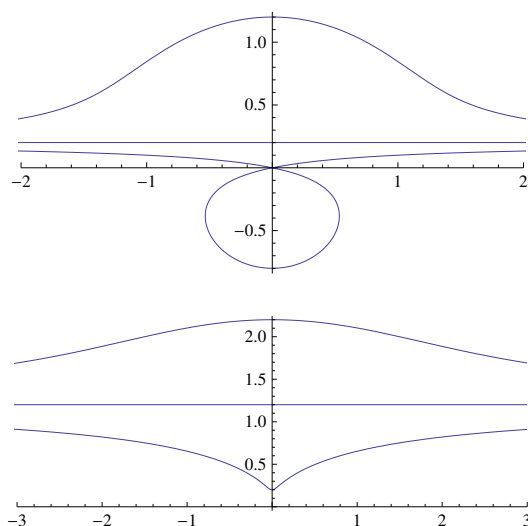
$$\rho = \frac{b}{\sen \theta} + d.$$

²En principio habría que considerar también la ecuación con $-d$ en lugar de $+d$, pero ambas describen los mismos puntos, pues la expresión

$$(\rho \cos \theta, \rho \sen \theta) = \left(\left(\frac{b}{\sen \theta} + d \right) \cos \theta, \left(\frac{b}{\sen \theta} + d \right) \sen \theta \right)$$

cambia el signo de d cuando θ se cambia por $-\theta$.

Las gráficas siguientes muestran dos concoides con $d = 1$, la primera con $b = 0.2$ y la segunda con $b = 1.2$. En ambas figuras está dibujada también la recta $y = b$.



Vemos que cada concoide tiene dos ramas, una se recorre cuando θ varía en $]0, \pi[$ y la otra cuando varía en $]\pi, 2\pi[$. Los ángulos $\theta = 0, \pi, 2\pi$ cumplen $\sin \theta = 0$ y no dan lugar a ningún punto de la concoide. Si trazamos cualquier recta que pase por el origen, corta a la concoide en dos puntos A que distan una unidad del punto B donde OA corta a la recta $y = b$.

Volviendo a nuestro argumento, si la otra recta tiene s ecuación $px + qy = u$, con $p, q, u \in K'$, los puntos A y B determinados por la regla marcada colineales con O son un punto de corte A entre s y la concoide, por lo que sus coordenadas (x, y) cumplirán

$$(x^2 + y^2)(y - b)^2 = y^2, \quad px + qy = u,$$

y el punto B donde OA corta a la recta $y = b$, cuyas coordenadas son $(bx/y, b)$.

Al despejar x o y en $px + qy = u$ (según sea $p \neq 0$ o $q \neq 0$) y sustituir en la ecuación de la concoide, obtenemos un polinomio de grado³ 4 en x o en y , luego una coordenada de A está en un cuerpo K'' tal que $|K'' : K'| \leq 4$ y la otra depende linealmente de ella con coeficientes en K' , luego está en el mismo cuerpo K'' . Las coordenadas de B son $(bx/y, b)$, luego también están en K'' , y lo mismo vale para las coordenadas de A y B en el sistema de referencia inicial.

Resumimos lo que hemos demostrado:

³Notemos que no se cancelan los términos de grado 4, pues éstos son $x^2y^2 + y^4$, que se convierten en $(\lambda^2 + 1)y^4$ o bien en $(\lambda^2 + \lambda^4)x^4$ según que la sustitución sea $x = \lambda y + \mu$ o bien $y = \lambda x + \mu$, con $\lambda \in R$ no nulo, luego $\lambda^2 + 1, \lambda^2 + \lambda^4 > 0$. No obstante, dicho polinomio no tiene por qué ser irreducible.

Teorema 6.25 *Si en una construcción con regla marcada un punto P se obtiene por la condición 5 a partir de puntos anteriores cuyas coordenadas están en un cuerpo $K \subset R$, entonces existen cuerpos $K \subset K' \subset K'' \subset R$ de modo que $|K' : K| \leq 2$ y $|K'' : K'| \leq 4$ y las coordenadas de P están en K'' .*

Combinando esto con la prueba del teorema 6.3 (que cubre los casos de los puntos obtenidos por las condiciones relativas a la regla no marcada y el compás), ahora es inmediato el teorema siguiente:

Teorema 6.26 *Si un punto P es constructible con regla marcada a partir de un conjunto Γ de puntos cuyas coordenadas estén en un cuerpo $K \subset R$, entonces existe una cadena de extensiones*

$$K = K_0 \subset K_1 \subset \cdots \subset K_n \subset R$$

de modo que $|K_{i+1} : K_i| \leq 4$ y las coordenadas (α_1, α_2) de P están en K_n . En particular, $|K(\alpha_i) : K|$ sólo puede ser divisible entre los primos 2 y 3.

Por ejemplo, ahora es claro que el endecágono regular no es constructible con regla marcada, pues si $\eta = \cos(2\pi/11)$, entonces $|\mathbb{Q}(\eta) : \mathbb{Q}| = 5$.

Vamos a ver que las fórmulas de Cardano-Ferrari nos permiten probar el recíproco, con lo que obtenemos una caracterización algebraica de la constructibilidad con regla marcada. Para ello necesitamos dos observaciones previas:

El problema [IGE 6.3] muestra que es posible trisecar cualquier ángulo con una regla marcada. En particular, el ángulo de 20° es constructible con regla marcada, luego el de 40° también, luego el eneágono regular es constructible con regla marcada. De hecho, la construcción de [IGE 6.3] (partiendo de un ángulo de 60° , que sabemos construir) se prolonga de forma obvia para acabar construyendo un eneágono regular. En particular, $\cos(2\pi/18)$ es un ejemplo concreto de número constructible con regla marcada que no es constructible con regla y compás. De hecho, sabemos que es algebraico de grado 3 sobre \mathbb{Q} .

Por otro lado, la prueba del teorema [IGE 6.17] demuestra, un poco más en general, el teorema siguiente:

Teorema 6.27 *Si $a \in \mathcal{C}_R^{rm}(\Gamma)$, sus raíces cúbicas están en $\mathcal{C}_R^{rm}(\Gamma)$.*

En particular, es posible duplicar un cubo con regla y compás (pues esto equivale a construir $\sqrt[3]{2}$, que es otro ejemplo de número constructible con regla marcada y no con regla y compás).

Volviendo al teorema 6.26 (y al problema de encontrarle un recíproco), supongamos que K/k es una extensión de grado 3 de subcuerpos de \mathbb{R} . Entonces $K = k(\alpha)$, para cierto $\alpha \in K$, que es raíz de un polinomio de grado 3 con coeficientes en k . Si el polinomio tiene discriminante $D > 0$, el teorema [Al 7.19] nos da una expresión para α , a partir de la cual podemos formar una cadena de cuerpos

$$k = K_0 \subset K_1 \subset K_2 \subset K_3 \subset \mathbb{R}$$

con $K_1 = k(\sqrt{D})$, $K_2 = K_1(\sqrt[3]{-q/2 + \sqrt{D}})$, $K_3 = K_2(\sqrt[3]{-q/2 - \sqrt{D}})$. Así $\alpha \in K_3$ y cada cuerpo resulta de adjuntar al anterior una raíz cuadrada o cúbica.

Si por el contrario $D < 0$, el teorema [Al 7.20] nos da una cadena

$$k = K_0 \subset K_1 \subset K_2 \subset \mathbb{R},$$

donde $K_1 = k(\sqrt{-p/3})$ (y la raíz es real porque $D < 0$ implica $p < 0$) y $K_2 = K_1(\cos(\theta/3))$, para cierto θ tal que $\cos \theta \in k$. De nuevo, $\alpha \in K_2$.

Si K/k tiene grado 4, entonces según la fórmula de Ferrari podemos formar una cadena $k = K_0 \subset \dots \subset K_m \subset \mathbb{R}$ en una de las condiciones anteriores tal que $P \in K_m$ (siempre podemos tomar como P una raíz real de la cúbica auxiliar). Si α no está en K_m , entonces es raíz de una de las ecuaciones [Al (7.20)], luego la segunda raíz es también real, luego Q, R tienen que ser reales, y podemos prolongar K_m con $K_{m+1} = K_m(\sqrt{P^2 - r})$ y $K_{m+2} = K_m(\sqrt{\Delta^*}) \subset \mathbb{R}$, donde Δ^* es el discriminante de la ecuación correspondiente de [Al (7.20)]. Con esto tenemos probada una parte del teorema siguiente:

Teorema 6.28 *Sea $K \subset \mathbb{R}$ un cuerpo y $\alpha \in \mathbb{R}$. Las afirmaciones siguientes son equivalentes:*

1. α es constructible con regla marcada a partir de puntos con coordenadas en K .
2. Existe una cadena de cuerpos $K = K_0 \subset K_1 \subset \dots \subset K_n \subset \mathbb{R}$ de modo que $\alpha \in K_n$ y $|K_{i+1} : K_i| \leq 4$.
3. Existe una cadena de cuerpos $K = K_0 \subset K_1 \subset \dots \subset K_n \subset \mathbb{R}$ de modo que $\alpha \in K_n$ y $K_{i+1} = K_i(\beta_i)$, donde β_i está en uno de los casos siguientes:
 - (a) $\beta_i = \sqrt{a}$, con $a \in K_i$, $a > 0$.
 - (b) $\beta_i = \sqrt[3]{a}$, con $a \in K_i$.
 - (c) $\beta_i = \cos(\theta/3)$, para cierto θ tal que $\cos \theta \in K_i$.

DEMOSTRACIÓN: 1) \Rightarrow 2) es el teorema 6.26. 2) \Rightarrow 3) es consecuencia inmediata de las observaciones precedentes al enunciado. Por último, 3) \Rightarrow 1) se deduce del hecho de que las raíces cuadradas y cúbicas son constructibles con regla marcada, así como que todo ángulo se puede trisecar con regla marcada. ■

El problema [IGE 6.4] muestra que es posible construir un heptágono regular con regla marcada, con lo que concluimos que todos los polígonos de hasta 10 lados son constructibles (pero ya hemos observado que el de 11 lados no lo es). Ahora podemos caracterizar los polígonos regulares constructibles con regla marcada:

Teorema 6.29 *El polígono regular de n lados es constructible con regla marcada si y sólo si $\phi(n) = 2^a 3^b$.*

DEMOSTRACIÓN: El polígono regular de n lados es constructible si y sólo si lo es $\eta = \cos(2\pi/n)$. Sea $\zeta = \cos(2\pi/n) + i \operatorname{sen}(2\pi/n)$. Entonces $\mathbb{Q}(\eta) \subset \mathbb{Q}(\zeta)$, y la extensión ciclotómica $\mathbb{Q}(\zeta)/\mathbb{Q}$ es abeliana de grado $\phi(n)$. Además el cuerpo $\mathbb{Q}(\eta) = \mathbb{Q}(\zeta) \cap \mathbb{R}$ tiene grado $\phi(n)/2$ sobre \mathbb{Q} . Si η es constructible, el teorema anterior nos da que $\phi(n)/2 = 2^a 3^b$, luego $\phi(n) = 2^{a+1} 3^b$.

Recíprocamente, si $\phi(n) = 2^a 3^b$, entonces, como $G(\mathbb{Q}(\eta)/\mathbb{Q})$ es abeliano, posee subgrupos de todos los órdenes que dividen a su orden, luego podemos formar una cadena de subgrupos

$$G(\mathbb{Q}(\eta)/\mathbb{Q}) = G_0 \triangleright G_1 \triangleright \cdots \triangleright G_n = 1,$$

de modo que $|G_i : G_{i+1}| = 2, 3$. Por el teorema de Galois existe una cadena de cuerpos $\mathbb{Q} = K_0 \subset \cdots \subset K_n = \mathbb{Q}(\eta)$ tal que $|K_{i+1} : K_i| = 2, 3$, y el teorema anterior nos da entonces que η es constructible con regla marcada. ■

Teniendo en cuenta las propiedades de la función de Euler, es inmediato que la caracterización del teorema anterior equivale a la siguiente:

Teorema 6.30 *El polígono regular de n lados es constructible con regla marcada si y sólo si $n = 2^a 3^b p_1 \cdots p_r$, donde los p_i son primos distintos dos a dos de la forma $2^u 3^v + 1$.*

Los primos indicados en el teorema se llaman *primos de Pierpont*. Como en el caso de los primos de Fermat o de Mersenne, no se sabe si hay infinitos, pero, al igual que sucede con los primos de Mersenne y al contrario que con los primos de Fermat, se conocen muchos ejemplos de primos de Pierpont, incluso de tamaños muy grandes. Los primeros primos de Pierpont son:

$$2, 3, 5, 7, 13, 17, 19, 37, 73, 97, 109, 163, 193, 257, 433, 487, \dots$$

Así pues, como ya habíamos observado, el menor polígono regular no constructible con regla marcada es el endecágono.

En la prueba del teorema 6.29 hemos usado que las extensiones ciclotómicas son abelianas para obtener una serie del grupo de Galois que, a través de la teoría de Galois, nos diera la cadena de cuerpos que requiere el teorema 6.28. Sin embargo, hay un teorema nada trivial de la teoría de grupos que hace esta hipótesis innecesaria, y nos da una caracterización mucho más simple de la constructibilidad con regla marcada.

Se trata del teorema de Burnside [TG 6.40], que afirma que todo grupo finito cuyo orden es divisible a lo sumo entre dos primos es resoluble. Aceptando esto podemos probar:

Teorema 6.31 *Un número real α es constructible con regla marcada si y sólo si el cuerpo de escisión K de $\operatorname{polmín}(\alpha, \mathbb{Q})$ cumple que $|K : \mathbb{Q}| = 2^a 3^b$.*

DEMOSTRACIÓN: Si α es constructible, el teorema 6.28 nos da que $|\mathbb{Q}(\alpha) : \mathbb{Q}|$ es divisible únicamente entre 2 y 3, y lo mismo vale para todos sus conjugados (aunque no sean reales), luego lo mismo vale para $|K : \mathbb{Q}|$.

Recíprocamente, si $|K : \mathbb{Q}| = 2^a 3^b$, entonces, por el teorema de Burnside, el grupo $G = G(K/\mathbb{Q})$ es resoluble, luego tiene una serie cíclica cuyos factores tienen todos orden 2 o 3. El teorema de Galois nos da una cadena de cuerpos intermedios

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_n = K$$

de modo que $|K_{i+1} : K_i| = 2, 3$ y lo mismo vale para la cadena

$$\mathbb{Q} = K_0 \cap \mathbb{R} \subset K_1 \cap \mathbb{R} \subset \cdots \subset K_n \cap \mathbb{R}$$

y $\alpha \in K_n \cap \mathbb{R}$, luego por el teorema 6.28 es constructible con regla marcada. ■

6.6 Construcciones con regla marcada y compás

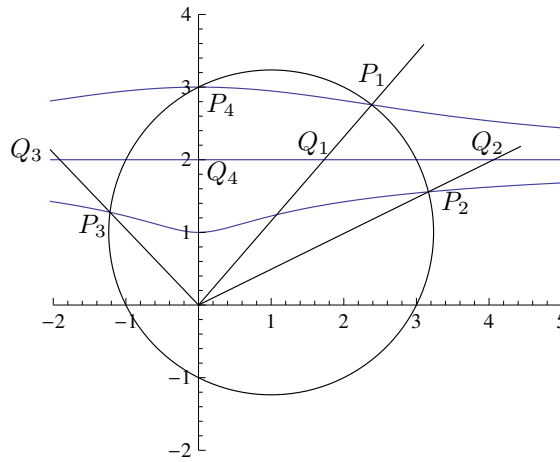
Aunque hemos admitido el uso del compás en la definición de construcción con regla marcada, hemos visto que no es esencial, en el sentido de que (a costa de admitir unos pocos puntos más de partida para evitar la necesidad de construcciones con elecciones arbitrarias de puntos) todo lo que podemos construir con los usos que le hemos admitido a una regla marcada y un compás, pueden construirse sólo con la regla marcada. Si sólo admitimos la regla marcada, es decir, si no podemos dibujar circunferencias (aunque los puntos en que las circunferencias que podríamos construir cortan a otras rectas o circunferencias sí que son constructibles) entonces no hay más usos posibles de la regla marcada, y por eso hemos hablado de “construcciones con regla marcada” y no de “construcciones con regla marcada y compás”.

Ahora bien, si admitimos el compás, caben más usos de la regla marcada, ya que podemos usarla para encontrar puntos a una distancia unitaria colineales con un punto dado y que estén uno en una recta y otro en una circunferencia dadas, o bien cada uno en una circunferencia dada (lo que se conoce como *neusis* recta-circunferencia o *neusis* circunferencia-circunferencia). Si admitimos estos dos usos adicionales tenemos lo que propiamente podemos llamar “construcciones con regla marcada y compás”, y sucede que el conjunto de puntos constructibles de este modo es estrictamente mayor que el de los puntos constructibles sólo con regla marcada.

Ejemplo Si trazamos la recta $y = 2$ y la circunferencia de centro $(1, 1)$ y radio $\sqrt{5}$, podemos usar la regla marcada para encontrar cuatro pares de puntos situados uno en la recta y otro en la circunferencia que disten una unidad. Un par es trivial: la circunferencia pasa por el punto $P_4 = (0, 3)$, que dista una unidad de $Q_4 = (0, 2)$. Analíticamente, los puntos P_i cumplen la ecuación de la circunferencia:

$$(x - 1)^2 + (y - 1)^2 = 5 \quad \leftrightarrow \quad x^2 + y^2 = 2x + 2y + 3$$

y los de la conoide de parámetros $b = 2$ y $d = 1$, que es $(x^2 + y^2)(y - 2)^2 = y^2$.



Sustituyendo obtenemos

$$(2x + 2y + 3)(y - 2)^2 = y^2 \quad \Leftrightarrow \quad 2(y - 2)^2x + (2y + 3)(y - 2)^2 = y^2$$

$$\Leftrightarrow \quad 2(y - 2)^2(x - 1) = y^2 - (2y + 5)(y - 2)^2.$$

Elevamos al cuadrado y sustituimos $(x - 1)^2$ usando la ecuación de la circunferencia:

$$4(y - 2)^4(5 - (y - 1)^2) = (y^2 - (2y + 5)(y - 2)^2)^2.$$

Al operar queda:

$$y^6 - 7y^5 + 14y^4 - 2y^3 - 10y^2 - 12y + 18 = 0.$$

Ahora bien, sabemos que una raíz de este polinomio tiene que ser $y = 3$. Dividiendo entre $y - 3$ llegamos a que las segundas coordenadas de los puntos P_1, P_2, P_3 tienen que cumplir la ecuación

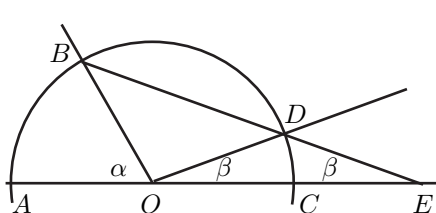
$$y^5 - 4y^4 + 2y^3 + 4y^2 + 2y - 6 = 0.$$

Este polinomio es irreducible por el criterio de Eisenstein,⁴ luego la segunda coordenada de cualquiera de los tres puntos P_1, P_2, P_3 cumple $|\mathbb{Q}(y) : \mathbb{Q}| = 5$, luego ninguno de los tres puntos es constructible sólo con regla marcada, por el teorema 6.26. Vemos así que existen puntos constructibles con regla marcada y compás que no son constructibles con regla marcada. ■

⁴Más aún, el teorema [Al 7.16] implica que su grupo de Galois es Σ_5 , y [Al 7.27] implica que la ecuación no es resoluble por radicales. Vemos así que con regla marcada y compás es posible construir números no expresables en términos de radicales a partir de números racionales. Esto no es posible sólo con regla marcada.

Trisección de un ángulo Recordemos (problema [IGE 3.6]) que Arquímedes encontró este procedimiento para trisecar un ángulo agudo mediante una *neusis* recta-circunferencia:

Partimos del ángulo \widehat{AOB} , donde podemos exigir que $\overline{OA} = \overline{OB} = 1$. Trazamos la circunferencia de centro O y radio 1 y usamos la regla marcada para trazar los puntos D y E , el primero sobre la circunferencia, el segundo sobre OA , ambos colineales con



el punto B y a una distancia unitaria. Entonces el triángulo \widehat{ODE} es isósceles, y usando que el ángulo en E mide la semidiferencia de los arcos que abarca, vemos que $\beta = \frac{\alpha - \beta}{2}$, luego $\alpha = 3\beta$, luego \widehat{COD} es la trisección de \widehat{AOB} . ■

El caracol de Pascal El uso de la regla marcada en el método de trisección de Arquímedes consiste en tomar una circunferencia de radio 1 y un punto B en ella, y calcular un punto E que se encuentra en una recta dada (la recta OC) de modo que la distancia de E al otro punto D donde la recta BE corta a la circunferencia sea igual a 1. La curva formada por tales puntos E recibe el nombre de *caracol de Pascal*.

Si tomamos concretamente la circunferencia de centro $(0, 0)$ y radio 1 y el punto $B = (-1, 0)$, dado un punto $E = (x, y)$, la recta BE está formada por los puntos de la forma

$$(-1, 0) + \lambda(x - 1, y).$$

Calculamos λ para que el punto esté en la circunferencia y resulta $\lambda = \frac{2(x+1)}{(x+1)^2 + y^2}$, de donde D tiene coordenadas

$$(A, B) = \left(-1 + \frac{2(x+1)^2}{(x+1)^2 + y^2}, \frac{2(x+1)y}{(x+1)^2 + y^2} \right).$$

Para que E esté en el caracol tiene que cumplirse que $(A - x)^2 + (B - y)^2 = 1$. Desarrollamos usando que $A^2 + B^2 = 1$ y resulta

$$(x^2 + y^2)((x+1)^2 + y^2) = -2x((x+1)^2 + y^2) + 4(x+1)^2x + 4(x+1)y^2,$$

que operando se reduce a

$$(x^2 + y^2)^2 - 3(x^2 + y^2) - 2x = 0.$$

Ésta es la ecuación del caracol de Pascal. En coordenadas polares podemos simplificar un ρ y queda

$$\rho^3 - 3\rho - 2\cos\theta = 0.$$

Las soluciones de esta ecuación son⁵

$$\rho = 2 \cos \frac{\theta + 2k\pi}{3}, \quad k = 0, 1, 2.$$

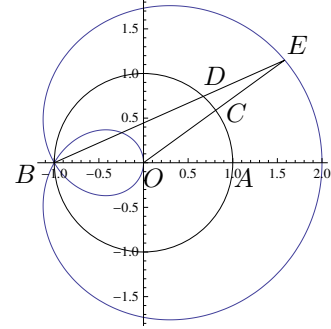
⁵Véase el teorema [A1 9.20].

Teniendo en cuenta que ρ tiene que ser positivo, es fácil ver que la relación

$$\rho = 2 \cos \frac{\theta}{3}, \quad \theta \in \left[-\frac{3\pi}{2}, \frac{3\pi}{2} \right]$$

recoge todos los puntos de la curva.

Si llamamos E al punto de intersección con el caracol de cualquier recta que pase por el origen O , el ángulo \widehat{OEB} es la tercera parte de \widehat{AOE} .



Sin embargo, el uso de la regla marcada con una circunferencia no exige que el punto B que se toma como referencia tenga que estar en la propia circunferencia. Vamos a obtener la ecuación general de la curva equivalente a la conchoide de Nicomedes cuando en lugar de partir de una recta, partimos de una circunferencia, que es una generalización del caracol de Pascal.

Partimos de un punto O con coordenadas en un cuerpo $K \subset R$ y de una circunferencia cuyo radio y las coordenadas de su centro estén también en K . El mismo argumento previo al cálculo de la conchoide muestra que, mediante un cambio de sistema de referencia y tal vez sustituyendo K por una extensión cuadrática, podemos suponer que $O = (0, 0)$ y que el centro de C está en el eje de abscisas, es decir, que tiene coordenadas $(b, 0)$.

Tomemos un punto arbitrario de R^2 y expresémoslo en coordenadas polares: $X = (x, y) = (\rho \cos \theta, \rho \sin \theta)$. Los puntos que están sobre la recta OX a una distancia d de X son $((\rho \pm d) \cos \theta, (\rho \pm d) \sin \theta)$. Buscamos los puntos X tales que uno de estos puntos está en C , lo cual equivale a que

$$((\rho \pm d) \cos \theta - b)^2 + (\rho \pm d)^2 \sin^2 \theta = r^2.$$

Operamos:

$$\begin{aligned} (\rho \pm d)^2 - 2b(\rho \pm d) \cos \theta + b^2 &= r^2, \\ \rho^2 \pm 2d\rho + d^2 - 2b(\rho \pm d) \cos \theta + b^2 - r^2 &= 0, \\ \rho^2 - 2bd\rho \cos \theta + b^2 - r^2 + d^2 &= \pm(2d\rho - 2bd \cos \theta). \end{aligned}$$

El doble signo hace que la ecuación sea equivalente a la que resulta de elevar al cuadrado sus miembros:

$$(\rho^2 - 2bd\rho \cos \theta + b^2 - r^2 + d^2)^2 = 4d(\rho - b \cos \theta)^2.$$

Si multiplicamos por ρ^2 a lo sumo añadimos el punto $X = (0, 0)$:

$$\rho^2(\rho^2 - 2bd\rho \cos \theta + b^2 - r^2 + d^2)^2 = 4d^2(\rho^2 - b\rho \cos \theta)^2.$$

En coordenadas cartesianas queda:

$$(x^2 + y^2)(x^2 + y^2 - 2bdx + b^2 + d^2 - r^2)^2 = 4d^2(x^2 + y^2 - bx)^2.$$

A partir de aquí hacemos $d = 1$, pues es el caso correspondiente a una regla con dos marcas a distancia unitaria. Si usamos la regla marcada para encontrar dos puntos colineales con O a distancia unitaria de modo que uno esté en la circunferencia que hemos tomado y otro en otra circunferencia de centro y radio en K , el punto determinado cumplirá la ecuación del caracol generalizado y la ecuación de la nueva circunferencia, que será de la forma

$$(x - u)^2 + (y - v)^2 = s^2, \quad (6.1)$$

con $u, v, s \in K$. Equivalentemente, $x^2 + y^2 = 2ux + 2vy + s^2 - u^2 - v^2$. En definitiva, todo punto (x, y) obtenido al aplicar la regla marcada a dos circunferencias cumple el sistema de ecuaciones siguiente:

$$(x^2 + y^2)(x^2 + y^2 - 2bx + b^2 + 1 - r^2)^2 = 4(x^2 + y^2 - bx)^2, \quad (6.2)$$

$$x^2 + y^2 = 2ux + 2vy + s^2 - u^2 - v^2. \quad (6.3)$$

Sustituyendo la segunda ecuación en la primera queda:

$$\begin{aligned} (2ux + 2vy + s^2 - u^2 - v^2)(2ux + 2vy + s^2 - u^2 - v^2 - 2bx + b^2 + 1 - r^2)^2 \\ = 4(2ux + 2vy + s^2 - u^2 - v^2 - bx)^2. \end{aligned} \quad (6.4)$$

Esta ecuación es de la forma

$$P_0(x)y^3 + P_1(x)y^2 + P_2(x)y + P_3(x) = 0,$$

donde $P_i(x)$ es un polinomio de grado $\leq i$. Por otra parte, (6.3) es de la forma $y^2 = 2vy + Q_2(x)$. Además

$$y^3 = 2vy^2 + Q_2(x)y = 2v(2vy + Q_2(x)) + Q_2(x)y = R_2(x)y + R'_2(x).$$

Si sustituimos estas expresiones para y^2 e y^3 en (6.4) obtenemos una ecuación de la forma

$$S_2(x)y = S_3(x), \quad (6.5)$$

donde cada polinomio tiene siempre grado menor o igual que su subíndice. Esto es equivalente a:

$$S_2(x)(y - v) = S_3(x) - S_2(x)v, \quad (6.6)$$

y elevando al cuadrado:

$$S_2^2(x)(y - v)^2 = (S_3(x) - S_2(x)v)^2. \quad (6.7)$$

Por último, usando (6.1) obtenemos

$$S_2^2(x)(s^2 - ((x - u)^2)) = S_3(x) - S_2(x)v^2, \quad (6.8)$$

con lo que llegamos a que x es raíz de un polinomio de $K[x]$ de grado menor o igual que 6. A su vez, (6.1) nos da que y está en una extensión cuadrática de $K(x)$.

Nota Hay que justificar que el polinomio que resulta de pasar todos los términos de (6.8) a un mismo miembro no es idénticamente nulo. Esto se debe a que partimos de que el sistema formado por (6.2) y (6.3) tiene un número finito de soluciones. Lo contrario significaría que las dos circunferencias que hemos tomado no permiten determinar ningún par de puntos con la regla marcada (por ejemplo, porque están muy alejadas entre sí, y no hay un par de puntos en cada una de ellas a una distancia unitaria, o porque son concéntricas con O y $|r - s| = 1$ con lo que hay infinitos puntos colineales con O a distancia unitaria, y entonces tampoco podemos usar la regla marcada, etc.). A partir de (6.3) y (6.4) se puede recuperar (6.2), luego el sistema sigue teniendo un número finito de soluciones. Igualmente, a partir de (6.5) y (6.3) se puede recuperar (6.2). La ecuación (6.6) es equivalente a (6.5), por lo que el sistema formado por (6.6) y (6.1) es equivalente al inicial. Al elevar (6.6) al cuadrado estamos incorporando las posibles soluciones del sistema formado por (6.1) y la ecuación

$$S_2(x)(y - v) = -S_3(x) + S_2(x)v. \quad (6.9)$$

Si probamos que este sistema también tiene a lo sumo un número finito de soluciones (tal vez ninguna), entonces el sistema formado por (6.1) y (6.7) tendrá un número finito de soluciones, y lo mismo valdrá para el sistema formado por (6.1) y (6.8), lo cual exige que la ecuación (6.8) no sea idénticamente nula.

En cuanto a (6.9), es de la forma $T_2(x)(y - v) - T_3(x) = 0$, donde los polinomios no son idénticamente nulos, ya que si lo fueran también sería idénticamente nula la ecuación (6.6), pues es $T_2(x)(y - v) + T_3(x) = 0$, y el sistema formado por (6.6) y (6.1) tendría infinitas soluciones.⁶ Si $T_2(x)$ es idénticamente nulo, entonces toda solución (x, y) de (6.6) cumple que x es raíz de $T_3(x)$, lo que sólo deja un número finito de posibilidades, y para cada una de ellas hay a lo sumo dos valores de y que cumplen (6.1). Supongamos, pues, que $T_2(x)$ no es idénticamente nulo. Si el sistema tuviera infinitas soluciones, éstas tomarían infinitos valores de x (pues para cada valor de x hay a lo sumo dos valores de y que cumplen la ecuación de la circunferencia), luego la ecuación

$$(x - u)^2 + \left(\frac{T_3}{T_2}\right)^2 = s^2 \quad \Leftrightarrow (x - u)^2 T_2^2 + T_3^2 - s^2 T_2^2 = 0$$

tendría infinitas soluciones (todos los valores de x que cumplen el sistema sin anular a T_2). Esto implica que el polinomio de la derecha es idénticamente nulo, luego el polinomio $s^2 - (x - u)^2 = (T_3/T_2)^2$ es un cuadrado en el cuerpo de fracciones algebraicas $K(x)$, pero, por el criterio de Gauss, las raíces del polinomio $T^2 - s^2 + (x - u)^2 \in K(x)[T]$ tienen que estar en $K[x]$, luego llegamos

⁶A partir de aquí podemos concluir inmediatamente aplicando el teorema de Bezout, que es un resultado de la geometría algebraica que afirma que un sistema de dos ecuaciones polinómicas $f(x, y) = 0$ y $g(x, y) = 0$ tiene a lo sumo tantas soluciones como el producto de sus grados, salvo que ambos polinomios tengan un factor común. Es fácil ver que el polinomio que define la circunferencia es irreducible en $K[x, y]$, luego para que el sistema tuviera infinitas soluciones tendría que dividir a (6.9), lo cual es imposible, porque (6.9) tiene grado 1 en y y la circunferencia tiene grado 2.

a que $s^2 - (x - u)^2 = (ax + b)^2$, para ciertos $a, b \in K$, y ahora es fácil ver que esto implica que $s = 0$, que no es el caso. ■

Si aplicamos la regla marcada para seleccionar dos puntos en una recta y una circunferencia, el punto de la recta cumplirá el sistema formado por su ecuación (de grado 1) y (6.2), y en este caso es fácil despejar una variable en la ecuación de la recta para sustituirla en (6.2) y concluir que una de las coordenadas del punto cumple una ecuación de grado menor o igual que 6 (y la segunda depende linealmente de la primera, luego está en cualquier cuerpo que contenga a la primera). En conclusión:

Teorema 6.32 *Si un punto P es constructible con regla marcada y compás a partir de un conjunto Γ de puntos cuyas coordenadas estén en un cuerpo $K \subset R$, entonces existe una cadena de extensiones*

$$K = K_0 \subset K_1 \subset \cdots \subset K_n \subset R$$

de modo que $|K_{i+1} : K_i| \leq 6$ y las coordenadas (α_1, α_2) de P están en K_n . En particular, $|K(\alpha_i) : K|$ sólo puede ser divisible entre los primos 2, 3 y 5.

Por ejemplo, ahora es inmediato que el polígono regular de 23 lados no es constructible con regla marcada y compás, pues si $\eta = \cos(2\pi/23)$, entonces $|\mathbb{Q}(\eta) : \mathbb{Q}| = 11$.

6.7 Construcciones con regla y transportador

Estudiamos ahora un caso intermedio entre las construcciones sólo con regla y las construcciones con regla y compás: vamos a usar una regla y un transportador de segmentos, que en la práctica podemos considerar que es lo mismo que un compás, pero restringimos su uso al de transportar un segmento dado a una recta dada de la que conocemos ya un punto, según el axioma C2. Explícitamente:

Definición 6.33 Sea E un plano euclídeo y Γ un conjunto de puntos de E . Diremos que un punto $P \in E$ es *constructible con regla y transportador* a partir de Γ si existe una sucesión finita de puntos P_1, \dots, P_n tal que $P = P_n$ y cada P_i cumple una de las condiciones siguientes:

1. $P_i \in \Gamma$.
2. P_i es la intersección de dos rectas que pasan cada una por dos puntos previos.
3. Existen cuatro puntos previos A, B, C, D en la sucesión de modo que P_i es un punto en la recta CD tal que $\overline{CP_i} \equiv \overline{AB}$.

Diremos que una recta es *constructible con regla y transportador* a partir de Γ si pasa por dos puntos constructibles a partir de Γ .

En esta sección no necesitamos suponer que el cuerpo R del espacio euclídeo sea euclídeo, sino que basta con que sea pitagórico.

Diremos que un elemento de $C = R[i]$ es *constructible con regla y transportador* si y sólo si es constructible a partir de $\{0, 1, i\}$ (aunque es claro que i puede construirse con ayuda de elecciones arbitrarias de puntos).

Exactamente los mismos argumentos que hemos empleado en las secciones anteriores nos permiten concluir que los resultados sobre construcciones con regla son válidos también si consideramos regla y transportador. En particular, el conjunto $\mathcal{C}_R^{rt} \subset R^2$ de los puntos constructibles con regla y transportador es un subcuerpo del cuerpo \mathbb{C} de los números constructibles con regla y compás, al igual que $R_{rt} = \mathcal{C}_R^{rt} \cap R$ es un subcuerpo de \mathbb{C}_r , de modo que $\mathcal{C}^{rt} = R_{rt}[i]$.

Vamos a estudiar el cuerpo R_{rt} . El teorema siguiente implica que no es \mathbb{Q} :

Teorema 6.34 *Si R es cualquier cuerpo ordenado pitagórico, el cuerpo R_{rt} también es pitagórico.*

DEMOSTRACIÓN: Si $a, b \in R_{rt}$, podemos levantar la perpendicular al eje real por a y transportar b sobre dicha perpendicular, para construir el punto $a+bi \in \mathcal{C}_R^{rt}$. A su vez podemos transportar el módulo $\sqrt{a^2 + b^2}$ de dicho número sobre el eje real, lo cual nos da que $\sqrt{a^2 + b^2} \in R_{rt}$. ■

El cuerpo R_{rt} admite una caracterización algebraica similar a la de \mathbb{C}_r :

Teorema 6.35 *Sea R un cuerpo pitagórico. Entonces $z \in R$ está en R_{rt} si y sólo si existe una cadena de cuerpos*

$$\mathbb{Q} = K_0 \subset K_1 \subset \dots \subset K_n \subset R$$

de modo que $z \in K_n$ y $K_{i+1} = K_i(\sqrt{1 + a_i^2})$, para cierto $a_i \in K_i$.

DEMOSTRACIÓN: Sea $z_1, \dots, z_n = z$ una sucesión de puntos según la definición de constructibilidad. Basta probar inductivamente que para cada z_i existe una cadena como la del enunciado que contiene a su parte real y a su parte imaginaria. A su vez, esto se reduce a probar que si $z \in R^2$ se obtiene de puntos de K^2 , con $K \subset R$, aplicando una de las condiciones de constructibilidad con recta y transportador, entonces existen extensiones

$$K \subset K' = K(\sqrt{1 + x^2}) \subset K'' = K'(\sqrt{1 + y^2})$$

con $x, y \in K$ tales que $z \in K''^2$.

Si z se obtiene como intersección de dos rectas que pasan cada una por dos puntos de K^2 , entonces el mismo argumento del teorema 6.3 prueba que $z \in K^2$.

Sólo falta considerar el caso en que z se obtiene transportando un segmento. Específicamente, tenemos cuatro puntos $(a, b), (c, d), (e, f), (g, h) \in K^2$

de modo que y y z es un punto de la recta $(e, f) + \lambda(g - e, h - f)$ que dista $\sqrt{(c - a)^2 + (d - b)^2}$ de (e, f) . Por lo tanto, sus coordenadas son

$$(e, f) + \frac{\sqrt{(c - a)^2 + (d - b)^2}}{\sqrt{(g - e)^2 + (h - f)^2}}(g - e, h - f).$$

Así pues, las coordenadas de z están en un cuerpo $K(\sqrt{u^2 + v^2}, \sqrt{p^2 + q^2})$, donde $u, v, p, q \in K$. Podemos suponer que los cuatro números son no nulos, y entonces, expresando $\sqrt{u^2 + v^2} = \pm u\sqrt{1 + (v/u)^2}$, tenemos que el cuerpo es de la forma $K(\sqrt{1 + x^2}, \sqrt{1 + y^2})$, con $x, y \in K$, como había que probar. ■

Ahora la prueba del teorema 6.6 se adapta de forma obvia para probar el teorema siguiente:

Teorema 6.36 *Si R es un cuerpo ordenado pitagórico, entonces:*

1. R_{rt} es el menor subcuerpo pitagórico de R . En particular, si $R \subset R'$ son cuerpos pitagóricos, entonces $R_{rt} = R'_{rt}$ y $\mathcal{C}_R^{rt} = \mathcal{C}_{R'}^{rt}$.
2. La extensión $\mathcal{C}_R^{rt}/\mathbb{Q}$ es de Galois.
3. Si R y R' son dos cuerpos pitagóricos cualesquiera, entonces $R_{rt} \cong R'_{rt}$ y $\mathcal{C}_R^{rt} \cong \mathcal{C}_{R'}^{rt}$.

Definición 6.37 Llamaremos $\mathcal{P} = \mathcal{C}_{\mathbb{R}}^{rt}$ al cuerpo de los números complejos constructibles con regla y transportador, y $\mathcal{P}_r = \mathcal{P} \cap \mathbb{R}$, de modo que $\mathcal{P} = \mathcal{P}_r[i]$.

Así $\mathbb{Q} \subset \mathcal{P} \subset \mathbb{C}$ y $\mathbb{Q} \subset \mathcal{P}_r \subset \mathbb{C}_r$. El teorema anterior prueba que \mathcal{P}_r es el menor cuerpo ordenado pitagórico, en el sentido de que cualquier otro contiene un subcuerpo isomorfo. Por lo tanto, \mathcal{P}_r es el menor cuerpo ordenado que permite dotar a un espacio afín de estructura euclídea. El plano pitagórico $\mathcal{P} = \mathcal{P}_r^2$ es el menor plano euclídeo que cumple los axiomas de la geometría euclídea sin contar el axioma de continuidad ni el de las circunferencias. Una figura es constructible con regla y transportador si y sólo si existe en dicho plano.

Así pues, tenemos que \mathcal{P}_r es el menor cuerpo pitagórico en el mismo sentido en que \mathbb{C}_r es el menor cuerpo euclídeo. El teorema siguiente implica que $\mathcal{P}_r \subsetneq \mathbb{C}_r$, con lo que obtenemos finalmente un ejemplo de cuerpo pitagórico que no es euclídeo.

Teorema 6.38 *La extensión \mathcal{P}_r/\mathbb{Q} es de Galois.*

DEMOSTRACIÓN: Claramente es una extensión algebraica separable. Falta probar que es normal. Para ello tomamos $a \in \mathcal{P}_r$ y hemos de probar que todos sus conjugados están en \mathcal{P}_r . Ciertamente están en \mathcal{P} , porque la extensión \mathcal{P}/\mathbb{Q} es normal. Falta probar que son reales.

Sea $\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_n \subset \mathbb{R}$ una cadena de cuerpos en las condiciones del teorema 6.35 tal que $a \in K_n$. Vamos a probar inductivamente que la clausura

normal \bar{K}_n de cada K_n sobre \mathbb{Q} está contenida en \mathbb{R} . Esto es trivialmente cierto para $\bar{K}_0 = \mathbb{Q}$. Si vale para K_i , sea $K_{i+1} = K_i(\sqrt{1+a_i^2})$, con $a_i \in K_i$. Sea $\alpha = \sqrt{1+a_i^2}$ y sea $\beta \in \mathbb{C}$ un conjugado de α . Entonces existe un automorfismo $\sigma \in G(\bar{K}_{i+1}/\mathbb{Q})$ tal que $\sigma(\alpha) = \beta$. Entonces $\bar{K}_i \subset \bar{K}_{i+1}$ y $\sigma|_{\bar{K}_i} \in G(\bar{K}_i/\mathbb{Q})$. Como $\alpha^2 = 1+a_i^2$, tenemos que

$$\beta = \sigma(\alpha) = \pm\sqrt{1+\sigma(a_i)^2},$$

donde $\sigma(a_i) \in \bar{K}_i \subset \mathbb{R}$, luego $1+\sigma(a_i)^2 > 0$, luego $\beta \in \mathbb{R}$. Concluimos que $\bar{K}_n \subset \mathbb{R}$ y contiene a todos los conjugados de a . ■

Así pues, todo conjugado de un elemento de \mathcal{P}_r está en \mathcal{P}_r , luego los elementos de \mathcal{P}_r son totalmente reales. Más aún:

Teorema 6.39 *El cuerpo \mathcal{P}_r está formado por los elementos totalmente reales de \mathcal{C}_r .*

DEMOSTRACIÓN: Acabamos de observar que todos los elementos de \mathcal{P}_r son elementos totalmente reales de \mathcal{C}_r . Recíprocamente, si $\alpha \in \mathcal{C}_r$ es totalmente real, consideramos una cadena de extensiones

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_n \subset \mathcal{C}_r \subset \mathbb{R}$$

el las condiciones del teorema 6.3, con $\alpha \in K_n$. La clausura normal de $\mathbb{Q}(\alpha)$ sobre \mathbb{Q} está contenida en \mathbb{R} , y si intersecamos todos los cuerpos K_j con dicha clausura normal obtenemos una cadena en las mismas condiciones, luego podemos suponer que todos los cuerpos K_j son totalmente reales.

Basta probar inductivamente que cada $K_j \subset \mathcal{P}_r$. Supuesto que se cumple para j , tenemos que $K_{j+1} = K_j(\sqrt{\alpha_j})$, donde necesariamente $\alpha_j > 0$. Más aún, α_j , que es totalmente real, tiene que ser totalmente positivo, pues si tuviera un conjugado $\beta < 0$, entonces $\sqrt{\alpha_j}$ sería conjugado de $\sqrt{\beta}$, que es imaginario, luego K_{j+1} no estaría formado por números totalmente reales. Como α_j es totalmente positivo, por el teorema [AL 5.79] tenemos que es suma de cuadrados en K_j , luego $\sqrt{\alpha_j} \in \mathcal{P}_r$, luego $K_{j+1} \subset \mathcal{P}_r$. ■

Como consecuencia:

Teorema 6.40 *$\mathcal{P}_r \subsetneq \mathcal{C}_r$, luego no todo punto constructible con regla y compás es constructible con regla y transportador.*

DEMOSTRACIÓN: Por ejemplo, $a = \sqrt{1+\sqrt{2}} \in \mathcal{C}_r$, pues tenemos la cadena de extensiones cuadráticas

$$\mathbb{Q} \subset \mathbb{Q}[\sqrt{2}] \subset \mathbb{Q}[a] \subset \mathbb{R},$$

pero $a \notin \mathcal{P}_r$, pues no es totalmente real, ya que $\sqrt{1-\sqrt{2}}$ es un conjugado imaginario. ■

Lo que hemos visto en la prueba del teorema anterior es que la extensión \mathcal{C}_r/\mathbb{Q} no es de Galois.

Una consecuencia es que el plano \mathcal{P}_r^2 no cumple el axioma de las circunferencias, pero sí los axiomas restantes de la geometría euclídea (menos el de continuidad), lo que prueba que el axioma de las circunferencias no puede demostrarse a partir de los demás.

Veamos una última consecuencia de estos hechos:

Teorema 6.41 *Todo polígono regular constructible con regla y compás es constructible con regla y transportador.*

DEMOSTRACIÓN: Si $\zeta = \cos \frac{2\pi}{n} + i \operatorname{sen} \frac{2\pi}{n}$, la extensión ciclotómica $\mathbb{Q}(\zeta)/\mathbb{Q}$ tiene grupo de Galois abeliana, luego todos sus subcuerpos son normales sobre \mathbb{Q} . En particular esto vale para $\mathbb{Q}(\cos \frac{2\pi}{n}) \subset \mathbb{R}$, de donde se sigue que $\eta = \cos \frac{2\pi}{n}$ es totalmente real. Si $\eta \in \mathcal{C}_r$, entonces de hecho $\eta \in \mathcal{P}_r$, luego es constructible con regla y transportador. ■

Capítulo VII

La geometría afín

En la exposición que hemos hecho de la geometría euclídea en términos sintéticos y analíticos hay unas asimetrías que impiden afirmar en sentido estricto que sean dos formas alternativas de presentar las mismas estructuras matemáticas. En efecto, en la sección 3.1 hemos probado que si E es una geometría tridimensional euclídea arquimediana, entonces E es esencialmente la misma estructura matemática que R^3 , para un cierto cuerpo ordenado pitagórico arquimediano R , que será \mathbb{R} si E cumple el axioma de continuidad. Sin embargo, en el capítulo III hemos visto también que cualquier espacio R^n con $n \geq 2$ cumple todos los axiomas de la geometría de Hilbert salvo los que determinan la tridimensionalidad del espacio, pero no hemos probado que todo espacio que cumpla los axiomas de Hilbert sin contar los relativos a la dimensión sea un espacio R^n .

El lector puede imaginar que no es especialmente difícil establecer axiomas que expresen que el espacio tiene dimensión n , para un cierto número natural n , y que la introducción de coordenadas que hemos hecho en la sección 3.1 se generaliza sin dificultades al caso de la geometría de n dimensiones. Así es, en efecto, pero ésta es la más leve de las asimetrías que de momento tenemos entre el enfoque analítico y sintético. Por ejemplo, hemos necesitado la propiedad arquimediana para construir el cuerpo R en el que tomar las coordenadas de los puntos, pero al mismo tiempo hemos visto que si R es cualquier cuerpo ordenado pitagórico, aunque no sea arquimediano, el espacio afín R^n cumple los axiomas de la geometría de Hilbert sin contar la propiedad arquimediana, pero no hemos demostrado que cualquier espacio que cumpla tales axiomas tenga que ser equivalente a un espacio R^n . Similarmente, sabemos que un espacio R^n cumple los axiomas de incidencia y ordenación aunque el cuerpo ordenado R no sea pitagórico, pero no hemos demostrado que cualquier espacio que cumpla los axiomas de incidencia y ordenación tenga que ser equivalente a un espacio R^n , para un cierto cuerpo ordenado R . Y, por último, sabemos que R^n cumple los axiomas de incidencia si R es un cuerpo cualquiera, pero no hemos demostrado que cualquier geometría que cumpla los axiomas de incidencia tenga que ser la geometría de un espacio R^n , para un cierto cuerpo R .

En este capítulo demostraremos, con ciertas salvedades, todas estas cosas que no hemos demostrado hasta ahora. Una salvedad obvia es que no es cierto que todo espacio que cumpla los axiomas de incidencia tenga que ser de la forma R^n para un cierto cuerpo R , por la sencilla razón de que es fácil ver que la prueba de que R^n cumple los axiomas de incidencia vale igualmente si R es un anillo de división, aunque no sea conmutativo. Por consiguiente, si queremos establecer un resultado de equivalencia entre la geometría analítica y la geometría sintética, hemos de incorporar el caso de los espacios afines sobre anillos de división. Aparte de esto hay otra pequeña discrepancia que discutiremos en su momento, pero salvo por ella, vamos a probar que, eligiendo adecuadamente unos axiomas de incidencia (es decir, unos axiomas que no involucren más que puntos, rectas, planos, “pasar por”, o “ser paralelo a”, pero no “estar entre” o “ser congruente”, etc.), podemos demostrar que todo espacio que los satisfaga es esencialmente un espacio K^n , para cierto anillo de división K . Así podremos afirmar que (salvo un caso excepcional muy particular referente a lo que llamaremos ‘planos no arguesianos’) el enfoque analítico de la geometría es equivalente al sintético a todos los niveles, en el sentido de que, admitiendo que los axiomas de incidencia son el sustrato mínimo que nos permite decir que “estamos haciendo geometría euclídea y no otra cosa”, estudiar geometrías de Hilbert que satisfagan más o menos axiomas equivale a estudiar espacios K^n sobre anillos de división que satisfagan más o menos propiedades.

7.1 Los axiomas de la geometría afín

Recordemos los axiomas de incidencia que presentamos en el capítulo I. Si incluimos entre ellos al axioma de las paralelas, son los siguientes:

Axioma A1	<i>Por cada par de puntos distintos pasa una única recta.</i>
Axioma A2	<i>Por cada punto pasa una única paralela a una recta dada.</i>
Axioma A3	<i>Por cada tres puntos no colineales pasa un único plano.</i>
Axioma A4	<i>Si una recta tiene dos puntos en común con un plano, está contenida en él.</i>
Axioma A5	<i>Todo plano pasa al menos por tres puntos no colineales.</i>

Notemos que hemos puesto el axioma de las paralelas en el lugar del axioma A2 del capítulo I, según el cual las rectas tienen al menos dos puntos, porque, según mostramos al final de la sección 1.1, dicho axioma puede demostrarse a partir del axioma de las paralelas.

Estos axiomas no afirman nada sobre la dimensión del espacio, salvo que es mayor o igual que 2, ya que son satisfechos en todo espacio K^n (con las definiciones usuales de recta y plano) para cualquier cuerpo (o, más en general, cualquier anillo de división) K y todo $n \geq 2$. Sabemos que la tridimensionalidad del espacio se puede expresar con los axiomas siguientes:

Axioma A5	<i>Existen cuatro puntos no coplanares.</i>
Axioma A6	<i>Si dos planos tienen un punto en común, entonces tienen dos puntos en común.</i>

En cambio, si queremos unos axiomas que expresen que el espacio tiene dos dimensiones, aunque podemos hacerlo postulando que no existen puntos no coplanares, o que todos los planos son iguales, lo más simple es eliminar la noción de plano de los axiomas, y así obtenemos los axiomas de incidencia de la geometría plana:

Axioma A1	<i>Por cada par de puntos distintos pasa una única recta.</i>
Axioma A2	<i>Por cada punto pasa una única paralela a una recta dada.</i>
Axioma A3	<i>Existen tres puntos no colineales.</i>

Es importante observar que la eliminación del concepto de plano supone una modificación no trivial del concepto de “paralela”. En efecto, en el axioma A2 general hay que entender que, por definición, dos rectas son paralelas si son iguales o bien son coplanares y disjuntas, mientras que en el correspondiente a la geometría plana la noción se reduce a que dos rectas son paralelas si y sólo si son iguales o disjuntas, y es ahí realmente, al negar la posibilidad de que dos rectas se crucen, donde queda expresada la bidimensionalidad del espacio.

Sucede que el bloque de los cinco axiomas de incidencia adimensionales que conocemos no es suficiente como punto de partida para una geometría afín (es decir, que considere únicamente las propiedades de incidencia y paralelismo) que no especifique la dimensión del espacio, en el sentido de que hay propiedades que “deberían poder demostrarse” y, en cambio, no pueden probarse a partir de ellos.

Un caso muy claro es el de la transitividad del paralelismo. En la geometría plana, es obvio que si una recta r_1 es paralela a otra r_2 y ésta a su vez es paralela a r_3 , entonces r_1 es paralela a r_3 , pues, si no son iguales, tienen que ser disjuntas, ya que si r_1 y r_3 se cortaran en un punto P , entonces ambas serían paralelas a r_2 por P , en contradicción con el axioma de las paralelas.

En el caso tridimensional la comprobación es más delicada, porque exige comprobar que r_1 y r_3 están contenidas en un mismo plano. No obstante, se cumple igualmente, y lo hemos probado en 1.37. Como se recalca en la nota posterior, la prueba sólo requiere los axiomas A1–A6, pero si prescindimos de A6 necesitamos los axiomas de ordenación para construir una variedad afín tridimensional que contenga las tres rectas (véase la sección 1.3) y usar que A6 se cumple en ella.

Sucede lo mismo con la existencia de planos paralelos a uno dado. Hemos demostrado la existencia de planos paralelos sin usar el axioma de las paralelas, mediante el concepto de perpendicularidad, que involucra los axiomas del grupo C. Alternativamente, si usamos el axioma de las paralelas nos basta con los axiomas del grupo A, más los del grupo B sólo si no suponemos los de tridimensionalidad:

Teorema 7.1 *Si r y s son rectas que se cortan en un punto P y r' , s' son sus paralelas por un punto P' , entonces el plano que contiene a r' , s' es paralelo al plano que contiene a r , s .*

DEMOSTRACIÓN: Sea Π el plano que contiene a r, s y sea Π' el plano que contiene a r', s' . Si $P' \in \Pi$, entonces $r', s' \subset \Pi$, luego $\Pi = \Pi'$ y la conclusión es trivial. Supongamos que $P' \notin \Pi$.

Si no suponemos los axiomas de tridimensionalidad, entonces observamos que r' y s' están contenidas, al igual que Π , en la variedad afín tridimensional $E = A^3(\Pi, P')$, luego también $\Pi' \subset E$, que una de las condiciones que exige la definición de planos paralelos.

Falta probar que $\Pi \cap \Pi' = \emptyset$. En caso contrario se cortan en una recta t , sea por el axioma A6, sea porque ambos están contenidos en E (teorema 1.33). Dicha recta no puede ser paralela simultáneamente a r' y a s' (o éstas serían paralelas entre sí). No perdemos generalidad si suponemos que t no es paralela a r' y, como ambas están en el mismo plano, se cortan en un punto X . Sea Π^* el plano que contiene a r, r' . Entonces $X \in r' \cap \Pi \subset \Pi^* \cap \Pi = r$, luego $X \in r' \cap r$, pero esto sólo puede ser si $r = r'$, en cuyo caso $P \in \Pi$, en contra de lo supuesto. ■

Vemos que los axiomas de ordenación sólo se requieren en la prueba para construir la variedad $A^3(\Pi, P')$, lo cual es innecesario si suponemos los axiomas de tridimensionalidad, con lo que el paralelismo de planos se reduce a que sean iguales o disjuntos. En el tratamiento que vamos a presentar aquí de la geometría afín será fundamental el teorema siguiente:

Teorema 7.2 Sean A, B, C tres puntos no colineales y A', B' dos puntos tales que la recta $A'B'$ sea paralela a AB . Entonces la recta paralela a AC por A' y la recta paralela a BC por B' se cortan en un punto C' .

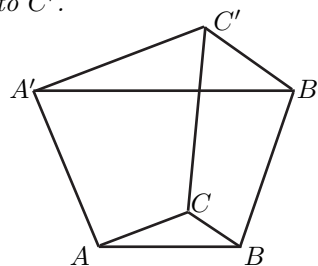
DEMOSTRACIÓN: Las dos rectas paralelas que indica el enunciado no pueden ser paralelas entre sí, pues por transitividad lo serían AC y BC , y no es el caso. Basta probar que ambas están contenidas en un mismo plano, pues entonces para no ser paralelas tienen que cortarse.

Sea $\Pi = ABC$. Si $A' \in \Pi$, entonces $A'B' \subset \Pi$, luego $B' \in \Pi$, y las rectas paralelas que indica el enunciado están también contenidas en Π , lo que termina la prueba en este caso.

Supongamos que $A' \notin \Pi$ y sean $\Pi_1 = ABA'$, $\Pi_2 = ACA'$, $\Pi_3 = BCB'$. Notemos que, como $A'B'$ es paralela a AB , se cumple que $B' \in \Pi_1$. Más aún, todos los planos están contenidos en $A^3(\Pi, A')$, por lo que Π_2 y Π_3 , que se cortan en C , se cortan de hecho en una recta r .

La paralela a AC por A' está contenida en Π_2 y, como AC corta a r en C , dicha paralela tiene que cortar a r en un punto C' . Sea $\Pi' = A'B'C'$. Por el teorema anterior, aplicado a las rectas AB y AC , cuyas paralelas por A' son $A'B'$ y $A'C'$, tenemos que Π' es paralelo a Π , y como son planos distintos, tienen intersección vacía. En particular, la recta $B'C'$ no corta a BC y ambas están contenidas en Π_3 , luego $B'C'$ es la paralela a BC por B' y, en efecto, corta en C' a la paralela a AC por A' . ■

Como en los otros casos que acabamos de analizar, esta demostración es válida si entendemos que parte de los axiomas de la geometría plana (en cuyo



caso el resultado es trivial) o de los axiomas de la geometría tridimensional, pero si no queremos usar A6, entonces necesitamos los axiomas de ordenación para construir la variedad $A^3(\Pi, A')$ y usar que en ella se cumple A6.

Sucede que el teorema anterior y la transitividad del paralelismo son los únicos resultados afines por naturaleza que requieren el apoyo de los axiomas de ordenación para ser demostrados a partir de los axiomas de incidencia adimensionales A1–A5. Alternativamente, si los incorporamos a nuestro sistema de axiomas, los axiomas de ordenación dejan de ser necesarios:

Definición 7.3 Una *geometría afín* es un conjunto no vacío \mathbb{E} al que llamaremos *espacio*, junto con una familia de subconjuntos de \mathbb{E} a los que llamaremos *rectas*, en la que hay definida una relación de equivalencia a la que llamaremos *paralelismo*, y de modo que se cumplan los axiomas siguientes:

Axioma A1 *Por cada par de puntos pasa una única recta, y cada recta contiene al menos dos puntos.*

Axioma A2 *Por cada punto pasa una única paralela a una recta dada.*

Axioma A3 Si A, B, C son tres puntos no colineales y A', B' son puntos tales que la recta AB es paralela a $A'B'$, entonces la paralela a AC por A' y la paralela a BC por B' se cortan en un punto C' .

Axioma A4 Existen tres puntos no colineales.

Observaciones Es importante señalar que estos axiomas no presuponen ninguna definición de ‘paralela’, sino que ‘paralela’ es un término no definido, al igual que ‘punto’ o ‘recta’. No podríamos definir el paralelismo de rectas porque para ello se necesita el concepto de ‘plano’, y éste no aparece entre los conceptos no definidos de la geometría afín. Sucede que, en lugar de definir ‘paralela’ a partir del concepto de ‘plano’, podremos definir ‘plano’ a partir del concepto de ‘paralela’.

No obstante, observemos que estos axiomas implican que dos rectas paralelas r y s son iguales o disjuntas, pues si tienen en común un punto P , entonces r y s son paralelas a r por P , luego tiene que ser $r = s$. Ahora bien, no podemos afirmar que dos rectas sean paralelas simplemente por que sean disjuntas.

Notemos también que hemos tenido que rescatar (como parte de A1) el antiguo axioma A2 de la sección 1.1, que afirma que cada recta pasa al menos por dos puntos. Habíamos visto que el axioma de las paralelas permite demostrar este hecho, pero la prueba requiere la definición de rectas paralelas como rectas iguales o coplanares disjuntas, con la que ahora no contamos.

Otro hecho que conviene tener presente es que los axiomas de la geometría afín no son sólo los axiomas A1–A4 anteriores, sino que entre sus postulados se encuentra también el hecho no trivial de que el paralelismo es una relación de equivalencia.

Hemos visto que tanto los axiomas de incidencia de la geometría plana como los de la geometría tridimensional implican los axiomas de la geometría afín, por lo que todo lo que deduzcamos de éstos valdrá también en los otros dos casos.

■

Vamos a usar a menudo estas dos consecuencias elementales de los axiomas:

Teorema 7.4 Sean A, B, C tres puntos no colineales. Entonces:

1. La paralela a BC por A corta a la paralela a AC por B .
2. Si B' está en AB , la paralela a BC por B' corta a AC .

DEMOSTRACIÓN: 1) Basta aplicar A3 con $A' = B$ y $B' = A$.

2) Aplicamos A3 con $A' = A$ y el B' dado. ■

Otro hecho obvio es que A4 implica que toda recta tiene al menos un punto exterior.

7.2 Variedades afines

En esta sección veremos que con los axiomas de la geometría afín podemos hacer lo que hasta ahora no sabíamos hacer en general sin recurrir a los axiomas de ordenación: definir el concepto de variedad afín. En la sección 1.3 construimos únicamente las variedades afines tridimensionales porque era el único caso que necesitábamos, pero el método es generalizable a variedades de dimensión arbitraria. Aquí vamos a considerar el caso general:

Definición 7.5 Si \mathbb{E} es una geometría afín, una *variedad afín* en \mathbb{E} es un conjunto $E \subset \mathbb{E}$ que cumpla las propiedades siguientes:

1. Toda recta que pasa por dos puntos de E está contenida en E .
2. La paralela a una recta contenida en E por un punto de E está contenida en E .

Observemos que, esta definición la cumplen trivialmente el conjunto vacío, los conjuntos con un solo punto, las rectas y el espacio \mathbb{E} .

También es inmediato que la intersección de cualquier familia no vacía de variedades afines es una variedad afín, por lo que podemos definir la *variedad afín generada por un conjunto* $X \subset \mathbb{E}$ como la intersección de todas las variedades afines que contienen a X . La representaremos por $\langle X \rangle$ y, cuando $X = \{P_0, \dots, P_n\}$ sea un conjunto finito, escribiremos también $\langle P_0, \dots, P_n \rangle$.

Es claro que $\langle \emptyset \rangle = \emptyset$, $\langle P \rangle = \{P\}$ y que si $P \neq Q$, entonces $\langle P, Q \rangle = PQ$.

Obviamente, podemos definir así las variedades afines siempre que tengamos a nuestra disposición los conceptos de recta y paralelismo, independientemente de las propiedades que postulemos sobre ellos, pero los axiomas de la geometría afín nos permiten demostrar el siguiente resultado fundamental:

Teorema 7.6 Sea E una variedad afín que contenga al menos un punto A y sea $B \in \mathbb{E} \setminus E$. Entonces $\langle E, B \rangle$ es la unión de todas las rectas paralelas a AB que pasan por un punto de E .

DEMOSTRACIÓN: Naturalmente, por $\langle E, B \rangle$ hay que entender $\langle E \cup \{B\} \rangle$, es decir, la menor variedad afín que contiene a E y a B . Llamemos $E_{A,B}$ al conjunto descrito en el enunciado. De la propia definición de variedad se sigue trivialmente que $E_{A,B} \subset \langle E, B \rangle$. Basta probar que $E_{A,B}$ es una variedad afín, pues evidentemente contiene a E y a B , y esto nos da la inclusión opuesta.

Tomemos dos puntos $P, Q \in E_{A,B}$ y veamos que $PQ \subset E_{A,B}$. Sean X_P y X_Q los puntos donde las paralelas a AB por P y Q , respectivamente, cortan a E . Distinguiamos varios casos:

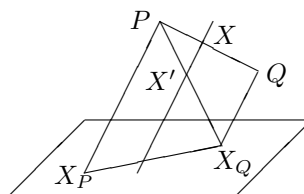
a) Si $X_P = X_Q$ entonces $PQ = r_P = r_Q$ es paralela a AB y corta a E en X_P , luego $PQ \subset E_{A,B}$, por definición. A partir de aquí suponemos que $X_P \neq X_Q$.

b) Si $P = X_P$ y $Q = X_Q$, entonces $P, Q \in E$, luego $PQ \subset E \subset E_{A,B}$.

c) Si $P = X_P$ y $Q \neq X_Q$, entonces $PX_Q \subset E$ y la segunda parte de 7.4 nos da que la paralela a QX_Q por cualquier punto de PQ corta a PX_Q o, equivalentemente, que la paralela a AB por cualquier punto de PQ corta a E , lo cual implica que $PQ \subset E_{A,B}$.

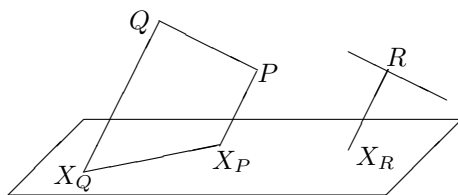
Obviamente, el caso $P \neq X_P$, $Q = X_Q$ es análogo, luego sólo queda considerar el caso siguiente:

d) Si $P \neq X_P$ y $Q \neq X_Q$, aplicamos dos veces la segunda parte de 7.4: la paralela a QX_Q (o sea, a AB) por cualquier punto $X \in PQ$ corta a PX_Q en un punto X' , y la paralela a PX_P (o sea, a AB) por X' (que es XX') corta a $X_PX_Q \subset E$, luego la paralela a AB por X corta a E y, por consiguiente $X \in E_{A,B}$.



Ahora tomamos una recta $PQ \subset E_{A,B}$ y un punto $R \in E_{A,B}$, y tenemos que probar que la paralela r a PQ por R está contenida en $E_{A,B}$. Definimos X_P y X_Q igual que en el caso anterior, y ahora tenemos además el punto X_R donde la paralela a AB por R corta a E . Distinguiamos casos:

a) Si $X_P = X_Q$ entonces PQ es paralela a AB , y lo mismo le sucede a r , luego r es la paralela a AB por R , que sabemos que corta a E en X_R , luego $r \subset E_{A,B}$. Podemos suponer, pues, que $X_P \neq X_Q$.



b) Si $P \neq X_P$, por la primera parte de 7.4 tenemos que la paralela a PQ por X_P corta a la paralela a PX_P (o sea, a AB) por Q en un punto Q' que también está en $E_{A,B}$, porque la paralela a AB por Q' es la misma que la paralela por Q , y corta a E en X_Q . Esto significa que si cambiamos P y Q por X_P y Q' seguimos teniendo dos puntos de $E_{A,B}$ y r sigue siendo paralela a X_PQ' , pero ahora $X_P \in E$, luego no perdemos generalidad si suponemos que $P = X_P$.

c) Si $P = X_P$ y $Q = X_Q$, llamamos s a la paralela a $X_P X_Q$ por X_R , que está contenida en E y es paralela a r . Si $X_R = R$ entonces $s = r \subset E$, y en caso contrario basta usar la primera parte de 7.4: si $X \neq R$ está en r , la paralela a RX_R (o sea, a AB) por X corta a la paralela a XR (o sea, a r) por X_R (que es s), luego $X \in E_{A,B}$.

d) Si $P = X_P$ y $Q \neq X_Q$, llamamos $s \subset E$ a la paralela a PX_Q por X_R y tomamos en ella cualquier otro punto $Y \neq X_R$. El axioma A3 nos da que la paralela a PQ por X_R (que es r) corta a la paralela a QX_Q (o sea, a AB) por Y en un punto $Y' \in r$ que está en $E_{A,B}$ (porque la paralela a AB por Y' corta a E en Y , luego r pasa por dos puntos de $E_{A,B}$ (X_R e Y'), luego por la parte ya probada $r \subset E_{A,B}$. ■

En particular vemos que $E_{A,B}$ no depende de la elección de $A \in E$, pues el resultado es siempre $\langle E, B \rangle$. Más aún:

Teorema 7.7 *Sea E una variedad afín, sea $B \in \mathbb{E} \setminus E$ y sea $C \in \langle E, B \rangle \setminus E$. Entonces $B \in \langle E, C \rangle \setminus E$ y $\langle E, B \rangle = \langle E, C \rangle$.*

DEMOSTRACIÓN: Fijemos un punto $A \in E$. Como $C \in E_{A,B}$, la paralela a AB por C corta a E en un punto $X_C \neq C$, ya que $C \notin E$. Si $X_C = A$ entonces $B \in AC \subset \langle E, C \rangle$. Supongamos ahora que $X_C \neq A$, con lo que A, C, X_C son tres puntos no colineales. Aplicamos A3 para concluir que la paralela a AC por B corta a la paralela a AX_C por A (que es la propia $AX_C \subset E$). Esto prueba que $B \in E_{A,C} = \langle E, C \rangle$.

De la propia definición de variedad generada por un conjunto se siguen ahora las inclusiones $\langle E, B \rangle \subset \langle E, C \rangle$ y $\langle E, C \rangle \subset \langle E, B \rangle$. ■

Definición 7.8 Una variedad afín E es *finitamente generada* si admite un generador finito. En tal caso, si el menor número de generadores de E es $n + 1$, diremos que E tiene *dimensión n* . Representaremos por $\dim E$ a la dimensión de E .

Un conjunto de $n + 1$ puntos es *afínmente independiente* si no está contenido en una variedad afín de dimensión menor que n (y por lo tanto genera una variedad afín de dimensión n).

Técnicamente, esta definición hace que el conjunto vacío tenga asignada una dimensión de -1 , pero tiene que ser así para que podamos decir que los puntos (o los conjuntos con un solo punto) son las variedades afines de dimensión 0 y que las rectas sean las variedades afines de dimensión 1.

También es claro que los conjuntos con uno o dos puntos son siempre afínmente independientes. Por el contrario, tres puntos son afínmente independientes si y sólo si no son colineales.

Teorema 7.9 *Un conjunto de puntos P_0, \dots, P_n es afínmente independiente si y sólo si ninguno de ellos pertenece a la variedad afín generada por los demás.*

DEMOSTRACIÓN: La variedad afín generada por todos menos uno tiene a lo sumo dimensión $n - 1$, luego una implicación es obvia. Supongamos ahora que los puntos no son afinmente independientes, con lo que están contenidos en una variedad afín de dimensión menor que n . Sea $m < n$ la menor dimensión de una variedad afín que los contenga. Así, $P_0, \dots, P_n \in \langle Q_0, \dots, Q_m \rangle$. Entonces no puede ser que $P_0, \dots, P_n \in \langle Q_0, \dots, Q_{m-1} \rangle$, luego, reordenando los puntos si es preciso, podemos suponer que $P_n \notin \langle Q_0, \dots, Q_{m-1} \rangle$. También tenemos que $Q_m \notin \langle Q_0, \dots, Q_{m-1} \rangle$, pues m es el menor número de elementos de un generador de la variedad. El teorema anterior implica entonces que $\langle Q_0, \dots, Q_m \rangle = \langle Q_0, \dots, Q_{m-1}, P_n \rangle$.

Similarmente, no puede ocurrir que $P_0, \dots, P_{n-1} \in \langle P_n, Q_0, \dots, Q_{m-2} \rangle$, porque entonces todos los P_i estarían en una variedad de dimensión $< m$, luego podemos suponer que $P_{n-1} \notin \langle P_n, Q_0, \dots, Q_{m-2} \rangle$ y el teorema anterior implica que

$$\langle Q_0, \dots, Q_m \rangle = \langle P_n, Q_0, \dots, Q_{m-1} \rangle = \langle P_{n-1}, P_n, Q_0, \dots, Q_{m-2} \rangle.$$

Como $m < n$, tras un número finito de pasos llegamos a que

$$P_0 \in \langle Q_0, \dots, Q_m \rangle = \langle P_{n-m}, \dots, P_n \rangle \subset \langle P_1, \dots, P_n \rangle,$$

luego uno de los puntos está en la variedad generada por los restantes. ■

El mismo argumento prueba el teorema siguiente:

Teorema 7.10 *Si P_0, \dots, P_n y Q_0, \dots, Q_n son dos conjuntos de puntos afinmente independientes y $P_0, \dots, P_n \in \langle Q_0, \dots, Q_n \rangle$, entonces se da la igualdad $\langle P_0, \dots, P_n \rangle = \langle Q_0, \dots, Q_n \rangle$.*

Ahora ya podemos demostrar las propiedades básicas de las variedades afines y los conjuntos afinmente independientes:

- *Todo subconjunto de un conjunto afinmente independiente es afinmente independiente.*

Es inmediato a partir de 7.9.

- *Toda variedad afín finitamente generada admite un generador afinmente independiente.*

En efecto, basta considerar un generador con el menor número de elementos posible. Si tiene $n+1$ elementos, entonces la variedad tiene dimensión n por definición de dimensión, y el generador es afinmente independiente porque ninguno de sus puntos puede pertenecer a la variedad generada por los restantes, ya que entonces se podría eliminar y habría otro generador con menos elementos.

- *Por $n + 1$ puntos afinmente independientes pasa una única variedad afín de dimensión n .*

Pues si P_0, \dots, P_n son afinmente independientes, una variedad afín de dimensión n es de la forma $E = \langle Q_0, \dots, Q_n \rangle$, con los Q_i afinmente independientes (por el punto anterior) y, si contiene a los P_i , el teorema anterior nos da que necesariamente $E = \langle P_0, \dots, P_n \rangle$.

- Si X es un conjunto finito afinmente independiente y $P \in \mathbb{E} \setminus \langle X \rangle$, entonces $X \cup \{P\}$ es afinmente independiente.

Si X tiene $n + 1$ elementos y $X \cup \{P\}$ estuviera contenido en una variedad afín E de dimensión $\leq n$, la dimensión tendría que ser exactamente n porque X es afinmente independiente, pero entonces $E = \langle X \rangle$, luego tendríamos que $P \in \langle X \rangle$ y 7.9 implica que $X \cup \{P\}$ no es independiente.

- Si $E_1 \subset E_2$ son variedades afines y E_2 es finitamente generada, entonces E_1 también lo es. Además $\dim E_1 \leq \dim E_2$ y la igualdad se da si y sólo si $E_1 = E_2$.

Tomamos un punto $P_0 \in E_1$, si $\langle P_0 \rangle \neq E_1$, tomamos $P_1 \in E_1 \setminus \langle P_0 \rangle$, con lo que P_0, P_1 son independientes. Si $\langle P_0, P_1 \rangle \neq E_1$, tomamos otro punto $P_2 \in E_1 \setminus \langle P_0, P_1 \rangle$, con lo que P_0, P_1, P_2 son independientes. Si $\dim E_2 = d$, entonces E_2 (y en particular E_1) no puede contener más de $d + 1$ puntos afinmente independientes, luego tras un número finito de pasos llegamos a que $E_1 = \langle P_0, \dots, P_n \rangle$, con $n = \dim E_1$, y si $n = d$ entonces $E_1 = E_2$ porque sólo hay una variedad de dimensión n que pueda contener a P_0, \dots, P_n .

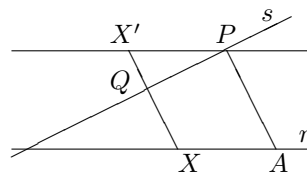
Llamamos *planos* a las variedades afines de dimensión 2.

Teorema 7.11 *Dos rectas distintas son paralelas si y sólo si están contenidas en un mismo plano y no tienen puntos en común.*

DEMOSTRACIÓN: Si r y s son paralelas distintas, sabemos que no tienen puntos en común. Tomamos $A, B \in r$ y $C \in s$, de modo que $r \subset \langle A, B, C \rangle$ y a su vez $s \subset \langle A, B, C \rangle$, por ser paralela a una recta contenida en el plano por un punto del plano.

Recíprocamente, supongamos que r y s están contenidas en un plano Π . Tomamos $A, B \in r$ y $P \in s$, de modo que $\Pi = \langle A, B, P \rangle = AB_{A,P}$, es decir, que Π es la unión de las rectas paralelas a AP que pasan por puntos de AB . Sea Q otro punto cualquiera de s , de modo que la paralela a AB por Q corta a AB en un punto X .

Si $X = A$ entonces $s = AP$ corta a $r = AB$, luego podemos suponer que $X \neq A$. Por la primera parte del teorema 7.4, la paralela a AX (o sea, a r) por P corta a la paralela a AP por X (que es XQ) en un punto X' . Si $X' = Q$, entonces $s = PX'$ es paralela a r .



En caso contrario aplicamos el axioma A3 a los puntos P, Q, X' . Como la recta QX' es la misma que QX , concluimos que la paralela a PQ por Q (que es $PQ = s$) corta a la paralela a PX' por X (que es r). Así pues, en todos los casos r y s son paralelas o se cortan. ■

Ahora es inmediato que, con la definición de plano que hemos dado, una geometría afín cumple todos los axiomas de incidencia adimensionales A1–A5, así como la definición de paralelas como rectas iguales o coplanares disjuntas. Por otra parte, a partir de estos axiomas se prueba fácilmente que los planos cumplen la definición de plano que hemos dado aquí, es decir, que si consideramos a una geometría que cumpla los axiomas adimensionales A1–A5 como espacio afín, los planos que definimos en ella son los que ya venían dados desde un principio.

Recapitulando, los axiomas de la geometría afín son más fuertes que los axiomas A1–A5 del capítulo I, puesto que contienen, por ejemplo, la transitividad del paralelismo, pero hemos visto que todos ellos pueden demostrarse si añadimos los axiomas de ordenación. Lo que hemos hecho ha sido “trasladar” parte de la información que estaba contenida en los axiomas de ordenación a unos axiomas de incidencia ligeramente más fuertes.

Sin embargo, también hemos visto que todos los axiomas de la geometría afín pueden demostrarse (sin necesidad de los axiomas de ordenación) a partir de los axiomas de incidencia de la geometría plana o de la geometría tridimensional.

Más aún, los axiomas de incidencia de la geometría plana equivalen a los axiomas de la geometría afín más un axioma que afirme que no existen puntos no coplanares o, equivalentemente, que dos rectas son paralelas si y sólo si son disjuntas. Igualmente, los axiomas de incidencia de la geometría tridimensional son equivalentes a los axiomas de la geometría afín más los axiomas A5 y A6 (aunque de hecho A5 puede reemplazar a A4).

Veamos por último que A6 puede probarse a partir de los axiomas de la geometría afín relativizado a variedades de dimensión 3:

Teorema 7.12 *Si dos planos están contenidos en una variedad afín de dimensión 3 y tienen un punto en común, entonces tienen una recta en común.*

DEMOSTRACIÓN: Sea E una variedad afín de dimensión 3 y sean Π_1, Π_2 dos planos contenidos en E . Supongamos que su intersección consta de un único punto P . Pongamos que $\Pi_2 = \langle P, Q, R \rangle$. Como $Q \notin \Pi_1$, tenemos que $\langle \Pi_1, Q \rangle$ tiene dimensión 3, luego $E = \langle \Pi_1, Q \rangle = (\Pi_1)_{P,Q}$.

Por consiguiente, la paralela a PQ por R corta a Π_1 en un punto X . Dicha paralela está contenida en Π_2 , luego $X \in \Pi_1 \cap \Pi_2$, luego $X = P$, pero entonces la paralela a PQ por R es PQ , luego P, Q, R son colineales, contradicción. ■

Definimos dos planos paralelos como dos planos que, o bien son iguales, o bien son disjuntos y están contenidos en una variedad afín tridimensional.

Teorema 7.13 *Por cada punto pasa un único plano paralelo a un plano dado.*

DEMOSTRACIÓN: Sea Π un plano arbitrario y P un punto, que podemos suponer exterior a él. La construcción del teorema 7.1 vale ahora literalmente en cualquier geometría afín (pues los axiomas de separación sólo hacían falta para construir las variedades tridimensionales) y prueba que existe un plano Π' paralelo a Π por P , contenido en $E = \langle \Pi, P \rangle$. Supongamos que hubiera otro plano paralelo $\Pi'' \subset E$.

Sea $Q \in \Pi$ y sea $Q' \in \Pi'$ un punto que no esté en $\Pi' \cap \Pi''$. Consideramos el plano $\Pi^* = \langle Q, Q', P \rangle \subset E$. Entonces Π^* corta a Π en una recta r que contiene a Q , corta a Π' en la recta $r' = PQ'$, que no es la intersección $\Pi' \cap \Pi''$, luego corta a Π'' en una recta r'' que contiene a P y que tampoco es $\Pi' \cap \Pi''$. Entonces r' y r'' son rectas disjuntas de r y todas están en el plano Π^* , luego r' y r'' son dos paralelas a r por P , contradicción. ■

Podríamos obtener más resultados sobre variedades afines en general, pero no nos va a ser necesario.

7.3 Introducción de coordenadas

Ahora estamos en condiciones de probar que, salvo en un caso especial que discutiremos en su momento, los espacios afines finitamente generados coinciden con el concepto de espacio afín de la geometría analítica (en un contexto ligeramente más general en el que admitimos que los escalares varíen en un anillo de división en lugar de en un cuerpo).

7.3.1 Homotecias y traslaciones

El teorema 5.8 caracteriza las homotecias y las traslaciones de un espacio afín como las biyecciones que transforman cada recta en una paralela. Lo más destacable es que no hace falta suponer que las biyecciones sean afines, por lo que estamos en condiciones de tomar esta caracterización como definición.¹

Definición 7.14 Dado un espacio afín E , llamaremos $\text{HT}(E)$ al conjunto de todas las aplicaciones biyectivas de E en sí mismo que transforman cada recta en una paralela. Es claro que $\text{HT}(E)$ es un grupo con la composición de aplicaciones. Un elemento de $\text{HT}(E)$ es una *homotecia* si tiene al menos un punto fijo y es una *traslación* si es la identidad o bien no tiene puntos fijos.

Probamos en primer lugar un resultado general:

Teorema 7.15 *Si dos elementos de $\text{HT}(E)$ coinciden en dos puntos distintos cualesquiera, entonces son iguales.*

DEMOSTRACIÓN: Sean P y Q dos puntos de E y $f \in \text{HT}(E)$. Veamos que la imagen de cualquier otro punto R está determinada por las de P y Q . Supongamos primeramente que R no está en la recta PQ . Entonces $f[PR]$ es una recta paralela a PR , luego es exactamente la recta paralela a PR que pasa por $f(P)$. Similarmente $f[QR]$ es la recta paralela a QR que pasa por $f(Q)$. Como PR y QR no son paralelas entre sí, las rectas $f[PR]$ y $f[QR]$ tampoco lo son. En particular no son coincidentes, pero tienen en común el punto $f(R)$. Por lo tanto $f(R)$ es necesariamente la intersección de la paralela a PR por

¹Éste es un buen ejemplo del enfoque que estamos siguiendo aquí: tomamos como definición un hecho que intuitivamente no es trivial, pero que permite deducir las propiedades de las homotecias y las traslaciones a partir únicamente de los axiomas de incidencia.

$f(P)$ con la paralela a QR por $f(Q)$, lo cual sólo depende de $f(P)$ y $f(Q)$. Con esto hemos probado que dos elementos cualesquiera de $\text{HT}(E)$ que coincidan sobre PQ han de coincidir sobre E salvo quizá la recta PQ . Tomando una recta paralela a PQ , sobre la que ya tenemos que coinciden, el razonamiento anterior prueba que también coinciden en PQ . ■

En particular cada homotecia distinta de la identidad tiene un único punto fijo, al que llamaremos *centro* de la homotecia.

Probamos ahora una caracterización de las homotecias y traslaciones mediante una propiedad más débil que nos será útil después:

Teorema 7.16 *Sea $f : E \rightarrow E$ una aplicación tal que para todo par de puntos P y Q , $f(Q)$ está sobre la recta paralela a PQ que pasa por $f(P)$. Entonces f es constante o bien $f \in \text{HT}(\mathbb{E})$.*

DEMOSTRACIÓN: Supongamos que existen dos puntos P y Q tales que $f(P) = f(Q) = X$. Entonces, para todo punto R exterior a la recta PQ , tenemos que $f(R)$ ha de estar en la recta paralela a PR por X y en la recta paralela a QR por X . Como estas rectas no son paralelas entre sí, se cortan sólo en X , luego $f(R) = X$ para todo punto fuera de PQ . Sustituyendo P y Q por dos puntos en una recta paralela a PQ , el razonamiento anterior nos da que $f(R) = X$ para todo punto de PQ , luego f es constante.

Dicho de otro modo, si f no es constante, entonces es inyectiva. En tal caso consideremos dos puntos cualesquiera P y Q y sea R' un punto exterior a la recta $f(P)f(Q)$. Por hipótesis las rectas PQ y $f(P)f(Q)$ son paralelas, luego están contenidas en un mismo plano Π . Sea Π' el plano paralelo a $f(P)f(Q)R'$ por P .

Observemos que todos los planos están en la variedad afín $E = \langle \Pi, R' \rangle$, luego Π' corta a Π en una recta paralela a $f(P)f(Q)$ por P , es decir, en la recta PQ . El plano $Pf(P)R'$ corta a Π' en una r recta paralela a $f(P)R'$, y el plano $Qf(Q)R'$ corta a Π' en una s recta paralela a $f(Q)R'$ por Q . Como ambas están contenidas en Π' y no pueden ser paralelas (porque $f(P)R'$ y $f(Q)R'$ no lo son) necesariamente se cortan en un punto R . Por hipótesis $f(R)$ está en la paralela a PR por $f(P)$ y en la paralela a QR por $f(Q)$, luego está en $r \cap s$ y por lo tanto $f(R) = R'$. Esto prueba que todos los puntos exteriores a la recta $f(P)f(Q)$ tienen antiimagen por f . Cambiando de puntos probamos lo mismo para los puntos de la recta exceptuada. Por lo tanto f es biyectiva.

Finalmente, dada una recta r y un punto $P \in r$, la imagen de cualquier punto $Q \in r$ ha de estar en la paralela a r por $f(P)$, llamémosla r' . Recíprocamente, todo punto de r' es de la forma $f(Q)$, para un cierto punto Q que ha de estar en r , o en caso contrario $f(Q)$ estaría en la paralela a PQ por $f(P)$, que sería una recta distinta a r' . Por lo tanto $f[r] = r'$ y así $f \in \text{HT}(E)$. ■

Definición 7.17 Una *traza* de $f \in \text{HT}(E)$ es una recta r tal que $f[r] = r$.

Tres hechos obvios sobre trazas son los siguientes:

1. Las trazas de la identidad son todas las rectas.
2. Si $f \in \text{HT}(E)$, entonces cada punto P que no sea un punto fijo para f está en una única traza, a saber en la recta $Pf(P)$.
3. Si dos trazas se cortan en un punto, entonces éste es un punto fijo.

Un *haz de rectas paralelas* es una clase de equivalencia de rectas paralelas, bien de todo el espacio bien en un plano prefijado. Un *haz de rectas concurrentes* es el conjunto de todas las rectas (del espacio o de un plano) que pasan por un mismo punto.

Teorema 7.18 *Las trazas de una homotecia distinta de la identidad son el haz de rectas que pasan por su centro. Las trazas de una traslación distinta de la identidad forman un haz de rectas paralelas al que llamaremos dirección de la traslación.*

DEMOSTRACIÓN: Si f es una homotecia de centro P , entonces toda recta PQ es una traza, pues $f[PQ] = Pf(Q)$ ha de ser una recta paralela a PQ , luego ha de ser PQ . Recíprocamente, si QR es una traza pero no coincide con PQ , que también es una traza, entonces $f(Q) = Q$, luego $f(Q) = Q$, luego $Q = P$. En cualquier caso la recta QR pasa por P .

Si f es una traslación, las trazas de f son claramente las rectas de la forma $Pf(P)$. Hemos de ver que dos cualesquiera de ellas son paralelas. Sea $Qf(Q)$ una recta distinta. Como f no tiene puntos fijos, las dos rectas son disjuntas. Hemos de probar que están sobre un mismo plano. Ahora bien, $f[PQ]$ es una recta paralela a PQ , luego P , Q , $f(P)$ y $f(Q)$ están sobre un mismo plano, luego $Pf(P)$ y $Qf(Q)$ también. ■

Teorema 7.19 *Si dos traslaciones coinciden sobre un punto son iguales.*

DEMOSTRACIÓN: Sea f una traslación y P un punto arbitrario. Sea Q un punto fuera de $Pf(P)$. Entonces las trazas $Pf(P)$ y $Qf(Q)$ son paralelas, luego $Qf(Q)$ es exactamente la paralela a $Pf(P)$ por Q . Igualmente $f(P)f(Q)$ es la paralela a PQ por $f(P)$. Como $Pf(P)$ no es paralela a PQ , las rectas $Qf(Q)$ y $f(P)f(Q)$ no pueden ser paralelas, luego no son coincidentes, luego $f(Q)$ es exactamente la intersección de la paralela a $Pf(P)$ por Q y la paralela a PQ por $f(P)$. Esto prueba que dos traslaciones que coincidan sobre P coinciden también sobre un segundo punto Q , luego son iguales por 7.15. ■

Teorema 7.20 *Las traslaciones forman un subgrupo normal de $\text{HT}(E)$, al que representaremos por $\text{T}(E)$. Las traslaciones con una misma dirección forman también un subgrupo normal.*

DEMOSTRACIÓN: Es obvio que la inversa de una traslación es una traslación (si f no tiene puntos fijos, f^{-1} tampoco los tiene).

Si f y g son traslaciones, entonces $f \circ g$ es una traslación, pues si tiene un punto fijo $(f \circ g)(P) = P$, tenemos que $f(P) = g^{-1}(P)$, luego $f = g^{-1}$, ya que ambas son traslaciones y coinciden en un punto. Por consiguiente $f \circ g = 1$, que también es una traslación.

Esto prueba que $T(\mathbb{E})$ es un subgrupo de $HT(E)$. Veamos que es normal. Hemos de ver que si $f \in T(\mathbb{E})$ y $g \in HT(\mathbb{E})$, entonces $g^{-1}fg \in T(\mathbb{E})$. En efecto, si $(g^{-1}fg)(P) = P$, entonces $f(g^{-1}(P)) = g^{-1}(P)$, luego f tiene un punto fijo y es, por tanto, la identidad. Así $g^{-1}fg = 1 \in T(\mathbb{E})$.

Es fácil ver que la composición de dos traslaciones con la misma dirección es la identidad o una traslación con la misma dirección, así como que la inversa de una traslación no trivial tiene la misma dirección.

Si $f \neq 1$ se cumple que $g^{-1}fg$ es una traslación con la misma dirección que f . En efecto, dado $P \in E$, la recta $g^{-1}(P)$, $f(g^{-1}(P))$ es una traza de f , luego tiene la dirección de f . Su imagen por g es $P(g^{-1}fg)(P)$, que es paralela a la anterior y es una traza de $(g^{-1}fg)$, luego ambas traslaciones tienen la misma dirección. ■

Teorema 7.21 *Si existen traslaciones en dos direcciones distintas entonces $T(E)$ es abeliano.*

DEMOSTRACIÓN: Sean f y g dos traslaciones cualesquiera. Hemos de probar que $fg = gf$ o, equivalentemente, que $f^{-1}g^{-1}fg = 1$. Podemos suponer que $f \neq 1 \neq g$. Supongamos que f y g tienen direcciones distintas. Si $f^{-1}g^{-1}fg \neq 1$, entonces $g^{-1}fg$ tiene la misma dirección que f , al igual que f^{-1} y también la composición $f^{-1}g^{-1}fg$ (supuesta no trivial), pero del mismo modo se razona que su dirección es la de g , con lo que tenemos una contradicción.

Si f y g tienen la misma dirección, por hipótesis existe otra traslación h con dirección distinta. Por la parte ya probada $fh = hf$. Por otra parte, gh no puede tener la misma dirección que f , o de lo contrario $g^{-1}(gh) = h$ tendría la misma dirección. Así pues, $fgh = ghf$. Uniendo todo esto tenemos que $fgh = ghf = gfh$, luego $fg = gf$. ■

7.3.2 Vectores y escalares

La insólita hipótesis del teorema anterior nos hace reparar en que ninguno de los teoremas anteriores nos garantiza que el grupo $HT(E)$ sea no trivial. Garantizar la existencia de homotecias y traslaciones es un problema técnico del que nos ocuparemos en la subsección siguiente. Para sortear el obstáculo momentáneamente introduciremos dos axiomas de carácter provisional. El primero se refiere a traslaciones:

Axioma B1 *Dados dos puntos P y Q , existe una traslación T_{PQ} que transforma P en Q .*

Puesto que las traslaciones están determinadas por la imagen de un único punto, la traslación T_{PQ} es única. Ahora es obvio que existen traslaciones con direcciones distintas, por lo que el grupo de traslaciones es abeliano.

Con esto podemos recuperar el conjunto de los vectores libres. Llamaremos $\vec{E} = T(E)$. Usaremos notación aditiva, de modo que $\vec{0}$ será la identidad. La traslación T_{PQ} la representaremos por \overrightarrow{PQ} , de modo que todo vector de \vec{E} se puede expresar de forma única como \overrightarrow{OP} para cualquier punto prefijado O .

La imagen de un punto R por la traslación \overrightarrow{PQ} la representaremos $R + \overrightarrow{PQ}$. Las afirmaciones siguientes son reformulaciones de hechos ya probados:

1. $\overrightarrow{PP} = \vec{0}$,
2. $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$,
3. $(P + \vec{v}) + \vec{w} = P + (\vec{v} + \vec{w})$.

Ahora buscamos un cuerpo con el que convertir a \vec{E} en un espacio vectorial. Hemos visto que en un espacio afín (en el sentido algebraico) K^* es isomorfo al grupo de homotecias lineales $H(\vec{E})$. Más aún, si añadimos la aplicación nula a $H(\vec{E})$, el conjunto que obtenemos puede ser dotado de forma natural de estructura de cuerpo isomorfo a K . Basta definir la suma de dos homotecias f y g como la aplicación dada por $(f + g)(\vec{v}) = f(\vec{v}) + g(\vec{v})$. Fijando una base, esta suma y la composición se corresponden con la suma y el producto de matrices, pero las matrices de las homotecias lineales (más la aplicación nula) son las de la forma αI , donde $\alpha \in K$, luego efectivamente, con estas operaciones tenemos un cuerpo isomorfo a K .

Nuestro problema es esencialmente definir las homotecias lineales en nuestro contexto, pero una homotecia lineal se caracteriza por que envía cada vector a un múltiplo suyo y, vistos como traslaciones, esto significa que ambos tienen la misma dirección.

Definición 7.22 Una *homotecia lineal* en un espacio afín E es un homomorfismo de grupos $\alpha : \vec{E} \rightarrow \vec{E}$ tal que para todo $\vec{v} \in \vec{E}$, se cumple $\alpha\vec{v} = \vec{0}$ o bien $\alpha\vec{v}$ tiene la misma dirección que \vec{v} . Llamaremos K al conjunto de todas las homotecias lineales de E . A los elementos de K los llamaremos también *escalares*.

Notemos que el hecho de que una homotecia lineal α sea un homomorfismo de grupos equivale a que $\alpha(\vec{v} + \vec{w}) = \alpha\vec{v} + \alpha\vec{w}$.

El homomorfismo nulo, determinado por $0\vec{v} = \vec{0}$ para todo $\vec{v} \in \vec{E}$, es trivialmente una homotecia lineal. Lo mismo le sucede a la identidad, determinada por $1\vec{v} = \vec{v}$. Llamaremos -1 a la aplicación dada por $(-1)\vec{v} = -\vec{v}$. Es claro que también es una homotecia lineal.

Definición 7.23 Sean $\alpha, \beta \in K$, definimos $\alpha + \beta$ como la aplicación $\vec{E} \rightarrow \vec{E}$ dada por $(\alpha + \beta)\vec{v} = \alpha\vec{v} + \beta\vec{v}$. Definimos $\alpha\beta$ como la aplicación dada por $(\alpha\beta)\vec{v} = \beta(\alpha(\vec{v}))$.

Notar que, vistos como homotecias lineales, el producto de dos escalares es simplemente su composición.

Teorema 7.24 *El conjunto de escalares K es un anillo unitario con las operaciones de la definición anterior.*

DEMOSTRACIÓN: Claramente,

$$(\alpha + \beta)(\vec{v} + \vec{w}) = \alpha(\vec{v} + \vec{w}) + \beta(\vec{v} + \vec{w}) = \alpha\vec{v} + \alpha\vec{w} + \beta\vec{v} + \beta\vec{w} = (\alpha + \beta)\vec{v} + (\alpha + \beta)\vec{w},$$

luego $\alpha + \beta$ es un homomorfismo. Si $\vec{v} = \vec{0}$ entonces $(\alpha + \beta)\vec{v} = \vec{0}$. Si $\vec{v} \neq \vec{0}$ entonces $\alpha\vec{v}$ y $\beta\vec{v}$ tienen la misma traza que \vec{v} , luego su suma también, es decir, $(\alpha + \beta)\vec{v}$ tiene la misma dirección que \vec{v} , luego $\alpha + \beta \in K$.

Del mismo modo se prueba que $\alpha\beta \in K$. La asociatividad y la conmutatividad de la suma de siguen de las de la suma en \vec{E} .

El neutro es 0 y el simétrico de α es $(-1)\alpha$, pues

$$(\alpha + (-1)\alpha)\vec{v} = \alpha\vec{v} + -\alpha\vec{v} = \vec{0} = 0\vec{v},$$

luego $\alpha + (-1)\alpha = 0$.

Tampoco presenta ningún problema probar la propiedad distributiva (por la izquierda y la derecha), la asociatividad del producto y el hecho de que 1 es el elemento neutro del producto (por la izquierda y la derecha). ■

Para justificar la existencia de inversos mostraremos primero la relación entre las homotecias y las homotecias lineales:

Teorema 7.25 *Sea $\alpha \in K$ no nulo y P un punto dado. Existe una única homotecia f_α de centro P tal que $T_{\alpha\vec{v}} = f_\alpha^{-1}T_{\vec{v}}f_\alpha$, para todo $\vec{v} \in \vec{E}$.*

DEMOSTRACIÓN: Observar que usamos la doble notación $\vec{v} = T_{\vec{v}}$ según si vemos la traslación como elemento de \vec{E} (con notación aditiva) o como elemento de $\text{HT}(\vec{E})$ (con notación multiplicativa).

Veamos en primer lugar la unicidad. Sea Q un punto arbitrario y $\vec{v} = T_{PQ}$. Aplicando la fórmula del enunciado al punto P tenemos

$$P + \alpha\overrightarrow{PQ} = T_{\alpha\vec{v}}(P) = f_\alpha(T_{\vec{v}}(f_\alpha^{-1}(P))) = f_\alpha(T_{\vec{v}}(P)) = f_\alpha(Q).$$

Esto prueba que $f_\alpha(Q)$ está completamente determinado por α y P . Además la fórmula anterior nos indica cómo hemos de construir f_α .

Definimos f_α mediante

$$f_\alpha(Q) = P + \alpha\overrightarrow{PQ}. \quad (7.1)$$

Dados dos puntos distintos Q y R , tenemos que $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$, luego $P + \alpha\overrightarrow{PQ} + \alpha\overrightarrow{QR} = P + \alpha\overrightarrow{PR}$. Por (7.1) esto implica

$$f_\alpha(Q) + \alpha\overrightarrow{QR} = f_\alpha(R). \quad (7.2)$$

Si $\alpha\overrightarrow{QR} \neq \vec{0}$ entonces esta ecuación significa que la recta $f_\alpha(Q)f_\alpha(R)$ es una traza de $\alpha\overrightarrow{QR}$, luego también de \overrightarrow{QR} , luego es paralela a la recta QR , que es otra traza de la misma traslación. Si $\alpha\overrightarrow{QR} = \vec{0}$ tenemos que $f_\alpha(Q) = f_\alpha(R)$. En ambos casos $f_\alpha(R)$ está sobre la recta paralela a QR por $f_\alpha(Q)$. Podemos aplicar el teorema 7.16 para concluir que f_α es constante o bien una homotecia

(no puede ser una traslación no trivial porque tiene a P como punto fijo). Ahora bien, si fuera constante, la fórmula (7.2) implicaría que $\alpha\overrightarrow{QR} = \vec{0}$ para todo vector \overrightarrow{QR} , luego sería $\alpha = 0$, en contra de la hipótesis. Así pues, f_α es una homotecia de centro P .

Veamos que f_α cumple lo pedido. Puesto que $f_\alpha(P) = P$, podemos escribir (7.1) como

$$f_\alpha(Q) = f_\alpha(P) + \alpha\overrightarrow{PQ} = T_{\alpha\vec{v}}(f_\alpha(P)),$$

donde $\vec{v} = \overrightarrow{PQ}$ es un vector arbitrario, luego

$$Q = f_\alpha^{-1}\left(T_{\alpha\vec{v}}(f_\alpha(P))\right).$$

Esto significa que la traslación $f_\alpha T_{\alpha\vec{v}} f_\alpha^{-1}$ envía P a Q , luego

$$f_\alpha T_{\alpha\vec{v}} f_\alpha^{-1} = T_{PQ} = T_{\vec{v}}.$$

Despejando obtenemos la fórmula del enunciado: $T_{\alpha\vec{v}} = f_\alpha^{-1} T_{\vec{v}} f_\alpha$. ■

Teorema 7.26 *Cada homotecia se expresa de forma única como f_α , para un cierto escalar α no nulo.*

DEMOSTRACIÓN: Sea f una homotecia de centro P . Sea α la homotecia lineal dada por $T^\alpha = f^{-1}Tf$ (usamos notación exponencial para los escalares cuando usamos notación multiplicativa para las traslaciones). Es fácil ver que α es un homomorfismo: $(TT')^\alpha = f^{-1}TT'f = f^{-1}Tff^{-1}T'f = T^\alpha T'^\alpha$. Además α conserva las direcciones por el teorema 7.20. Por lo tanto α es un escalar, claramente no nulo. Cambiando la notación en la definición de α tenemos que $T_{\alpha\vec{v}} = f^{-1}T_{\vec{v}}f$ para todo vector \vec{v} , luego por la unicidad ha de ser $f = f_\alpha$.

La unicidad se sigue de (7.1), pues si $f_\alpha = f_\beta$, entonces $\alpha\overrightarrow{PQ} = \beta\overrightarrow{PQ}$ para todo vector \overrightarrow{PQ} , luego $\alpha = \beta$. ■

Con esto ya estamos en condiciones de probar:

Teorema 7.27 *El anillo de escalares K es un anillo de división.*

DEMOSTRACIÓN: Recordemos que un anillo de división es un cuerpo no necesariamente conmutativo. Sólo falta probar que todo escalar no nulo tiene un inverso. Más en general, observamos que la biyección $\alpha \mapsto f_\alpha$ entre K^* y el grupo de las homotecias de centro un punto fijo P conserva el producto, es decir, $f_{\alpha\beta} = f_\alpha f_\beta$. En efecto, se cumple

$$(f_\alpha f_\beta)^{-1} T_{\vec{v}} f_\alpha f_\beta = f_\beta^{-1} T_{\alpha\vec{v}} f_\beta = T_{\beta(\alpha\vec{v})} = T_{(\alpha\beta)\vec{v}}.$$

Por la unicidad $f_\alpha f_\beta = f_{\alpha\beta}$. Esto implica que K^* es un grupo isomorfo al grupo de homotecias de centro P . En particular todo escalar no nulo tiene un inverso. ■

Con esto tenemos probado que \vec{E} es un K -espacio vectorial por la derecha, es decir, con la asociatividad mixta en la forma $(\alpha\beta)\vec{v} = \beta(\alpha\vec{v})$, con lo que E es un espacio afín sobre K salvo por el hecho de que K no es necesariamente conmutativo. No podemos aspirar a demostrar la conmutatividad de K a partir de nuestros axiomas, pues es fácil ver que si K es cualquier anillo de división entonces K^n satisface los axiomas de espacio afín tomando como rectas las clases de congruencia módulo los subespacios vectoriales de dimensión 1, y el anillo de división que se obtiene según la construcción anterior es isomorfo a K , luego no es necesariamente conmutativo. La geometría afín básica es válida en el caso no conmutativo, pero hay algunas diferencias, como por ejemplo que las homotecias no son necesariamente afinidades.

Para calcular la dimensión de \vec{E} necesitamos un axioma adicional B2 que introducimos también con carácter provisional. En la subsección siguiente demostraremos los axiomas B1 y B2 a partir del grupo A.

Axioma B2 *Si \vec{v} y \vec{w} son dos vectores distintos y no nulos con la misma dirección, entonces existe un $\alpha \in K$ tal que $\vec{w} = \alpha\vec{v}$.*

Los casos exceptuados se cumplen trivialmente: si $\vec{v} = \vec{w}$ sirve $\alpha = 1$ y si $\vec{w} = \vec{0}$ sirve $\alpha = 0$. Por lo tanto sólo hay que exigir que \vec{v} sea no nulo. La estructura vectorial de \vec{E} implica que α es único.

En primer lugar daremos una forma equivalente del axioma B2. Si B1 garantizaba la existencia de traslaciones, el axioma B2 postula el hecho análogo sobre homotecias:

Axioma B2(P) (para un punto P) *Para todo par de puntos Q, R tales que P, Q, R estén alineados y sean distintos, existe una homotecia de centro P que envía Q a R .*

Teorema 7.28 *El axioma B2 para un punto P equivale al axioma B2 para todo punto P y también al axioma B2.*

DEMOSTRACIÓN: Supongamos el axioma B2 para un punto P y sean \vec{v}, \vec{w} en las condiciones del axioma B2 (global). Sean $Q = P + \vec{v}$, $R = P + \vec{w}$. Como los dos vectores tienen la misma dirección, los tres puntos son colineales. Como son distintos y no nulos, los tres puntos P, Q, R son distintos. Por B2P existe una homotecia de centro P , que será de la forma f_α , tal que $f_\alpha(Q) = R$, es decir, $P + \alpha\vec{PQ} = R$, luego $\alpha\vec{PQ} = \vec{PR}$, o también, $\alpha\vec{v} = \vec{w}$.

Supongamos ahora B2 y probemos B2P para un punto arbitrario. Sean P, Q, R tres puntos colineales distintos. Entonces \vec{PQ} y \vec{PR} son vectores no nulos distintos con la misma dirección. por B2 existe un escalar no nulo tal que $\vec{PR} = \alpha\vec{PQ}$. Equivalentemente, $P + \alpha\vec{PQ} = R$, luego $f_\alpha(Q) = R$. ■

Teorema 7.29 *Si $E = \langle P_0, \dots, P_n \rangle$, donde los puntos son afínmente independientes, y $\vec{v}_i = \vec{P_0P_i}$, entonces los vectores $\vec{v}_1, \dots, \vec{v}_n$ son una base de \vec{E} .*

DEMOSTRACIÓN: Sea $\overrightarrow{P_0X}$ un vector arbitrario de \vec{E} . La paralela a P_0P_n por X corta a $\langle P_0, \dots, P_{n-1} \rangle$ en un punto X_1 , de modo que

$$\overrightarrow{P_0X} = \overrightarrow{P_0X_1} + \overrightarrow{X_1X}.$$

El vector $\overrightarrow{X_1X}$ es nulo o tiene la misma dirección que $\vec{v}_n = \overrightarrow{P_0P_n}$, luego por B2 existe un escalar α_n tal que $\overrightarrow{X_1X} = \alpha_n \vec{v}_n$.

Análogamente, la paralela a P_0P_{n-1} por X_1 corta a $\langle P_0, \dots, P_{n-2} \rangle$ en un punto X_2 , de modo que

$$\overrightarrow{P_0X} = \overrightarrow{P_0X_1} + \alpha_n \vec{v}_n = \overrightarrow{P_0X_2} + \overrightarrow{X_2X_1} + \alpha_n \vec{v}_n.$$

A su vez, el vector $\overrightarrow{X_2X_1}$ es nulo o tiene la dirección de \vec{v}_{n-1} , luego existe un escalar α_{n-1} tal que

$$\overrightarrow{P_0X} = \overrightarrow{P_0X_1} + \alpha_n \vec{v}_n = \overrightarrow{P_0X_2} + \alpha_{n-1} \vec{v}_{n-1} + \alpha_n \vec{v}_n.$$

Tras n pasos llegamos a que $\overrightarrow{P_0X} = \alpha_1 \vec{v}_1 + \dots + \alpha_n \vec{v}_n$, luego $\vec{v}_1, \dots, \vec{v}_n$ son un sistema generador de \vec{E} .

Supongamos ahora que $\alpha_1 \vec{v}_1 + \dots + \alpha_n \vec{v}_n = \vec{0}$, con algún coeficiente no nulo. No perdemos generalidad si suponemos que $\alpha_n \neq 0$. Entonces

$$P_0 - \alpha_n \overrightarrow{P_0P_n} = P_0 + \alpha_1 \overrightarrow{P_0P_1} + \dots + \alpha_{n-1} \overrightarrow{P_0P_{n-1}}.$$

Ahora bien, el miembro izquierdo está en la recta P_0P_n , mientras que el miembro derecho está en $\langle P_0, \dots, P_{n-1} \rangle$, pues $P_0 + \alpha_1 \overrightarrow{P_0P_1}$ está en $\langle P_0, P_1 \rangle$, de donde $P_0 + \alpha_1 \overrightarrow{P_0P_1} + \alpha_2 \overrightarrow{P_0P_2}$ está en una recta que pasa por P_2 y un punto de $\langle P_0, P_1 \rangle$, luego está en $\langle P_0, P_1, P_2 \rangle$, etc.

Ahora bien, la recta P_0P_n no puede tener más punto en común con la variedad $\langle P_0, \dots, P_{n-1} \rangle$ que P_0 , pues si hubiera otro punto en la intersección, toda la recta estaría contenida en la variedad, luego $P_n \in \langle P_0, \dots, P_{n-1} \rangle$, y los P_i no serían linealmente independientes.

Por lo tanto, $P_0 - \alpha_n \overrightarrow{P_0P_n} = P_0$, lo que se traduce en que $\alpha_n = 0$, contradicción. Esto prueba que los vectores \vec{v}_i son linealmente independientes. ■

7.3.3 Los teoremas de Desargues y Papos-Pascal

En esta subsección veremos que los axiomas del grupo B se pueden demostrar a partir de los del grupo A. Más concretamente, resultan ser equivalentes a un caso particular del teorema de Desargues:

Teorema de Desargues Sean r_1, r_2 y r_3 tres rectas distintas paralelas o concurrentes en un punto P . Sean R, R' puntos en r_1 , sean S, S' puntos en r_2 y T, T' puntos en r_3 distintos de P si las rectas se cortan. Si RS es paralela a $R'S'$ y RT es paralela a $R'T'$ entonces ST es paralela a $S'T'$.

Observemos que si los puntos R, S, T están alineados, lo mismo sucede con R', S', T' , luego el teorema es trivialmente cierto. Además, si $Q = Q'$ entonces $R = R'$ y $T = T'$.

Vamos a probar este teorema a partir de los axiomas del grupo A para espacios de dimensión ≥ 3 , pero antes probaremos su equivalencia con los axiomas B1 y B2. Llamaremos DA al enunciado anterior cuando las rectas son paralelas y DP al enunciado anterior cuando las rectas se cortan en el punto P . Trabajamos en un espacio afín E sin suponer más que los axiomas del grupo A.

Teorema 7.30 *DA equivale al axioma B1 y DP equivale a B2P.*

DEMOSTRACIÓN: Suponiendo B1 (o B2P) tomamos como f la traslación (o la homotecia de centro P) determinada por $f(Q) = Q'$. Entonces las rectas r_1, r_2, r_3 son trazas de f . La recta $f(Q)f(R)$ es paralela a QR y pasa por $f(Q) = Q'$, luego es $Q'R'$, y como $f(R)$ ha de estar en esta recta y en r_2 , ha de ser $f(R) = R'$, e igualmente $f(S) = S'$. Por consiguiente $R'S' = f[RS]$ es paralela a RS .

Veamos ahora que DA implica B1. Tomemos dos puntos distintos Q y Q' . Vamos a construir una traslación que envíe uno sobre el otro. En primer lugar definimos una aplicación $T^{QQ'}$ definida sólo sobre los puntos exteriores a la recta QQ' . Dado un punto R en estas condiciones, tomamos la recta r paralela a QQ' por R (con lo que $r \neq QQ'$). Las rectas QQ', QR y r están en un mismo plano, luego éste contiene también a la recta paralela a QR por Q' , que cortará a r en un punto R' . Definimos $T^{QQ'}(R) = R'$.

El punto R' está caracterizado por el hecho de que RR' es paralela a QQ' y QR es paralela a $Q'R'$. Es claro entonces que $T^{RR'}(Q) = Q'$. Si probamos que $T^{QQ'}$ y $T^{RR'}$ coinciden en su dominio común, entonces definirán una aplicación sobre todo E , a la que llamaremos $T_{QQ'}$ y probaremos que es la traslación buscada.

Sea, pues, S un punto exterior a QQ' y RR' y sea $S' = T^{QQ'}(S)$. Entonces tenemos tres rectas paralelas distintas, QQ', RR' y SS' . Además QS es paralela a $Q'S'$ y QR es paralela a $Q'R'$. Por DA podemos concluir que RS es paralela a $R'S'$, lo que implica que $S' = T^{RR'}(S)$.

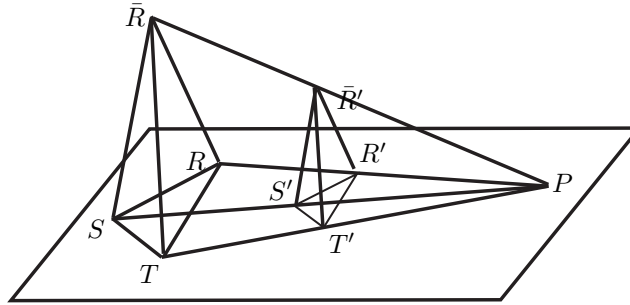
Para probar que $T_{QQ'}$ es una traslación basta probar que cumple las condiciones del teorema 7.16, pues ciertamente $T_{QQ'}$ no es constante y tiene dos trazas paralelas, a saber, QQ' y RR' , luego será una traslación. Tomamos dos puntos U, V y hemos de probar que $V' = T_{QQ'}(V)$ está en la recta paralela a UU' por $U' = T_{QQ'}(U)$. Si U está en QQ' y V está en RR' , podemos tomar un punto S exterior al plano que contiene a estas dos. La aplicación $T^{SS'}$ coincidirá con $T_{QQ'}$ en su dominio común. En cualquier caso, tomando bien, QQ' , bien RR' o bien SS' , encontramos una recta que no contiene a ninguno de los puntos U y V . Podemos suponer que es QQ' .

Si UV es paralela a QQ' , entonces UV contiene también a U' y V' , con lo que ciertamente V' está en la paralela a UV por U' . Si UV no es paralela a QQ' , entonces Q, U y V están dispuestos como antes Q, R y S , y el mismo razonamiento prueba que $U'V'$ es paralelo a UV .

La prueba de que DP implica el axioma $B2P$ es análoga, cambiando traslaciones por homotecias y líneas paralelas por líneas que pasan por P . ■

Nota Todos los resultados que hemos probado hasta ahora se adaptan sin dificultad (de hecho se simplifican) para el caso de un plano afín. La adaptación del teorema anterior requiere una consideración adicional. En un punto de la prueba hemos usado que siempre es posible tomar un punto exterior a dos rectas paralelas (o concurrentes en el caso $DP \Rightarrow B2P$). Para ello hemos tomado un punto exterior al plano que las contiene, lo cual no es viable si el espacio es plano. Si toda recta contiene al menos tres puntos no hay ningún problema, pues basta tomar una recta que corte a las dos dadas y tomar un tercer punto en ella. El único caso peculiar se da si las rectas tienen exactamente dos puntos, pero entonces el plano consta únicamente de cuatro puntos (y seis rectas). En este caso el axioma $B2P$ se cumple trivialmente y la existencia de traslaciones se prueba sin dificultad.

DEMOSTRACIÓN: (del teorema de Desargues para espacios afines de dimensión ≥ 3) Supongamos primero que los planos RST y $R'S'T'$ son distintos. Entonces son paralelos, pues las rectas $R'S'$ y $R'T'$ están contenidas en el plano paralelo a RST por R' , luego éste es $R'S'T'$. El plano que contiene a r_2 y r_3 corta a RST en ST y a $R'S'T'$ en $S'T'$, luego ambas rectas están contenidas en un mismo plano y son paralelas, pues están contenidas en planos paralelos.



Supongamos ahora que $RST = R'S'T'$. Tomemos un punto \bar{R} fuera de este plano. Sea \bar{r} la recta paralela a las r_i por \bar{R} si éstas son paralelas o bien $\bar{r} = \bar{R}P$ si concurren en P .

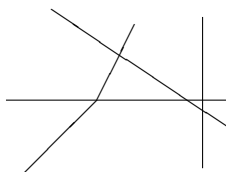
La recta paralela a $R\bar{R}$ por R' está contenida en el plano que contiene a r_1 y a \bar{r} , luego ha de cortar a \bar{r} en un punto \bar{R}' . Los planos $\bar{R}RS$ y $\bar{R}'R'S'$ son distintos y podemos aplicar el caso ya probado del teorema de Desargues para concluir que $\bar{R}S$ es paralela a $\bar{R}'S'$. Del mismo modo concluimos que $\bar{R}T$ es paralela a $\bar{R}'T'$. A su vez esto nos permite aplicar la parte ya probada a los puntos $\bar{R}ST$ y $\bar{R}'S'T'$, lo que nos da que ST es paralela a $S'T'$, como queríamos probar. ■

La demostración anterior es esencialmente tridimensional (en el sentido de que requiere el axioma A5, pero no A6). El ejemplo siguiente muestra que no es posible demostrar el teorema de Desargues a partir de los axiomas de la geometría plana:

Ejemplo Vamos a construir un plano afín que no satisface el teorema de Desargues. Tomamos $\Pi = \mathbb{R}^2$ como conjunto de puntos, aunque las rectas de Π no serán las usuales. Una recta usual en \mathbb{R}^2 está formada por los puntos (x, y) que cumplen una ecuación del tipo $ax + by = c$. Si $b = 0$ tenemos una recta vertical de la forma $x = n$, mientras que si $b \neq 0$ la ecuación es equivalente a una de la forma $y = mx + n$, donde m y n están unívocamente determinados por la recta. Tomamos como rectas de Π las rectas usuales de la forma $x = n$ (verticales), las de la forma $y = mx + n$ con $m \leq 0$ (lo cual incluye a las horizontales) y, para cada $m, n \in \mathbb{R}$, $m > 0$, los conjuntos de pares (x, y) que cumplen

$$y = \begin{cases} 2mx + n & \text{si } y \geq 0 \\ mx + n & \text{si } y < 0 \end{cases}$$

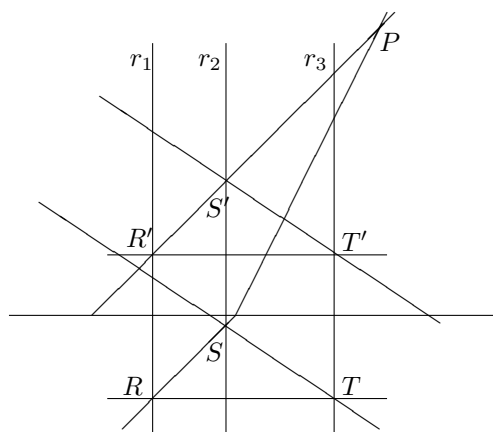
cada uno de estos conjuntos es una línea quebrada formada por la unión de dos semirrectas con origen común en el punto $(n, 0)$, una contenida en el semiplano superior $y \geq 0$ y otra en el inferior. La figura muestra algunas de las rectas de Π .



La comprobación de los tres axiomas del plano afín no presenta ninguna dificultad. El menos evidente es la existencia de una única recta que pasa por dos puntos (a, b) y (c, d) cuando $b < 0$, $d > 0$ y $a < b$. Tal recta ha de ser necesariamente de las quebradas, y basta ver que las ecuaciones $ma + n = b$, $2mc + n = d$ tienen solución única (m, n) con $m > 0$.

La figura de la página siguiente muestra un contraejemplo al teorema de Desargues: vemos dos triángulos en posición de Desargues con dos pares de lados paralelos, mientras que el tercero se corta en P . Dejamos los detalles al lector. ■

Ejercicio: Probar que en Π no hay traslaciones cuyas trazas sean las rectas verticales. Mostrar un contraejemplo del teorema de Desargues en el que las rectas r_i sean concurrentes.



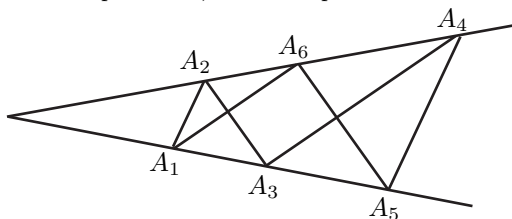
Definición 7.31 Un plano afín es *arguesiano* si en él se cumple el teorema de Desargues.

Si un plano afín Π es sumergible en un espacio afín E (en el sentido de que $\Pi \subset E$ y las rectas de E contenidas en Π son exactamente las rectas de Π) entonces Π es arguesiano, pues E cumple el teorema de Desargues. Recíprocamente, un plano afín arguesiano es isomorfo a un plano afín de la forma K^2 , para un cierto anillo de división K , que puede sumergirse en K^3 , que es un espacio afín. Así pues, los planos arguesianos son exactamente los planos sumergibles en espacios.

Ahora mostraremos un enunciado puramente geométrico equivalente a la conmutatividad del anillo de división asociado a un espacio afín (o un plano arguesiano). Ya hemos observado que el grupo K^* es isomorfo al grupo de las homotecias con un mismo centro P . Así pues, K será un cuerpo si y sólo si este grupo es abeliano.

Definición 7.32 Un *hexágono* es un conjunto de seis puntos ordenados $A_1, A_2, A_3, A_4, A_5, A_6$. Llamaremos *lados* del hexágono a las seis rectas $A_1A_2, A_2A_3, A_3A_4, A_4A_5, A_5A_6, A_6A_1$. Dos lados de un hexágono son *contiguos* si tienen un vértice en común, y en caso contrario son *opuestos*.

Teorema 7.33 (Teorema de Papos–Pascal) Sean r y s dos rectas secantes y consideremos un hexágono tal que los vértices A_1, A_3, A_5 están sobre r y A_2, A_4, A_6 están sobre s (todos ellos distintos del punto de intersección). Si dos pares de lados opuestos son paralelos, el tercer par también lo es.



Vamos a probar que este teorema se cumple exactamente en los espacios afines cuyo anillo de división asociado es un cuerpo. Puesto que todos sus elementos están contenidos en un plano, de hecho se cumple en los espacios afines de dimensión arbitraria.

DEMOSTRACIÓN: Supongamos en primer lugar el teorema de Papos y veamos que K es conmutativo. Sean r y s dos rectas que se corten en un punto P . Sea A_1 un punto sobre r distinto de P . Sea f una homotecia arbitraria de centro P (no trivial). Entonces $f(A_1) = A_3 \neq A_1$ está en r y eligiendo la homotecia apropiada podemos hacer que sea cualquier punto de r prefijado. Tomemos otra homotecia no trivial g de centro P y un punto A_2 en s distinto de P . Sea $A_6 = g(A_2) \neq A_2$, que también estará en s y es un punto arbitrario si elegimos g .

Sean $A_5 = g(f(A_1)) = g(A_3)$ y $A_4 = f(g(A_2)) = f(A_6)$. Claramente A_1A_6 es paralela a $f(A_1)f(A_6) = A_3A_4$ y A_2A_3 es paralela a $g(A_2)g(A_3) = A_6A_5$. Por el teorema de Papos también A_1A_2 es paralela a A_4A_5 .

La conmutatividad $gf = fg$ equivale a que $f(g(A_1)) = g(f(A_1))$, es decir, $f(g(A_1)) = A_5$, pero $f(g(A_1))$ es el punto donde la paralela a A_1A_2 por $f(g(A_2)) = A_4$ corta a r , es decir, se trata efectivamente de A_5 .

Recíprocamente, dado un hexágono en las condiciones del teorema de Papos, digamos con A_1A_6 paralela a A_3A_4 y con A_2A_3 paralela a A_6A_5 , consideramos la homotecia f de centro P que cumple $f(A_1) = A_3$ y la homotecia g de centro P que cumple $g(A_2) = A_6$. Invertiendo el razonamiento y usando que $fg = gf$ concluimos que A_1A_2 es paralela a A_4A_5 . ■

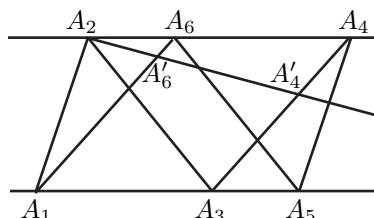
Para terminar probamos que los dos teoremas que acabamos de estudiar no son independientes:

Teorema 7.34 (Hessenberg) *En todo espacio afín en el que se cumple el teorema de Papos-Pascal se cumple también el teorema de Desargues.*

Empezamos demostrando que el teorema de Papos-Pascal vale igualmente cuando las rectas sobre las que se encuentran los seis puntos son paralelas en lugar de secantes.

Suponemos que las rectas A_1A_2 y A_4A_5 son paralelas, al igual que A_1A_6 y A_3A_4 , y tenemos que probar que A_2A_3 y A_5A_6 también lo son.

En primer lugar observamos que si dos de los puntos A_1, A_3, A_5 son iguales, la conclusión es inmediata. Por ejemplo, si se cumple $A_3 = A_5$, tiene que ser $A_2 = A_6$. Si es $A_1 = A_5$, entonces $A_2 = A_4$ y la conclusión coincide con la hipótesis del paralelismo de A_1A_6 y A_3A_4 . Igualmente se razona si $A_1 = A_3$.



Así pues, podemos suponer que A_1, A_3, A_5 son distintos dos a dos y, por simetría, también que A_2, A_4, A_6 son distintos dos a dos.

Sea r la paralela a A_2A_3 que pasa por A_5 . Tiene que cortar a A_1A_6 en un punto A'_6 , pues en caso contrario A_2A_3 sería paralela a A_1A_6 , luego a A_3A_4 y sería $A_2 = A_4$. Basta probar que $A'_6 = A_6$. Supongamos lo contrario.

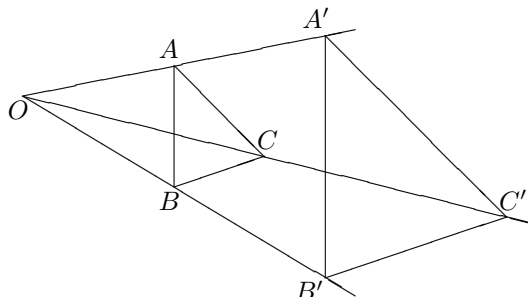
No puede ser $A'_6 = A_2$, pues entonces $A_2 \in A_1A_6$, luego sería $A_2 = A_6$. Sea $s = A_2A'_6 \neq A_2A_6$. Como A_3A_4 es paralela a A_1A_6 y s corta a la segunda, también corta a la primera, digamos en $A'_4 \neq A_4$. Igualmente, s tiene que cortar a A_1A_3 en un punto O .

Sucede entonces que A_1, A_3, A_5 y A_2, A'_4, A'_6 cumplen las hipótesis del teorema de Papos-Pascal para rectas secantes. Ahora sabemos que $A_5A'_6$ es paralela a A_2A_3 y que $A_1A'_6$ es paralela a $A_3A'_4$, luego concluimos que $A_5A'_4$ es paralela a A_1A_2 , luego $A_5A'_4 = A_5A_4$, y el punto de corte de esta recta con A_3A_4 es $A_4 = A'_4$, contradicción. ■

A continuación demostramos el teorema de Desargues para rectas concurrentes. Como ya sabemos que el teorema de Desargues se cumple en todo espacio afín de dimensión mayor que 2, podemos restringirnos al caso de un plano afín.

Tenemos la situación que muestra la figura, en la que suponemos que las tres rectas concurrentes en O son distintas dos a dos, que los tres pares de puntos

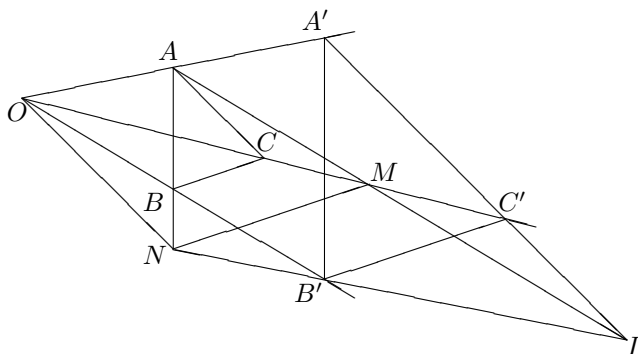
son distintos de O y que las rectas AB y $A'B'$ son paralelas, al igual que AC y $A'C'$, y tenemos que demostrar que también lo son BC y $B'C'$.



No puede suceder que $O \in AB$, pues entonces OA sería igual a OB , en contra de lo supuesto. Igualmente $O \notin AC$ y $O \notin BC$. Distinguimos dos casos:

Caso 1 OB no es paralela a AC .

Entonces la paralela a OB por A no es AC , luego no es paralela a $A'C'$, luego la corta en un punto L . Sea M el punto de corte de OC con AL (existe porque OC corta a OB , luego también a su paralela AL). Por último, sea N el punto de corte de $B'L$ con AB . Existe tal punto, pues en caso contrario $B'L$ sería la paralela AB por B' , pero ésta es $A'B'$, luego sería $L = A'$, pues ambos son la intersección de esta recta con $A'C'$, pero entonces $AL = AA'$ no sería paralela a OB , contradicción.



Aplicamos el teorema de Pappus-Pascal al hexágono formado por los puntos N, B', L y O, A, A' . (Las rectas que los contienen pueden ser paralelas o secantes, pero contamos con el teorema en ambos casos.) Sabemos que $AN = AB$ es paralela a $A'B'$ y que AL es paralela a OB' , luego concluimos que ON es paralela a $A'L = A'C'$, luego también a AC .

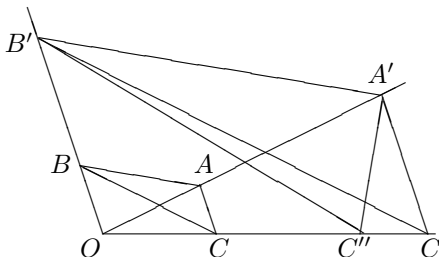
Ahora aplicamos el teorema de Pappus-Pascal al hexágono formado por los puntos N, B, A y O, M, C . Sabemos que $AM = AL$ es paralela a OB y que ON es paralela a AC , luego MN es paralela a BC .

Finalmente consideramos el hexágono formado por O, M, C' y N, B', L . Sabemos que ON es paralela a LC' y que $ML = AL$ es paralela a OB' , luego MN es paralela a $B'C'$.

Así pues, BC es paralela a MN y ésta a $B'C'$, y tenemos la conclusión.

Caso 2 OB es paralela a AC .

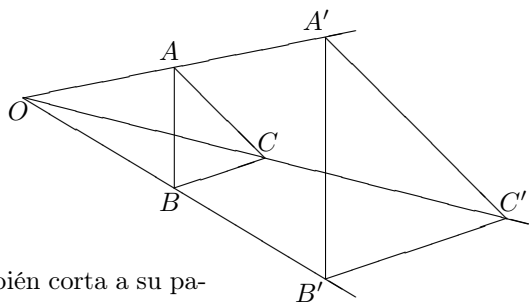
Supongamos, por reducción al absurdo, que BC no es paralela a $B'C'$. Entonces, una de las dos rectas, BC o $B'C'$, no es paralela a OA . Por simetría, no perdemos generalidad si suponemos que es BC . La paralela a BC por B' tiene que cortar a OC en un punto C'' , porque OC corta a BC , luego también a sus paralelas. Además $C'' \neq C'$, porque estamos suponiendo que $B'C'$ no es paralela a BC .



Ahora los puntos B, A, C y B', A', C'' están en las condiciones del caso 1 en lugar de A, B, C y A', B', C' , porque OA no es paralela a BC . La conclusión es que AC es paralela a $A'C''$, pero entonces $A'C'$ es paralela a $A'C''$, luego $C' = C''$, contradicción. ■

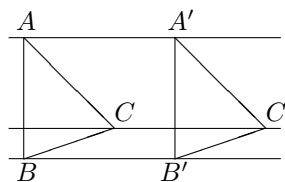
Falta probar el teorema de Desargues para rectas paralelas. Para ello demostramos primero el recíproco del teorema de Desargues para rectas concurrentes:

Ahora suponemos que los lados de ambos triángulos son paralelos dos a dos (pero sobre rectas distintas) y que las rectas AA' y BB' se cortan en el punto O , y vamos a probar que la recta CC' también pasa por O .



Como OC corta a BC , también corta a su paralela $B'C'$ en un punto C'' . Entonces A, B, C y A', B', C'' están en las condiciones del teorema de Desargues ya demostrado, con BC paralela a $B'C'' = B'C'$ y AB paralela a $A'B'$, luego concluimos que AC es paralela a $A'C''$, luego $A'C''$ es la paralela a AC por A' , pero ésta es $A'C'$, luego $C'' = C'$, pues ambos son el punto de corte de $A'C' = A'C''$ con $B'C'$, luego $CC' = OC''$ pasa por O . ■

Seguidamente probamos lo mismo, pero suponiendo que las rectas AA' y BB' son paralelas, y vamos a ver que CC' también es paralela a ambas. En efecto, si CC' no fuera paralela a AA' , ambas rectas se cortarían en un punto O , pero entonces el teorema que acabamos de probar implicaría que BB' también pasaría por O , cuando estamos suponiendo que es paralela a AA' . ■



Y por último probamos el teorema de Desargues para rectas paralelas: si las tres rectas AA' , BB' y CC' son paralelas distintas dos a dos, AB es paralela a $A'B'$ y AC es paralela a $A'C'$, entonces BC es paralela a $B'C'$.

En efecto, si BC no es paralela a $B'C'$, podemos considerar la paralela a BC por C' , que cortará a BB' en un punto B'' , pero entonces A, B, C y A', B', C'' están en las condiciones del resultado precedente, luego BB'' es paralela a AA' , luego BB'' es la paralela a CC' por B , luego es BB' , luego $B' = B''$ y tenemos que BC sí que es paralela a $B'C'$, contradicción. ■

Esto termina la prueba del teorema de Hessenberg.

7.4 Los axiomas de ordenación

Veamos ahora el efecto de incorporar los axiomas de ordenación a los de la geometría afín. Ante todo, es fácil ver que el ejemplo de plano no arguesiano que hemos dado en la sección anterior cumple también los axiomas de ordenación con la definición obvia de la relación “estar entre”, por lo que los axiomas de incidencia y ordenación en la geometría plana no son suficientes para demostrar el teorema de Desargues o introducir coordenadas. Por consiguiente, aquí vamos a considerar únicamente el caso de un espacio afín E (de dimensión finita) arguesiano, y esto último se cumple siempre que la dimensión es mayor que 2. No perdemos generalidad si suponemos que $E = K^n$, para un cierto anillo de división K y un número natural $n \geq 2$.

Otro hecho que ya hemos discutido es que, al considerar los axiomas de ordenación, los axiomas de la geometría afín pueden relajarse hasta los axiomas A1–A5 considerados al principio de la sección 7.1, pues los hechos adicionales que contienen los axiomas afines pueden probarse con los axiomas de ordenación.

En la sección 3.2 vimos que si K es un cuerpo ordenado, entonces en K^n podemos definir una relación “estar entre” que verifica los axiomas de ordenación, y es fácil ver que la prueba vale igualmente si K es un anillo de división ordenado. Ahora vamos a demostrar el recíproco: si K^n cumple los axiomas de ordenación con cierta relación “estar entre”, entonces el anillo de división K admite una estructura de anillo ordenado.

Ante todo, observemos que no perdemos generalidad si suponemos que $n = 2$, pues si K^n cumple los axiomas de ordenación, entonces cualquiera de sus planos los cumple también (con la restricción de la restricción “estar entre” de K^n), luego a través de una asignación de coordenadas podemos transportar a K^2 dicha relación “estar entre”.

Trabajamos, pues, con los axiomas de incidencia y ordenación de la geometría plana. Recordemos que en este contexto a cada recta le podemos asociar dos relaciones de orden mutuamente inversas, descritas en el teorema 1.20. Por otra parte, ahora contamos con el axioma de las paralelas, que no considerábamos en el capítulo I, y que nos permite introducir el concepto siguiente:

Definición 7.35 Sean r y r' dos rectas distintas y sea t una recta que las corte a ambas. Llamaremos *proyección paralela* entre ambas rectas a la aplicación $\pi : r \rightarrow r'$ que a cada punto $X \in r$ le asigna el punto donde la paralela a t por X corta a r' .

Es claro que las proyecciones paralelas entre rectas son biyectivas (estamos usando que las rectas se encuentran en el mismo plano, pues la proyección paralela no estaría bien definida si las rectas se cruzaran en el espacio).

Sólo necesitaremos tener en consideración una consecuencia de los axiomas de ordenación:

Teorema 7.36 *Dadas dos rectas r y r' y fijada una de sus dos ordenaciones posibles en cada una de ellas, las proyecciones paralelas entre ambas conservan o invierten la orientación.*

DEMOSTRACIÓN: Sea t una recta que corte a r y a r' en puntos A y A' . Como dos rectas paralelas definen la misma proyección paralela, en el caso en que r y r' sean secantes, podemos suponer que $A = A'$ es el punto de corte.

La recta s divide al plano en dos semiplanos S_+ y S_- , cada uno de los cuales contiene a una de las semirrectas que los puntos A y A' determinan en r y r' , respectivamente. Llamemos r_+ y r'_+ a las semirrectas contenidas en S_+ y r_- y r'_- a las contenidas en S_- . Intercambiando las ordenaciones si es preciso, podemos suponer que los puntos de r_+ son mayores que los de r_- y que, igualmente, los puntos de r'_+ son mayores que los de r'_- .

Vamos a probar que, con estos ajustes, la proyección paralela π determinada por t conserva el orden (lo cual significa que lo conserva o lo invierte con las ordenaciones originales). Para ello tomamos dos puntos $X, Y \in r$, tales que $X < Y$. Llamemos $X' = \pi(X)$, $Y' = \pi(Y)$. Distinguimos varios casos:

Si $X < A < Y$, entonces $X \in r_-$, $Y \in r_+$, y es claro que la proyección transforma r_- en r'_- y r_+ en r'_+ (pues el segmento que une cada punto con su imagen es paralelo a t , luego t no lo corta, luego el punto y la imagen están en el mismo semiplano respecto de t). Por lo tanto $X' < A' < Y'$.

Si $A < X < Y$, entonces $A - X - Y$. Si $A = A'$, el axioma B4 aplicado a A , Y , Y' implica que la paralela a t por X tiene que cortar al segmento $\overline{A'Y'}$, y el punto de corte es X' , luego $A' - X' - Y'$, luego $X' < Y'$.

Si $A \neq A'$ aplicamos primero B4 a los puntos A, Y, Y' , con lo que la paralela a t por X corta a $\overline{AY'}$, y luego a los puntos A, A', Y' , con lo que dicha paralela corta a $\overline{A'Y'}$, y concluimos igualmente.

El caso $X < Y < A$ es análogo, y los casos en los que $X = A$ o $Y = A$ tampoco ofrecen ninguna dificultad. ■

Es inmediato comprobar que un anillo de división K admite estructura de anillo ordenado si y sólo si tiene un *cono positivo*, que es un conjunto $P \subset K$ cerrado para sumas y productos tal que $K = P \cup (-P)$ y $P \cap (-P) = \{0\}$.

Concretamente, si existe tal conjunto P , es claro que la relación en K dada por $a \leq b$ si y sólo si $b - a \in P$ es una relación de orden compatible con la estructura algebraica de K .

Para conectar esto con los axiomas de ordenación introducimos el concepto siguiente:

Definición 7.37 Un anillo de división K está *débilmente ordenado* si sobre él hay definida una relación de orden total de modo que

1. para cada $a \in K$, la aplicación $x \mapsto x + a$ conserva o invierte el orden.
2. para cada $a \in K$, $a \neq 0$, la aplicación $x \mapsto ax$ conserva o invierte el orden.

Cuando en este contexto digamos que los elementos a, b, c, \dots están ordenados deberá entenderse que $a < b < c, \dots$ o bien $\dots < c < b < a$.

Es fácil ver que todo anillo de división ordenado está débilmente ordenado. Vamos a ver que el recíproco es cierto salvo en un caso: si K es el cuerpo de dos elementos entonces es obvio que está débilmente ordenado, pues sólo admite dos ordenaciones mutuamente inversas y por lo tanto cualquier biyección conserva o invierte el orden. Sin embargo no puede ser ordenado.

Teorema 7.38 *Si K es un anillo de división con más de dos elementos y está débilmente ordenado, entonces está ordenado.*

DEMOSTRACIÓN: En primer lugar, la característica de K no puede ser 2. En efecto, sean $0, 1, a$ tres elementos distintos en K . Si están ordenados en la forma $0, 1, a$, entonces sumamos 1 y obtenemos la ordenación $1, 0, a+1$, mientras que si sumamos a obtenemos la ordenación $a, a+1, 0$. Estas dos ordenaciones implican que $1, 0, a+1, a$, lo que contradice la ordenación inicial.

Si la ordenación es $a, 0, 1$ entonces al sumar 1 y a obtenemos $a+1, 1, 0$ y $0, a, a+1$. Las dos primeras implican $a, 0, 1, a+1$, que contradice a la tercera.

Dado $a \in K^*$ consideramos las posibles ordenaciones de $0, a/2$ y $-a/2$. Si se da $-a/2, 0, a/2$ entonces sumando $\pm a/2$ obtenemos $-a, -a/2, 0$ y $0, a/2, a$, lo que implica $-a, -a/2, 0, a/2, a$. En particular $-a, 0, a$.

Si por el contrario $0, -a/2, a/2$ obtenemos $-a/2, -a, 0$ y $a/2, 0, a$, de donde se sigue igualmente $-a, 0, a$.

Cambiando la ordenación de K por su inversa si es necesario, K sigue siendo un cuerpo débilmente ordenado con $0 < 1$. Acabamos de probar que si P es el conjunto de los elementos ≥ 0 , entonces $-P$ es el conjunto de los elementos ≤ 0 , luego se cumple que $K = P \cup (-P)$ y que $P \cap (-P) = \{0\}$.

Dados $a, b \in P$ no nulos, intercambiando a y b podemos suponer que $0, a, b$. Sumamos a , con lo que obtenemos $a, 2a, a+b$. Sumando a en $-a, 0, a$ obtenemos $0, a, 2a$, luego $0, a, 2a, a+b$, luego $a+b \in P$.

Finalmente, si $a, b \in P$ son no nulos y $b \neq 1$, tenemos la ordenación $0, 1, b$ o bien $0, b, 1$, luego también $0, a, ab$ o bien $0, ab, a$. En ambos casos $ab \in P$. El caso $b = 1$ es trivial. ■

Teorema 7.39 *Sea E un espacio afín sobre un anillo de división K cuyas rectas tengan al menos tres puntos. Entonces E cumple los axiomas de ordenación (con cierta definición de la relación “estar entre”) si y sólo si K admite una estructura de anillo ordenado.*

DEMOSTRACIÓN: Si K es un anillo de división ordenado, ya sabemos que es posible definir en E una relación “estar entre” que cumple los axiomas de ordenación. Recíprocamente, si E cumple los axiomas de ordenación, ya hemos razonado que no perdemos generalidad si suponemos que E es un plano afín.

Podemos biyectar K con los puntos de una recta mediante la correspondencia $\alpha \mapsto P + \alpha\vec{v}$, donde P y \vec{v} son un punto y un vector prefijados. Probaremos que este orden es un orden débil en K y, por el teorema anterior, él o su inverso (el orden que hace $1 > 0$) es una ordenación de K . Antes veremos que el orden no depende de la elección de P o \vec{v} .

En primer lugar cambiamos \vec{v} por otro vector $\vec{w} \neq \vec{0}$. Consideremos la homotecia dada por $f_\alpha(Q) = P + \alpha\vec{PQ}$. En particular

$$P + \alpha\vec{v} = f_\alpha(P + \vec{v}), \quad P + \alpha\vec{w} = f_\alpha(P + \vec{w}).$$

Esto implica que las rectas que pasan por $P + \alpha\vec{v}$ y $P + \alpha\vec{w}$ forman un haz de rectas paralelas o, equivalentemente, cada $P + \alpha\vec{w}$ se obtiene de $P + \alpha\vec{v}$ por proyección paralela. Por hipótesis el orden se conserva (o invierte) y en ambos casos el orden en K (que hace $1 > 0$) es el mismo.

Si reemplazamos el punto P por otro punto Q , tomamos un vector \vec{v} de dirección distinta a PQ . Entonces las rectas $P + \alpha\vec{v}$ y $Q + \alpha\vec{v}$ son paralelas, al igual que las rectas que pasan por cada par de puntos de esta forma. Por lo tanto $Q + \alpha\vec{v}$ se obtiene de $P + \alpha\vec{v}$ por proyección paralela y concluimos del mismo modo que ambas rectas inducen el mismo orden en K .

Para probar que la ordenación de K es un orden débil observamos que, fijados P y \vec{v} , la aplicación $\alpha \mapsto \alpha + \delta$ en K se corresponde con la aplicación $P + \alpha\vec{v} \mapsto P + (\alpha + \delta)\vec{v} = (P + \delta\vec{v}) + \alpha\vec{v}$, luego la ordenación de K se convierte en la inducida fijando $P + \alpha\vec{v}$ y \vec{v} que, según hemos visto, es la misma o la opuesta.

Similarmente, la aplicación $\alpha \mapsto \delta\alpha$ en K se corresponde con la aplicación $P + \alpha\vec{v} \mapsto P + \alpha(\delta\vec{v})$, luego el orden de K se transforma en el inducido por P y $\delta\vec{v}$, que es el mismo o el inverso. ■

Puede probarse que existen anillos de división ordenados no conmutativos, y esto implica que los axiomas de ordenación no bastan para demostrar que el anillo de escalares tenga que ser conmutativo o, equivalentemente, para demostrar el teorema de Papos-Pascal.

7.5 Los axiomas de congruencia

Vamos a considerar ahora un espacio afín arbitrario (no excluimos a los planos no arguesianos) que satisfaga los axiomas de ordenación y de congruencia.

Si suponemos además la propiedad arquimediana, en el capítulo III hemos probado que el espacio dado estructura de espacio afín sobre un cuerpo, en el sentido algebraico. En realidad lo hemos probado para dimensiones 2 y 3, pero ahí está precisamente el caso no trivial, pues sabemos que el teorema de Desargues se cumple en todo espacio afín de dimensión mayor que 2 y, en cuanto al teorema de Papos-Pascal, si un espacio de dimensión arbitraria cumple los

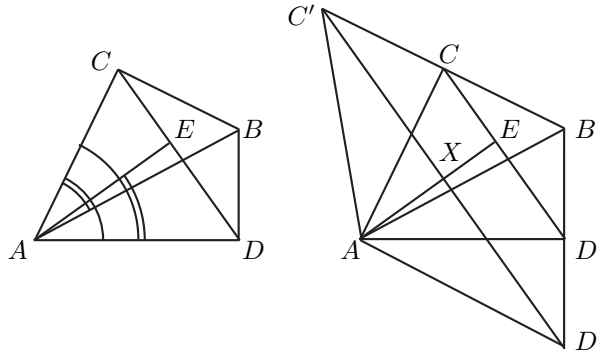
grupos de axiomas A, B, C (incluyendo el axioma de las paralelas) más la propiedad arquimediana, lo mismo valdrá para cualquiera de sus planos, luego todos ellos cumplirán el teorema de Papos-Pascal, pero como éste es un teorema sobre figuras planas, el hecho de que se cumpla en todo plano equivale a que se cumpla en general.

En esta sección vamos a demostrar que la propiedad arquimediana no es necesaria realmente, de modo que toda geometría que cumpla los grupos de axiomas A, B, C (incluyendo el axioma de las paralelas) satisface el teorema de Papos-Pascal y, por consiguiente el de Desargues. Por el mismo argumento anterior, basta considerar el caso de la geometría plana.

Suponemos, pues, los grupos de axiomas A, B y C del capítulo I para la geometría plana, incluyendo el axioma de las paralelas.

Necesitamos algunos resultados previos, que se basan en el teorema siguiente:

Teorema 7.40 Sean \widehat{ABC} y \widehat{ABD} dos triángulos rectángulos con la misma hipotenusa \overline{AB} y de modo que C y D estén en lados opuestos respecto a ella, y sea E el pie de la perpendicular a CD por A . Entonces $C-E-D$ y se cumplen las congruencias $\widehat{BAC} \equiv \widehat{DAE}$ y $\widehat{BAD} \equiv \widehat{CAE}$.



DEMOSTRACIÓN: Como C y D están en lados opuestos respecto de AB , se cumple que B está dentro del ángulo \widehat{DAC} , luego $\widehat{BAC} \leq \widehat{DAC}$. Por lo tanto, existe un punto E' en \overline{CD} tal que $\widehat{DAE'} \equiv \widehat{BAC}$. Como las sumas de ángulos se pueden cancelar, esto implica a su vez que $\widehat{E'AC} \equiv \widehat{DAB}$.

Basta probar que $E' = E$, para lo cual basta ver que $r = AE'$ es perpendicular a CD .

Sea D' el simétrico de B respecto de D , es decir, el único punto sobre la recta BD que cumple $\overline{DD'} \equiv \overline{BD}$. Similarmente, sea C' el simétrico de B respecto de C . Las hipótesis sobre los ángulos rectos implican que $\overline{AC'} \equiv \overline{AB} \equiv \overline{AD'}$, así como que $\widehat{BAC} \equiv \widehat{C'AC}$, $\widehat{BAD} \equiv \widehat{D'AD}$.

De aquí se sigue que

$$\widehat{C'AE'} \equiv \widehat{C'AC} + \widehat{CAE'} \equiv \widehat{BAC} + \widehat{DAB} \equiv \widehat{E'AD} + \widehat{DAD'} \equiv \widehat{E'AD'}.$$

Así, los triángulos $\widehat{C'AE'}$ y $\widehat{E'AD'}$ son congruentes y tienen un lado en común, pero no pueden ser coincidentes, pues claramente $C' \neq D'$ (por ejemplo, porque AB los separa). La unicidad del transporte de triángulos exige que C' y D' estén en lados opuestos respecto a $r = AE'$, luego r corta a $\overline{C'D'}$ en un punto X . Los triángulos $\widehat{C'AX}$ y $\widehat{D'AX}$ tienen iguales dos lados y el ángulo que forman, luego también $\overline{XC'} \equiv \overline{XD'}$ o, lo que es lo mismo, X es el punto medio de $\overline{C'D'}$ y los ángulos $\widehat{AXC'}$ y $\widehat{AXD'}$ son iguales, luego son rectos. En otras palabras, r es perpendicular a $C'D'$.

Finalmente consideramos el triángulo $\widehat{C'BD'}$. Como C y D son los puntos medios de dos sus lados, es claro que CD es paralela a $C'D'$, luego r también es perpendicular a CD . ■

En el contexto en el que estamos trabajando no tenemos definida la longitud de un segmento ni la amplitud de un ángulo, pero podemos considerar como “longitudes” las clases de congruencia de segmentos (incluyendo por convenio una clase nula que representaremos por 0) y como “amplitudes” las clases de congruencia de ángulos (incluyendo una clase de ángulos nulos que representaremos por 0 y otra de ángulos llanos que representaremos por π). Llamaremos $\pi/2$ a la clase de los ángulos rectos.

Si l es una clase de segmentos no nula y α es una clase de ángulos distinta de $0, \pi/2, \pi$, podemos definir la clase que llamaremos $l|\cos \alpha|$ como la definida del modo siguiente:

Tomamos cualquier ángulo \widehat{AOB} en la clase α , cambiamos B si es preciso para exigir que \overline{OB} esté en la clase l , llamamos X al pie de la perpendicular a OA por B (que no será O , porque el ángulo no es recto), y llamamos $l|\cos \alpha|$ a la clase de \overline{OX} .

Esta definición no depende de la elección del ángulo en la clase α , porque si tomamos otro distinto, el triángulo \widehat{OBX} y su análogo serán dos triángulos rectángulos con otro ángulo congruente aparte del recto, luego con los tres ángulos iguales, e hipotenusas congruentes, luego serán congruentes, y en particular \overline{OX} será congruente con su análogo.

Si $\alpha = \pi/2$ al realizar la construcción obtenemos $X = O$, y definimos $l|\cos \alpha| = 0$. Si $\alpha = 0$ o $\alpha = \pi$ definimos $l|\cos \alpha| = l$. Por último si $l = 0$ definimos $l|\cos \alpha| = 0$.

Observemos ahora que si $\alpha \neq \pi/2$ y $l|\cos \alpha| = l'|\cos \alpha|$, entonces $l = l'$.

En efecto, esto es obvio si $\alpha = 0$ o $\alpha = \pi$, luego descartamos estos casos. Al realizar la construcción de $l|\cos \alpha|$ y $l'|\cos \alpha|$ formamos dos triángulos rectángulos con los ángulos iguales y que resultan tener un cateto igual, luego también tienen las hipotenusas iguales, lo que significa que $l = l'$.

El resultado fundamental sobre esta trigonometría rudimentaria es el siguiente:

$$l|\cos \alpha||\cos \beta| = l|\cos \beta||\cos \alpha|.$$

Para probarlo empezamos observando que el resultado es trivial si alguno de los ángulos es $0, \pi/2$ o π , así que descartamos estos casos. Para calcular $l|\cos \alpha|$

formamos un triángulo rectángulo \widehat{ABC} con hipotenusa $\overline{AB} = l$, $\widehat{BAC} = \alpha$ y $\overline{AC} = l|\cos\alpha|$. A su vez, para calcular $l|\cos\beta|$ podemos formar otro triángulo rectángulo \widehat{BAD} que tenga la misma hipotenusa, $\widehat{BAD} = \beta$ y $\overline{AD} = l|\cos\beta|$. Además podemos exigir que C y D estén separados por AB , con lo que la situación es justo la que ilustra la figura del teorema anterior.

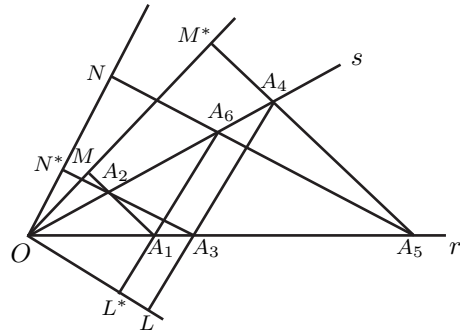
Ahora consideramos el punto E dado por dicho teorema, que determina dos triángulos rectángulos, \widehat{ACE} y \widehat{ADE} . El primero tiene hipotenusa $l|\cos\alpha|$ y $\widehat{CAE} = \beta$, por lo que $\overline{AE} = l|\cos\alpha||\cos\beta|$, pero si razonamos con el segundo triángulo, la conclusión es $\overline{AE} = l|\cos\beta||\cos\alpha|$. ■

Con esto ya podemos demostrar el teorema de Pappus-Pascal 7.33:

DEMOSTRACIÓN: Tenemos dos rectas r y s que se cortan en un punto O . Sobre r tenemos los puntos A_1, A_3, A_5 y sobre s los puntos A_2, A_4, A_6 , todos distintos de O .

Además suponemos que A_1A_2 es paralela a A_4A_5 y que A_3A_4 es paralela a A_1A_6 . Tenemos que probar que A_2A_3 es paralela a A_5A_6 .

Consideramos la perpendicular común a A_1A_6 y a A_3A_4 que pasa por O . Llamemos L al punto donde corta a A_3A_4 y L^* al punto donde corta a A_1A_6 . Similarmente, trazamos la perpendicular por O a las otras dos paralelas, y llamamos M al punto de corte con A_1A_2 y M^* al punto de corte con A_4A_5 . Por último tomamos la perpendicular a A_5A_6 que pasa por O y llamamos N al punto de corte. Definimos también los ángulos siguientes:



$$\begin{aligned} \lambda &= \widehat{LOA_4} = \widehat{L^*OA_6}, & \lambda' &= \widehat{LOA_3} = \widehat{L^*OA_1}, \\ \mu &= \widehat{MOA_2} = \widehat{M^*OA_4}, & \mu' &= \widehat{MOA_1} = \widehat{M^*OA_5}, \\ \nu &= \widehat{NOA_6}, & \nu' &= \widehat{NOA_5}. \end{aligned}$$

Adoptamos además el convenio de usar una letra minúscula para nombrar la longitud del segmento de extremos O y la mayúscula correspondiente, como $l = \overline{OL}$, etc. Por la propia definición de los “cosenos” que hemos considerado antes de la prueba obtenemos las relaciones siguientes, excepto la última:

1. $a_3|\cos\lambda'| = a_4|\cos\lambda| = l$.
2. $a_1|\cos\mu'| = a_2|\cos\mu| = m$.
3. $a_5|\cos\nu'| = a_6|\cos\nu| = n$.
4. $a_1|\cos\lambda'| = a_6|\cos\lambda| = l^*$.
5. $a_5|\cos\mu'| = a_4|\cos\mu| = m^*$.
6. $a_3|\cos\nu'| = a_2|\cos\nu|$.

Vamos a demostrar la sexta igualdad, para lo cual usamos repetidas veces la conmutatividad que hemos probado para los cosenos:

$$a_3 |\cos \nu'| |\cos \lambda'| = a_3 |\cos \lambda'| |\cos \nu'| =_1 a_4 |\cos \lambda| |\cos \nu'|,$$

luego

$$\begin{aligned} a_3 |\cos \nu'| |\cos \lambda'| |\cos \mu| &= a_4 |\cos \lambda| |\cos \nu'| |\cos \mu| = a_4 |\cos \mu| |\cos \lambda| |\cos \nu'| \\ &=_{5} a_5 |\cos \mu'| |\cos \lambda| |\cos \nu'| = a_5 |\cos \nu'| |\cos \mu'| |\cos \lambda| =_{3} a_6 |\cos \nu| |\cos \mu'| |\cos \lambda| \\ &= a_6 |\cos \lambda| |\cos \nu| |\cos \mu'| =_{4} a_1 |\cos \lambda'| |\cos \nu| |\cos \mu'| = a_1 |\cos \mu'| |\cos \lambda'| |\cos \nu| \\ &=_{2} a_2 |\cos \mu| |\cos \lambda'| |\cos \nu| = a_2 |\cos \nu| |\cos \lambda'| |\cos \mu|. \end{aligned}$$

Observemos que μ no puede ser recto, pues forma parte del triángulo rectángulo $\widehat{OMA_2}$ salvo si $OM = OA_2 = s$, en cuyo caso $\mu = 0$ o $\mu = \pi$. Igualmente se razona que λ' tampoco es recto, y esto implica que sus cosenos son simplificables, y al eliminarlos obtenemos 6).

Sean N^* y N^{**} los pies de las perpendiculares a ON por A_2 y A_3 respectivamente. Basta probar que $N^* = N^{**}$, pues entonces A_2A_3 es perpendicular a ON , luego paralela a A_5A_6 .

Por definición $a_2 |\cos \nu| = n^*$ y $a_3 |\cos \nu'| = n^{**}$, luego por 6) tenemos que $n^* = n^{**}$, es decir, que $\overline{ON^*} \equiv \overline{ON^{**}}$, luego basta probar que N^* y N^{**} están ambos al mismo lado o bien ambos en lados opuestos que N respecto de O , ya que entonces ambos estarán al mismo lado de O y podremos concluir que son iguales. Para ello usamos el teorema 7.36, que nos asegura que las proyecciones paralelas conservan o invierten las relaciones de orden de las rectas, luego en cualquier caso conservan la relación “estar entre”. Vamos a considerar diversas proyecciones paralelas entre rectas que se cortan en O y que, por tanto, dejan fijo a O .

Por ejemplo, O , A_1 y A_5 se transforman por una proyección paralela en A_2 y A_4 , luego A_1 y A_5 están al mismo lado de O si y sólo si lo están A_2 y A_4 . Igualmente, A_1 y A_3 están al mismo lado de O si y sólo si lo están A_6 y A_4 .

De aquí podemos deducir que A_3 y A_5 están al mismo lado de O si y sólo si lo están A_2 y A_6 . En efecto, hay que distinguir cuatro casos:

Si A_1 y A_5 están al mismo lado y A_1 y A_3 también, entonces resulta que A_1, A_3, A_5 están todos al mismo lado, luego A_2, A_4, A_6 también, y se cumple la equivalencia.

Si A_1 y A_5 están al mismo lado, pero A_1 y A_3 están en lados opuestos, entonces A_2 y A_4 están al mismo lado, pero A_6 y A_4 están en lados opuestos. Entonces A_3 y A_5 están en lados opuestos y A_2 y A_6 también.

Las otras dos posibilidades se tratan igualmente. Por último, considerando proyecciones paralelas, N^* y N están al mismo lado de O si y sólo si lo están A_2 y A_6 , si y sólo si lo están A_3 y A_5 , si y sólo si lo están N^{**} y N , como había que probar. ■

Así pues, los axiomas de congruencia descartan tanto los planos no arguesianos como los anillos de escalares no conmutativos.

7.6 Planos afines finitos

Según hemos visto en la sección 7.1, los planos afines pueden definirse como pares (X, \mathcal{B}) donde X es un conjunto de puntos y \mathcal{B} es una familia de conjuntos de puntos (a los que llamamos rectas) de modo que se cumplen los axiomas siguientes, donde hay que entender que dos rectas son paralelas si son iguales o disjuntas:

Axioma A1 *Por cada par de puntos distintos pasa una única recta.*
Axioma A2 *Por cada punto pasa una única paralela a una recta dada.*
Axioma A3 *Existen tres puntos no colineales.*

Los planos afines admiten una caracterización combinatoria muy simple:

Definición 7.41 Si $1 \leq l < m < n$, un *sistema de Steiner* de tipo $S(l, m, n)$ es un par (X, \mathcal{B}) , donde X es un conjunto de n puntos y \mathcal{B} es una familia de subconjuntos de X , llamados *bloques*, de modo que cada bloque tiene m puntos y cada conjunto de l puntos está contenido en un único bloque.

Teorema 7.42 *Los planos afines son los sistemas de Steiner de tipo $S(2, n, n^2)$, para $n \geq 2$.*

DEMOSTRACIÓN: Es claro que todo sistema de Steiner de tipo $S(2, n, n^2)$ es un plano afín, pues por definición se cumple A1, cada recta contiene n puntos de un total de n^2 , luego hay tres puntos no colineales y, por último, por cada punto P pasan exactamente $n+1$ rectas, ya que hay n^2-1 pares $\{P, Q\}$, donde Q recorre los puntos distintos de P , y cada recta que pasa por P contiene a $n-1$ de estos pares, luego hay $(n^2-1)/(n-1) = n+1$ rectas que pasan por P . Si L es una recta tal que $P \notin L$, hay exactamente n rectas que pasan por P y cortan a L (pues L tiene n puntos), luego hay una única recta que pasa por P y no corta a L . Si $P \in L$, la única recta paralela a L por P es obviamente L .

Supongamos ahora que (X, \mathcal{B}) es un plano afín finito. Entonces:

1. Si L_1 es paralela a L_2 y L_2 es paralela a L_3 , entonces L_1 es paralela a L_3 .
 En efecto, podemos suponer que $L_1 \neq L_2 \neq L_3$. Si $L_1 \neq L_3$ y se cortan en el punto P , entonces L_1 y L_3 son paralelas a L_2 por P , en contradicción con A2.
2. Las rectas no son vacías, pues si una recta es vacía, es paralela a todas las demás, pero por A3 hay dos rectas distintas que se cortan en un punto P , y entonces ambas rectas serían paralelas a la recta vacía por el punto P , en contradicción con A2.
3. Las rectas tienen al menos dos puntos, pues si una recta fuera $L = \{P\}$, por A3 existen puntos Q y R no colineales con P y podemos tomar la paralela s a la recta PQ por R , pero entonces $L = \{P\} \subset PQ$, luego ambas son paralelas a s y pasan por P , en contradicción con A2.

4. Todas las rectas tienen el mismo número de puntos $n \geq 2$.

Si L_1 y L_2 son dos rectas distintas, podemos tomar un punto $P_1 \in L_2 \setminus L_1$ y otro $P_2 \in L_2 \setminus L_1$. Para cada $P \in L_1$, la paralela L' a P_1P_2 que pasa por P tiene que cortar a L_2 en un punto P' (pues L' no puede ser paralela a L_2 , ya que la paralela a L_2 por P es L_1 y no puede ser $L_1 = L'$ porque L_1 no es paralela a P_1P_2). La aplicación $P \mapsto P'$ biyecta L_1 con L_2 , pues si $P, Q \in L_1$ son distintos, las paralelas a P_1P_2 por P y por Q tienen que ser distintas (pues $PQ = L_1$ no es paralela a P_1P_2), luego disjuntas, luego $P' \neq Q'$. Esto prueba que la aplicación es inyectiva, luego $|L_1| \leq |L_2|$. No es difícil probar que la aplicación es biyectiva, pero no lo necesitamos, pues, por el mismo motivo, tiene que ser $|L_2| \leq |L_1|$, luego todas las rectas tienen el mismo número de puntos.

5. Hay n^2 puntos.

Sean O, P_1, Q_1 tres puntos no colineales, de modo que $L_1 = OP_1$ y $L_2 = OQ_1$ son dos rectas distintas y tienen n puntos cada una, digamos

$$L_1 = \{O, P_1, \dots, P_{n-1}\}, \quad L_2 = \{O, Q_1, \dots, Q_{n-1}\}.$$

La recta paralela a L_2 por P_i y la recta paralela a L_1 por Q_j no son paralelas entre sí, porque entonces L_1 y L_2 serían paralelas. Por lo tanto se cortan en un punto P_{ij} . Todo punto P es de la forma P_{ij} , pues la paralela a L_2 por P corta a L_1 en cierto P_i y la paralela a L_1 por P corta a L_2 en cierto Q_j , de modo que $P = P_{ij}$. Por último, los puntos P_{ij} son distintos dos a dos, pues si $P_{ij} = P_{kl}$, la paralela a L_2 por este punto tiene que cortar a L_1 en P_i y en P_k , luego $P_i = P_k$, luego $i = k$, e igualmente concluimos que $j = l$.

Es claro entonces que (X, \mathcal{B}) es un sistema de Steiner de tipo $S(2, n, n^2)$. ■

En particular hemos demostrado que todas las rectas de un plano afín finito tienen el mismo número de puntos.

Definición 7.43 El número n de puntos de cualquier recta en un plano afín se llama *orden* del plano.

En la prueba del teorema anterior hemos visto que un plano afín de orden n tiene n^2 puntos, que el paralelismo es una relación de equivalencia (hemos visto que es una relación transitiva y, trivialmente, es reflexiva y simétrica) y que por cada punto pasan $n + 1$ rectas. También es fácil ver que tiene $n(n + 1)$ rectas.

En efecto, como hay n^2 puntos, hay $n^2(n^2 - 1)/2$ pares de puntos $\{P, Q\}$, cada uno de los cuales está contenido en una única recta, y cada recta, con n puntos, contiene $n(n - 1)/2$ pares, luego el número de rectas tiene que ser

$$\frac{n^2(n^2 - 1)/2}{n(n - 1)/2} = n(n + 1).$$

Por último, cada recta tiene exactamente n paralelas (incluida ella misma), pues las paralelas a una recta deben cubrir todo el espacio y, como hay n^2 puntos y cada recta tiene n puntos, cada clase de rectas paralelas tiene que constar de n rectas.

Un *isomorfismo* entre dos sistemas de Steiner (X, \mathcal{B}) y (X', \mathcal{B}') es una biyección $f : X \rightarrow X'$ que hace corresponder los bloques de \mathcal{B} con los de \mathcal{B}' .

Obviamente, dos sistemas de Steiner isomorfos son del mismo tipo. En particular, los isomorfismos en este sentido entre dos planos afines finitos no son sino las colineaciones en el sentido de la definición 5.15 (pues un poco más abajo hemos demostrado que las colineaciones transforman rectas en rectas, que es lo que exige la definición de isomorfismo entre sistemas de Steiner).

Si en lugar de considerar planos afines consideramos espacios afines de dimensión $d \geq 3$, tenemos que todos ellos son arguesianos, (pues el teorema de Desargues es realmente un teorema en dimensión ≥ 3), luego, según hemos visto en la sección 7.3, son espacios afines sobre un anillo de división en el sentido algebraico. Ahora bien, si el espacio es finito, el anillo de división es finito, luego es un cuerpo por el teorema de Wedderburn [Al 9.4]. Por lo tanto, todo espacio afín finito de dimensión $d \geq 3$ es un espacio afín de dimensión d sobre un cuerpo finito en el sentido algebraico. Por ello podemos afirmar que los espacios afines de dimensión mayor que 2 están perfectamente clasificados. En cambio, no todos los planos afines finitos pueden ser descritos algebraicamente, sino que hay planos afines finitos no arguesianos (esto lo probaremos al final de la sección 8.8).

Más detalladamente, si k es un cuerpo finito, de q elementos (donde q es necesariamente una potencia de primo, por [Al 9.2]), cualquier plano afín E sobre k en el sentido algebraico —tomando como rectas las subvariedades afines de dimensión 1— es un plano afín arguesiano de orden q en el sentido axiomático, y entre dos cualesquiera de ellos existe una biyección afín, la cual es trivialmente un isomorfismo en el sentido que acabamos de definir (es decir, que transforma rectas en rectas). Recíprocamente, en la sección 7.3 hemos visto que todo plano afín arguesiano en el sentido axiomático es isomorfo a un plano afín en el sentido algebraico.

Por lo tanto, dado que, para cada potencia de primo q existe, salvo isomorfismo, un único cuerpo con q elementos [Al 9.2], podemos afirmar que existe, salvo isomorfismo, un único plano afín arguesiano de orden q . Pero esto no impide que puedan existir otros planos afines no arguesianos de orden q , o incluso que puedan existir planos afines no arguesianos de órdenes que no sean potencias de primo.

Es evidente que existe un plano afín de orden 2, único salvo isomorfismo, pues necesariamente tiene que tener 4 puntos y los 6 pares de puntos posibles tienen que ser rectas. Veamos un par de ejemplos no triviales de unicidad:

Teorema 7.44 *Existe, salvo isomorfismo, un único plano afín de orden 3 o, equivalentemente, un único sistema de Steiner de tipo $S(2, 3, 9)$.*

DEMOSTRACIÓN: Sea un sistema de Steiner del tipo indicado, a cuyos puntos los podemos llamar simplemente $\{1, \dots, 9\}$. Renumerándolos, podemos suponer que $\{1, 2, 3\}$ es una recta. Tiene que haber una recta paralela a ésta que pase por el punto 4. Renumerando de nuevo, podemos suponer que es $\{4, 5, 6\}$. Igualmente, tiene que haber una recta paralela a ambas que pase por 7, luego necesariamente es $\{7, 8, 9\}$.

La recta que pasa por 1, 4 no es paralela a $\{1, 2, 3\}$, luego tampoco a $\{7, 8, 9\}$. Reordenando sus puntos podemos suponer que la corta en 7, con lo que $\{1, 4, 7\}$ es otra recta. Igualmente, la paralela a esta recta por 2 tiene que cortar a $\{4, 5, 6\}$ en un punto distinto de 4 y a $\{7, 8, 9\}$ en un punto distinto de 7. Renumerando 5, 6 y 8, 9 si es preciso, podemos suponer que es $\{2, 5, 8\}$. Entonces, la paralela a $\{1, 4, 7\}$ por 3 tiene que ser paralela a la anterior, luego es necesariamente $\{3, 6, 9\}$. En resumen, tenemos que las filas y las columnas del cuadrado siguiente son rectas:

$$\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array}$$

La recta 15 no es paralela a $\{1, 4, 7\}$, luego tampoco a $\{3, 6, 9\}$, pero no puede cortarla ni en un punto de $\{1, 2, 3\}$ ni en un punto de $\{4, 5, 6\}$ (pues ya corta a estas rectas en 1 y en 5). Necesariamente es $\{1, 5, 9\}$. Por el mismo argumento llegamos a que todas las transversales del cuadrado (es decir, los conjuntos de tres puntos tales que no hay dos en la misma fila ni en la misma columna) son también rectas, con lo que tenemos un total de 12 rectas. En la prueba del teorema anterior hemos visto que por cada punto pasan $n + 1 = 4$ rectas, y esto ya sucede con las 12 rectas que hemos encontrado, así que no puede haber más.

En resumen, hemos probado que, dado cualquier sistema de Steiner de tipo $S(2, 3, 9)$, es posible biyectar sus puntos con $\{1, \dots, 9\}$ de modo que sus rectas se correspondan con las 12 rectas que hemos determinado, con lo que el sistema dado es isomorfo al que hemos construido. ■

Teorema 7.45 *Existe, salvo isomorfismo, un único plano afín de orden 4 o, equivalentemente, un único sistema de Steiner de tipo $S(2, 4, 16)$.*

DEMOSTRACIÓN: En un plano afín de orden 4 hay 16 puntos, luego 120 pares (desordenados) de puntos distintos, y cada recta contiene a 6 de dichos pares, luego hay 20 rectas en total. Tomemos dos rectas L_1 y M_1 que se corten en un punto O . Cada paralela a L_1 corta a M_1 en uno de sus 4 puntos, luego L_1 tiene exactamente cuatro paralelas L_1, L_2, L_3, L_4 . Igualmente M_1 tiene exactamente cuatro paralelas M_1, M_2, M_3, M_4 . Además de estas 8 rectas, tiene que haber otras 12, digamos R_1, \dots, R_{12} , que no son paralelas a ninguna de las anteriores.

Sea $\{P_{ki}\} = L_k \cap M_i$. Así tenemos 16 puntos distintos, luego son todos los puntos del plano. Definimos una matriz T de tipo 4×12 de modo que $t_{ij} = k$ si y sólo si $R_j \cap M_i = \{P_{ki}\}$ o, equivalentemente, si $P_{ki} \in R_j$.

Así, cada columna $t_{1j}, t_{2j}, t_{3j}, t_{4j}$ es una permutación de 1, 2, 3, 4, pues la recta R_j cortará a las rectas M_i en cuatro puntos distintos, todos de la forma P_{ki} , para distintos valores de k .

Renumerando las rectas M_i , no perdemos generalidad si suponemos que la primera columna es precisamente 1, 2, 3, 4. Por otra parte, por cada punto de M_i pasan 5 rectas (esto lo hemos probado en la primera parte de la demostración del teorema 7.42), de las cuales una es M_i , otra es una L_k y las otras tres son tres de las R_j , luego en cada fila de T aparece tres veces cada índice 1, 2, 3, 4. Reordenando las rectas R_j podemos suponer que T es de la forma:

$$T = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 4 \\ 2 & & & & & & & & & & & \\ 3 & & & & & & & & & & & \\ 4 & & & & & & & & & & & \end{pmatrix}.$$

Más aún, T no puede contener una submatriz de la forma

$$\begin{pmatrix} a & a \\ b & b \end{pmatrix},$$

pues esto significaría que las rectas R_j y $R_{j'}$ correspondientes a las columnas contendrían ambas a los puntos $P_{ai}, P_{bi'}$. Por lo tanto, permutando los índices de las rectas R_j con un mismo valor de t_{1j} , podemos suponer que

$$T = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 4 \\ 2 & 3 & 4 & 1 & 3 & 4 & 1 & 2 & 4 & 1 & 2 & 3 \\ 3 & & & & & & & & & & & \\ 4 & & & & & & & & & & & \end{pmatrix}.$$

Ahora es fácil comprobar que la única forma de completar la matriz es

$$T = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 4 \\ 2 & 3 & 4 & 1 & 3 & 4 & 1 & 2 & 4 & 1 & 2 & 3 \\ 3 & 4 & 2 & 4 & 1 & 3 & 2 & 4 & 1 & 3 & 1 & 2 \\ 4 & 2 & 3 & 3 & 4 & 1 & 4 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}.$$

Esta matriz determina completamente las rectas del plano. Por ejemplo, de la penúltima columna se sigue que $R_{11} = \{P_{41}, P_{22}, P_{13}, P_{34}\}$. Hemos probado que, dado cualquier plano de índice 4, numerando adecuadamente sus 16 puntos en la forma P_{ki} , sus rectas se corresponden con las 20 rectas que hemos descrito aquí, luego el plano es isomorfo al que acabamos de construir. ■

Planos afines y cuadrados latinos ortogonales Si existiera un plano afín de orden 6, sería necesariamente no arguesiano, pues 6 no es potencia de primo. Sin embargo, vamos a probar que no existen tales planos a través de una caracterización de la existencia de planos afines. En [Al C.4] probamos que a partir de un cuerpo de n elementos es posible construir $n - 1$ cuadrados latinos mutuamente ortogonales de orden n . Ahora podemos generalizar ligeramente este hecho:

Teorema 7.46 *Existe un plano afín de orden n si y sólo si existe una familia de $n - 1$ cuadrados latinos mutuamente ortogonales de orden n .*

DEMOSTRACIÓN: Según [Al C.11, C.12], la existencia de una familia de $n - 1$ cuadrados latinos mutuamente ortogonales de orden n es equivalente a la existencia de un conjunto X de n^2 “puntos” y una familia de $(n + 1)n$ “rectas” (conjuntos de n puntos) divididas en $n + 1$ familias de n rectas “paralelas” (con intersección vacía) de modo que dos rectas de familias distintas se cortan en un único punto. Obviamente, esto lo cumplen las rectas de un plano afín de orden n , luego tenemos una implicación.

Recíprocamente, un conjunto X con una familia de rectas en las condiciones indicadas es un plano afín de orden n . En efecto, dados dos haces de rectas paralelas, cada una de las n rectas de uno de los haces corta a cada una de las n rectas del otro en un único punto, distinto para cada par, luego en total las intersecciones determinan n^2 puntos distintos, que son todos los de X . Esto significa que por cada punto de X pasa exactamente una recta de cada haz de rectas paralelas y esto es lo que afirma el axioma A2.

En particular, por cada punto P pasan $n + 1$ rectas en total. Dos rectas que pasan por P no se cortan en más puntos, luego las $n + 1$ rectas contienen $(n - 1)(n + 1) = n^2 - 1$ puntos distintos de P , luego contienen todos los puntos de X distintos de P . Esto significa que por dos puntos distintos de P pasa una única recta, como requiere el axioma A1.

El axioma A3 es inmediato, porque una recta contiene n puntos y en total hay n^2 , luego hay puntos fuera de ella. ■

En particular, como en la sección [Al C.3] hemos probado que no existen cuadrados latinos ortogonales de orden 6, concluimos que no existen planos proyectivos de orden 6.

En realidad, hasta la fecha no se conoce ningún plano afín cuyo orden no sea potencia de primo, y es posible probar que no pueden existir en algunos casos (como el caso de orden 6 que acabamos de ver), pero tampoco ha demostrado nadie que no existan.

Capítulo VIII

La geometría proyectiva

Según hemos visto en el capítulo anterior, una buena parte de las propiedades geométricas del espacio que percibimos, las que no dependen de ordenación ni de distancias, son, desde un punto de vista matemático, las propiedades algebraicas de los espacios afines. Esta visión algebraica de la geometría nos da una perspectiva más objetiva de la misma, independiente de la forma en que tenemos de percibirla por nuestra estructura psicológica. Por ejemplo, desde un punto de vista algebraico, \mathbb{R}^3 y \mathbb{R}^4 no son, ciertamente, una misma cosa, pero tienen mucho en común. Desde un punto de vista intuitivo, en cambio, la diferencia es que \mathbb{R}^3 es “todo” y \mathbb{R}^4 es “nada”. Quien entienda la geometría en términos algebraicos se desenvolverá en \mathbb{R}^4 casi con la misma facilidad que en \mathbb{R}^3 , mientras que en términos intuitivos es imposible concebir, digamos, dos planos con un único punto en común.

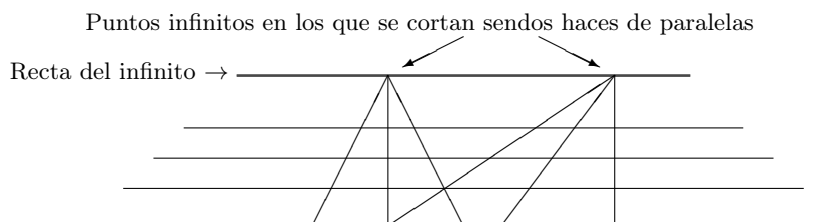
Sin embargo, en este capítulo veremos que la estructura afín no es la forma más natural de describir la geometría desde un punto de vista algebraico, sino que aún está sesgada por la forma de nuestra intuición. Existe otra estructura algebraica todavía más simple, en el sentido de más simétrica y cuyas propiedades se siguen de principios formalmente más sencillos, que describe más eficientemente el espacio intuitivo, junto con una amplia familia de espacios algebraicos de interés.

Estamos hablando de la geometría proyectiva, cuya conexión con la geometría afín consiste en imaginar que el espacio que “vemos” no contiene todos los puntos, sino que hay además unos “puntos infinitos” que no podemos ver porque están infinitamente lejos de cualquier punto finito. Más concretamente, cada recta tiene un único punto infinito, y cuando tenemos en cuenta estos puntos infinitos resulta que dos rectas contenidas en un mismo plano se cortan siempre. Cuando “las vemos” paralelas es porque se cortan “en el infinito”.

Los puntos infinitos de la geometría proyectiva son como los números imaginarios en álgebra. Quien se plantee si existen o no existen realmente no está entendiendo bien la situación. Técnicamente, lo que vamos a demostrar es que a cada espacio afín es posible añadirle unos puntos adicionales (a los que podemos llamar puntos infinitos) de modo que el espacio resultante cumple unos

axiomas alternativos a los de la geometría afín, que son los axiomas de la geometría proyectiva. Estos axiomas implican, entre otras cosas, que no existen rectas paralelas, sino que dos rectas coplanares cualesquiera se cortan siempre en un punto, pero esto no debe entenderse como que la geometría proyectiva contradice a la geometría afín y que, por tanto, es una geometría sin contenido intuitivo. Por el contrario, todos los resultados de la geometría proyectiva tienen una interpretación intuitiva (al menos en dimensiones menores o iguales que tres) como propiedades de los espacios afines, sin más que interpretar los haces de rectas que pasan por un mismo punto infinito como haces de rectas paralelas. Veremos que el punto de vista proyectivo es más natural y ventajoso en muchas situaciones.

Pese a lo dicho, sucede que en realidad los puntos infinitos de un plano proyectivo no son tan “invisibles” ni “ideales” como podría parecer, sino que podemos visualizarlos sin más que mirar el plano oblicuamente, es decir, en perspectiva. Entonces aparece una “línea del horizonte” de forma que todas las rectas de un mismo haz de rectas paralelas confluyen a un punto de dicha línea:



Dicha recta del horizonte es la “recta infinita” que contiene a todos los puntos infinitos del plano. En realidad la representación en perspectiva de un plano nos oculta todavía un punto de la recta infinita, a saber, el punto al cual confluyen las rectas que en la representación aparecen como horizontales. Ello se debe a que la recta infinita, al igual que las “rectas ordinarias”, tiene también su propio punto infinito adicional. Vamos a ver cómo convertir estas ideas en una teoría matemática rigurosa.

8.1 Los axiomas de la geometría proyectiva

Planos proyectivos Recordemos del capítulo anterior que un plano afín puede definirse como un conjunto de puntos E en el que hemos seleccionado una familia de subconjuntos no vacíos llamados rectas, de modo que se cumplen los axiomas siguientes:

- Axioma A1** *Por cada par de puntos distintos pasa una única recta.*
Axioma A2 *Por cada punto pasa una única paralela a una recta dada.*
Axioma A3 *Existen tres puntos no colineales.*

Aquí hay que entender que dos rectas son paralelas si son iguales o no tienen puntos en común.

Definimos ahora un *plano proyectivo* como un conjunto P en el que hemos seleccionado una familia de subconjuntos no vacíos, a los que llamaremos también rectas, de modo que se cumplen los axiomas siguientes:

Axioma P1	<i>Por cada par de puntos distintos pasa una única recta.</i>
Axioma P2	<i>Dos rectas cualesquiera tienen al menos un punto en común.</i>
Axioma P3	<i>Existen tres puntos no colineales.</i>
Axioma P4	<i>Toda recta tiene al menos tres puntos.</i>

Como en el caso afín, llamaremos AB a la única recta que pasa por los puntos distintos A y B . Notemos además que dos rectas distintas en un plano proyectivo tienen un único punto en común. Tienen uno por P2, y no pueden tener más de uno por P1.

Vemos que el axioma P2 niega la existencia de paralelas. Recordemos también que de los axiomas de la geometría plana afín se puede deducir que toda recta tiene al menos dos puntos, luego P4 es una versión reforzada de este hecho.

Así pues, la única diferencia esencial entre los axiomas de la geometría plana afín y la geometría plana proyectiva es que P2 niega la existencia de paralelas que afirma A2. Sin embargo, tal y como hemos explicado en la introducción, sería erróneo interpretar esto como que la geometría proyectiva es incompatible con la geometría afín y, en particular, con nuestra geometría intuitiva. Al contrario, la relación entre ambas es la que muestra el teorema siguiente:

Teorema 8.1 *Sea P un plano proyectivo y sea r_∞ una cualquiera de sus rectas. Sea $E = P \setminus r_\infty$ y definamos las rectas de E como las intersecciones con E de las rectas de P distintas de r_∞ . Entonces E es un plano afín.*

DEMOSTRACIÓN: Observemos que si r es una recta de P distinta de r_∞ , entonces $r \cap r_\infty = \{p_r^\infty\}$, para un único punto p_r^∞ , al que llamaremos “punto infinito de la recta r ”. Entonces, $r \cap E = r \setminus \{p_r^\infty\}$, es decir, las rectas afines resultan de quitar a cada recta proyectiva distinta de r_∞ su punto infinito.

Dados dos puntos $A, B \in E$, la intersección $AB \cap E$ es una recta de E que pasa por ellos, y es única, porque si $r \cap E$ es otra recta de E que pasa por ellos, por P1 tenemos que $r = AB$, luego $r \cap E = AB \cap E$. Esto prueba que E cumple el axioma A1.

Para probar A2 tomamos una recta $r \cap E$ y un punto $A \in E$. Si $A \in r \cap E$, es claro que $r \cap E$ es su única paralela por A . En caso contrario, consideremos la recta $s = Ap_r^\infty$. Tenemos que $r \neq s$, porque $A \in s \setminus r$, luego $r \cap s = \{p_r^\infty\}$. Es claro entonces que $s \cap E$ es paralela a $r \cap E$ y pasa por A .

Si $t \cap E$ es otra recta paralela a $r \cap E$ que pasa por A , entonces, puesto que $r \cap t \neq \emptyset$ y la intersección no puede ser ningún punto de E , tiene que ser $r \cap t = \{p_r^\infty\}$. Por lo tanto, $t = Ap_r^\infty = s$, luego $t \cap E = s \cap E$. Esto prueba que la paralela es única.

Sean A, B dos puntos en r_∞ (existen por P3). Por P4 existe otro punto $C \in P \setminus r_\infty = E$. Las rectas AC y BC tienen cada una un tercer punto $A', B' \in E$, de modo que los puntos $A', B', C \in E$ no son colineales. Así pues, se cumple A3. ■

Recíprocamente, vamos a ver que todo plano afín puede sumergirse en un plano proyectivo:

Definición 8.2 Sea E un plano afín. Para cada recta r de E , llamaremos P_r al haz de rectas paralelas a r . Sea r_∞ el conjunto de todos los haces P_r . Llamaremos *compleción proyectiva* de E al conjunto $E^\infty = E \cup r_\infty$. A los elementos de E los llamaremos *puntos finitos* de E^∞ , mientras que los de r_∞ serán los *puntos infinitos*. Llamaremos *rectas* de E^∞ a los conjuntos de la forma $\bar{r} = r \cup \{P_r\}$, donde r es una recta de E , más el conjunto r_∞ , al que llamaremos *recta infinita* de E^∞ . Por contraposición, a las rectas \bar{r} las llamaremos *rectas finitas* de E^∞ .

En definitiva, la compleción proyectiva se obtiene añadiendo un punto infinito a cada recta, de modo que a dos rectas les añadimos el mismo punto infinito si y sólo si son paralelas.

Teorema 8.3 *La compleción de un plano afín es un plano proyectivo.*

DEMOSTRACIÓN: Es claro que por dos puntos finitos pasa una única recta, concretamente la extensión de la recta afín que pasa por ellos. Si A es un punto finito y P_r un punto infinito, entonces la recta s paralela a r por A se extiende a una recta finita \bar{s} que pasa por A y por $P_s = P_r$. Recíprocamente, una recta que pase por estos dos puntos ha de ser finita, luego de la forma \bar{t} , de modo que t pasa por A y es paralela a r , luego es la misma \bar{s} . Finalmente, si P_r y P_s son dos puntos infinitos, la única recta que pasa por ellos es la recta infinita r_∞ . Esto prueba P1

Veamos que dos rectas cualesquiera se cortan en un punto. Si una es la recta infinita es claro, pues por construcción todas las rectas contienen un punto infinito. Si las dos son finitas, digamos \bar{r} y \bar{s} , entonces o bien r y s tienen un punto en común, en cuyo caso \bar{r} y \bar{s} también, o por el contrario r y s son paralelas, en cuyo caso \bar{r} y \bar{s} tienen en común el punto infinito $P_r = P_s$. Esto prueba P2.

Es claro que tres puntos no colineales en el plano afín de partida siguen siendo no colineales en la compleción, luego se cumple P3

Toda recta afín r tiene al menos dos puntos, y su compleción \bar{r} tiene además un tercer punto infinito, luego todas las rectas finitas tienen al menos tres puntos. Para probar P4 falta ver que la recta infinita tiene al menos tres puntos, pero tomando tres puntos finitos no colineales A, B, C obtenemos tres rectas no paralelas dos a dos, a saber, AB, AC, BC , y cada una de ellas se extiende con un punto infinito distinto. ■

Esto hace posible (y conveniente) pensar que un plano afín intuitivo está sumergido en un plano proyectivo cuyos puntos infinitos no podemos ver, pero sobre los cuales podemos razonar indirectamente a través de la parte afín del plano que sí que vemos (aunque, como ya hemos indicado, los puntos infinitos también podemos “verlos” mirando el plano oblicuamente).

Conviene observar que los axiomas que hemos fijado para los planos proyectivos son equivalentes a éstos:

Axioma P1	<i>Por cada par de puntos distintos pasa una única recta.</i>
Axioma P2'	<i>Dos rectas distintas cualesquiera tienen un único punto en común.</i>
Axioma P3'	<i>Existen cuatro puntos no colineales tres a tres.</i>

En efecto, P2' implica obviamente P2 y es consecuencia de P1 y P2, porque si dos rectas tuvieran dos puntos en común, serían la misma recta por P1. Por lo tanto, unos puntos y rectas cumplen P1 y P2 si y sólo si cumplen P1 y P2'. Veamos ahora que el primer grupo de axiomas implica P3'.

Sean A, B, C tres puntos no colineales, cuya existencia está garantizada por el axioma P3. Entonces las rectas $l_1 = AB$ y $l_2 = BC$ son distintas. Por P4 la recta l_1 contiene un tercer punto X distinto de A y B . Sea $l_3 = XC$, que es una recta distinta de l_1 (porque pasa por C) y de l_2 (porque pasa por X). De nuevo por P4, la recta l_3 contiene un tercer punto D distinto de X y C . Es fácil ver entonces que los cuatro puntos A, B, C, D no son colineales tres a tres. ■

Recíprocamente, es obvio que P3' implica P3. Vamos a ver que también implica P4.

Tomamos cuatro puntos A, B, C, D no colineales tres a tres y sea l una recta cualquiera. Entonces a lo sumo pasa por dos de los cuatro puntos. No perdemos generalidad si suponemos que no pasa por A . Entonces las rectas AB, AC y AD son distintas (o habría tres puntos colineales) y tienen que cortar a l en tres puntos distintos (pues si dos cortan a l en el mismo punto, éste será distinto de A y las dos rectas tendrán dos puntos en común). Por lo tanto, l tiene al menos tres puntos. ■

Con esto queda probado que los dos conjuntos de axiomas son equivalentes.

Dualidad Los axiomas P1 y P2' son duales en el sentido de que uno se obtiene del otro intercambiando las palabras “punto” y “recta”. La afirmación dual de P3' es fácil de probar:

Existen cuatro rectas no concurrentes tres a tres.

En efecto, basta tomar cuatro puntos A, B, C, D no colineales tres a tres y considerar las rectas AB, BC, CD, AD . ■

De aquí se sigue que si P es un plano proyectivo, podemos formar otro plano proyectivo P^* tomando como puntos las rectas de P y tomando como rectas los haces de rectas de P que pasan por un mismo punto. En efecto, la comprobación de que P^* cumple cada uno de los axiomas P1, P2', P3' se reduce al hecho de que P cumple las afirmaciones duales correspondientes. A este plano proyectivo se le llama *plano dual* de P .

Nos encontramos así con una característica que la geometría proyectiva que no tiene la geometría afín, y es que si todo plano proyectivo cumple una afirmación (en la que intervengan únicamente los conceptos de “punto”, “recta” e incidencia), la afirmación dual que resulta de intercambiar los conceptos de “punto” y “recta” también es válida en todo plano proyectivo.

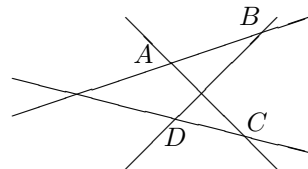
Esto se debe a que un plano proyectivo P cumplirá una afirmación si y sólo si el plano dual P^* cumple la afirmación dual. Más adelante generalizaremos este concepto de dualidad a espacios proyectivos de cualquier dimensión y a afirmaciones que incluyan conceptos más sofisticados que los meros conceptos de “punto” y “recta”.

Espacios proyectivos Veamos ahora unos axiomas que dejen abierta la dimensión del espacio. La idea básica es que en un espacio proyectivo de dimensión mayor que 2 ya no es cierto que todo par de rectas tengan que cortarse, sino que las rectas pueden cruzarse igual que en los espacios afines. Lo que queremos es que dos rectas contenidas en el mismo plano se corten siempre. Podemos recoger esta idea introduciendo el concepto de “plano” entre los conceptos primitivos de la geometría proyectiva o bien mediante un axioma lo suficientemente ingenioso como para expresar lo mismo sin recurrir al concepto de plano. De momento seguimos este segundo camino y luego explicaremos la alternativa:

Definición 8.4 Un *espacio proyectivo* es un conjunto P , a cuyos elementos llamaremos *puntos*, junto con una familia de subconjuntos a los que llamaremos *rectas* y que cumplan los axiomas siguientes:

Axioma P1	<i>Por cada par de puntos distintos pasa una única recta.</i>
Axioma P2	<i>Si A, B, C, D son cuatro puntos distintos tales que las rectas AB y CD se cortan, entonces AC y BD también se cortan.</i>
Axioma P3	<i>Existen tres puntos no colineales.</i>
Axioma P4	<i>Toda recta tiene al menos tres puntos.</i>

El axioma P2 se conoce también como *axioma de Veblen-Young*. La idea subyacente es que la hipótesis fuerza a que las rectas AC y BD estén en el mismo plano, aun sin mencionar la palabra ‘plano’.



Observemos que el axioma P2 de la definición de plano proyectivo implica trivialmente el axioma de Veblen-Young, por lo que todo plano proyectivo es un espacio proyectivo en este sentido general.

Definición 8.5 Si P es un espacio proyectivo, una *variedad proyectiva* en P es un subconjunto E tal que si $P, Q \in E$ son dos puntos distintos, entonces $PQ \subset E$.

Obviamente, el conjunto vacío, los subconjuntos de un solo punto, las rectas y el espacio P son variedades proyectivas.

También es inmediato que la intersección de toda familia no vacía de variedades proyectivas es una variedad proyectiva, por lo que podemos definir la *variedad proyectiva generada por un conjunto $X \subset P$* como la intersección $\langle X \rangle$ de todas las variedades proyectivas en P que contienen a X . Si $X = \{A_0, \dots, A_n\}$ escribiremos $\langle X \rangle = \langle A_0, \dots, A_n \rangle$.

Es claro que $\langle \emptyset \rangle = \emptyset$, $\langle A \rangle = \{A\}$, y que si $A \neq B$, entonces $\langle A, B \rangle = AB$.

El lector debería comparar estas definiciones con las correspondientes para espacios afines que dimos en el capítulo anterior. Ya puede apreciarse que la definición de variedad proyectiva es más simple que la de variedad afín porque no hay que tratar aparte el caso de las rectas paralelas. La simplificación es más patente al comparar el teorema siguiente con su análogo afín:

Teorema 8.6 *Sea E una variedad proyectiva no vacía y sea $A \in P \setminus E$. Entonces $\langle E, A \rangle$ es la unión de todas las rectas que pasan por A y un punto de E .*

DEMOSTRACIÓN: Llamemos E_A al conjunto descrito en el enunciado. Tenemos que probar que $\langle E, A \rangle = E_A$, y es claro que $E_A \subset \langle E, A \rangle$. Para probar la otra inclusión basta demostrar que E_A es una variedad proyectiva.

Para ello tomamos dos puntos distintos $B, C \in E_A$ y $D \in BC$. Tenemos que demostrar que $D \in E_A$.

Por hipótesis B y C están en dos rectas AX_B y AX_C , con $X_B, X_C \in E$. Si $X_B = X_C$ entonces $BC = AX_B \subset E_A$. Supongamos, pues que $X_B \neq X_C$. También podemos suponer que $B \neq A \neq C$.

Si $B = X_B$ y $C = X_C$ entonces $BC = X_B X_C \subset E \subset E_A$. Por lo tanto, podemos suponer que $B \neq X_B$ o $C \neq X_C$. Por simetría, no perdemos generalidad si suponemos $B \neq X_B$.

Veamos que BC corta a $X_B X_C$. Esto es obvio si $C = X_C$. En caso contrario aplicamos el axioma P2: como las rectas BX_B y CX_C se cortan en A , concluimos que BC y $X_B X_C$ se cortan en un punto $Q \in E$.

Finalmente, como $QD = AB$ y AX_B se cortan en B , el axioma P2 implica que AD corta a $QX_B \subset E$, luego $R \in E_A$. ■

Teorema 8.7 *Sea E una variedad proyectiva en un espacio proyectivo P , sea $B \in P \setminus E$ y sea $C \in \langle E, B \rangle \setminus E$. Entonces $B \in \langle E, C \rangle \setminus E$ y $\langle E, B \rangle = \langle E, C \rangle$.*

DEMOSTRACIÓN: Tenemos que C está en una recta que pasa por B y un punto de E , luego B está en una recta (la misma) que pasa por C y un punto de E , luego $B \in \langle E, C \rangle \setminus E$. Esto implica que $\langle E, B \rangle \subset \langle E, C \rangle$, e igualmente tenemos la otra inclusión. ■

Definición 8.8 Una variedad proyectiva E es *finitamente generada* si admite un generador finito. En tal caso, si el menor número de generadores de E es $n+1$, diremos que E tiene *dimensión n* . Representaremos por $\dim E$ a la dimensión de E .

Un conjunto de $n+1$ puntos es *proyectivamente independiente* si no está contenido en una variedad proyectiva de dimensión menor que n (y por lo tanto genera una variedad proyectiva de dimensión n).

A partir de aquí, todos los resultados análogos para variedades proyectivas a los comprendidos desde la definición 7.8 hasta el teorema 7.11 (no incluido) se enuncian de forma obvia y demuestran exactamente con las mismas pruebas.

En particular, cada variedad proyectiva de dimensión n contiene $n+1$ puntos proyectivamente independientes, y por $n+1$ puntos proyectivamente independientes pasa una única variedad de dimensión n . Llamaremos planos a las variedades proyectivas de dimensión 2. El equivalente (si no análogo) a 7.11 es el hecho siguiente:

Teorema 8.9 *En un espacio proyectivo, dos rectas distintas se cortan si y sólo si son coplanares.*

DEMOSTRACIÓN: Si r y s se cortan en un punto A , entonces $r = AB$ y $s = AC$, para ciertos puntos B y C , y entonces ambas están contenidas en el plano ABC .

Recíprocamente, si $r = AB$ y $s = CD$ están contenidas en un plano, podemos suponer que los dos puntos que determinan una de las rectas no están en la otra (pues si no, ya tenemos que las rectas se cortan). Por la unicidad del plano que pasa por tres puntos no colineales, el plano que las contiene es r_C , luego D está en una recta que pasa por C y por un punto $P \in r$, pero dicha recta tiene que ser $CD = s$, luego $P \in r \cap s$. ■

Esto implica que los planos proyectivos (en el sentido de ‘variedades proyectivas de dimensión 2’) cumplen los axiomas de plano proyectivo.

Ahora es fácil ver que la definición siguiente de espacio proyectivo es equivalente a la que hemos dado:

Definición 8.10 Un *espacio proyectivo* es un conjunto P , a cuyos elementos llamaremos *puntos*, junto con dos familias de subconjuntos a los que llamaremos *rectas* y *planos* y que cumplan los axiomas siguientes:

Axioma P1	<i>Por cada par de puntos distintos pasa una única recta.</i>
Axioma P2	<i>Dos rectas se cortan si y sólo si son coplanares.</i>
Axioma P3	<i>Por tres puntos no colineales pasa un único plano.</i>
Axioma P4	<i>Si una recta pasa por dos puntos de un plano, está contenida en el plano.</i>
Axioma P5	<i>Todo plano contiene tres puntos no colineales.</i>
Axioma P6	<i>Toda recta tiene al menos tres puntos.</i>

Nota En lo sucesivo consideraremos únicamente espacios proyectivos de dimensión finita n . A las variedades proyectivas de dimensión $n-1$ las llamaremos *hiperplanos*. ■

Veamos una variante de 8.9. Más adelante daremos un teorema general sobre la dimensión de la intersección de dos variedades proyectivas:

Teorema 8.11 *Si H es un hiperplano en un espacio proyectivo y E es una variedad proyectiva no vacía de dimensión d no contenida en él, entonces la intersección $E \cap H$ es una variedad proyectiva de dimensión $d-1$.*

DEMOSTRACIÓN: Sea $A \in E \setminus H$. Entonces $P = \langle H, A \rangle$. Razonamos por inducción sobre d . Si $d = 0$ es trivial. Si $d = 1$, es decir, si E es una recta, tomamos otro punto $B \in E$, y entonces B está en una recta que pasa por A y por un punto de H , pero dicha recta tiene que ser r , luego r corta a H en un punto.

Si E tiene dimensión $d + 1$, podemos expresarla como $E = \langle E', A \rangle$, donde E' es una variedad proyectiva de dimensión d . Si $E' \subset H$ es claro que $E' = E \cap H$ y tenemos ya la conclusión. En caso contrario, por hipótesis de inducción, tenemos que $E'' = E' \cap H$ es una variedad de dimensión $d - 1$. Sea $B \in E' \setminus E''$. En particular $B \notin H$ y $B \neq A$. La recta AB corta a H en un punto C (por el caso $d = 1$ ya probado).

Tenemos que $C \in AB \subset E$ y $C \notin E'$, porque AB corta a E' en el punto B (que no es C , porque $B \notin H$), luego la variedad $E^* = \langle E'', C \rangle \subset E \cap H$ y tiene dimensión d , luego claramente $E^* = E \cap H$. ■

En particular, en un espacio proyectivo tridimensional, dos planos cualesquiera se cortan en una recta.

Ahora generalizamos el teorema 8.1 a dimensiones arbitrarias:

Teorema 8.12 *Sea P un espacio proyectivo de dimensión $n \geq 2$ y H_∞ un hiperplano en P . Entonces $E = P \setminus H_\infty$ es un espacio afín de dimensión n tomando como rectas las intersecciones $r \cap E$, donde r es una recta de P no contenida en H_∞ . Las variedades afines d -dimensionales de E son las de la forma $V \cap E$, donde V es una variedad proyectiva d -dimensional de P no contenida en H_∞ .*

DEMOSTRACIÓN: Cada recta proyectiva r no contenida en H_∞ corta a H_∞ en un único punto p_r^∞ , luego $r \cap E = r \setminus \{p_r^\infty\}$. Recíprocamente, para cada recta de E existe una única recta \bar{r} de P tal que $r = \bar{r} \cap E$, pues r tiene al menos dos puntos, luego dos rectas proyectivas que la contengan tienen que ser iguales. Esto hace que podamos hablar del punto infinito de una recta de r como el único punto p_r^∞ que hace que $r \cup \{p_r^\infty\}$ sea una recta de P .

Definimos el paralelismo en E estableciendo que dos rectas de E son paralelas si y sólo si tienen el mismo punto infinito. Vamos a comprobar que E cumple la definición 7.3. Es inmediato que el paralelismo que hemos definido es una relación de equivalencia.

La comprobación de A1, A2 y A4 es análoga a la que ya hemos hecho en la prueba de 8.1. Veamos A3. Tenemos tres puntos no colineales $A, B, C \in E$. Sea Π el plano proyectivo que los contiene. Entonces $\Pi \cap H_\infty$ es una recta r_∞ , la cual contiene a los puntos infinitos $p_\infty, q_\infty, r_\infty$ de las rectas AB, AC y BC , que son distintos dos a dos, porque las rectas (restringidas a E) no son paralelas.

Tomamos ahora los puntos A', B' tales que $A'B'$ es paralela a AB , luego pasa por p_∞ . Sea Π' el plano que contiene a $A'B'$ y a q_∞ , que también contiene a p_∞ , luego a r_∞ . La recta paralela a AC por A' es $A'q_\infty$, y la paralela a BC por B' es $B'r_\infty$. Como ambas están contenidas en Π' , tienen que cortarse, pero no lo hacen en el infinito, porque sus puntos infinitos son distintos, luego se cortan en E , como había que probar.

Ahora observamos que si V es una variedad proyectiva no contenida en H_∞ , entonces $V \cap E$ cumple las dos condiciones que definen las variedades afines.

Veamos por inducción sobre n que si A_0, \dots, A_n son puntos de E afínmente independientes, entonces son proyectivamente independientes y la variedad afín que generan es la intersección con E de la variedad proyectiva que generan. Si $n = 1$ es trivial. Si vale para $n - 1$, sea $V_{n-1}^p = \langle A_0, \dots, A_{n-1} \rangle \subset P$, de modo que $V_{n-1}^a = V_{n-1}^p \cap E$ es la variedad afín generada por los n puntos. Como son afínmente independientes, $A_n \notin V_{n-1}^p \cap E$, luego de hecho $A_n \notin V_{n-1}^p$, luego A_0, \dots, A_n son proyectivamente independientes.

Si V_n^p y V_n^a son, respectivamente, la variedad proyectiva y la variedad afín que generan, sabemos que V_n^a es la unión de todas las rectas paralelas a A_0A_n que cortan a V_{n-1}^a . Si llamamos p_∞ al punto infinito de A_0A_n , se trata de la unión de todas las rectas que pasan por p_∞ y un punto de V_{n-1}^a (quitando p_∞).

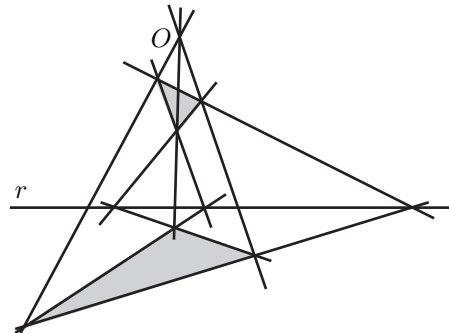
Observemos que A_0A_n está contenida en V_n^p , pero no en V_{n-1}^p , luego se cumple que $p_\infty \in V_n^p \setminus V_{n-1}^p$ y, por consiguiente, V_n^p es la unión de todas las rectas que pasan por p_∞ y un punto de V_{n-1}^p . Es fácil ver entonces que $V_n^a = V_n^p \cap E$. ■

Veremos también que todo espacio afín de dimensión finita se puede obtener de un espacio proyectivo en la forma descrita por el teorema anterior. El lector puede demostrarlo imitando la definición 8.2, pero luego daremos una prueba analítica.

Conviene señalar que la prueba del teorema anterior muestra que el “hiperplano del infinito” no es ningún hiperplano especial en un espacio proyectivo, sino que cualquier hiperplano H de un espacio proyectivo P es el hiperplano del infinito del espacio afín $P \setminus H$.

En las secciones siguientes veremos cómo describir analíticamente un espacio proyectivo asignando coordenadas a sus puntos. Como en el caso afín, esto sólo es posible si el espacio cumple el teorema de Desargues, el cual se cumple siempre en espacios de dimensión mayor o igual que 3, pero no necesariamente en dimensión 2. Vamos a enunciar y demostrar la versión proyectiva dicho teorema:

Definición 8.13 Un *triángulo* en un espacio proyectivo P es un conjunto de tres puntos no colineales, a los que llamaremos *vértices*. Los *lados* del triángulo serán las rectas que pasan por cada par de vértices. Diremos que dos triángulos ABC y $A'B'C'$, con sus vértices en este orden, tienen a un punto O como *centro de perspectiva* si las rectas AA' , BB' y CC' concurren en O . Diremos que tienen a una recta r como *eje de perspectiva* si cada par de lados AB y $A'B'$, AC y $A'C'$, BC y $B'C'$ se cortan en un punto de r .



Así, si pensamos que el plano afín que “vemos” es un plano proyectivo salvo una recta infinita que no vemos, podemos decir que si dos triángulos tienen un centro de perspectiva finito, sus vértices están sobre rectas concurrentes, mientras que si tienen un centro de perspectiva infinito, sus vértices están sobre rectas paralelas. Si tienen un eje de perspectiva infinito entonces sus lados son paralelos dos a dos.

Observación Podemos pensar la figura precedente como una imagen tridimensional en perspectiva, de modo que la recta r es la recta del horizonte, con un triángulo por debajo de nuestra vista “en el suelo” y otro por encima, en el aire. Así entendida, los lados de los triángulos son paralelos dos a dos. ■

Teorema 8.14 (Teorema de Desargues) *En un espacio proyectivo de dimensión mayor que dos, dos triángulos tienen un centro de perspectiva si y sólo si tienen un eje de perspectiva.*

DEMOSTRACIÓN: Sean dos triángulos ABC y $A'B'C'$. Si dos vértices coinciden, por ejemplo, $A = A'$, el resultado es obvio. Por ejemplo, las rectas BB' y CC' (o dos rectas que pasen por estos puntos, si coinciden) se cortan en un punto O , al igual que la recta OA , luego O es un centro de perspectiva para los triángulos. Del mismo modo se prueba que tienen un eje de perspectiva. Supongamos, pues, $A \neq A'$, $B \neq B'$, $C \neq C'$. Un razonamiento similar nos permite suponer que los lados son diferentes: $AB \neq A'B'$, $AC \neq A'C'$, $BC \neq B'C'$

Sea Π el plano que pasa por A, B, C y Π' el plano que pasa por A', B', C' . Supongamos primeramente que $\Pi \neq \Pi'$.

Si los triángulos tienen un eje de perspectiva r , consideramos los planos Π_1 , Π_2 y Π_3 que contienen respectivamente a los lados AB y $A'B'$, AC y $A'C'$, BC y $B'C'$. Como $\Pi \neq \Pi'$, el plano Π_1 corta a Π exactamente en la recta AB , mientras que los otros dos lo hacen en AC y BC . En particular los tres planos son distintos. Claramente $\Pi_1 \cap \Pi_2 = AA'$, $\Pi_1 \cap \Pi_3 = BB'$ y $\Pi_2 \cap \Pi_3 = CC'$.

La recta AA' no puede estar contenida en Π_3 , pues entonces cortaría a Π en un punto de BC , luego A sería colineal con BC . Así pues, $AA' \cap \Pi_3$ es un único punto O . Por lo tanto $O = \Pi_1 \cap \Pi_2 \cap \Pi_3$, luego las tres rectas AA' , BB' y CC' han de concurrir en O , que es, por consiguiente, un centro de perspectiva de los triángulos.

Supongamos ahora que los triángulos tienen un centro de perspectiva O , que necesariamente ha de ser distinto de los seis vértices. Notemos que $\Pi' \subset \langle \Pi, O \rangle$, que es una variedad proyectiva tridimensional en la que Π y Π' son hiperplanos, luego la intersección $\Pi \cap \Pi'$ tiene que ser una recta s .

El plano OAB corta a Π' en $A'B'$. Similarmente, el plano OAC corta a Π' en $A'C'$ y el plano OBC corta a Π' en $B'C'$. Así pues, AB y $A'B'$ están en un mismo plano, luego se cortan en un punto, que necesariamente está en s . Lo mismo sucede con los otros pares de lados, luego s es un eje de perspectiva.

Ahora supongamos que los dos triángulos están en un mismo plano Π y que tienen un eje de perspectiva s . La hipótesis de tridimensionalidad nos permite tomar un plano Π' que corte a Π en s . Sea O un punto fuera de Π y Π' (una recta que pase por un punto de Π y otro de Π' tiene un tercer punto fuera de ambos planos).

Sean A'' , B'' y C'' los puntos donde las rectas OA' , OB' y OC' cortan a Π' . Es fácil ver que estos tres puntos no son colineales. Los triángulos $A''B''C''$ y $A'B'C'$ tienen a O como centro de perspectiva. Por la parte ya probada tienen un eje de perspectiva, y concretamente es la recta s .

Entonces $A''B''$ y $A'B'$ se cortan en un punto de s , que ha de ser también el punto donde se cortan AB y $A'B'$. Por lo tanto $A''B''$ y AB se cortan en un punto de s y lo mismo vale para los demás lados. Por consiguiente s es un eje de perspectiva para ABC y $A''B''C''$. Por la parte ya probada ambos triángulos tienen un centro de perspectiva O' . No puede ser $O' = O$, pues entonces $A = A'$, $B = B'$, $C = C'$.

Sea O'' el punto donde OO' corta a π . Veamos que O'' es un centro de perspectiva para los triángulos originales. Tenemos que los puntos O' , A y A'' son colineales. El plano que contiene a estos tres y a O corta a π en una recta que contiene a O'' , a A y a A' , luego los tres están alineados. Similarmente sucede con los otros vértices.

Supongamos que los triángulos tienen un centro de perspectiva O . De nuevo por la hipótesis de tridimensionalidad, podemos tomar un punto O' exterior a Π . Sea \bar{A} un tercer punto en la recta $O'A$. Sea \bar{A}' la intersección de $O\bar{A}$ con $O'A'$ (ambas rectas están en el plano OAA'). El punto O es el centro de perspectiva de los triángulos $\bar{A}BC$ y $\bar{A}'B'C'$, y éstos no están sobre el mismo plano, luego por la parte ya probada ambos triángulos tienen un eje de perspectiva r , que concretamente es la intersección de los planos de los dos triángulos. O' no está en r , pues la recta $O\bar{A}$ corta a $\bar{A}BC$ en \bar{A} (si estuviera contenida entonces A estaría en BC). Sea r' la intersección con Π del plano que contiene a O y a r . Veamos que r' es un eje de perspectiva para los triángulos originales. Las rectas $\bar{A}B$ y $\bar{A}'B'$ se cortan en un punto P de r . Sea P' el punto donde $O'P$ corta a Π . Se trata de un punto de r' . El plano $O'PB$ contiene también a P' , a \bar{A} y a A , luego su intersección con Π es la recta AB , que también contiene a P' . Del mismo modo se prueba que $A'B'$ pasa por P' , luego los lados AB y $A'B'$ se cortan en r' . Lo mismo vale para AC y $A'C'$. Respecto a BC y BC' , se cortan en un punto de r que ha de estar también en Π , luego en r' ■

De aquí se deducen las distintas versiones afines del teorema de Desargues que hemos demostrado separadamente en los capítulos precedentes así como en [IGE 4.59]. Por ejemplo, si dos triángulos tienen sus lados paralelos dos a dos (o concurrentes), eso se interpreta como que tienen un centro de perspectiva (que sea finito o infinito es irrelevante), luego por el teorema de Desargues tienen un eje de perspectiva, lo cual significa que las rectas que unen sus vértices son concurrentes (sin descartar que el punto de concurrencia sea infinito, en cuyo caso las rectas son paralelas).

Es inmediato comprobar que si P es un plano proyectivo arguesiano (es decir, si cumple el teorema de Desargues) y r es una recta en P , entonces $P \setminus r$ es un espacio afín arguesiano. Esto implica que existen planos proyectivos no arguesianos, pues basta considerar la compleción proyectiva de un plano afín no arguesiano. Por otra parte, todo plano proyectivo contenido en un espacio proyectivo de dimensión superior es obviamente arguesiano.

8.2 Espacios proyectivos analíticos

En la sección siguiente demostraremos que, con la única excepción de los planos proyectivos no arguesianos, los puntos de un espacio proyectivo se pueden describir en términos de coordenadas con valores en un cierto anillo de división. El primer paso es describir el análogo proyectivo de los espacios afines en sentido algebraico, y de eso nos ocupamos ahora. Para no mezclar confusamente los conceptos sintéticos que ya conocemos con los analíticos que vamos presentar aquí, estudiaremos los espacios proyectivos analíticos de forma independiente a la presentación axiomática que acabamos de ver, y en la sección siguiente relacionaremos ambos puntos de vista.

Aunque en la práctica nos interesarán únicamente los espacios proyectivos sobre cuerpos, trabajaremos con un anillo de división K para conectar adecuadamente la geometría proyectiva con los resultados generales que hemos obtenido para la geometría afín, lo cual no complica en nada el desarrollo de la teoría.

Definición 8.15 Si K es un anillo de división y V es un K -espacio vectorial a derecha de dimensión $n + 1 \geq 2$, definimos el *espacio proyectivo de soporte V* como el conjunto $P(V)$ formado por todos los subespacios vectoriales de V de dimensión 1. A sus elementos los llamaremos *puntos proyectivos*. Llamaremos $P^n(K) = P(K^{n+1})$.

Así pues, un punto $A \in P(V)$ es de la forma $\langle v \rangle$, donde $v \in V$ es un vector no nulo. Esta definición es un tecnicismo pensado para que los espacios proyectivos tengan las propiedades deseadas. Vamos a exponer lo más rápidamente posible las ideas geométricas subyacentes.

Para empezar, observamos en un espacio proyectivo podemos hablar de rectas, planos, etc.:

Variedades proyectivas Si W es un subespacio vectorial de V , es claro que $P(W) \subset P(V)$ (todo subespacio vectorial de dimensión 1 de W es también un subespacio de V). A los subconjuntos de $P(V)$ de la forma $P(W)$ los llamaremos *variedades proyectivas* de $P(V)$.

Observemos que si E es una variedad proyectiva de $P(V)$, existe un único espacio vectorial $W \subset V$ tal que $E = P(W)$.

En efecto: un caso especial se da si $W = 0$, en cuyo caso $E = P(0) = \emptyset$. Tenemos así que el conjunto vacío es una variedad proyectiva, y es claro que sólo puede expresarse como $P(W)$ para $W = 0$, pues si existe $v \in W$ no nulo, entonces $A = \langle v \rangle \in P(W) \neq \emptyset$.

Si $E = P(W) \neq \emptyset$, entonces necesariamente $W = \bigcup E$. En efecto, si $v \in \bigcup E$, existe un $A \in E$ tal que $v \in A$, pero A es entonces un subespacio vectorial de W de dimensión 1, luego $v \in A \subset E$.

Recíprocamente, si $v \in W$ no es el vector nulo, tenemos que $v \in \langle v \rangle \in E$, luego $v \in \bigcup E$. Como W contiene en cualquier caso un vector $v \neq 0$, tenemos que $0 \in \langle v \rangle \in E$, luego también $0 \in \bigcup E$, luego $W \subset \bigcup E$. ■

Por lo tanto, podemos definir el *espacio soporte* de una variedad proyectiva $E \subset P(V)$ como el único subespacio vectorial $W \subset V$ tal que $E = P(W)$.

Es inmediato que $W_1 \subset W_2$ si y sólo si $P(W_1) \subset P(W_2)$.

A su vez, podemos definir la dimensión de una variedad proyectiva como

$$\dim P(W) = \dim W - 1.$$

De este modo, tomando $W = 0$ vemos que $\dim \emptyset = -1$. Si $\dim W = 1$, entonces $P(W) = \{W\}$ es un punto arbitrario de $P(V)$ (técnicamente, es un conjunto que contiene un único punto de $P(V)$), luego concluimos que los puntos de $P(V)$ son las variedades proyectivas de dimensión 0.

Llamaremos *rectas* y *planos* de $P(V)$ a las variedades proyectivas de dimensión 1 y 2 respectivamente, es decir, las que tienen espacios soporte de dimensión 2 y 3. Si $\dim P(V) = n$ (es decir, si $\dim V = n + 1$), llamaremos *hiperplanos* de $P(V)$ a las variedades proyectivas de dimensión $n - 1$.

Es fácil ver que la intersección de una familia no vacía de variedades proyectivas es una variedad proyectiva, concretamente,

$$\bigcap_{i \in I} P(W_i) = P\left(\bigcap_{i \in I} W_i\right).$$

Por lo tanto, dado $X \subset P(V)$, podemos definir la variedad proyectiva $\langle X \rangle$ generada por X como la intersección de todas las variedades proyectivas que contienen a X . Si $X = \{A_0, \dots, A_n\}$ es un conjunto finito, escribiremos también $\langle X \rangle = \langle A_0, \dots, A_n \rangle$.

Como $X \subset P(W)$ equivale a $\bigcup X \subset W$, es claro que

$$\langle X \rangle = P(\langle \bigcup X \rangle).$$

Si E_1 y E_2 son dos variedades proyectivas en $P(V)$, definimos su suma

$$E_1 + E_2 = \langle E_1 \cup E_2 \rangle$$

como la menor variedad proyectiva que las contiene a ambas. Puesto que, claramente $P(W) \subset P(W')$ si y sólo si $W \subset W'$, es inmediato que

$$P(W_1) + P(W_2) = P(W_1 + W_2).$$

He aquí un resultado fundamental:

Teorema 8.16 *Si E_1 y E_2 son variedades proyectivas en un espacio $P(V)$, entonces*

$$\dim(E_1 + E_2) = \dim E_1 + \dim E_2 - \dim(E_1 \cap E_2).$$

DEMOSTRACIÓN: Sea $E_i = P(W_i)$, donde W_1, W_2 son subespacios vectoriales de V . Entonces tenemos la relación

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2),$$

luego aplicando P obtenemos que

$$\dim P(W_1 + W_2) - 1 = \dim P(W_1) - 1 + \dim P(W_2) - 1 - \dim P(W_1 \cap W_2) + 1,$$

y como los unos se simplifican, llegamos a la relación del enunciado. ■

Este teorema ya nos da una primera idea de que los espacios proyectivos se comportan “como debe comportarse” un espacio proyectivo. Por ejemplo, si E_1 y E_2 son dos rectas proyectivas disjuntas, entonces $\dim(E_1 \cap E_2) = -1$, y el miembro derecho de la fórmula del teorema anterior suma 3, luego dos rectas en un plano proyectivo se cortan necesariamente.

Otra consecuencia destacable es que si H es un hiperplano en $P(V)$ y E es una variedad de dimensión d no contenida en H , entonces $E + H = P(V)$, luego el teorema anterior implica que $\dim(H \cap E) = d - 1$.

Observemos en la geometría afín no es posible dar una fórmula análoga a la anterior, pues, por ejemplo, si E_1 y E_2 son rectas coplanares, entonces $\dim(E_1 + E_2) = 2$, $\dim E_1 = \dim E_2 = 1$, pero $\dim(E_1 \cap E_2)$ puede ser 0 o -1 según si las rectas son secantes o paralelas.

Homografías Antes de seguir explorando la geometría de los espacios proyectivos conviene introducir las aplicaciones que conservan la estructura de forma natural:

Si $f : V_1 \rightarrow V_2$ es un isomorfismo de espacios vectoriales, la *homografía* inducida por f es la aplicación $H(f) : P(V_1) \rightarrow P(V_2)$ definida mediante $H(f)(\langle v \rangle) = \langle f(v) \rangle$.

Es inmediato comprobar que si U es un subespacio vectorial de V , entonces $H(f)[P(U)] = P(f[U])$, de modo que las homografías transforman variedades proyectivas en variedades proyectivas de la misma dimensión.

Llamaremos *grupo lineal general proyectivo* de un espacio proyectivo $X = P(V)$ al grupo de todas las homografías de P en sí mismo. Lo representaremos por $\text{LGP}(X)$ o $\text{LGP}(V)$.

Es claro que $H : \text{LG}(V) \rightarrow \text{LGP}(V)$ es un epimorfismo de grupos. Su núcleo está formado por todas las aplicaciones lineales que dejan invariante a cada recta de V que pasa por 0. Tales aplicaciones son exactamente las homotecias lineales de V . Así pues, $\text{LGP}(V) \cong \text{LG}(V)/H(V)$.

Si fijamos una base de V , tenemos un isomorfismo $\text{LG}(V) \cong \text{LG}(n, K)$ entre el grupo de automorfismos de V y el grupo de matrices regulares de n filas y columnas con coeficientes en K , a través del cual $H(V)$ se corresponde con el subgrupo de las matrices escalares, que según [A1 4.39] es el centro de $\text{LG}(n, K)$. Por lo tanto, $\text{LGP}(V)$ es isomorfo al cociente de $\text{LG}(V)$ sobre su centro, y podemos definir, para $n \geq 2$, los grupos lineales generales proyectivos

$$\text{LGP}(n, K) = \text{LG}(n, K)/Z(\text{LG}(n, K)).$$

(Notemos que $\text{LGP}(n, K)$ se corresponde con el grupo de homografías de un espacio proyectivo de dimensión $n - 1$. Los grupos $\text{LGP}(1, K)$ son todos triviales.)

Ahora vamos a ver que los puntos de un espacio proyectivo pueden describirse a través de coordenadas, como sucede en los espacios afines, para lo cual tenemos que definir el concepto de sistema de referencia:

Puntos independientes Diremos que unos puntos $A_0, \dots, A_m \in P(V)$ son *proyectivamente independientes* si no están contenidos en ninguna variedad proyectiva de dimensión menor que m .

Si $A_i = \langle v_i \rangle$, esto equivale a que los vectores v_0, \dots, v_m sean linealmente independientes en V . En efecto, se cumple que $A_0, \dots, A_m \in E = P(W)$ si y sólo si $v_0, \dots, v_m \in W$, luego si los vectores son independientes, necesariamente $\dim W \geq m + 1$, luego $\dim E \geq m$ y, viceversa, si son dependientes, entonces $W = \langle v_0, \dots, v_m \rangle$ cumple $\dim W \leq m$, luego $E = P(W)$ es una variedad de dimensión menor que m que contiene a los $m + 1$ puntos.

Así, un único punto es siempre proyectivamente independiente, y también lo son dos puntos distintos. En cambio, tres puntos son proyectivamente independientes si y sólo si no son colineales, cuatro puntos son proyectivamente independientes si y sólo si no son coplanares, etc.

La caracterización en términos de la independencia lineal de los vectores v_i implica inmediatamente que todo subconjunto de un conjunto proyectivamente independiente es proyectivamente independiente, así como que todo conjunto proyectivamente independiente puede completarse hasta uno con $n + 1$ elementos, donde n es la dimensión del espacio.

Ahora podemos demostrar otra propiedad geométrica “razonable” de los espacios proyectivos:

Teorema 8.17 *Por $m + 1$ puntos proyectivamente independientes pasa una única variedad proyectiva de dimensión m .*

DEMOSTRACIÓN: En efecto, si A_0, \dots, A_m son proyectivamente independientes, digamos con $A_i = \langle v_i \rangle$, entonces $W = \langle v_0, \dots, v_m \rangle$ tiene dimensión $m + 1$, luego $E = P(W)$ es una variedad de dimensión m que pasa por todos los puntos dados. Si $E' = P(W')$ es otra, necesariamente $\langle v_0, \dots, v_m \rangle \subset W'$, y como ambos subespacios tienen dimensión $m + 1$, tienen que ser iguales, luego $E' = E$. ■

En particular, por dos puntos proyectivos pasa una única recta proyectiva.

Notemos que toda variedad proyectiva E de dimensión m contiene $m + 1$ puntos proyectivamente independientes, pues su espacio soporte será de la forma $W = \langle v_0, \dots, v_m \rangle$ con los vectores v_i linealmente independientes, y basta tomar los puntos $A_i = \langle v_i \rangle$.

Sistemas de referencia En un espacio afín de dimensión n , un conjunto de $n + 1$ puntos afinmente independientes determina un sistema de referencia. En cambio, en los espacios proyectivos hace falta un punto más:

Definición 8.18 Si $P(V)$ es un espacio proyectivo de dimensión n , diremos que unos puntos A_0, \dots, A_{n+1} son un *sistema de referencia proyectivo* si cualquier subconjunto de $n + 1$ puntos es proyectivamente independiente.

Vamos a caracterizar este concepto en términos de la estructura vectorial subyacente:

Teorema 8.19 *Unos puntos A_0, \dots, A_{n+1} en un espacio proyectivo $P(V)$ de dimensión n son un sistema de referencia proyectivo si y sólo si $A_i = \langle v_i \rangle$ para ciertos vectores $v_i \in V$ de modo que v_0, \dots, v_n son linealmente independientes y $v_{n+1} = v_0 + \dots + v_n$. Además, la $n + 2$ -tupla (v_0, \dots, v_{n+1}) está unívocamente determinada salvo múltiplos por un mismo escalar no nulo.*

DEMOSTRACIÓN: Sea $A_i = \langle v_i \rangle$, con $v_i \in V$ no nulo. Que A_0, \dots, A_n sean proyectivamente independientes equivale a que v_0, \dots, v_n sean una base de V . En particular podemos expresar

$$v_{n+1} = \alpha_0 v_0 + \dots + \alpha_n v_n,$$

para unos únicos $\alpha_i \in K$. Que algún α_i sea 0 equivale a que los $n + 1$ vectores que resultan de eliminar el v_i correspondiente sean linealmente dependientes, luego que los puntos dados sean un sistema de referencia equivale a que todos los α_i sean no nulos. Entonces $A_i = \langle \alpha_i v_i \rangle$, luego cambiando v_i por $\alpha_i v_i$ tenemos que $v_{n+1} = v_0 + \dots + v_n$. El recíproco es obvio, si $A_i = \langle v_i \rangle$ de modo que v_0, \dots, v_n son linealmente independientes y $v_{n+1} = v_0 + \dots + v_n$ es claro que cualesquiera $n + 1$ de los $n + 2$ vectores son linealmente independientes, luego los puntos son un sistema de referencia proyectivo.

Observemos que se da la unicidad del enunciado: si $v_{n+1} = v_0 + \dots + v_n$, la única posibilidad para que otros vectores que generen los mismos puntos, es decir, que sean de la forma $\lambda_i v_i$, cumplan $\lambda_{n+1} v_{n+1} = \lambda_0 v_0 + \dots + \lambda_n v_n$ es que todos los λ_i sean iguales, puesto que se cumple $\lambda_{n+1} v_{n+1} = \lambda_{n+1} v_0 + \dots + \lambda_{n+1} v_n$ y las coordenadas de $\lambda_{n+1} v_{n+1}$ en la base v_0, \dots, v_n son únicas. ■

Ahora es claro que a cualquier conjunto de $n + 1$ puntos proyectivamente independientes se le puede añadir (no de forma única) un punto más que determine un sistema de referencia (basta tomar un vector que genere cada punto y tomar el punto determinado por la suma de todos ellos).

El teorema siguiente ilustra la necesidad del punto adicional:

Teorema 8.20 *Toda biyección entre dos sistemas de referencia de dos espacios proyectivos analíticos se extiende a una única homografía entre ellos.*

DEMOSTRACIÓN: Sean

$$\langle v_0 \rangle, \dots, \langle v_n \rangle, \langle v_0 + \dots + v_n \rangle \quad \text{y} \quad \langle v'_0 \rangle, \dots, \langle v'_n \rangle, \langle v'_0 + \dots + v'_n \rangle$$

dos sistemas de referencia en dos espacios proyectivos $P(V)$ y $P(V')$. Consideremos el isomorfismo $u : V \rightarrow V'$ determinado por $u(v_i) = v'_i$. Es claro que la homografía que induce transforma un sistema en otro.

Recíprocamente, si $H(u')$ es una homografía que transforma un sistema en otro, entonces $u'(v_i) = \alpha_i v'_i$, e igualando las imágenes del último generador,

$$\alpha_0 v'_1 + \cdots + \alpha_n v'_n = \alpha v'_0 + \cdots + \alpha v'_n,$$

luego $\alpha_i = \alpha$ para todo i , es decir, $u' = \alpha u$, luego $H(u') = H(u)$. ■

A la hora de definir una biyección afín de un plano afín en sí mismo, tenemos libertad para fijar las imágenes de tres puntos, con tal de que sean afinmente independientes, al igual que las imágenes asignadas, mientras que, según acabamos de ver en el caso de un plano proyectivo tenemos libertad para fijar las imágenes de cuatro puntos, con tal de que formen sean un sistema de referencia proyectivo y las imágenes asignadas también.

Ejercicio: Encontrar una homografía de un plano proyectivo en sí mismo que deje fijos a tres puntos no colineales sin ser la identidad.

Coordenadas homogéneas Finalmente estamos en condiciones de describir los puntos de un espacio proyectivo en términos de coordenadas:

Definición 8.21 Sea $P(V)$ un espacio proyectivo de dimensión n . La *asignación de coordenadas* asociada a un sistema de referencia dado (con sus puntos en un orden dado) es la única homografía $h : P(V) \rightarrow P^n(K)$ que hace corresponder el sistema dado con el *sistema de referencia canónico* de $P^n(K)$, formado por los puntos

$$\langle (1, 0, \dots, 0) \rangle, \dots, \langle (0, \dots, 0, 1) \rangle, \langle (1, \dots, 1) \rangle.$$

Si $A \in P(V)$, cualquiera de las $n+1$ -tuplas que forman $h(A)$ recibe el nombre de vector de *coordenadas homogéneas* de A respecto del sistema de referencia prefijado. De este modo, cada punto de $P(V)$ está determinado por diversos vectores de coordenadas homogéneas, de modo que no pueden ser todas nulas y, dado uno de ellos, los demás son simplemente sus múltiplos.

Más explícitamente, si el sistema de referencia es

$$\langle v_0 \rangle, \dots, \langle v_n \rangle, \langle v_0 + \cdots + v_n \rangle,$$

entonces unas coordenadas homogéneas de un punto $A = \langle v \rangle$ son las coordenadas de v en la base v_0, \dots, v_n . Las demás coordenadas homogéneas se obtienen cambiando los vectores v_i , y sabemos que sólo pueden cambiarse (sin alterar el sistema de referencia prefijado) por múltiplos por un mismo escalar no nulo.

Ejemplo Así, la mejor forma de concebir un plano proyectivo $P(V)$ no es pensar en subespacios vectoriales de \mathbb{R}^2 , sino pensar que un punto es, por ejemplo, $(2, 5, 4)$, entendiendo que nos referimos al punto $\langle 2v_0 + 5v_1 + 4v_2 \rangle$, donde

$$A_0 = \langle v_0 \rangle, \quad A_1 = \langle v_1 \rangle, \quad A_2 = \langle v_2 \rangle, \quad A_3 = \langle v_1 + v_2 + v_3 \rangle$$

es un sistema de referencia prefijado en $P(V)$.

Pero es importante tener presente que $(1, 5/2, 2)$ es otro vector de coordenadas para el mismo punto, pues las coordenadas homogéneas no son únicas, sino que están determinadas salvo multiplicación por un mismo escalar no nulo. ■

Como en el caso afín, podemos expresar en términos de coordenadas los distintos conceptos geométricos. Por ejemplo, los hiperplanos de $P(V)$ son las variedades de la forma $P(W)$, donde W es un hiperplano de V , y los hiperplanos de V son los conjuntos cuyas coordenadas respecto de una base v_0, \dots, v_n satisfacen una ecuación lineal

$$a_0x_0 + \dots + a_nx_n = 0,$$

con algún a_i no nulo, luego los hiperplanos de $P(V)$ son los puntos cuyas coordenadas homogéneas respecto de un sistema de referencia dado satisfacen una ecuación como la anterior.

Las variedades proyectivas de dimensiones inferiores están determinadas por sistemas de ecuaciones lineales (con término independiente nulo).

Similarmente, si tenemos una homografía $H(f) : P(V) \rightarrow P(V)$, fijado un sistema de referencia proyectivo $A_i = \langle v_i \rangle$, los vectores v_0, \dots, v_n forman una base de V , respecto a la cual el automorfismo f tendrá asociada una matriz $M \in \text{LG}(n+1, K)$, de modo que si $x = (x_0, \dots, x_n)$ son las coordenadas de un vector $v \in V$, entonces $x' = xM$ son las coordenadas de $f(v)$, pero x y x' son también las coordenadas homogéneas de $A = \langle v \rangle$ y $H(f)(A) = \langle f(v) \rangle$. En definitiva:

Teorema 8.22 *Si $F : P(V) \rightarrow P(V)$ es una homografía, fijado un sistema de referencia proyectivo, existe una matriz $M \in \text{LG}(n+1, K)$ tal que la expresión coordenada de F es $X' = XM$, donde X son unas coordenadas homogéneas de un punto $A \in P(V)$ y X' son unas coordenadas homogéneas de $F(A)$.*

Es claro que toda matriz regular determina una homografía, aunque matrices distintas pueden determinar la misma homografía.

Relación con la geometría afín Ahora podemos entender la motivación última de la definición de los espacios proyectivos analíticos.

Observemos en primer lugar que si H es un hiperplano cualquiera en $P(V)$, podemos tomar en él n puntos proyectivamente independientes A_1, \dots, A_n y completarlos hasta un sistema de referencia proyectivo con dos puntos más A_0 y A_{n+1} . Respecto a dicho sistema de referencia, sus coordenadas son

$$(1, 0, \dots, 0), \quad (0, \dots, 0, 1), \quad (1, \dots, 1).$$

Por lo tanto, A_1, \dots, A_n están tanto en H como en el hiperplano de ecuación $x_0 = 0$, luego H es el hiperplano de ecuación $x_0 = 0$ en dicho sistema de referencia. Así pues, cualquier hiperplano admite ecuación $x_0 = 0$ respecto de un sistema de referencia adecuado (y obviamente lo mismo vale con $x_i = 0$ para cualquier i).

Fijado un sistema de referencia en estas condiciones, podemos definir una aplicación $\Phi : P(V) \setminus H \rightarrow K^n$ de modo que el punto A de coordenadas homogéneas (x_0, \dots, x_n) tiene por imagen $\Phi(A) = (x_1/x_0, \dots, x_n/x_0)$.

Claramente $\Phi(A)$ no depende de la elección de las coordenadas homogéneas del punto, y Φ es biyectiva.

Esto nos permite dotar a $P(V) \setminus H$ de estructura de espacio afín. El teorema 8.31 prueba que esta estructura no depende de la elección del sistema de referencia proyectivo. De momento vamos a ver un ejemplo concreto para formarnos una imagen intuitiva de la situación.

Ejemplo Consideremos un plano proyectivo P sobre un cuerpo K de característica 0 en el que hemos fijado un sistema de referencia proyectivo. Así, cada punto está determinado por sus coordenadas homogéneas (x, y, z) , que son únicas salvo múltiplos escalares.

Llamaremos *recta infinita* r_∞ a la recta (hiperplano) de ecuación $z = 0$ y llamamos $E = P \setminus r_\infty$, de modo que los puntos de E son los que admiten coordenadas homogéneas con $z \neq 0$, lo cual equivale a que admitan un (único) vector de coordenadas homogéneas de la forma $(x, y, 1)$. Las coordenadas afines de dicho punto son (x, y) .

Una recta r de P distinta de r_∞ está formada por los puntos cuyas coordenadas homogéneas cumplen una ecuación

$$ax + by + cz = 0,$$

con $ab \neq 0$ (pues de lo contrario la ecuación equivaldría a $z = 0$). Los puntos de $r \cap E$ son los que admiten unas coordenadas homogéneas con $z = 1$ o, equivalentemente, aquellos cuyas coordenadas afines cumplen la ecuación

$$ax + by + c = 0.$$

Pero ésta es la ecuación de una recta en un plano afín. Así pues, hemos probado que si r es una recta proyectiva distinta de la recta “infinita” que hemos quitado, entonces $r \cap E$ es una recta afín. ¿Y cuántos puntos hemos perdido al restringirnos a E ? Hemos perdido los puntos que cumplen la ecuación proyectiva con $z = 0$, pero hay un único punto así, el punto $(-b, a, 0)$. El sistema de ecuaciones tiene infinitas soluciones, pero todas son múltiplos unas de otras, por lo que todas corresponden al mismo punto de la recta r_∞ .

Recíprocamente, es fácil ver que toda recta afín se extiende a una recta proyectiva (sin más que añadirle una z al término independiente de su ecuación) que tiene los puntos de la recta original más un punto en la recta infinita.

En definitiva, las rectas de E se corresponden biunívocamente con las rectas de P distintas de r_∞ , y la correspondencia consiste en añadir o quitar a cada recta un único punto infinito.

Por ejemplo, las rectas afines de ecuaciones $y = 2x + 5$ e $y = 2x + 7$ son paralelas, pero sus compleciones proyectivas son las rectas de ecuaciones

$$2x - y + 5z = 0, \quad 2x - y + 7z = 0,$$

y ambas tienen un mismo punto infinito en común, a saber, el punto $(1, 2, 0)$. Es fácil ver que esto es general: dos rectas afines son paralelas si y sólo si se cortan en su punto infinito. ■

Teorema 8.23 *Si $P = P(V)$ es un espacio proyectivo de dimensión n y H es un hiperplano, entonces las variedades afines m -dimensionales de $E = P \setminus H$ son las de la forma $U \cap E$, donde U es una variedad proyectiva m -dimensional.*

DEMOSTRACIÓN: Sea $H = P(W_\infty)$ y sea $U = P(W)$. Entonces la intersección $H \cap U = P(W_\infty \cap W)$ tiene dimensión $m - 1$, luego $W_\infty \cap W$ tiene dimensión m . Pongamos que $W_\infty \cap W = \langle v_1, \dots, v_m \rangle$ y sea $v_0 \in W \setminus W_\infty$ tal que v_0, \dots, v_m sea una base de W . Los puntos de U son, pues, los de la forma $A = \langle \lambda_0 v_0 + \dots + \lambda_m v_m \rangle$, con $\lambda_i \in K$.

En principio, la estructura afín de E depende de una asignación de coordenadas respecto de un sistema de referencia prefijado en P . Sean $x_i \in K^{n+1}$ los vectores de coordenadas de los v_i respecto de la base de V asociada al sistema de referencia elegido. Entonces las coordenadas de $\lambda_0 v_0 + \dots + \lambda_m v_m$ (es decir, unas coordenadas homogéneas de A) son

$$x = \lambda_0 x_0 + \dots + \lambda_m x_m.$$

Observemos que, como $v_i \in W_\infty$, para $i > 0$, las coordenadas x_i tienen su primera componente nula, mientras que la primera componente de x_0 podemos tomarla igual a 1. Por lo tanto, la primera componente de x es 0 si y sólo si $\lambda_0 = 0$. Por consiguiente, los puntos de $U \cap E$ son los que tienen un vector de coordenadas homogéneas x con $\lambda_0 \neq 0$. Podemos, pues, dividir entre λ_0 y suponer que $\lambda_0 = 1$. En definitiva, las coordenadas afines de los puntos de $U \cap E$ son de la forma

$$\bar{x} = \bar{x}_0 + \lambda_1 \bar{x}_1 + \dots + \lambda_m \bar{x}_m,$$

donde \bar{x}_i es el vector que resulta de eliminar la primera componente de x_i . Concluimos que las coordenadas de los puntos de $U \cap E$ son

$$\bar{x}_0 + \langle \bar{x}_1, \dots, \bar{x}_m \rangle.$$

Este conjunto es una variedad afín de dimensión m en K^n , luego $U \cap E$ es una variedad afín de dimensión m en E . El razonamiento puede seguirse en sentido inverso para concluir que toda variedad afín de E procede de una variedad proyectiva U de la misma dimensión. ■

Observaciones El teorema anterior muestra que las variedades afines de E no dependen del sistema de referencia de P usado para definir la estructura afín. Más aún, ahora es inmediato que $m + 1$ puntos de E son afinmente independientes en E si y sólo si son proyectivamente independientes en P , pues que sean afinmente independientes equivale a que no estén contenidos en ninguna variedad afín de dimensión menor que m , lo cual equivale claramente a que no estén contenidos en ninguna variedad proyectiva de dimensión menor que m , que a su vez equivale a que sean proyectivamente independientes.

A su vez esto implica que cada variedad afín m -dimensional U de E se extiende a una única variedad proyectiva m -dimensional \bar{U} de P , pues dos de ellas contendrían un mismo conjunto de $m+1$ puntos afinmente independientes, luego proyectivamente independientes, luego ambas variedades serían la misma.

La correspondencia $U \mapsto U \cap E$ entre variedades proyectivas y afines conserva claramente la inclusión, luego lo mismo vale para su inversa $U \mapsto \bar{U}$.

Observemos también en qué se traduce el paralelismo de rectas en E : dos rectas distintas r y s de E son paralelas si y sólo si están contenidas en un plano Π y son disjuntas, pero entonces $\bar{r}, \bar{s} \subset \bar{\Pi}$ y el teorema 8.16 implica que \bar{r} y \bar{s} tienen un punto en común, que necesariamente tiene que estar en H , porque $r \cap s = \emptyset$. Así pues, dos rectas de E son paralelas si y sólo si tienen el mismo punto infinito.

Ahora es claro que todo espacio afín E (en el sentido algebraico) se puede identificar con un subconjunto $P \setminus H$, donde P es un espacio proyectivo de la misma dimensión y H un hiperplano, pues a través de un sistema de coordenadas E se puede identificar con K^n y éste a su vez con $P^n(K) \setminus H$, donde H es el hiperplano de ecuación $x_0 = 0$.

Sin embargo, es importante advertir que el espacio que resulta de añadir a un espacio afín E un hiperplano H de puntos infinitos es homogéneo, en el sentido de que nada en él distingue a los puntos finitos de los infinitos. Es fácil ver que en un espacio proyectivo $P(V)$ existen homografías que transforman cualquier hiperplano en cualquier otro, luego no hay ninguno que se distinga de los demás, y sólo podemos distinguir entre “puntos finitos” y “puntos infinitos” seleccionando arbitrariamente un hiperplano cualquiera como hiperplano infinito. Un mismo punto será finito o infinito según qué hiperplano hayamos elegido como hiperplano infinito. Dicho de otro modo, la distinción entre “finito” e “infinito” es relativa a la elección de un sistema de referencia. ■

Teorema 8.24 *Sea $P = P(V)$ un espacio proyectivo, sea H un hiperplano y $E = P \setminus H$. Sea $\text{LGP}_H(P)$ el subgrupo de $\text{LGP}(P)$ formado por las homografías F tales que $F[H] = H$. Entonces, la restricción define un isomorfismo de grupos $\text{LGP}_H(P) \rightarrow \text{GA}(E)$.*

DEMOSTRACIÓN: Fijemos un sistema de referencia en P , que a su vez determina un sistema de referencia en E . Según vimos en el capítulo III, la expresión en coordenadas de una biyección afín $f \in \text{GA}(E)$ es $Y = A + XM$, donde $A \in K^n$, $M \in \text{LG}(n, K)$, X es el vector de coordenadas de un punto cualquiera $A \in E$ e Y es el vector de coordenadas de $f(A)$. En términos de coordenadas homogéneas esto equivale a que

$$(1, Y) = (1, X) \left(\begin{array}{c|c} 1 & A \\ \hline 0 & M \end{array} \right).$$

Por lo tanto, la homografía F cuya matriz en el sistema de referencia fijado es la que aparece en la fórmula anterior (que claramente es inversible, pues su

determinante es el mismo que el de M) cumple que su restricción a E es f , lo cual en particular requiere que $f[H] = H$. Así pues, toda biyección afín se extiende a una homografía.

Recíprocamente, si F es una homografía que fija a H y M' es una matriz de F respecto al sistema de referencia prefijado, la condición de que $F[H] = H$ equivale a que $(0, X)M'$ tiene que tener la primera componente nula, sea cual sea $X \in K^n$, lo cual sólo puede ocurrir si la primera columna de M' es de la forma $(a, 0, \dots, 0)$, con $a \neq 0$. Si dividimos toda la matriz M' entre a obtenemos una nueva matriz para la misma homografía y ahora $a = 1$, luego M' es de la forma

$$\left(\begin{array}{c|c} 1 & A \\ \hline 0 & M \end{array} \right),$$

para cierta matriz $M \in \text{LG}(n, K)$ y cierto $A \in K^n$, de donde se sigue que la restricción de F a E admite la expresión coordenada $Y = A + XM$, luego se trata de una biyección afín.

Con esto tenemos que la restricción induce una aplicación suprayectiva (que claramente es un homomorfismo de grupos) $\text{LGP}_H(P) \rightarrow \text{GA}(E)$. Falta probar que es inyectiva, pero si $F|_E$ es la identidad y $A \in H$, tomamos cualquier punto $B \in E$, de modo que la recta AB tiene todos sus puntos en E salvo A (y una recta en un espacio afín tiene al menos dos puntos). Por lo tanto F deja fijos al menos dos puntos de AB , luego deja fija a AB , luego deja fijo a A (porque $\{A\} = AB \cap H$ y F fija a ambas variedades). Esto prueba que F es también la identidad en H , luego $F = 1$. ■

Como consecuencia, a su vez:

Teorema 8.25 *Si $P = \mathbb{P}(V)$ es un espacio proyectivo, H es un hiperplano y $E = P \setminus H$, entonces el grupo $\text{HT}(E)$ de las homotecias y traslaciones de E está formado por las restricciones a E de las homografías que fijan a todos los puntos de H .*

DEMOSTRACIÓN: Según el teorema 5.8, si $\dim E \geq 2$ el grupo $\text{HT}(E)$ está formado por los $F \in \text{LGP}_H(P)$ que transforman cada recta r de E en una paralela. Así, si $F \in \text{HT}(E)$ y $A \in H$, tomamos $B \in E$, de modo que $F[AB \cap E]$ es paralela a $AB \cap E$, lo cual equivale a que $F[AB] = F(A)F(B)$ tenga el mismo punto infinito que AB , es decir, a que $F(A) = A$. El recíproco se prueba análogamente.

Falta considerar el caso en que $\dim E = 1$, pero en este caso todas las biyecciones afines son homotecias o traslaciones y la conclusión es trivial (notemos que en este caso H consta de un único punto, por lo que fijar a H es lo mismo que fijar a todos sus puntos). En efecto, si $f \in \text{GA}(E)$ tiene un punto fijo O , entonces

$$f(P) = O + \vec{f}(\overrightarrow{OP}) = O + k\overrightarrow{OP},$$

donde (k) es la matriz de \vec{f} en cualquier sistema de referencia (notemos que necesariamente $k \neq 0$), luego f es una homotecia. Si f no tiene puntos fijos, entonces su expresión en coordenadas respecto de un sistema de referencia (O, \vec{v})

será $y = ax + b$, para ciertos $a, b \in K$, con $a \neq 0$, y es fácil ver que la condición de que no tenga puntos fijos equivale a que $a = 1$, luego la expresión en coordenadas es $y = x + b$, luego $f(P) = P + b\vec{v}$ y f es una traslación. ■

Más aún, cualquiera que sea la dimensión de E , tenemos que el grupo $T(E)$ de las traslaciones de E está formado por la identidad y los elementos de $HT(E)$ que no tienen puntos fijos en E , es decir, por los elementos de $LGP(P)$ que fijan a todos los puntos de H y no tienen puntos fijos en E .

Veamos a continuación la versión proyectiva del teorema fundamental de la geometría afín:

Definición 8.26 Si P y P' son dos espacios proyectivos de la misma dimensión $n \geq 2$, una *colineación* entre ellos es una aplicación $f : P \rightarrow P'$ que transforma puntos colineales en puntos colineales.

Si $f : V_1 \rightarrow V_2$ es una biyección semilineal entre dos espacios vectoriales, la *semihomografía* inducida por f es la aplicación $H(f) : P(V_1) \rightarrow P(V_2)$ definida mediante $H(f)(\langle v \rangle) = \langle f(v) \rangle$.

Teorema 8.27 (Teorema fundamental de la geometría proyectiva) *Las colineaciones entre dos espacios proyectivos de la misma dimensión $n \geq 2$ coinciden con las semihomografías entre ambos espacios.*

DEMOSTRACIÓN: Es inmediato que las semihomografías son colineaciones. Recíprocamente, supongamos que $f : P \rightarrow P'$ es una colineación. Dividimos la prueba en varios pasos:

1. Si $X = \{A_0, \dots, A_m\}$ de un conjunto de $m + 1$ puntos proyectivamente independientes en P , entonces $f[\langle X \rangle] \subset \langle f[X] \rangle$.

Por inducción sobre m . Para $m = 1$ se trata de la hipótesis del teorema. Supongámoslo cierto para conjuntos de m puntos. Si $A \in \langle X \rangle$, entonces la recta AA_m corta a $\langle A_0, \dots, A_{m-1} \rangle$, que es un hiperplano del espacio proyectivo $\langle X \rangle$. Sea B el punto de corte. Por hipótesis de inducción

$$f(B) \in f[\langle A_0, \dots, A_{m-1} \rangle] \subset \langle f(A_0), \dots, f(A_{m-1}) \rangle \subset \langle f[X] \rangle.$$

A su vez, $f(A)$ está en la recta $f(B)f(A_m)$, que está en $\langle f[X] \rangle$.

2. La imagen de un conjunto proyectivamente independiente es un conjunto proyectivamente independiente.

En efecto, el conjunto X puede completarse hasta un conjunto X' de $n + 1$ puntos proyectivamente independientes. Por el apartado anterior tenemos que $f[P] \subset \langle f[X'] \rangle$, luego, teniendo en cuenta las dimensiones, tenemos que $\langle f[X'] \rangle = P'$, luego $f[X']$ es proyectivamente independiente y $f[X]$ también. De aquí se sigue obviamente:

3. La inclusión en 1) es de hecho una igualdad. En particular la imagen de un hiperplano es un hiperplano.

Si fijamos un hiperplano H en P y $H' = f[H]$, entonces f induce una biyección $P \setminus H \rightarrow P' \setminus H'$ en las hipótesis del teorema fundamental de la geometría afín, luego es una biyección semiafín. Sea σ el automorfismo de K asociado. El argumento del teorema 8.24 se generaliza para probar que $f|_{P \setminus H}$ se puede extender a una semihomografía $\bar{f}: P \rightarrow P'$. En efecto, fijamos sistemas de referencia en P y P' , que a su vez determinan sistemas de referencia de $P \setminus H$ y $P' \setminus H'$. Observemos ahora cómo es la expresión en coordenadas de una aplicación semiafín: si un punto Q tiene coordenadas afines $X = (x_1, \dots, x_n)$, esto significa que

$$Q = O + x_1 \vec{v}_1 + \dots + x_n \vec{v}_n,$$

donde O es el origen del sistema de referencia y $\vec{v}_1, \dots, \vec{v}_n$ su base asociada. Entonces

$$f(Q) = f(O) + \sigma(x_1) \vec{f}(\vec{v}_1) + \dots + \sigma(x_n) \vec{f}(\vec{v}_n),$$

y si $A \in K^n$ es la n -tupla de coordenadas de $f(O)$ en el sistema de referencia de $P' \setminus H'$ y M es la matriz cuyas filas son las coordenadas de los vectores $\vec{f}(\vec{v}_i)$, entonces las coordenadas de $f(Q)$ son $Y = A + \sigma(X)M$, donde llamamos $\sigma(X) = (\sigma(x_1), \dots, \sigma(x_n))$. En términos de coordenadas homogéneas es

$$(1, Y) = \sigma(1, X) \left(\begin{array}{c|c} 1 & A \\ \hline 0 & M \end{array} \right).$$

Si llamamos M' a la última matriz, esta expresión vale igualmente para cualquier $n+1$ -tupla de coordenadas homogéneas (x_0, \dots, x_n) de Q , pues las coordenadas afines son $(1, x_1/x_0, \dots, x_n/x_0)$, y la fórmula anterior nos da que unas coordenadas homogéneas de $f(Q)$ son

$$(1, \sigma(x_1)/\sigma(x_0), \dots, \sigma(x_n)/\sigma(x_0))M',$$

luego, multiplicando por $\sigma(x_0)$, otras coordenadas homogéneas son $Y = \sigma(X)M'$.

Pongamos que $P = P(V)$, $P' = P(V')$ y sean e_0, \dots, e_n y e'_0, \dots, e'_n las bases de V y V' tales que las coordenadas de un vector no nulo \vec{v} en V (resp. V') son las coordenadas homogéneas del punto $\langle v \rangle$ de P (resp. P'). Consideramos la aplicación $V \rightarrow V'$ que transforma el vector de coordenadas X en el vector de coordenadas $\sigma(X)M'$. Es inmediato comprobar que es una biyección semilineal, de automorfismo σ , luego induce una semihomografía $\bar{f}: P \rightarrow P'$ que coincide con f en $P \setminus H$, pero entonces coincide con f en todo P , pues, fijado un punto $O \in P \setminus H$, para todo $A \in H$, tomamos un tercer punto B en OA y así, tanto $f(A)$ como $\bar{f}(A)$ están en H' y en la imagen de la recta OB , que es una recta, luego ambos son el único punto de intersección entre la recta y el hiperplano. ■

En particular, si el cuerpo K no tiene automorfismos no triviales, como sucede si K es un cuerpo euclídeo arquimediano, según el teorema 5.18, tenemos que las homografías entre espacios proyectivos coinciden con las colineaciones.

En general, si V es un espacio vectorial de dimensión $n \geq 2$ sobre un cuerpo K , podemos definir el grupo $\text{L}\Gamma(V)$ de todas las biyecciones semilineales

de V en sí mismo. Es fácil comprobar que $\text{LG}(V) \leq \text{L}\Gamma(V)$. Fijada una base $v_1, \dots, v_n \in V$, para cada $\sigma \in \text{Aut}(K)$, podemos definir $f_\sigma : V \rightarrow V$ mediante

$$f_\sigma(\alpha_1 v_1 + \dots + \alpha_n v_n) = \sigma(\alpha_1) v_1 + \dots + \sigma(\alpha_n) v_n.$$

Es inmediato que $f_\sigma \in \text{L}\Gamma(V)$, así como que la aplicación $\text{Aut}(K) \rightarrow \text{L}\Gamma(V)$ dada por $\sigma \mapsto f_\sigma$ es un monomorfismo de grupos. Más aún, si $g \in \text{L}\Gamma(V)$ es semilineal con automorfismo σ , es inmediato comprobar que $h = g \circ f_\sigma^{-1}$ es lineal, por lo que $g = h f_\sigma \in \text{LG}(V)H$, donde H es la imagen de $\text{Aut}(K)$ en $\text{L}\Gamma(V)$. Como la única biyección lineal de H es la identidad, de hecho $\text{LG}(V) \cap H = 1$, por lo que $\text{L}\Gamma(V)$ es el producto semidirecto de $\text{LG}(V)$ por H o, equivalentemente:

$$\text{L}\Gamma(V) \cong [\text{LG}(V)]\text{Aut}(K),$$

Si $P = P(V)$ es un espacio proyectivo, el grupo $\text{LFP}(V)$ de las semihomografías de P es el cociente de $\text{L}\Gamma(V)$ sobre el subgrupo de las homografías lineales, que inducen la identidad en P , de modo que

$$\text{LFP}(V) = [\text{LGP}(V)]\text{Aut}(K).$$

La razón doble Vamos a estudiar con algo más de detalle las rectas proyectivas. Si X es una recta proyectiva sobre un anillo de división K , un sistema de referencia en X consiste simplemente en tres puntos distintos P, Q, R . Estos puntos determinan una única asignación de coordenadas $\Phi : X \rightarrow \mathbb{P}^1(K)$ tal que las coordenadas homogéneas de P, Q, R son, respectivamente $(1, 0), (0, 1), (1, 1)$.

Vamos a adoptar el convenio de tomar como punto infinito de $\mathbb{P}^1(K)$ el punto $\infty = (1, 0)$, y entonces podemos expresar $\mathbb{P}^1(K) = K \cup \{\infty\}$, donde cada $\alpha \in K$ se identifica con el punto de coordenadas homogéneas $(\alpha, 1)$. Recíprocamente, el punto de coordenadas (α, β) es ∞ si $\beta = 0$ o bien $\alpha/\beta \in K$ en caso contrario.

Si un punto de X tiene coordenadas homogéneas (α, β) , diremos que α/β es su coordenada afín respecto del sistema de referencia considerado P, Q, R , entendiendo que la coordenada afín de P es ∞ (mientras que las de Q y R son 0 y 1 , respectivamente). Equivalentemente, si llamamos $E = X \setminus \{P\}$, entonces la coordenada afín de un punto es su coordenada respecto del sistema de referencia afín (Q, \overrightarrow{QR}) .

Definición 8.28 Sean P, Q, R, S cuatro puntos colineales distintos en un espacio proyectivo X sobre un anillo de división K y sea r la recta que los contiene. Definimos la *razón doble* entre ellos como

$$\mathfrak{R}(P, Q, R, S) = \frac{\alpha}{\beta} \in K \setminus \{0, 1\}$$

donde (α, β) son las coordenadas homogéneas de S respecto al sistema de referencia P, Q, R de r .

Observemos que la razón doble no puede tomar los valores 0, 1 (ni ∞) porque los puntos con estas coordenadas son Q, R, P y suponemos que S es distinto de éstos.

La razón doble de cuatro puntos no depende de ninguna elección de ningún sistema de referencia, es decir, depende de un sistema de referencia de la recta X determinado por los propios puntos, de modo que no tiene sentido plantearse si una razón doble puede tomar un valor respecto de un sistema de referencia y otro distinto respecto de otro.

Más aún, la razón doble es invariante por homografías, pues si $H : X \rightarrow X'$ es una homografía entre dos espacios proyectivos y P, Q, R, S son cuatro puntos distintos contenidos en una recta r , entonces $H(P), H(Q), H(R), H(S)$ son cuatro puntos distintos en la recta $H[r]$, y si S tiene coordenadas (α, β) respecto de P, Q, R , entonces $H(S)$ tiene coordenadas (α, β) respecto de $H(P), H(Q), H(R)$, porque si Φ es la asignación de coordenadas en r , entonces la homografía $H^{-1} \circ \Phi : H[r] \rightarrow \mathbb{P}^1(K)$ transforma $H(P), H(Q), H(R)$ en $\infty, 0, 1$, respectivamente, luego es la asignación de coordenadas correspondiente al sistema de referencia $H(P), H(Q), H(R)$. Concluimos que

$$\mathcal{R}(P, Q, R, S) = \frac{\alpha}{\beta} = \mathcal{R}(H(P), H(Q), H(R), H(S)).$$

El teorema siguiente explica el nombre de “razón doble”:

Teorema 8.29 Sean P_1, P_2, P_3, P_4 cuatro puntos distintos de una recta proyectiva cuyas coordenadas cartesianas en un sistema de referencia dado sean x_1, x_2, x_3, x_4 respectivamente. Entonces¹

$$\mathcal{R}(P_1, P_2, P_3, P_4) = \frac{x_3 - x_1}{x_4 - x_1} : \frac{x_3 - x_2}{x_4 - x_2}.$$

DEMOSTRACIÓN: Al menos tres de las coordenadas tienen que ser finitas, vamos a suponer que lo son x_1, x_2 y x_3 . Los otros tres casos posibles se tratan análogamente. Consideramos la aplicación $H : \mathbb{P}^1(K) \rightarrow \mathbb{P}^1(K)$ dada por

$$H(x) = \frac{x_3 - x_1}{x - x_1} : \frac{x_3 - x_2}{x - x_2}.$$

En términos de coordenadas homogéneas es

$$H(x, 1) = \left(\frac{x_3 - x_1}{x - x_1} : \frac{x_3 - x_2}{x - x_2}, 1 \right) = ((x_3 - x_1)(x - x_2), (x_3 - x_2)(x - x_1))$$

o también

$$\begin{aligned} H(x, y) &= ((x_3 - x_1)\left(\frac{x}{y} - x_2\right), (x_3 - x_2)\left(\frac{x}{y} - x_1\right)) \\ &= ((x_3 - x_1)(x - x_2y), (x_3 - x_2)(x - x_1y)). \end{aligned}$$

¹Si expresamos la fórmula como

$$\mathcal{R}(P_1, P_2, P_3, P_4) = \frac{(x_3 - x_1)(x_4 - x_2)}{(x_4 - x_1)(x_3 - x_2)},$$

vale igualmente si uno de los x_i es ∞ entendiendo que los dos factores en los que aparece se cancelan.

En particular,

$$H(\infty) = H(1, 0) = (x_3 - x_1, x_3 - x_2) = \frac{x_3 - x_1}{x_3 - x_2}.$$

Así vemos que se trata de la homografía de matriz

$$M = \begin{pmatrix} x_3 - x_1 & x_3 - x_2 \\ -(x_3 - x_1)x_2 & -(x_3 - x_2)x_1 \end{pmatrix}.$$

Notemos que es ciertamente una homografía porque el determinante de M es $(x_3 - x_1)(x_3 - x_2)(x_2 - x_1) \neq 0$. Se comprueba inmediatamente que $H(x_1) = \infty$, $H(x_2) = 0$, $H(x_3) = 1$. La composición del sistema de coordenadas dado con la homografía H es una homografía que lleva P_1, P_2, P_3, P_4 a $\infty, 0, 1, H(x_4)$, respectivamente, luego la razón doble es $H(x_4)$, lo que nos da la fórmula indicada. ■

De aquí se sigue inmediatamente que la razón doble presenta unas simetrías que no son evidentes a partir de su definición:

$$\begin{aligned} \mathcal{R}(P_1, P_2, P_3, P_4) &= \mathcal{R}(P_3, P_4, P_1, P_2) \\ &= \mathcal{R}(P_2, P_1, P_3, P_4)^{-1} = \mathcal{R}(P_1, P_2, P_4, P_3)^{-1}. \end{aligned}$$

El teorema anterior nos da una expresión alternativa para la razón doble de cuatro puntos:

Teorema 8.30 *Sea P un espacio proyectivo, H un hiperplano y consideremos $P_1, P_2, P_3, P_4 \in E = P \setminus H$ cuatro puntos colineales distintos. Entonces*

$$\mathcal{R}(P_1, P_2, P_3, P_4) = \frac{\overrightarrow{P_1P_3}}{\overrightarrow{P_1P_4}} : \frac{\overrightarrow{P_2P_3}}{\overrightarrow{P_2P_4}},$$

donde los vectores se calculan respecto de la estructura afín de E . Si uno de los cuatro puntos es el punto infinito de la recta que los contiene (es decir, $P_i \in H$), la fórmula sigue siendo cierta si entendemos que los términos que contienen a P_i se cancelan, es decir:

$$\begin{aligned} \mathcal{R}(\infty, P_2, P_3, P_4) &= \frac{\overrightarrow{P_2P_4}}{\overrightarrow{P_2P_3}}, & \mathcal{R}(P_1, \infty, P_3, P_4) &= \frac{\overrightarrow{P_1P_3}}{\overrightarrow{P_1P_4}}, \\ \mathcal{R}(P_1, P_2, \infty, P_4) &= \frac{\overrightarrow{P_2P_4}}{\overrightarrow{P_1P_4}}, & \mathcal{R}(P_1, P_2, P_3, \infty) &= \frac{\overrightarrow{P_1P_3}}{\overrightarrow{P_2P_3}}. \end{aligned}$$

Observemos que (sobre un cuerpo de característica distinta de 2) la relación $\mathcal{R}(\infty, P_2, P_3, P_4) = -1$ equivale a que $\overrightarrow{P_2P_4} = -\overrightarrow{P_2P_3}$ o, equivalentemente, a que P_2 sea el punto medio de P_3 y P_4 .

Veamos finalmente que si P es un espacio proyectivo y H es un hiperplano, la estructura de espacio afín de $E = P \setminus H$ no depende del sistema de coordenadas del que nos valemos para identificar a E con K^n :

Teorema 8.31 Sea P un espacio proyectivo sobre un anillo de división K , sea H un hiperplano de P y $E = P \setminus H$. Entonces, la estructura afín de E está determinada por el espacio vectorial \vec{E} formado por el grupo de las traslaciones en E con el producto determinado por que si $k \in K$ es distinto de $0, 1$ y $T \in \vec{E}$ no es la identidad, entonces kT es el único elemento de \vec{E} tal que, para cualquier punto $A \in E$, se cumple que $(kT)(A) \in AT(A)$ y

$$\mathcal{R}(\infty, A, T(A), (kT)(A)) = k,$$

donde $\infty = AT(A) \cap H$.

DEMOSTRACIÓN: Fijamos un sistema de referencia en P , el cual determina una asignación de coordenadas $\Phi : P \rightarrow \mathbb{P}^n(K)$, que a su vez determina una biyección $\phi|_E : E \rightarrow K^n$ que determina una estructura de espacio afín en E . Tenemos que probar que dicha estructura es la descrita en el enunciado, de modo que no depende realmente del sistema de referencia elegido.

En todo espacio afín, E , el grupo aditivo del espacio vectorial \vec{E} (y su acción sobre E) se identifica de forma natural con el grupo $T(E)$ de las traslaciones, y sabemos que éste está formado por la identidad y las (restricciones de las) homografías de P cuyos puntos fijos son los de H .

Si T es una traslación distinta de la identidad, que identificamos con un vector $\vec{v} \neq \vec{0}$ (de manera que $T(A) = A + \vec{v}$), entonces la traslación kT correspondiente a $k\vec{v}$ es la dada por

$$(kT)(A) = A + k\vec{v} = A + k \overrightarrow{AT(A)}.$$

Esto significa que los puntos $A, T(A), (kT)(A)$ son colineales y, si llamamos r a la recta que los contiene, la coordenada de $(kT)(A)$ en el sistema de referencia $(A, \overrightarrow{AT(A)})$ de r es k , luego, si llamamos ∞ al punto infinito de r , de la propia definición de razón doble se sigue que

$$\mathcal{R}(\infty, A, T(A), (kT)(A)) = k.$$

Esta condición determina el punto $(kT)(A)$, y una traslación está determinada por la imagen de un único punto, luego esta condición determina unívocamente el producto kT . ■

Esto nos lleva a la definición siguiente:

Definición 8.32 Un espacio afín proyectivo sobre un anillo de división K es un par (E, E_∞) , donde E es un espacio proyectivo sobre K y E_∞ es un hiperplano en E al que llamaremos *hiperplano infinito* de E .

Los puntos de E_∞ se llaman *puntos infinitos*, mientras que los puntos de $E_{\text{fin}} = E \setminus E_\infty$ se llaman *puntos finitos*.

Las *variedades afines* de E son el conjunto vacío y las variedades proyectivas de E no contenidas en E_∞ .

Las *biyecciones afines* de E son los elementos del grupo $\text{GA}(E) = \text{LGP}_{E_\infty}(E)$ formado por las homografías de E que fijan a E_∞ .

Los *vectores* de E son los elementos del grupo $\vec{E} = \text{T}(E)$ de las *traslaciones* de E , es decir, las biyecciones afines cuyos puntos fijos son exactamente los puntos infinitos, y que admite una estructura de espacio vectorial sobre K . Si $\vec{v} \in \vec{E}$ y $P \in E_{\text{fin}}$, escribiremos $P + \vec{v}$ en lugar de $\vec{v}(P)$.

Las *homotecias* de E son las biyecciones afines cuyos puntos fijos son los puntos infinitos más un cierto punto finito.

Un *sistema de referencia afín* en E es un sistema de referencia proyectivo $O = P_0, \dots, P_{n+1}$ tal que $P_1, \dots, P_n \in E_\infty$. Así, respecto a tal sistema, los puntos finitos son aquellos cuyas coordenadas homogéneas cumplen $x_0 \neq 0$.

Las *coordenadas afines* de un punto finito P respecto de un sistema de referencia afín son las coordenadas $(x_1/x_0, \dots, x_n/x_0)$, donde (x_0, \dots, x_n) es cualquier vector de coordenadas homogéneas de P respecto del sistema dado.

Los resultados de esta sección prueban que el par $(E_{\text{fin}}, \vec{E}, +)$ es un espacio afín sobre K en el sentido de la definición 3.10 (relajando la condición de que K sea un cuerpo a que sea un anillo de división), y que todos los conceptos que acabamos de “redefinir” son equivalentes a los que ya teníamos definidos, no en el sentido de que sean lo mismo, sino en que ambos se determinan mutuamente. Por ejemplo, cada biyección afín en el sentido de 3.16 se extiende a una única biyección afín en el sentido de 8.32 y, recíprocamente, toda biyección afín en este sentido se restringe a una única biyección afín en el sentido anterior.

Además, dado cualquier espacio afín F , existe una biyección afín $F \rightarrow E_{\text{fin}}$ (en el sentido de 3.16), para cierto espacio afín proyectivo E , por lo que no perdemos generalidad si trabajamos únicamente con espacios afines proyectivos. En otras palabras, esto significa que no perdemos generalidad si suponemos que todo espacio afín lleva “incorporados” puntos infinitos que lo convierten en un espacio proyectivo, y que todas las definiciones de la geometría afín pueden extenderse para tener en cuenta también los puntos infinitos. Por ejemplo, toda variedad afín tiene sus puntos infinitos, toda biyección afín actúa sobre los puntos infinitos, etc.

8.3 Geometría proyectiva analítica y sintética

Vamos a relacionar finalmente los dos conceptos de espacio proyectivo que hemos presentado, el sintético y el analítico. Necesitamos un par de hechos adicionales sencillos sobre los espacios proyectivos algebraicos:

- Unos puntos A_0, \dots, A_{n+1} son proyectivamente independientes si y sólo si lo son A_0, \dots, A_n y $A_{n+1} \notin \langle A_0, \dots, A_n \rangle$.

En efecto, si $A_i = \langle v_i \rangle$, su independencia proyectiva equivale a que los vectores v_i sean linealmente independientes, lo cual equivale a que v_0, \dots, v_n sean linealmente independientes y $v_{n+1} \notin \langle v_0, \dots, v_n \rangle$, lo cual equivale a su vez a que A_0, \dots, A_n sean independientes y $A_{n+1} \notin \langle A_0, \dots, A_n \rangle$.

- Si E es una variedad proyectiva y $A \in P(V) \setminus E$, entonces $\langle E, A \rangle$ es la unión de todas las rectas proyectivas que pasan por A y por un punto de E .

En efecto, claramente $\langle E, A \rangle$ contiene a todas las rectas indicadas. Por otro lado, si $E = P(W)$, $A = \langle v \rangle$, entonces $\langle E, A \rangle = P(W + \langle v \rangle)$ y todo $X \in \langle E, A \rangle$ es de la forma $X = \langle w + \alpha v \rangle \subset \langle w, v \rangle$, luego $X \in P(\langle w, v \rangle)$, que es una recta que pasa por A y por $\langle w \rangle \in P(W) = E$.

Con esto ya estamos en condiciones de probar el resultado básico:

Teorema 8.33 *Si K es un anillo de división y V es un K -espacio vectorial por la derecha de dimensión $n + 1 \geq 3$, entonces $P(V)$ (tomando como rectas los conjuntos de la forma $P(W)$, donde W es un subespacio de V de dimensión 2) cumple los axiomas de la geometría proyectiva. Las variedades proyectivas en el sentido sintético coinciden con las analíticas.*

DEMOSTRACIÓN: Por 8.17, por dos puntos distintos pasa una única recta, que es el axioma P1.

Para probar P2 tomamos cuatro puntos distintos A, B, C, D y suponemos que las rectas AB y CD se cortan en un punto X . Entonces, por 8.16, si $E = AB + CD$, tenemos que $\dim E = 1 + 1 - 0 = 2$, pero entonces, como $A, B, C, D \in E$, es claro que $AC + BD \subset E$, luego $\dim(AC + BD) \leq 2$ y, por consiguiente,

$$\dim(AC \cap BD) = \dim AC + \dim BD - \dim(AC + BD) \geq 1 + 1 - 2 = 0,$$

luego $AC \cap BD \neq \emptyset$.

Si $v_1, v_2, v_3 \in V$ son vectores linealmente independientes, entonces los puntos $A_i = \langle v_i \rangle$ son no colineales, pues si, por ejemplo, $A_3 \in A_1A_2$, eso significaría que $v_3 \in \langle v_1, v_2 \rangle$. Esto prueba P3.

Para probar P4 tomamos una recta $P(W)$, donde $W = \langle v, w \rangle$ es un plano en V , y tenemos que probar que W contiene al menos tres subespacios vectoriales de dimensión 1. Basta tomar $\langle v \rangle, \langle w \rangle, \langle v + w \rangle$.

Ahora una simple inducción sobre n prueba que A_0, \dots, A_n son puntos proyectivamente independientes en el sentido sintético si y sólo si lo son en el sentido analítico, y que en tal caso la variedad proyectiva que generan en ambos sentidos es la misma.

En efecto, esto es trivial para $n = 0$ y, si se cumple para n , tenemos que si A_0, \dots, A_{n+1} son proyectivamente independientes en un sentido, eso equivale a que A_0, \dots, A_n sean proyectivamente independientes en ese mismo sentido y a que $A_{n+1} \notin \langle A_0, \dots, A_n \rangle$, pero la variedad es la misma en ambos sentidos, por hipótesis de inducción. Por último, la variedad $\langle A_0, \dots, A_{n+1} \rangle$ en ambos sentidos está formada por la unión de las rectas que pasan por A_{n+1} y un punto de $\langle A_0, \dots, A_n \rangle$, luego es la misma en ambos sentidos. ■

Ahora es obvio también que todo espacio afín E se puede sumergir en un espacio proyectivo. Si E es un plano lo hemos probado en el teorema 8.3, y si

E tiene dimensión ≥ 3 o bien es un plano arguesiano, entonces E es un espacio afín algebraico sobre un anillo de división K , y en la sección anterior hemos visto que puede sumergirse en $P^n(K)$.

Observemos que todo espacio proyectivo $P(V)$ de dimensión ≥ 2 es arguesiano. Si tiene dimensión ≥ 3 lo es por el teorema 8.14, y si es un plano porque se puede sumergir en $\text{Pr}^3(K)$, y como éste cumple el teorema de Desargues, sus planos también lo cumplen. Ahora demostramos que, recíprocamente, en todo espacio proyectivo arguesiano cada sistema de referencia determina una asignación de coordenadas:

Teorema 8.34 *Sea P un espacio proyectivo arguesiano de dimensión $n \geq 2$ y sea A_0, \dots, A_{n+1} un sistema de referencia proyectivo. Entonces existe un anillo de división K y una biyección $\Phi : P \rightarrow P^n(K)$ que hace corresponder las variedades proyectivas de P con las de $P^n(K)$ de modo que los puntos del sistema de referencia dado tienen por imágenes los del sistema de referencia canónico de $P^n(K)$.*

DEMOSTRACIÓN: Sea $H = \langle A_1, \dots, A_n \rangle$, que es un hiperplano en P . Entonces $E = P \setminus H$ es un espacio afín arguesiano. En el capítulo anterior hemos visto que entonces E admite una estructura de espacio afín algebraico sobre un cierto anillo de división K . Es fácil ver entonces que existe un (único) sistema de referencia afín en E de origen $O = A_0$ respecto al cual los ejes coordenados son las rectas $OA_i \cap E$, para $i = 1, \dots, n$, y de modo que el punto A_{n+1} tenga coordenadas $(1, \dots, 1)$. Sea $\Phi : P \setminus H \rightarrow P^n(K)$ la aplicación que a cada punto $A \in E$ de coordenadas (x_1, \dots, x_n) le asigna el punto $\langle (1, x_1, \dots, x_n) \rangle$.

Claramente, la imagen de Φ es $E' = P^n(K) \setminus H_\infty$, donde H_∞ es el hiperplano de ecuación $x_0 = 0$. Además Φ hace corresponder las variedades afines de E con las variedades afines de E' , y transforma rectas paralelas en rectas paralelas.

Extendemos Φ a los puntos de H del modo siguiente: dado $A \in H$, consideramos la recta $OH \cap E$, que se transforma en una recta de E' , la cual se extiende a una única recta de $P^n(K)$ añadiéndole un único punto infinito posible, que es el que tomamos como $\Phi(A)$.

Si $A, B \in H$ son puntos distintos, tenemos que las rectas $OA \cap E$ y $OB \cap E$ no son paralelas, luego sus imágenes tampoco lo son, luego $\Phi(A) \neq \Phi(B)$, luego la extensión de Φ sigue siendo inyectiva. Es claro que es suprayectiva, pues podemos calcular una antiimagen para cada punto de H siguiendo el proceso inverso al que hemos usado para definir la extensión de Φ .

Si U es cualquier variedad proyectiva de P , (que podemos suponer no vacía), consideramos primero el caso en que U no está contenida en H . Entonces $U \cap E$ es una variedad afín en E de la misma dimensión, y $\Phi[U \cap E]$ es una variedad afín E' , que será de la forma $U' \cap E'$, para cierta variedad proyectiva U' en $P^n(K)$ de la misma dimensión. Es fácil ver que $\Phi[U] = U'$. Por ejemplo, si $A \in U \setminus E$, entonces la recta $OA \cap E$ está contenida en $U \cap E$, luego $\Phi[OA \cap E]$ es una recta en E' contenida en U' , luego su punto infinito $\Phi(A)$ también está en U' . Esto prueba que $\Phi[U] \subset U'$, y la inclusión opuesta se demuestra del mismo modo.

Si $U \subset H$, consideramos la variedad $U^* = \langle U, O \rangle$, cuya dimensión es una unidad mayor, de modo que, por el caso ya probado $\Phi[U^*]$ es una variedad en $\mathbb{P}^n(K)$ y, como Φ es biyectiva, tenemos que $\Phi[U] = U^* \cap H_\infty$ es una variedad proyectiva de la misma dimensión que U .

Similarmente se demuestra que las antiimágenes de las variedades de $\mathbb{P}^n(K)$ son variedades de P .

El punto A_0 tiene coordenadas afines nulas, luego se transforma en el punto $\langle(1, 0, \dots, 0)\rangle$ de $\mathbb{P}^n(K)$. También es claro que A_{n+1} se transforma en el punto $\langle(1, \dots, 1)\rangle$. Por otra parte, $OA_1 \cap E$ es el primer eje coordenado de E , luego es el punto infinito de la recta que pasa por $(0, \dots, 0)$ y por $(1, 0, \dots, 0)$, luego $\Phi(A_1)$ es el punto infinito de la recta que pasa por $(1, 0, \dots, 0)$ y por $(1, 1, 0, \dots, 0)$, que es claramente $(0, 1, 0, \dots, 0)$ (tiene que estar en la variedad de ecuaciones $x_2 = \dots = x_n = 0$ y también en el hiperplano $x_0 = 0$). Lo mismo vale para los demás A_i , para $i = 2, \dots, n$. ■

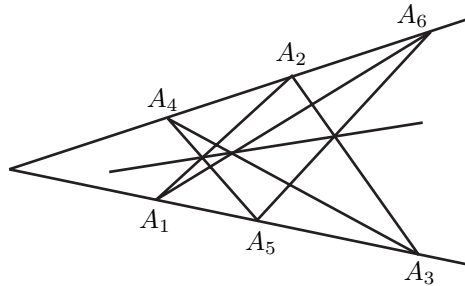
Así pues, si descartamos los planos no arguesianos, no perdemos generalidad si estudiamos únicamente los espacios proyectivos de la forma $\mathbb{P}(V)$, donde V es cualquier espacio vectorial sobre un anillo de división K . En la práctica nos interesará únicamente el caso en que K es un cuerpo, lo cual equivale a que se cumpla el teorema de Papos-Pascal:

Teorema 8.35 *Sea $P = \mathbb{P}(V)$ un espacio proyectivo de dimensión $n \geq 2$ sobre un anillo de división K . Se cumple que K es conmutativo si y sólo si P cumple el teorema de Papos-Pascal:*

Sean r y s dos rectas secantes y consideremos un hexágono tal que los vértices A_1, A_3, A_5 están sobre r y A_2, A_4, A_6 están sobre s (todos ellos distintos del punto de intersección). Entonces los puntos de intersección de cada par de lados opuestos:

$$A_1A_2 \cap A_4A_5, \quad A_2A_3 \cap A_5A_6, \quad A_3A_4 \cap A_1A_6$$

están sobre una misma recta.



Si el lector no reconoce el enunciado del teorema 7.33 es que todavía no ha asimilado la esencia de la geometría proyectiva. El teorema 7.33 es el caso particular de la versión que estamos presentando en el que dos de los pares de lados son paralelos, es decir, se cortan en dos puntos de la recta infinita, lo que obliga (según el enunciado de la versión proyectiva) a que el tercer par se corte también en un punto de dicha recta, luego los dos lados correspondientes resultan ser también paralelos.

DEMOSTRACIÓN: Supongamos que dos de los vértices son iguales, por ejemplo, que $A_2 = A_6$. Entonces $A_2A_3 \cap A_5A_6$ es el propio A_2 , y la recta $A_1A_2 = A_1A_6$ contiene trivialmente a los tres puntos de intersección. Así pues, podemos suponer que los seis vértices son distintos dos a dos.

Si dos de las tres intersecciones coinciden, la conclusión se cumple trivialmente, luego podemos suponer que las tres intersecciones son distintas. Tomemos la recta r que pasa por dos de ellas, digamos por $A_1A_2 \cap A_4A_5$ y $A_2A_3 \cap A_5A_6$. Dicha recta no puede contener a ninguno de los vértices, pues si contuviera, por ejemplo, a A_1 , como también tiene a $A_1A_2 \cap A_4A_5$, que no puede ser A_1 , también $A_2 \in r$, e igualmente $A_6 \in r$, pero entonces $A_1 \in A_2A_6$, y llegamos a que $r = s$, contradicción.

Consideramos el plano Π que contiene a las dos rectas (y a r). Tenemos que es de la forma $P(W)$, para un cierto W de dimensión 2 sobre K , luego $E = \Pi \setminus r$ es un plano afín isomorfo a K^2 que contiene a los seis vértices del hexágono y en el que dos pares de lados son paralelos (pues se cortan en puntos de la recta infinita r). Si K es un cuerpo, por 7.33 tenemos que K^2 cumple el teorema de Papos-Pascal, luego los lados del tercer par también son paralelos, luego se cortan en r , que es lo que había que probar.

Recíprocamente, si P cumple el teorema de Papos-Pascal y H es cualquier hiperplano, entonces $P \setminus H$ es un espacio afín que también cumple el teorema, y a través de una asignación de coordenadas concluimos que K^n también lo cumple, luego por 7.33 concluimos que K es un cuerpo. ■

8.4 Dualidad

Vamos a estudiar a fondo la noción de dualidad que hemos discutido brevemente al principio de este capítulo. Allí habíamos observado que las afirmaciones sobre planos proyectivos

Dados dos puntos distintos, hay una única recta que los contiene.

y

Dadas dos rectas distintas, hay un único punto contenido en ellas.

son duales, en el sentido de que una se obtiene de la otra intercambiando las palabras “punto” y “recta” a la vez que invertimos las inclusiones, y hemos razonado que si una afirmación es cierta para todo plano proyectivo, lo mismo sucede con su afirmación dual, y la razón de fondo es que a cada plano proyectivo P le podemos asociar el plano dual P^* que tiene por puntos a las rectas de P , y por rectas a los haces de rectas de P que pasan por un mismo punto.² El hecho de que los duales de los axiomas sean teoremas se traduce en que P^* cumple los

²Si en lugar de considerar que las rectas con conjuntos de puntos consideramos que un plano proyectivo está determinado por un conjunto de puntos, un conjunto de rectas y una relación de incidencia, entonces podemos definir el plano dual P^* como el que tiene por puntos a las rectas de P y por rectas a los puntos de P , con la relación de incidencia inversa.

axiomas de plano proyectivo, y así, si una afirmación es cierta para todo plano proyectivo, se cumple en P^* , lo cual equivale a que su afirmación dual se cumpla en P , luego la afirmación dual es cierta también en todo plano proyectivo.

Este principio sigue siendo válido si incorporamos como axioma el teorema de Desargues. En efecto, al dualizar un triángulo obtenemos un triángulo (un triángulo son tres puntos no colineales y las tres rectas que pasan por cada par de ellos, luego su dual son tres rectas no concurrentes y los tres puntos donde se cortan, o sea, otro triángulo) y el dual de dos triángulos con un centro de perspectiva son dos triángulos con un eje de perspectiva. Esto implica que el dual del teorema de Desargues es él mismo. Más concretamente, las dos implicaciones recíprocas son duales entre sí. Equivalentemente, un plano proyectivo es arguesiano si y sólo si lo es su dual, por lo que si una afirmación la cumplen todos los planos proyectivos arguesianos, lo mismo vale para su afirmación dual.

Los espacios tridimensionales también satisfacen un principio de dualidad, pero ahora las afirmaciones duales se obtienen intercambiando las palabras 'punto' y 'plano' e invirtiendo las inclusiones (las rectas son autoduales). Por ejemplo, dos afirmaciones duales son

Dados tres puntos no colineales existe un único plano que los contiene.

Tres planos que no contengan una misma recta se cortan en un único punto.

Vamos a interpretar algebraicamente este principio de dualidad, lo que nos permitirá generalizarlo a espacios de dimensión arbitraria. Nos basamos en las propiedades de los espacios vectoriales duales.

Definición 8.36 Sea V un espacio vectorial de dimensión $n + 1$ sobre un cuerpo K . Sea V^* su espacio vectorial dual. Si W es un subespacio de V definimos

$$W^\circ = \{f \in V^* \mid f(v) = 0 \text{ para todo } v \in W\},$$

que es claramente un subespacio de V^* .

Sea w_1, \dots, w_m una base de W y completémosla hasta una base w_1, \dots, w_{n+1} de V . Sea w_1^*, \dots, w_{n+1}^* la base dual. Es fácil ver entonces que

$$W^\circ = \langle w_{m+1}^*, \dots, w_{n+1}^* \rangle.$$

Por consiguiente $\dim W + \dim W^\circ = n + 1$.

Similarmente, si W es un subespacio de V^* definimos

$$W^\circ = \{v \in V \mid f(v) = 0 \text{ para todo } f \in W\},$$

que es un subespacio de V . La relación entre las dimensiones se prueba análogamente. Ahora es claro que si W es un subespacio de V o de V^* se cumple $W^{\circ\circ} = W$ (una inclusión es obvia y las dimensiones dan la igualdad). De aquí

se sigue que las correspondencias $W \mapsto W^\circ$ entre los subespacios de V y V^* son biyectivas y mutuamente inversas. Otro hecho obvio es que $W_1 \subset W_2$ si y sólo si $W_2^\circ \subset W_1^\circ$.

Consideremos ahora $X = P(V)$ y sea $X^* = P(V^*)$. Sea $u : V \rightarrow V^*$ un isomorfismo. Llamemos $\mathcal{P}(X)$ al conjunto de todas las variedades proyectivas de X . Definimos $p : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ como la aplicación dada por

$$p(P(W)) = P(u[W]^\circ).$$

Es claro que p es biyectiva, así como que invierte el orden: $L_1 \subset L_2$ si y sólo si $p(L_2) \subset p(L_1)$. Observemos que p no se altera si sustituimos u por un múltiplo suyo, luego p depende sólo de la homografía H inducida por u .

Definición 8.37 Una *correlación* en un espacio proyectivo X es una biyección $p : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ entre las variedades proyectivas de X que invierte el orden.

Hemos probado que cada homografía $H : X \rightarrow X^*$ induce una correlación entre las variedades lineales de X . A las correlaciones inducidas por homografías las llamaremos *correlaciones proyectivas*. Veremos enseguida que la existencia de correlaciones implica el principio de dualidad. Antes probamos las propiedades básicas de las correlaciones:

Teorema 8.38 Sea X un espacio proyectivo de dimensión n y consideremos en él una correlación $p : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$. Entonces

1. $L_1 \subset L_2$ si y sólo si $p(L_2) \subset p(L_1)$.
2. $\dim L + \dim p(L) = n$.
3. $p(L_1 + L_2) = p(L_1) \cap p(L_2)$, $p(L_1 \cap L_2) = p(L_1) + p(L_2)$.

DEMOSTRACIÓN: 1) es parte de la definición de correlación.

2) Dada una variedad L podemos formar una sucesión de variedades

$$L_{-1} \subset L_0 \subset L_1 \subset \cdots \subset L_n = X,$$

donde $\dim L_i = i$ para todo i y $L = L_i$ para algún i . Al aplicar p obtenemos una sucesión decreciente de variedades, luego necesariamente $\dim p(L_i) = n - i$.

3) es consecuencia inmediata de que p es biyectiva e invierte el orden, pues necesariamente ha de enviar la menor variedad que contiene a dos dadas en la mayor variedad contenida en dos dadas, y viceversa. ■

Ahora es claro que un espacio proyectivo X de dimensión n satisface el *principio de dualidad* siguiente:

Dada una afirmación sobre X , la afirmación que resulta de intercambiar las variedades de dimensión m por las variedades de dimensión $n - m$ e invertir las inclusiones es equivalente a la dada.

Por ejemplo, partamos de que toda recta contiene tres puntos y vamos a probar que toda variedad de dimensión $n-2$ está contenida en tres hiperplanos. Fijamos una correlación p . Dada una variedad L de dimensión $n-2$, entonces $p(L)$ es una recta, luego contiene tres puntos, que podemos expresar en la forma $p(L_1)$, $p(L_2)$, $p(L_3)$, donde L_1 , L_2 y L_3 son tres hiperplanos distintos. Por consiguiente $L \subset L_1 \cap L_2 \cap L_3$. ■

Veamos ahora cómo dualizar enunciados sobre planos proyectivos que hablen de razones dobles. Para ello demostramos el teorema siguiente:

Teorema 8.39 *Sea X un plano proyectivo, sea r una recta en X y P un punto que no esté en r . Sea $p : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ una correlación proyectiva y sea $l = p(P)$. Entonces, la aplicación $H : r \rightarrow l$ dada por $H(A) = p(PA)$ es una homografía.*

DEMOSTRACIÓN: Pongamos que $X = \mathbb{P}(V)$ y sea $u : V \rightarrow V^*$ el isomorfismo que induce la correlación p . Sea $r = \mathbb{P}(\langle w_1, w_2 \rangle)$ y $P = \mathbb{P}(\langle v \rangle)$. El hecho de que P no esté en r implica que los subespacios $\langle v, w_i \rangle$ tienen dimensión 2 (y son distintos entre sí), luego los duales $\langle u(v), u(w_i) \rangle^\circ$ tienen dimensión 1 (y también son distintos entre sí). Pongamos que $\langle u(v), u(w_i) \rangle^\circ = \langle z_i \rangle$, de modo que z_i está determinado salvo múltiplos. Observemos que $s = u(w_1)(z_2)$ y $t = u(w_2)(z_1)$ son escalares no nulos, ya que si, por ejemplo, $s = 0$, tendríamos que $z_2 \in \langle u(v), u(w_1) \rangle^\circ = \langle z_1 \rangle$, lo cual es imposible. Llamando $z'_1 = -(s/t)z_1$ tenemos que

$$u(w_1)(z_2) + u(w_2)(z'_1) = s - (s/t)t = 0,$$

luego cambiando z_1 por z'_1 podemos exigir que $u(w_1)(z_2) + u(w_2)(z_1) = 0$.

Observemos que $z_1, z_2 \in \langle u(v) \rangle^\circ$ y este subespacio tiene dimensión 2, luego de hecho $\langle u(v) \rangle^\circ = \langle z_1, z_2 \rangle$. Sea $f : \langle w_1, w_2 \rangle \rightarrow \langle u(v) \rangle^\circ$ el único isomorfismo que cumple $f(w_i) = z_i$. Así, para cada punto $A \in r$, tenemos que $A = \langle w \rangle$, con $w = \alpha w_1 + \beta w_2$, luego

$$u(v)(f(w)) = \alpha u(v)(z_1) + \beta u(v)(z_2) = 0,$$

$$u(w)(f(w)) = u(\alpha w_1 + \beta w_2)(\alpha z_1 + \beta z_2) =$$

$$\alpha^2 u(w_1)(z_1) + \beta^2 u(w_2)(z_2) + \alpha\beta(u(w_1)(z_2) + u(w_2)(z_1)) = 0.$$

Esto significa que $f(w) \in \langle u(v), u(w) \rangle^\circ$, luego de hecho

$$\langle f(w) \rangle = \langle u(v), u(w) \rangle^\circ = p(PA) = H(A),$$

y esto significa que H es la homografía inducida por f . ■

La dualización en un plano proyectivo de cuatro puntos colineales son cuatro rectas concurrentes, por lo que el concepto dual de la razón doble de cuatro puntos colineales tiene que ser una “razón doble” entre cuatro rectas concurrentes, que podemos definir así:

Definición 8.40 Si l_1, l_2, l_3, l_4 son cuatro rectas de un plano proyectivo X concurrentes en un punto P , definimos su *razón doble* como

$$\mathcal{R}(l_1, l_2, l_3, l_4) = \mathcal{R}(l_1 \cap r, l_2 \cap r, l_3 \cap r, l_4 \cap r),$$

donde r es cualquier recta de X que no pase por P .

Que la razón doble no depende de la elección de r es evidente a partir del teorema 8.51 que probaremos en la sección siguiente, pero también se deduce del teorema de dualidad que vamos a probar a continuación:

Teorema 8.41 Sea X un plano proyectivo y $p : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ una correlación proyectiva. Entonces, si l_1, l_2, l_3, l_4 son cuatro rectas concurrentes en un punto P , se cumple que

$$\mathcal{R}(l_1, l_2, l_3, l_4) = \mathcal{R}(p(l_1), p(l_2), p(l_3), p(l_4)).$$

DEMOSTRACIÓN: Sea r una recta que no pase por P y vamos a probar que la razón doble de las rectas calculada con r es igual al miembro derecho, que no depende de r , con lo que en particular habremos probado que la razón doble no depende de la elección de r . Para ello basta observar que, según el teorema anterior, la aplicación $H : r \rightarrow p(P)$ dada por $A \mapsto p(PA)$ es una homografía. Si llamamos $A_i = l_i \cap r$, tenemos que $PA_i = l_i$, luego

$$\begin{aligned} \mathcal{R}(l_1, l_2, l_3, l_4) &= \mathcal{R}(A_1, A_2, A_3, A_4) = \\ \mathcal{R}(H(A_1), H(A_2), H(A_3), H(A_4)) &= \mathcal{R}(p(l_1), p(l_2), p(l_3), p(l_4)). \end{aligned}$$

■

Por lo tanto, una afirmación sobre puntos y rectas en un plano proyectivo en la que intervengan razones dobles es equivalente a la afirmación dual que resulta de intercambiar puntos y rectas cambiando las razones dobles de puntos por razones dobles de rectas y viceversa.

Hay una forma equivalente de expresar las correlaciones proyectivas que permite introducir una propiedad de simetría. Un isomorfismo $u : V \rightarrow V^*$ induce una forma bilineal regular $F : V \times V \rightarrow K$ mediante $F(v, w) = u(v)(w)$. Recíprocamente, cada forma bilineal regular F induce un isomorfismo u por la relación anterior. De este modo tenemos una biyección entre isomorfismos y formas regulares, y cada correlación proyectiva está inducida por una forma regular.

Con más detalle, si F es la forma bilineal inducida por u y W es un subespacio de V , entonces

$$u[W]^\circ = \{v \in V \mid F(w, v) = 0 \text{ para todo } w \in W\}.$$

Representaremos este conjunto por W^F . De este modo, la correlación inducida por una forma bilineal F viene dada por $p(\mathcal{P}(W)) = \mathcal{P}(W^F)$. Hemos probado que

$$\dim W + \dim W^F = n + 1.$$

Ahora es fácil ver que existen correlaciones proyectivas con una propiedad adicional que las hace más sencillas de manejar:

Definición 8.42 Una correlación proyectiva p se dice *simétrica* si $p \circ p$ es la identidad.

Es fácil ver que la correlación inducida por una forma bilineal F es simétrica si y sólo si F es *ortosimétrica*, es decir, si $F(v, w) = 0$ equivale a $F(w, v) = 0$, para todo par de vectores v y w .

En efecto, una forma bilineal regular en un espacio V es ortosimétrica si y sólo si $W^{FF} = W$ para todo subespacio W de V , pues si F es ortosimétrica entonces $W \subset W^{FF}$, y como las dimensiones son iguales se da la igualdad. Recíprocamente, si F cumple esta condición y $F(v, w) = 0$ entonces $w \in \langle v \rangle^F$ y $v \in \langle v \rangle^{FF} \subset \langle w \rangle^F$, luego $F(w, v) = 0$. A su vez esto equivale a la simetría de la correlación inducida.

La ortosimetría equivale a dos propiedades alternativas posibles más fuertes:

Teorema 8.43 Sea F una forma bilineal ortosimétrica regular en un espacio vectorial V . Entonces F está en uno de los dos casos siguientes:

1. F es simétrica, es decir, $F(v, w) = F(w, v)$ para todo $v, w \in V$.
2. F es antisimétrica, es decir, $F(v, w) = -F(w, v)$ para todo $v, w \in V$. En tal caso $\dim V$ ha de ser par.

DEMOSTRACIÓN: Para cada $y \in V$ llamemos $u(y)$ y $v(y)$ a las formas de V^* dadas por $u(y)(x) = F(x, y)$ y $v(y)(x) = F(y, x)$. El hecho de que F sea regular implica que u y v son dos isomorfismos entre V y V^* . La hipótesis de ortosimetría implica que $u(y)$ y $v(y)$ tienen el mismo núcleo, luego existe un escalar t_y tal que $v(y) = t_y u(y)$.

Si y_1 e y_2 son vectores linealmente independientes, entonces

$$v(y_1 + y_2) = t_{y_1+y_2} u(y_1 + y_2) = t_{y_1} u(y_1) + t_{y_2} u(y_2),$$

de donde se sigue que $t_{y_1} = t_{y_1+y_2} = t_{y_2}$. Es fácil ver entonces que el escalar t_y no depende de y (si V tiene dimensión 1 se llega inmediatamente a la misma conclusión). Por consiguiente $v(y) = t u(y)$, para todo y .

Distinguimos dos casos. Si existe un vector $a \in V$ tal que $F(a, a) \neq 0$, entonces $u(a) = v(a) \neq 0$, de donde $t = 1$, luego

$$F(x, y) = u(y)(x) = v(y)(x) = F(y, x),$$

es decir, F es simétrica. Si $F(a, a) = 0$ para todo vector a , entonces

$$0 = F(v + w, v + w) = F(v, w) + F(w, v),$$

luego $F(v, w) = -F(w, v)$ para todo par de vectores, luego F es antisimétrica. Falta ver que en este caso la dimensión del espacio es par.

Sea e_1, \dots, e_n una base de V y e_1^*, \dots, e_n^* su base dual. La matriz A de u en estas bases tiene en el lugar (i, j) a $u(e_i)(e_j) = F(e_j, e_i)$. La antisimetría nos da entonces que $A^t = -A$. Tomando determinantes $|A| = (-1)^n|A|$, luego n ha de ser par, pues $|A| \neq 0$. ■

Esta clasificación es satisfactoria porque ahora basta conocer la matriz de F en una base para saber si es o no ortosimétrica (según si la matriz es simétrica o antisimétrica).

En particular resulta que las correlaciones proyectivas simétricas de los planos proyectivos son las inducidas por formas bilineales simétricas (no pueden ser antisimétricas porque están definidas sobre un espacio tridimensional).

Las correlaciones están determinadas por sus restricciones a los puntos:

Definición 8.44 Dado un espacio proyectivo X , llamaremos $\mathcal{H}(X)$ al conjunto de todos los hiperplanos de X . Así, cada correlación se restringe a una biyección $p : X \rightarrow \mathcal{H}(X)$. A tales biyecciones se las llama *dualidades*, y si la correlación es simétrica se llaman *polaridades*. A su vez, cabe distinguir entre *polaridades simétricas* o *antisimétricas* según si la correlación simétrica que las induce está inducida a su vez por una forma bilineal simétrica o antisimétrica.

Por el teorema 8.38, cada dualidad determina la correlación que la define.

Una polaridad p empareja cada punto P de X con un hiperplano H . Diremos que P es el *polo* de H y que H es el hiperplano *polar* de P .

La expresión en coordenadas de una polaridad es fácil de describir: un sistema de coordenadas de X determina una base del espacio soporte V , y si A es la matriz de una forma bilineal que induce la polaridad, entonces el hiperplano polar de un punto Q de coordenadas homogéneas U es el hiperplano de ecuación $UAX^t = 0$.

El ejemplo más simple lo tenemos cuando A es la matriz identidad. Fijado un sistema de coordenadas, el hiperplano polar del punto (a_1, \dots, a_{n+1}) es el determinado por la ecuación $a_1x_1 + \dots + a_{n+1}x_{n+1} = 0$, y viceversa.

Las dualidades pueden restringirse a subvariedades:

Teorema 8.45 Sea $p : X \rightarrow \mathcal{H}(X)$ una dualidad en un espacio proyectivo X y sea $V \subset X$ una variedad proyectiva que no esté contenida en $p(P)$, para ningún $P \in V$. Entonces la aplicación $p|_V : V \rightarrow \mathcal{H}(V)$ dada por $p|_V(P) = p(P) \cap V$ es una dualidad en V , y es una polaridad simétrica o antisimétrica si lo es p .

DEMOSTRACIÓN: Es claro que $p|_V$ es una dualidad. Supongamos ahora que p es una polaridad. Fijemos un sistema de referencia de X en el que V esté determinado por las ecuaciones $x_0 = \dots = x_m = 0$ y sea A la matriz de la forma bilineal que induce p en la base asociada a dicho sistema de referencia. Entonces, si $P \in V$ tiene coordenadas $(0, \dots, 0, a_{m+1}, \dots, a_n)$, la ecuación de $p(P)$ es $(0, \dots, 0, a_{m+1}, \dots, a_n)Ax^t = 0$ y un punto $(0, \dots, 0, x_{m+1}, \dots, x_m) \in V$ estará en $p(P) \cap V$ si y sólo si

$$(0, \dots, 0, a_{m+1}, \dots, a_n)A(0, \dots, 0, x_{m+1}, \dots, x_m)^t = 0,$$

que equivale a

$$(a_{m+1}, \dots, a_n)A'(x_{m+1}, \dots, x_m)^t = 0,$$

donde A' es la matriz formada por las últimas filas y columnas de A . No puede ocurrir que $(a_{m+1}, \dots, a_n)A' = 0$, porque entonces $V \subset p(P)$, en contra de lo supuesto. Por lo tanto, A' es una matriz regular y, respecto de cierto sistema de coordenadas de V , tenemos que $p|_V$ es la polaridad inducida por la forma bilineal de matriz A' , que será simétrica o antisimétrica si lo es A . ■

Notemos que si H es una homografía en X , entonces H se extiende a una biyección $H : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ que conserva las inclusiones ($H(L) = H[L]$). El comportamiento de las homografías así extendidas es muy similar al de las correlaciones, salvo que no invierten las inclusiones. El teorema siguiente muestra la compatibilidad entre ambas nociones:

Teorema 8.46 *Sea H una homografía en un espacio proyectivo X y sean p y q dos correlaciones proyectivas en X . Entonces $H \circ p$ y $p \circ H$ son correlaciones proyectivas y $p \circ q$ es una homografía.*

DEMOSTRACIÓN: Los tres hechos se prueban de forma similar. Veamos tan sólo que $p \circ q$ es una homografía. Fijemos un sistema de referencia proyectivo y sean A y B las matrices de p y q en dicho sistema. Sea U la $n + 1$ -tupla de coordenadas homogéneas de un punto Q . Entonces $p(Q)$ es el hiperplano de ecuación $UAX^t = 0$. La imagen por q de este hiperplano es UAB^{-1} , luego $p \circ q$ es la homografía inducida por el isomorfismo de matriz AB^{-1} . ■

Si p y q son correlaciones no necesariamente proyectivas, entonces $p \circ q$ es una colineación y por el teorema fundamental de la geometría proyectiva es una semihomografía. Si el cuerpo de escalares no tiene automorfismos no triviales, como el caso de \mathbb{R} , podemos concluir que se trata de una homografía. En particular, si p es una correlación arbitraria y q es proyectiva, tenemos que $p \circ q = H$ es una homografía, luego $p = q \circ H^{-1}$ es también proyectiva. Así:

Teorema 8.47 *Toda correlación en un espacio proyectivo sobre un cuerpo sin automorfismos no triviales es proyectiva.*

El principio de dualidad se aplica a enunciados que contienen homografías. Para verlo, dada una homografía $h \in \text{LGP}(X)$ y una dualidad p , definimos $h^p : X \rightarrow X$ como la homografía dada por $h^p(Q) = p(h[p^{-1}(Q)])$.

Ejemplo Diremos que una homografía h tiene un *centro de perspectiva* O si fija a los hiperplanos que pasan por el punto O , y diremos que h tiene un *eje de perspectiva* H si fija a todos los puntos del hiperplano H . Es claro que se cumple la afirmación siguiente:

Toda homografía con un eje de perspectiva tiene un centro de perspectiva.

La razón es que al restringir la homografía al complementario de su eje tiene que ser una traslación o una homotecia, luego tiene un centro de perspectiva en h si es una traslación o en E si es una homotecia. Veamos el teorema dual:

Toda homografía con un centro de perspectiva tiene un eje de perspectiva.

Fijamos una correlación proyectiva p . Basta observar que si h tiene a O por centro de perspectiva, entonces h^p tiene a $p(O)$ como eje de perspectiva: en efecto, si $Q \subset p(O)$, entonces $O \subset p^{-1}(Q)$, luego $h[p^{-1}(Q)] = p^{-1}(Q)$, luego $Q = h^p(Q)$. Por consiguiente h^p tiene un centro de perspectiva E , y un razonamiento similar muestra que $p^{-1}(E)$ es un eje de perspectiva de h . ■

En general podemos dualizar cualquier afirmación que haga referencia a homografías considerando a éstas autoduales.

Por ejemplo, el teorema 8.20 para planos proyectivos es:

Dadas dos cuádruplas de puntos no colineales tres a tres en un plano proyectivo, existe una homografía que transforma una en otra.

y su teorema dual es:

Dadas dos cuádruplas de rectas no concurrentes tres a tres en un plano proyectivo, existe una homografía que transforma una en otra.

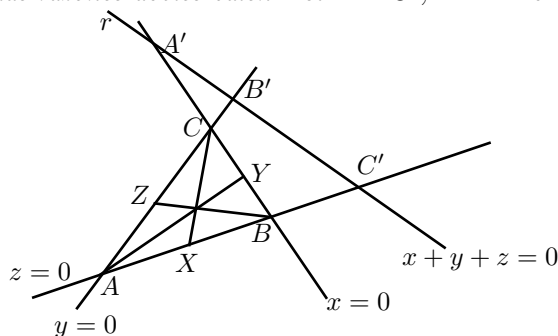
En particular, fijado un sistema de referencia en un plano proyectivo y dada una cuádrupla de rectas no concurrentes tres a tres, existe una homografía que las transforma en las rectas de ecuaciones $x = 0, y = 0, z = 0, x + y + z = 0$, o, equivalentemente, dada una cuádrupla de rectas no concurrentes tres a tres, siempre existe un sistema de referencia respecto al cual sus ecuaciones son las indicadas.

Como aplicación vamos a ver una prueba proyectiva del teorema de Ceva [IGE 4.49] y por dualidad deduciremos³ el teorema de Menelao [IGE 4.48]. El teorema de Ceva es un caso particular del teorema siguiente:

Teorema 8.48 Sean A, B, C tres puntos no colineales en un plano proyectivo que no estén en una recta r . Sean $A' = BC \cap r$, $B' = AC \cap r$, $C' = AB \cap r$ y sean X, Y, Z tres puntos distintos de A, B, C situados, respectivamente, en las rectas AB, BC, AC . Entonces, las cevianas CX, AY, BZ son concurrentes si y sólo si

$$\mathcal{R}(A, B, X, C') \cdot \mathcal{R}(B, C, Y, A') \cdot \mathcal{R}(C, A, Z, B') = -1,$$

entendiendo que las razones dobles valen 1 si $X = C', Y = A'$ o $Z = B'$.



³Las pruebas que vamos a dar son de Julio Benítez.

DEMOSTRACIÓN: Los tres lados del triángulo y la recta r son cuatro rectas no concurrentes tres a tres, luego por la observación precedente podemos tomar un sistema de referencia respecto al cual las cuatro rectas tengan las ecuaciones indicadas en la figura.

Es claro entonces que

$$\begin{aligned} A &= (1, 0, 0), & B &= (0, 1, 0), & C &= (0, 0, 1), \\ A' &= (0, 1, -1), & B' &= (-1, 0, 1), & C' &= (1, -1, 0). \end{aligned}$$

Llamemos

$$\alpha = \mathcal{R}(A, B, X, C'), \quad \beta = \mathcal{R}(B, C, Y, A'), \quad \gamma = \mathcal{R}(C, A, Z, B').$$

Observemos que $X = (1, \lambda, 0)$, para cierto $\lambda \neq 0$, ya que está sobre la recta $z = 0$ y es distinto de A, B . Si tomamos como sistema de referencia de la recta AB los puntos

$$A = (1, 0, 0), \quad B = (0, \lambda, 0), \quad X = (1, \lambda, 0),$$

entonces $C' = (-1, 1, 0) = -1(1, 0, 0) + \lambda^{-1}(0, \lambda, 0)$, luego

$$\alpha = \mathcal{R}(A, B, X, C') = \frac{-1}{\lambda^{-1}} = -\lambda$$

(supuesto que $X \neq C'$ y trivialmente, con el convenio que hemos adoptado, en caso contrario). Así pues, $X = (1, -\alpha, 0)$, e igualmente $Y = (0, 1, -\beta)$, $Z = (-\gamma, 0, 1)$. Como consecuencia, las ecuaciones de las tres cevianas son

$$\alpha x + y = 0, \quad \beta y + z = 0, \quad x + \gamma z = 0.$$

Claramente entonces, las tres son concurrentes si y sólo si el sistema formado por estas tres ecuaciones tiene una solución no trivial, es decir, si y sólo si

$$\begin{vmatrix} \alpha & 1 & 0 \\ 0 & \beta & 1 \\ 1 & 0 & \gamma \end{vmatrix} = \alpha\beta\gamma + 1 = 0,$$

que es la condición del enunciado. ■

Si aplicamos este teorema al caso en que A, B, C son los vértices de un triángulo en un plano afín y r es la recta del infinito (de modo que A', B', C' son los puntos infinitos de los lados del triángulo), usando el teorema 8.29, la condición necesaria y suficiente para que las tres cevianas⁴ sean concurrentes (tal vez paralelas, si el punto de concurrencia es infinito) es

$$\frac{\overrightarrow{AX}}{\overrightarrow{BX}} \cdot \frac{\overrightarrow{BY}}{\overrightarrow{CY}} \cdot \frac{\overrightarrow{CZ}}{\overrightarrow{AZ}} = -1,$$

⁴No excluimos que las cevianas puedan ser paralelas a los lados, en caso de que alguno de los puntos X, Y, Z sea infinito.

(entendiendo que un factor es 1 si el punto X, Y, Z correspondiente es infinito), que equivale a

$$\frac{\overrightarrow{AX}}{\overrightarrow{XB}} \cdot \frac{\overrightarrow{BY}}{\overrightarrow{YC}} \cdot \frac{\overrightarrow{CZ}}{\overrightarrow{ZA}} = 1,$$

(entendiendo ahora que un factor es -1 si el punto correspondiente es infinito), que es la condición dada en [IGE 4.49].

Para obtener el teorema de Menelao consideramos el teorema dual de 8.48:

Sean a, b, c tres rectas no concurrentes en un plano proyectivo y R un punto que no esté en ninguna de ellas. Sean $A = b \cap c$, $B = a \cap c$, $C = a \cap b$, $a' = RA$, $b' = RB$, $c' = RC$ y sean x, y, z tres rectas distintas de a, b, c que pasen por C, A, B , respectivamente. Entonces, los puntos $X = c \cap x$, $Y = a \cap y$, $Z = b \cap z$ son colineales si y sólo si

$$\mathcal{R}(a, b, x, c') \cdot \mathcal{R}(b, c, y, a') \cdot \mathcal{R}(c, a, z, b') = -1,$$

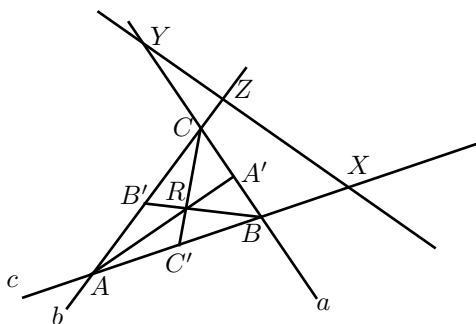
entendiendo que las razones dobles valen 1 si $x = c'$, $y = a'$ o $z = b'$.

Esto es equivalente a:

Teorema 8.49 *Sean A, B, C tres puntos no colineales en un plano proyectivo y R un punto que no esté en ninguna de las rectas $a = BC$, $b = AC$, $c = AB$. Sean $A' = AR \cap a$, $B' = BR \cap b$, $C' = CR \cap c$ y sean X, Y, Z puntos situados sobre las rectas AB, BC, AC , respectivamente, distintos de A, B, C . Entonces X, Y, Z son colineales si y sólo si*

$$\mathcal{R}(A, B, X, C') \cdot \mathcal{R}(B, C, Y, A') \cdot \mathcal{R}(C, A, Z, B') = -1,$$

entendiendo que las razones dobles valen 1 si $X = C'$, $Y = A'$ o $Z = B'$.



DEMOSTRACIÓN: Nos encontramos en las hipótesis del dual de 8.48 tomando $x = XC$, $y = YA$, $z = ZB$, pero, por definición de la razón doble entre rectas,

$$\mathcal{R}(a, b, x, c') = \mathcal{R}(a \cap c, b \cap c, x \cap c, c' \cap c) = \mathcal{R}(B, A, X, C') = \mathcal{R}(A, B, X, C')^{-1},$$

e igualmente, las otras dos razones dobles del enunciado dual son las inversas de las del enunciado y basta tener en cuenta que $(-1)^{-1} = -1$. ■

Ahora, si tenemos un triángulo ABC en un plano afín y tres puntos X, Y, Z (tal vez infinitos) sobre cada uno de sus lados y distintos de sus vértices, siempre podemos tomar un punto R que no esté en ninguno de los lados, el cual define puntos A', B', C' según el teorema anterior. Observemos que el teorema de Ceva nos da que

$$\frac{\overrightarrow{AC'}}{\overrightarrow{C'B}} \cdot \frac{\overrightarrow{BA'}}{\overrightarrow{A'C}} \cdot \frac{\overrightarrow{CB'}}{\overrightarrow{B'A}} = 1.$$

Por otra parte, según el teorema anterior (y teniendo en cuenta el teorema 8.30), los puntos X, Y, Z son colineales si y sólo si

$$\frac{\overrightarrow{AX}/\overrightarrow{AC'}}{\overrightarrow{BX}/\overrightarrow{BC'}} \cdot \frac{\overrightarrow{BY}/\overrightarrow{BA'}}{\overrightarrow{CY}/\overrightarrow{CA'}} \cdot \frac{\overrightarrow{CZ}/\overrightarrow{CB'}}{\overrightarrow{AZ}/\overrightarrow{AZ'}} = -1,$$

(entendiendo que si X o C' es infinito los términos correspondientes se cancelan, y análogamente en los otros dos factores), y esto equivale a

$$\frac{\overrightarrow{AX}/\overrightarrow{XB}}{\overrightarrow{AC'}/\overrightarrow{C'B}} \cdot \frac{\overrightarrow{BY}/\overrightarrow{YC}}{\overrightarrow{BA'}/\overrightarrow{A'C}} \cdot \frac{\overrightarrow{CZ}/\overrightarrow{ZA}}{\overrightarrow{CB'}/\overrightarrow{Z'A}} = -1$$

(donde ahora, si X o C' es infinito, el cociente en que aparece vale -1) y, por la igualdad de Ceva, equivale a

$$\frac{\overrightarrow{AX}}{\overrightarrow{XB}} \cdot \frac{\overrightarrow{BY}}{\overrightarrow{YC}} \cdot \frac{\overrightarrow{CZ}}{\overrightarrow{ZA}} = -1$$

(entendiendo que un factor vale -1 si X, Y o Z es infinito), que es la condición del teorema de Menelao [IGE 4.48].

8.5 Perspectividades

Por el teorema fundamental de la geometría proyectiva, las homografías de un espacio proyectivo en sí mismo se corresponden “casi” con las colineaciones. Ahora vamos a eliminar el “casi” encontrando una caracterización puramente geométrica de las homografías.

Definición 8.50 Sea P un espacio proyectivo de dimensión ≥ 2 , sean H y H' dos hiperplanos distintos en P y sea O un punto exterior a ambos. Llamaremos *proyección perspectiva* de centro O entre ambos hiperplanos a la aplicación $\pi_O : H \rightarrow H'$ que a cada punto $A \in H$ le hace corresponder el punto donde la recta OA corta a H' . La intersección entre los dos hiperplanos se llama *eje* de la proyección.

La aplicación que transforma un plano modelo en su imagen en un cuadro es una proyección perspectiva. La geometría proyectiva es el marco idóneo para estudiar este tipo de aplicaciones, pues si quisiéramos hacerlo desde un punto de vista afín tendríamos que distinguir constantemente casos particulares a causa del paralelismo. Por ejemplo, no todo punto tendría imagen por una proyección perspectiva.

Teorema 8.51 *Toda proyección perspectiva es una homografía.*

DEMOSTRACIÓN: Consideremos una proyección $\pi_O : H \rightarrow H'$ en un espacio P de dimensión n . El eje de la proyección es una variedad de dimensión $n - 2$, que con el punto O genera un hiperplano H_∞ . Sea $E = P \setminus H_\infty$. Las trazas de H y H' en E son dos hiperplanos paralelos y las rectas que pasan por O forman un haz de rectas paralelas. Es fácil ver entonces que la restricción a E de π_O es simplemente la restricción a H de una traslación, luego se extiende a una afinidad de E (la traslación) que a su vez se extiende a una homografía F en P que fija a H_∞ . Por otra parte, π_O fija a los puntos de $H \cap H'$, luego F extiende a π_O , que es, por consiguiente, una homografía. ■

En la práctica es más cómodo sustituir una proyección perspectiva por una homografía que la extienda. Ello nos lleva al concepto de perspectividad:

Definición 8.52 Una *perspectividad* en un espacio proyectivo P de dimensión ≥ 2 es una colineación $\pi : P \rightarrow P$ distinta de la identidad que tiene un centro de perspectiva O y un eje de perspectiva H . Si O está en H la perspectividad se llama *elación*, mientras que si O no está en H se llama *homología*.

Los conceptos de centro y eje de perspectiva los hemos definido en la sección anterior para homografías, pero las definiciones valen para biyecciones cualesquiera: un centro de perspectiva de π es un punto O tal que π fija a todas las rectas que pasan por O , y un eje de perspectiva es un hiperplano H tal que π fija a todos los puntos de H . Además hemos demostrado que una homografía tiene un centro de perspectiva O si y sólo si tiene un eje de perspectiva H , pero no estamos suponiendo que las perspectividades sean homografías. Sucede que podemos demostrarlo:

Teorema 8.53 *Toda perspectividad es una homografía.*

DEMOSTRACIÓN: Sea $\pi : P \rightarrow P$ una perspectividad. El razonamiento empleado en la prueba del teorema fundamental 8.27 muestra que π transforma variedades en variedades, en particular rectas en rectas. Sea H el eje de π y sea $E = P \setminus H$. Entonces $\pi|_E : E \rightarrow E$ es una colineación distinta de la identidad y, más aún, cada recta comparte con su imagen su punto infinito, luego $\pi|_E$ transforma rectas paralelas en rectas paralelas. Por el teorema 5.8 sabemos que $\pi|_E$ es una homotecia o una traslación de E , y en particular es una afinidad, que se extiende a una homografía $h \in \text{LGP}(P)$. De hecho, $h|_E$ es una homotecia o una traslación. Tenemos que h fija a los puntos de H , luego $\pi = h$ es una homografía. ■

Observemos que realmente hemos probado que toda colineación con un eje de perspectiva es una homografía, y por consiguiente tiene un centro de perspectiva. Por dualidad podemos probar también que una colineación con un centro de perspectiva es también una homografía.

En efecto, por el mismo argumento que hemos empleado, toda colineación π transforma variedades en variedades. Fijada una dualidad p , podemos considerar $\pi^p(Q) = p(\pi[p^{-1}(Q)])$, que es también colineación, pues si A, B, C son

tres puntos colineales, entonces $p^{-1}(A)$, $p^{-1}(B)$, $p^{-1}(C)$ son tres hiperplanos que se cortan en una variedad de dimensión $n - 2$ (donde n es la dimensión del espacio), luego lo mismo vale para $\pi[p^{-1}(A)]$, $\pi[p^{-1}(B)]$, $\pi[p^{-1}(C)]$, luego $\pi^p(A)$, $\pi^p(B)$, $\pi^p(C)$ son colineales. Además, es claro que si O es un centro de π , entonces $p(O)$ es un eje de π^p , luego π^p es una homografía, y de ahí se sigue que π también lo es.

En definitiva, las perspectividades son las homografías con un eje de perspectiva H y que, por lo tanto, se restringen a homotecias o traslaciones del espacio afín $E = P \setminus H$, según sean homologías o elaciones. En ambos casos no pueden fijar a los puntos de otro hiperplano, luego el eje H está unívocamente determinado. Así mismo, O es la intersección de las trazas de la homotecia-traslación, luego también está unívocamente determinado. También es claro que existe una única perspectividad con un centro O y un eje H dados que envíe un punto A exterior a H a cualquier otro punto B exterior a H , distinto de A y colineal con O y A .

En la prueba del teorema 8.51 hemos visto que toda proyección perspectiva entre hiperplanos se extiende a una elación. Es fácil ver de modo similar que también puede extenderse a una homología. Recíprocamente, es fácil ver que una perspectividad π en un espacio P fija a todas las rectas que pasan por su centro (expresándolas como intersecciones de hiperplanos). De aquí que si tomamos un hiperplano H distinto del eje y que no contenga al centro, la restricción de π es una proyección perspectiva de H en $\pi[H]$.

Teorema 8.54 *Toda homografía de un espacio proyectivo de dimensión ≥ 2 es producto de homologías, en particular de perspectividades.*

DEMOSTRACIÓN: Sea $H(f) : P(V) \rightarrow P(V)$ la homografía determinada por el automorfismo $f : V \rightarrow V$. Ahora usamos el teorema [A1 6.22], según el cual f se descompone en producto de dilataciones. Recordemos que una dilatación en V de hiperplano H , dirección $w \in V \setminus H$ y razón $\alpha \neq 0$ es el automorfismo $u \in \text{LG}(V)$ determinado por que fija a los puntos de H y $u(w) = \alpha w$. Es claro que entonces la homografía $H(u)$ fija a cada punto del hiperplano $P(H)$ y al punto $O = \langle w \rangle$, luego se trata de una homología. Por lo tanto $H(f)$ es producto de homologías. ■

Observemos que este teorema serviría como definición geométrica de las homografías: podemos definir una homografía como una composición de perspectividades, donde la definición de perspectividad es puramente geométrica.

No toda homografía entre dos hiperplanos H y H' es una proyección perspectiva, pues si fijamos un punto O exterior a ambos y dos puntos A y B en H , sabemos que π_O es la única perspectividad que transforma A y B en $\pi_O(A)$ y $\pi_O(B)$, mientras que hay infinitas homografías con esta propiedad. Por otro lado, la composición de proyecciones perspectivas no es necesariamente una proyección perspectiva. Vamos a probar que toda homografía entre dos hiperplanos (posiblemente iguales) es composición de proyecciones perspectivas. De este modo, los invariantes proyectivos son exactamente los invariantes por tales proyecciones.

Teorema 8.55 Si H es un hiperplano en un espacio proyectivo P de dimensión ≥ 2 , entonces toda homología de H en sí mismo es composición de dos proyecciones perspectivas en P .

DEMOSTRACIÓN: Sean A y r el centro y el eje de una homología f en H y $B \in H \setminus r$ un punto distinto de A . Entonces $f(B)$ es un punto de $H \setminus r$ distinto de B . Sea H' un hiperplano de P distinto de H que pase por r . Sea O un punto exterior a ambos.

Consideremos la perspectividad $\pi_O : H \rightarrow H'$. Sean A' y B' las imágenes de A y B . Como los puntos A , B y $f(B)$ son colineales, los tres son coplanares con O , A' y B' . La recta $f(B)B'$ corta a OA en un punto O' distinto de O . Consideremos la perspectividad $\pi'_{O'} : H' \rightarrow H$ y veamos que $f = \pi_O \circ \pi'_{O'}$.

Si llamamos g a la composición, tenemos que g fija a cada punto de r , lo que significa que g transforma cada recta de $H \setminus r$ en una paralela. Por lo tanto se trata de una homotecia o una traslación, pero $g(A) = A$, luego es una homotecia de centro A . Como $g(B) = f(B)$, necesariamente $f = g$. ■

Teorema 8.56 Toda homografía entre dos hiperplanos H y H' de un espacio proyectivo P de dimensión ≥ 2 es producto de proyecciones perspectivas.

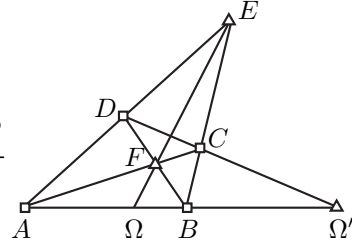
DEMOSTRACIÓN: Lo tenemos probado si $H = H'$. En otro caso, dada una homografía F , tomamos una proyección perspectiva cualquiera $\pi : H \rightarrow H'$ y resulta que $F \circ \pi^{-1}$ es producto de proyecciones perspectivas, luego F también. ■

8.6 Cuadriláteros completos

Vamos a introducir ahora una configuración de puntos de gran importancia en la geometría proyectiva. En realidad la conocemos ya, pues nos la hemos encontrado en el teorema 6.13.

Definición 8.57 Un *cuadrilátero completo* en un plano proyectivo es una figura formada por cuatro puntos no colineales tres a tres, llamados *vértices* junto con las seis rectas que unen cada par de ellos, llamadas *lados*. Dos lados se dicen *contiguos* u *opuestos* según si tienen o no un vértice en común. Las tres intersecciones de lados opuestos se llaman *puntos diagonales* del cuadrilátero.

La figura muestra un cuadrilátero completo con sus vértices marcados con cuadrados y sus puntos diagonales marcados con triángulos. Los vértices de un cuadrilátero completo (en un cierto orden) determinan un sistema de referencia. Consideremos el que asigna las coordenadas homogéneas



$$A = (1, 0, 0), \quad B = (0, 1, 0), \quad C = (0, 0, 1), \quad D = (1, 1, 1).$$

Entonces el lado AB es la recta $z = 0$ (porque esta ecuación define una recta y ciertamente A y B están en ella⁵). Similarmente DC es la recta $x - y = 0$, luego $\Omega' = (1, 1, 0)$. Similarmente, BC es $x = 0$ y AD es $y - z = 0$, luego $E = (0, 1, 1)$, y por último AC es $y = 0$ y BD es $z - x = 0$, luego $F = (1, 0, 1)$.

Esto muestra que los puntos diagonales de un cuadrilátero completo son siempre distintos entre sí y distintos de los vértices. El punto Ω que está también marcado en la figura está en las rectas EF , de ecuación $x + y - z = 0$ y AB , que ya sabemos que es $z = 0$, luego $\Omega = (1, -1, 0)$.

Este punto tiene interés por varios motivos. En primer lugar, es claro que los tres puntos diagonales del cuadrilátero completo están alineados si y sólo si $\Omega = \Omega'$, lo cual, comparando sus coordenadas, equivale a que $1 = -1$. Con ello hemos demostrado el teorema siguiente:

Teorema 8.58 *Sea P un plano proyectivo sobre un cuerpo K . Las afirmaciones siguientes son equivalentes:*

1. *Los puntos diagonales de todo cuadrilátero completo en P no son colineales.*
2. *Existe un cuadrilátero completo en P cuyos puntos diagonales no son colineales.*
3. *La característica de K es distinta de 2.*

La primera afirmación del teorema anterior se conoce como *Axioma de Fano*, y es una forma geométrica de descartar que la característica del cuerpo sea 2, lo cual es necesario, por ejemplo, al desarrollar la teoría de las secciones cónicas desde el punto de vista de la geometría proyectiva.

Volviendo al teorema 6.13, lo que afirma es que las rectas AB y CD son paralelas (lo cual equivale a que Ω' sea un punto infinito) si y sólo si Ω es el punto medio de A y B , lo cual (si Ω' es infinito) equivale a que $\mathcal{R}(\Omega', \Omega, A, B) = -1$. Conviene dar nombre a la primera parte de esta equivalencia:

Definición 8.59 *Dados cuatro puntos colineales A, B, Ω, Ω' , diremos que el par (A, B) separa harmónicamente al par (Ω, Ω') si A y B son dos vértices de un cuadrilátero completo, Ω' es un punto diagonal y Ω es la intersección con AB de la recta que une los otros dos puntos diagonales. Lo representaremos $\mathcal{H}(A, B; \Omega, \Omega')$.*

Ya casi tenemos demostrado el teorema siguiente:

Teorema 8.60 *Si A, B, Ω, Ω' son cuatro puntos colineales distintos, entonces $\mathcal{H}(A, B; \Omega, \Omega')$ si y sólo si $\mathcal{R}(A, B, \Omega, \Omega') = -1$.*

⁵En general es fácil ver que la ecuación de la recta que pasa por dos puntos es

$$\begin{vmatrix} x & y & z \\ a & b & c \\ a' & b' & c' \end{vmatrix} = 0.$$

DEMOSTRACIÓN: Supongamos $\mathcal{H}(A, B; \Omega, \Omega')$. Consideremos entonces un cuadrilátero completo de vértices A, B, C, D en las condiciones de la definición. Es fácil ver que al eliminar la componente z de las coordenadas homogéneas de los puntos de la recta AB obtenemos una homografía $AB \rightarrow \mathbb{P}^1(K)$, es decir, una asignación de coordenadas, respecto del cual

$$A = (1, 0), \quad B = (0, 1), \quad \Omega = (1, -1), \quad \Omega' = (1, 1).$$

Si si tomamos a A como punto infinito de la recta, las coordenadas afines de los otros puntos son $B = 0, \Omega' = 1, \Omega = -1$, luego por la propia definición de razón doble $\mathcal{R}(A, B, \Omega', \Omega) = -1$, y esto equivale a $\mathcal{R}(A, B, \Omega, \Omega') = -1$.

Recíprocamente, dados A, B, Ω , es fácil construir un punto P tal que $\mathcal{H}(A, B; \Omega, P)$. Por la parte ya probada $\mathcal{R}(A, B, \Omega, P) = -1 = \mathcal{R}(A, B, \Omega, \Omega')$, luego $P = \Omega'$. ■

Para probar 6.13 observamos que tenemos $\mathcal{H}(A, B; \Omega, \Omega')$, luego se cumple $\mathcal{R}(A, B, \Omega, \Omega') = -1$. Si Ω es el punto medio de A, B , y P_∞ es el punto infinito de AB , entonces

$$\mathcal{R}(A, B, \Omega, P_\infty) = \mathcal{R}(P_\infty, \Omega, A, B) = -1,$$

y de $\mathcal{R}(A, B, \Omega, \Omega') = \mathcal{R}(A, B, \Omega, P_\infty)$ se sigue que $\Omega' = P_\infty$, luego AB y CD son paralelas. El recíproco es inmediato. ■

De 8.60 se sigue inmediatamente que

$$\mathcal{H}(A, B; \Omega, \Omega') \leftrightarrow \mathcal{H}(B, A; \Omega, \Omega') \leftrightarrow \mathcal{H}(A, B; \Omega', \Omega) \leftrightarrow \mathcal{H}(\Omega, \Omega'; A, B),$$

Estas equivalencias expresan que la separación armónica de dos pares de puntos depende únicamente de los pares, A, B y Ω, Ω' , sin importar el orden de ambos ni el orden de sus componentes. Es claro que dados tres puntos colineales distintos A, B y Ω , existe un único punto Ω' distinto de los anteriores tal que $\mathcal{H}(A, B; \Omega, \Omega')$. A dicho punto se le llama el *conjugado armónico* de Ω respecto a A, B . Así, dos puntos cualesquiera A y B dividen la recta que los contiene en pares de conjugados.

Para terminar la sección aplicaremos los resultados sobre separación armónica al estudio de las involuciones, que definimos a continuación:

Definición 8.61 Una *involución* en una recta proyectiva X es una homografía $I : X \rightarrow X$ de orden 2, es decir, tal que $I \neq 1$ pero $I \circ I = 1$. Diremos que dos puntos P y Q de X son *conjugados* respecto a I si $I(P) = Q$.

Es claro que cada punto tiene un único conjugado respecto a una involución dada (que puede ser él mismo). Si P es el conjugado de Q , entonces Q es el conjugado de P .

Teorema 8.62 *Las involuciones de una recta coinciden con sus polaridades simétricas.*

DEMOSTRACIÓN: Sea A la matriz de una involución f en un sistema de referencia cualquiera. Entonces A^2 es un múltiplo de la matriz identidad. Planteando las ecuaciones a las que esto conduce llegamos fácilmente a que

$$A = \begin{pmatrix} u & v \\ w & -u \end{pmatrix}, \quad \text{donde } u^2 + vw \neq 0.$$

Consideremos la polaridad simétrica p que en el mismo sistema de referencia tiene matriz

$$B = \begin{pmatrix} v & -u \\ -u & -w \end{pmatrix}.$$

El punto polar de un punto de coordenadas (a, b) es el que tiene ecuación

$$(a, b) \begin{pmatrix} v & -u \\ -u & -w \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0,$$

es decir, $(av - bu)x + (-au - bw)y = 0$. El punto es, pues, $(au + bw, av - bu)$, que coincide con la imagen de (a, b) por f . Por consiguiente $f = p$. Recíprocamente, toda polaridad simétrica tiene matriz B en un sistema de referencia arbitrario, y la homografía de matriz A es una involución que coincide con la polaridad dada. ■

Ejercicio: Probar que la única polaridad antisimétrica de una recta proyectiva es la identidad.

Teorema 8.63 *Sea X una recta proyectiva.*

1. Si $H \in \text{LGP}(X)$ cumple $H(P) = Q$ y $H(Q) = P$ para dos puntos distintos P y Q , entonces H es una involución.
2. Si A y B son puntos fijos de una involución I , entonces cada par de puntos conjugados respecto de I son conjugados armónicos respecto de A y B .
3. Dados dos pares (A, A') , (B, B') (sin puntos comunes) existe una única involución I respecto a la cual son pares conjugados.
4. Toda homografía en X es producto de dos involuciones.

DEMOSTRACIÓN: 1) Sea $P = \langle v \rangle$ y $Q = \langle w \rangle$. Si $H = H(u)$, entonces $u(v) = \alpha w$ y $u(w) = \beta v$. La matriz de u en la base v, w es

$$A = \begin{pmatrix} 0 & \alpha \\ \beta & 0 \end{pmatrix},$$

y es claro que $A^2 = \alpha\beta I$, luego $H^2 = 1$.

Sean $C \neq C'$ un par de puntos conjugados respecto de I . Teniendo en cuenta que las homografías conservan la razón doble, aplicando I vemos que $\mathcal{R}(A, B, C, C') = \mathcal{R}(A, B, C', C) = \mathcal{R}(A, B, C, C')^{-1}$, lo cual sólo es posible si $\mathcal{R}(A, B, C, C') = -1$, lo cual equivale a que $\mathcal{H}(A, B; C, C')$.

3) Si uno de los pares tiene componentes distintas, digamos $A \neq A'$, entonces la homografía dada por $H(A) = A'$, $H(A') = A$ y $H(B) = B'$ es una involución por 1), y claramente es la única para la cual los pares dados son conjugados. Si $A = A' = \langle v \rangle$ y $B = B' = \langle w \rangle$, la homografía H inducida por $u(v) = -v$, $u(w) = w$ cumple $H(A) = A$ y $H(B) = B$, no es la identidad y claramente es una involución, y es la única que fija a A y B , pues, por 2), necesariamente tiene que enviar cada punto C a su conjugado armónico respecto de A y B .

4) Sea f una homografía. Podemos suponer que es distinta de la identidad. Sea A un punto no fijado por f . Sea $f(A) = A' \neq A$ y $f(A') = A''$. Sea I la involución que conjugue los pares (A, A'') y (A', A') . Entonces $J = f \circ I$ intercambia A y A' , luego es una involución y $f = J \circ I$. ■

Conviene observar que, aunque una homografía en una recta puede tener en general 0, 1 o 2 puntos fijos, lo cierto es que una involución no puede tener un único punto fijo, pues tomándolo como punto infinito obtenemos una afinidad de orden 2, de ecuación $f(x) = a + bx$, no puede ser $b = 1$ pues entonces no tendría orden 2 (aquí hay que suponer $\text{car } K \neq 2$), y entonces la ecuación $x = a + bx$ tiene como solución un segundo punto fijo.

Ejercicio: Dos involuciones se llaman *armónicas* si su producto es una involución. Probar que dos involuciones son armónicas si y sólo si conmutan.

Los cuadriláteros completos determinan una configuración de interés más general:

Definición 8.64 Diremos que seis puntos (no necesariamente distintos) en una recta r forman un *conjunto cuadrangular* y lo representaremos $\mathcal{Q}(ABC; A'B'C')$, si existe un cuadrilátero completo con vértices exteriores a r de modo que A, B, C sean las intersecciones con r de tres lados y A', B', C' sean las intersecciones con los lados opuestos correspondientes.

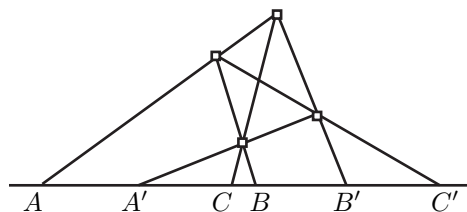
Observemos que las únicas coincidencias que pueden darse entre los puntos son $A = A', B = B', C = C'$, y si se dan las tres, es decir, si se cumple $\mathcal{Q}(ABC; ABC)$ es que los puntos diagonales del cuadrilátero están alienados, lo cual sólo puede suceder en cuerpos de característica 2.

Que se den dos coincidencias, por ejemplo $A = A'$ y $B = B'$ equivale a que estos puntos sean dos puntos diagonales del cuadrilátero, y por definición esto equivale a que (A, B) separe armónicamente a (C, C') . Así pues:

$$\mathcal{H}(AB; CD) \leftrightarrow \mathcal{Q}(ABC; ABD).$$

El resultado fundamental sobre esta configuración es el siguiente:

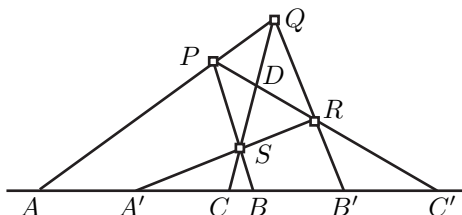
Teorema 8.65 Sean A, B, C , y A', B', C' dos ternas de puntos distintos dos a dos (los de cada terna) en una recta r . Supongamos además que $C \neq C'$. Las afirmaciones siguientes son equivalentes:



1. $\mathcal{Q}(ABC; A'B'C')$
2. Existe una (única) homografía $f : r \rightarrow r$ tal que $f(A) = B'$, $f(B) = A'$, $f(C) = C$, $f(C') = C'$.
3. Existe una (única) involución I en r respecto a la cual (A, A') , (B, B') y (C, C') son pares conjugados (entendiendo que son puntos fijos si las dos componentes coinciden).

DEMOSTRACIÓN: 1) \Rightarrow 2) Supongamos $\mathcal{Q}(ABC; A'B'C')$ y consideremos la composición de las proyecciones perspectivas con los centros indicados sobre las flechas:

$$r \xrightarrow{P} QS \xrightarrow{R} r.$$



Es inmediato que transforma $(A, B, C, C') \mapsto (Q, S, C, D) \mapsto (B', A', C, C')$, luego cumple lo requerido.

2) \Rightarrow 1) Tomamos puntos P y Q cualesquiera colineales con A , llamamos $S = PB \cap QC$ y $R = A'S \cap QB'$, y $C'' = PR \cap r$. Así obtenemos un cuadrilátero completo que prueba $\mathcal{Q}(ABC; A'B'C'')$. La homografía g en r que determina la parte ya probada cumple lo mismo que la f dada salvo que $g(C) = C''$, pero f y g coinciden en tres puntos, de hecho son iguales, y así $C'' = C'$.

2) \Rightarrow 3) Por el teorema anterior existe una involución en r que conjuga los pares (A', B') y (C, C') , y al componerla con la f dada obtenemos una homografía I que cumple $(A, B, C, C') \mapsto (A', B', C, C')$. Como $C \neq C'$ e I los intercambia, el teorema anterior nos da que I es una involución, que claramente cumple lo requerido.

3) \Rightarrow 2) Componiendo la involución I con la misma involución considerada en el apartado anterior obtenemos la homografía f requerida. ■

8.7 Espacios sobre cuerpos ordenados

Sabemos que una ordenación sobre un cuerpo se traduce en que en los espacios afines asociados cada recta tiene asociadas dos ordenaciones naturales, mutuamente inversas. En esta sección exploraremos las consecuencias que tiene en la geometría proyectiva el trabajar con un cuerpo ordenado.

Para interpretar intuitivamente lo que sigue hemos de pensar que los extremos de una recta afín se unen en el punto infinito, con lo que una recta proyectiva es en realidad una circunferencia. Así, del mismo modo que dos puntos en una circunferencia no determinan un segmento, sino que la dividen en dos arcos, igualmente dos puntos en una recta proyectiva no determinan un segmento, sino dos. Si los puntos son finitos, un segmento es el usual, y el otro está formado por los puntos exteriores al segmento más el punto infinito. Si tenemos

un punto finito y otro infinito, los dos segmentos que aparecen son las dos semirrectas con origen en el punto finito. Vemos, pues, que los conceptos afines de “segmento” y “semirrecta” corresponden a un mismo concepto proyectivo: una semirrecta no es más que un segmento con un extremo infinito.

A lo largo de toda la sección supondremos que los espacios proyectivos que consideramos están definidos sobre un cuerpo ordenado R (en particular de característica 0). Dada una recta proyectiva $X = P(V)$, dos puntos de X son de la forma $A = \langle v \rangle$ y $C = \langle w \rangle$, donde v y w son linealmente independientes. Definimos los conjuntos

$$S_+(v, w) = \{\langle u \rangle \mid u = \alpha v + \beta w \text{ con } \alpha\beta \geq 0\},$$

$$S_-(v, w) = \{\langle u \rangle \mid u = \alpha v + \beta w \text{ con } \alpha\beta \leq 0\}.$$

Es claro que todo punto de X distinto de A y C está en uno y sólo uno de los dos conjuntos anteriores. Ambos contienen infinitos puntos. Únicamente A y C están en ambos a la vez. Los llamaremos *segmentos* de extremos A y C .

Si cambiamos los generadores de P y Q por $v' = av$ y $w' = bw$, entonces según los signos de a y b sucederá

$$S_+(v, w) = S_+(v', w') \quad \text{y} \quad S_-(v, w) = S_-(v', w')$$

o bien

$$S_+(v, w) = S_-(v', w') \quad \text{y} \quad S_-(v, w) = S_+(v', w').$$

De este modo, cada par de puntos A y C divide la recta en dos segmentos, si bien es imposible distinguir uno de otro (la distinción depende de la elección de los generadores). Pese a esto, es posible distinguirlos en función de un sistema de referencia, pues entonces podemos elegir como generador de cada punto finito el vector que corresponde a las coordenadas homogéneas de la forma $(1, \alpha)$. Así, dados dos puntos $P(1, \alpha)$ y $Q(1, \beta)$, podemos definir

$$\begin{aligned} S_+(P, Q) &= \{a(1, \alpha) + b(1, \beta) \mid ab \geq 0\} = \left\{ \left(1, \frac{a\alpha}{a+b} + \frac{b\beta}{a+b} \right) \mid ab \geq 0 \right\} \\ &= \{(1, \lambda\alpha + (1-\lambda)\beta) \mid 0 \leq \lambda \leq 1\}, \end{aligned}$$

que es el segmento de extremos α y β en el sentido de la geometría afín, y $S_-(P, Q)$, que es el complementario del anterior.

Si consideramos un punto $P(1, \alpha)$ y el punto infinito $Q(0, 1)$, entonces

$$\begin{aligned} S_+(P, Q) &= \{a(1, \alpha) + b(0, 1) \mid ab \geq 0\} = \left\{ \left(1, \alpha + \frac{b}{a} \right) \mid ab > 0 \right\} \cup \{\infty\} \\ &= \{(1, \lambda) \mid \alpha \leq \lambda\} \cup \{\infty\}, \end{aligned}$$

que es una de las semirrectas de origen en α . Similarmente se comprueba que $S_-(P, Q)$ es la semirrecta complementaria.

Con esto hemos probado lo que habíamos afirmado: al quitarle el punto infinito a un segmento proyectivo obtenemos un segmento afín, o bien el complementario de un segmento afín, o bien una semirrecta.

Si no fijamos un sistema de referencia, para distinguir los dos segmentos con un mismo par de extremos hemos de usar un tercer punto:

Dados tres puntos distintos A, B, C en una recta, definimos \overline{ABC} como el segmento de extremos A y C que contiene a B . Veamos algunas propiedades:

Teorema 8.66 Sean A, B, C puntos de una recta proyectiva X .

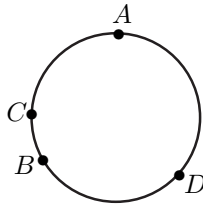
1. $\overline{ABC} = \overline{CBA}$.
2. Si $D \in \overline{ABC}$, $A \neq D \neq C$, entonces $\overline{ABC} = \overline{ADC}$.
3. Si $D \notin \overline{ABC}$ entonces $\overline{ABC} \cup \overline{ADC} = X$ y $\overline{ABC} \cap \overline{ADC} = \{A, C\}$.
4. Dado \overline{ABC} , existen puntos P y Q tales que $\overline{ABC} = \overline{APB} \cup \overline{BQC}$ y $\overline{APB} \cap \overline{BQC} = \{B\}$.
5. Si $D \in \overline{ABC}$, $D \neq A, B, C$ entonces existe P tal que $\overline{BPD} \subset \overline{ABC}$.
6. Una homografía que transforme A, B, C en A', B', C' , respectivamente, transforma \overline{ABC} en $\overline{A'B'C'}$.

DEMOSTRACIÓN: Las tres primeras propiedades son consecuencias inmediatas de la definición de segmento. La última es inmediata y las propiedades 4) y 5) se reducen a las propiedades de los segmentos afines. Por ejemplo, para probar 4) tomamos como punto infinito un punto del segmento complementario a \overline{ABC} y lo que hay que probar es que si $B \in \overline{AC}$ entonces $\overline{AC} = \overline{AB} \cup \overline{BC}$, $\overline{AB} \cap \overline{BC} = \{B\}$, lo cual es obvio. ■

En general, cualquier propiedad básica sobre segmentos se prueba sin dificultad reduciéndola a una recta afín mediante una elección oportuna del punto infinito.

Lo más parecido a la relación “estar entre” que podemos definir en una recta proyectiva es lo siguiente:

Definición 8.67 Sean A, B, C, D cuatro puntos distintos en una recta proyectiva. Diremos que el par AB separa al par CD , y lo representaremos por $AB \parallel CD$, si $D \notin \overline{ACB}$.



Esto significa que los dos segmentos de extremos A y B son \overline{ACB} y \overline{ADB} . En particular cualquiera de los segmentos de extremos A y B corta a cualquiera de los segmentos de extremos C y D . Por el contrario, si $AB \not\parallel CD$, entonces tenemos $C, D \in \overline{ACB}$, luego existe un P tal que $\overline{CPD} \subset \overline{ACB}$. Si tomamos un punto infinito en el segmento complementario de \overline{ACB} tenemos una inclusión de segmentos usuales con extremos distintos, luego es claro que \overline{CPD} es disjunto del segmento complementario de \overline{ACB} . Por lo tanto:

Teorema 8.68 *Cuatro puntos colineales distintos cumplen $AB \parallel CD$ si y sólo si los dos segmentos de extremos A y B cortan a cada uno de los segmentos de extremos C y D .*

Similarmente se prueba:

Teorema 8.69 *La relación $AB \parallel CD$ equivale a*

$$AB \parallel DC, \quad BA \parallel CD, \quad CD \parallel AB \quad \text{y} \quad AC \nparallel BD.$$

La separación de segmentos se caracteriza de forma muy simple mediante las razones dobles:

Teorema 8.70 *Cuatro puntos colineales distintos cumplen $AB \parallel CD$ si y sólo si $\mathcal{R}(A, B, C, D) < 0$.*

DEMOSTRACIÓN: Consideremos un sistema de referencia en la recta de modo que las coordenadas de A y B sean respectivamente O e ∞ . Entonces

$$\mathcal{R}(A, B, C, D) = \frac{\overrightarrow{AC}}{\overrightarrow{AD}}$$

y es claro que $AB \parallel CD$ si y sólo si los vectores \overrightarrow{AC} y \overrightarrow{AD} tienen sentidos opuestos, lo que equivale a que la razón anterior sea negativa. ■

Como consecuencia vemos que la separación armónica es un caso particular de separación, pues $\mathcal{H}(A, B; C, D)$ equivale a $\mathcal{R}(A, B, C, D) = -1$.

Teorema 8.71 *Si $\mathcal{H}(A, B; C, D)$ entonces $AB \parallel CD$.*

Terminamos la sección con unas propiedades de las involuciones. Conviene introducir la siguiente nomenclatura habitual:

Definición 8.72 Diremos que una homografía de una recta proyectiva es *elíptica*, *parabólica* o *hiperbólica* según si tiene 0, 1 o 2 puntos fijos.

En la sección anterior hemos visto que no existen involuciones parabólicas. Si suponemos que R es euclídeo, entonces existen sólo dos formas canónicas de matrices de involuciones. En efecto:

Teorema 8.73 *Toda involución de una recta proyectiva sobre un cuerpo euclídeo admite en un sistema de referencia adecuado una matriz de la forma*

$$M = \begin{pmatrix} 0 & 1 \\ \pm 1 & 0 \end{pmatrix}. \quad (8.1)$$

Si el signo es positivo la involución es hiperbólica y si es negativo es elíptica. Más en general, una involución es elíptica o hiperbólica según si el determinante de cualquiera de sus matrices es positivo o negativo.

DEMOSTRACIÓN: Sea I una involución en la recta $P(V)$, que será de la forma $I = H(f)$, para cierto automorfismo f de V . Sea (A, A') cualquier par de puntos conjugados por I distintos entre sí. Pongamos que $A = \langle v_1 \rangle$, de modo que $A' = \langle v_2 \rangle$, con $v_2 = f(v_1)$ y a su vez $f(v_2) = av_1$, para un cierto $a \in R$ no nulo. Sea $v'_2 = v_2/\sqrt{|a|}$. Entonces es claro que $f(v_1) = \sqrt{|a|}v'_2$ y $f(v'_2) = \pm\sqrt{|a|}v_1$. Si cambiamos f por $f/\sqrt{|a|}$ la involución inducida es la misma, pero ahora la matriz en la base v_1, v'_2 es de la forma indicada en el enunciado.

Los puntos fijos de la involución se corresponden con los valores propios de la matriz M , que son las raíces de polcar $M = x^2 + \Delta$, donde $\Delta = \pm 1$ es el determinante de M . Así pues, M tiene dos valores propios distintos o ninguno según si $\Delta < 0$ o $\Delta > 0$. Por último observamos que, aunque el determinante no es un invariante de las matrices de una misma homografía, dos matrices de una misma homografía en una misma base se diferencian en un factor constante $a \neq 0$, luego sus determinantes se diferencian en $a^2 > 0$, luego el signo del determinante sí es invariante, y podemos calcularlo a partir de cualquier matriz. ■

Nota El sistema de referencia construido en el teorema anterior está formado por $A = \langle v_1 \rangle, A' = \langle v'_2 \rangle$ y $A'' = \langle v_1 + v'_2 \rangle$, donde A y A' es cualquier par de puntos conjugados por I distintos entre sí, cuyas coordenadas son $(1, 0)$ y $(0, 1)$. Esto significa que el sistema de referencia puede tomarse de modo que, en coordenadas afines, 0 e ∞ sea cualquier par prefijado de puntos conjugados por I . Respecto de este sistema de referencia la expresión de la involución es $I(x) = \pm 1/x$, entendiendo que $I(0) = \infty$ y viceversa.

No obstante, si I es hiperbólica obtenemos una expresión más simple tomando cualquier sistema de referencia respecto al que los puntos fijos sean 0 e ∞ , pues entonces la involución es necesariamente $I(x) = -x$ (ya que esto define una involución que fija a 0 e ∞ y sólo hay una posible). ■

Veamos algunas aplicaciones de estos hechos:

Teorema 8.74 *En una recta proyectiva sobre un cuerpo euclídeo, toda involución elíptica es producto de dos involuciones hiperbólicas.*

DEMOSTRACIÓN: Tomemos un sistema de referencia respecto al que la involución elíptica venga dada por $I(x) = -1/x$. Entonces I es la composición de $J_1(x) = 1/x$ y $J_2(x) = -x$, y ambas son involuciones hiperbólicas. ■

Teorema 8.75 *En una recta proyectiva sobre un cuerpo euclídeo, cada par de puntos conjugados por una involución elíptica separa a cualquier otro par. Si la involución es hiperbólica ningún par separa a ningún otro.*

DEMOSTRACIÓN: Si la involución es hiperbólica, acabamos de ver que, respecto de un sistema de referencia adecuado, viene dada por $I(x) = -x$, y es claro entonces que si (A, A') y (B, B') son dos pares de puntos conjugados, los segmentos $\overline{AOA'}$ y $\overline{BOB'}$ son concéntricos, luego uno está contenido en el otro, y los extremos no se separan.

Supongamos ahora que existen dos pares de conjugados (A, A') y (B, B') que no se separan y veamos que la involución es hiperbólica, o sea, que su determinante es negativo. Sabemos que, respecto de un sistema de referencia que asigna a A y A' las coordenadas afines 0 e ∞ , la expresión de la involución es $I(x) = \pm 1/x$, pero si el signo es negativo es claro que $(0, \infty)$ separa a cualquier par $(x, -1/x)$, luego el signo tiene que ser positivo, lo que significa que I es hiperbólica. ■

Teorema 8.76 *En una recta proyectiva sobre un cuerpo euclídeo, dos involuciones distintas una de las cuales sea elíptica tienen un único par común de puntos conjugados.*

DEMOSTRACIÓN: Fijemos un sistema de referencia en el que una de las involuciones tenga matriz

$$M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Según hemos visto en la prueba de 8.62, la otra tendrá matriz

$$A = \begin{pmatrix} u & v \\ w & -u \end{pmatrix}, \quad \text{donde } u^2 + vw \neq 0.$$

Para que ambas coincidan sobre un mismo punto de coordenadas homogéneas (x, y) debe haber un $\lambda \neq 0$ tal que

$$(x, y)M = \lambda(x, y)A,$$

lo cual equivale a que $|M - \lambda A| = 0$. Explícitamente,

$$|M - \lambda A| = -u^2 - (v - \lambda)(w + \lambda) = \lambda^2 + (w - v)\lambda - (u^2 + vw) = 0.$$

La ecuación tendrá solución si su discriminante es no negativo:

$$D = (w - v)^2 + 4(u^2 + vw) = w^2 + v^2 - 2vw + 4u^2 + 4vw = (w + v)^2 + (2u)^2 \geq 0.$$

Así pues, existe el par común (A, A') , con $A \neq A'$, porque una involución es elíptica. No puede haber otro (B, B') , pues entonces las dos involuciones coincidirían en cuatro puntos, luego serían la misma. ■

Conviene destacar el siguiente caso particular:

Teorema 8.77 *En una recta proyectiva sobre un cuerpo euclídeo, una involución elíptica tiene un único par de puntos conjugados harmónicamente separados respecto a un par de puntos prefijado.*

Basta aplicar el teorema anterior a la involución hiperbólica que tiene por puntos fijos el par de puntos prefijados.

Si las involuciones son ambas hiperbólicas hemos de exigir un poco más:

Teorema 8.78 *En una recta proyectiva sobre un cuerpo euclídeo, dos involuciones hiperbólicas cuyos pares de puntos fijos no se separan tienen un único par común de puntos conjugados.*

DEMOSTRACIÓN: Sean I_1 e I_2 dos involuciones con puntos fijos A_1, B_1 y A_2, B_2 , respectivamente. Consideremos el sistema de referencia A_1, B_1, A_2 , de modo que los puntos fijos de I_1 tienen coordenadas $0, \infty$ y los de I_2 tienen coordenadas $1, x$. La hipótesis de que ambos pares no se separan equivale a que $x > 0$. La involución I_1 en coordenadas es $I_1(x) = -x$, y tenemos que encontrar un cierto y tal que $I_2(y) = -y$. Por el segundo apartado del teorema 8.63, esto equivale a que $\mathcal{R}(1, x, y, -y) = -1$ y, usando el teorema 8.29 la condición se reduce a $y^2 = x$, luego ciertamente $\pm\sqrt{x}$ determinan un par de puntos conjugados para ambas involuciones. La unicidad es inmediata. ■

8.8 Planos proyectivos finitos

Vamos a mostrar algunos ejemplos de planos proyectivos finitos, definidos como pares (X, \mathcal{B}) formados por un conjunto X de puntos y una familia de rectas (no vacías, por definición) que cumplen los axiomas considerados en la sección 8.1.

Los teoremas 8.1 y 8.3 afirman que todo de todo plano proyectivo puede obtenerse un plano afín sin más que eliminar los puntos de una recta y que, recíprocamente, todo plano afín se puede convertir en un plano proyectivo sin más que añadirle un punto infinito a cada recta. Esto hace que muchas propiedades de los planos proyectivos finitos puedan deducirse de las análogas de los planos afines finitos, aunque en ocasiones resulta más natural dar pruebas directas. Por ejemplo, del hecho de que las rectas de un mismo plano afín finito tienen todas el mismo número de puntos se sigue obviamente que lo mismo vale para las rectas de un plano proyectivo finito, pero una prueba directa es la siguiente:

Teorema 8.79 *Si X es un plano proyectivo finito, existe un número natural $t \geq 2$ tal que todo punto está exactamente en $t + 1$ rectas y toda recta contiene exactamente $t + 1$ puntos.*

DEMOSTRACIÓN: Vamos a usar repetidamente el hecho siguiente:

Si P es un punto y l una recta de modo que $P \notin l$, entonces el número de puntos de l es igual al número de rectas que pasan por P .

En efecto, a cada punto Q de l podemos asignarle la recta PQ , que pasa por P , y esta correspondencia es claramente biyectiva.

Sean A, B, C, D cuatro puntos no colineales tres a tres y sea $t + 1$ el número de rectas que pasan por A . Tiene que ser $t \geq 2$, pues al menos tenemos las rectas AB, AC, AD . Como A no está en la recta BC , ésta tiene exactamente $t + 1$ puntos, y a su vez, como D no está en BC , por D pasan exactamente $t + 1$ rectas. Del mismo modo razonamos que por B y C pasan exactamente $t + 1$ rectas.

Ahora, si l es una recta cualquiera, no pasará por alguno de los puntos A, B, C, D , luego por ella pasan exactamente $t + 1$ puntos, y si P es un punto cualquiera, siempre podemos tomar una recta l que no pase por P , y entonces por P pasan tantas rectas como puntos tiene l , es decir, exactamente $t + 1$ rectas. ■

Definición 8.80 El *orden* de un plano proyectivo finito es el número t dado por el teorema anterior, de modo que toda recta tiene $t + 1$ puntos y por cada punto pasan $t + 1$ rectas.

Con esta definición, al quitar una recta a un plano proyectivo de orden t obtenemos un plano afín de orden t .

Observemos ahora que todo plano proyectivo de orden t tiene $t^2 + t + 1$ puntos y otras tantas rectas.

En efecto, fijado un punto P , hay $t + 1$ rectas que pasan por P , y cada una de ellas contiene t puntos distintos de P , y estos t puntos son distintos en cada recta, por lo que hay un total de $(t + 1)t$ puntos distintos de P , luego $t^2 + t + 1$ puntos en total.

Como un plano y su dual tienen el mismo orden, el plano dual tiene $t^2 + t + 1$ puntos, pero éste es el número de rectas del plano dado. ■

Todos estos hechos son triviales para el caso de los planos proyectivos arguesianos, pues si un plano proyectivo finito P es arguesiano el teorema 8.34 nos garantiza que sus puntos y rectas pueden ponerse en correspondencia con los de un plano $P^n(k)$, donde K es un anillo de división, necesariamente finito, luego es un cuerpo finito por el teorema de Wedderburn [Al 9.4] y claramente entonces todas sus rectas tienen $|k| + 1$ puntos y por cada punto pasan $|k| + 1$ rectas. En otras palabras, el orden de P es simplemente $|k|$.

Los planos proyectivos finitos, al igual que los afines, pueden caracterizarse como sistemas de Steiner:

Teorema 8.81 *Los planos proyectivos finitos son los sistemas de Steiner de tipo $S(2, n + 1, n^2 + n + 1)$, para cierto $n \geq 2$.*

DEMOSTRACIÓN: Ciertamente, los puntos y las rectas de un plano proyectivo finito satisfacen la definición de sistema de Steiner del tipo indicado (teniendo en cuenta los resultados precedentes sobre el número de puntos y el número de puntos en cada recta).

Recíprocamente, si (X, \mathcal{B}) es un sistema de Steiner del tipo indicado, obviamente cumple el axioma P1 (por cada par de puntos pasa una única recta).

Por otro lado, precisamente por P1, las rectas que pasan por un punto P cubren todo el espacio y no se cortan más que en P . Como cada una contiene n puntos distintos de P y hay $n^2 + n$ puntos distintos de P , tiene que haber exactamente $n + 1$ rectas que pasan por P .

Si L_1 y L_2 son dos rectas distintas, podemos tomar $P \in L_1 \setminus L_2$. Como L_2 tiene $n + 1$ puntos, hay $n + 1$ rectas que pasan por P y que cortan a L_2 , pero

éstas son todas las rectas que pasan por P , es decir, que toda recta que pasa por P corta a L_2 , y en particular $L_1 \cap L_2$ es no vacío (y la intersección es un único punto por P1). Esto prueba el axioma P2'.

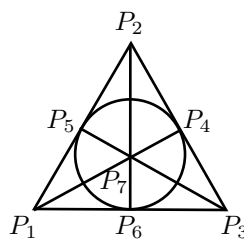
Finalmente, tomamos dos rectas L_1 y L_2 , tomamos en cada una de ellas dos puntos distintos del punto de intersección, y así tenemos cuatro puntos no colineales tres a tres, luego se cumple P3' y concluimos que (X, \mathcal{B}) es un plano proyectivo. ■

Es obvio que los isomorfismos entre planos proyectivos como sistemas de Steiner coinciden con las colineaciones entre ellos como planos proyectivos. Como en el caso afín, para cada potencia de primo q existe, salvo isomorfismo/colineación, un único plano proyectivo arguesiano de orden q , a saber, el de la forma $P^2(k)$, donde k es el cuerpo de q elementos, pero nada impide *a priori* que pueda haber otros planos no arguesianos, o incluso que pueda haber planos proyectivos no arguesianos de órdenes que no sean potencia de primo (aunque no se conoce ninguno).

El hecho de que sólo exista un plano afín de orden 2 implica que sólo existe un plano proyectivo de orden 2 (pues, dado cualquier plano proyectivo de orden 2, al quitarle una recta obtenemos el único plano afín de orden 2, y al devolverle la recta obtenemos, por una parte, el plano de partida y, por otra, la completación proyectiva del plano afín de orden 2, luego es la única posibilidad. No obstante, podemos dar una prueba directa de la unicidad que además nos da una descripción explícita del plano:

El plano de Fano El menor plano proyectivo es el de orden 2, que consta de 7 puntos y 7 rectas. Vamos a probar que es único y, según indicábamos en la introducción, recibe el nombre de *plano de Fano*.

Sea P un plano proyectivo cualquiera de orden 2. Llamemos P_1 a uno cualquiera de sus puntos y consideremos cualquier recta que no pase por él. Constará de 3 puntos, a los que podemos llamar P_2, P_3, P_4 . Cada recta que une P_1 con cada uno de estos puntos tendrá un tercer punto, con lo que tenemos tres nuevos puntos que podemos llamar P_5, P_6, P_7 . Así, el plano contiene las rectas:



$$\{P_2, P_3, P_4\}, \quad \{P_1, P_2, P_5\}, \quad \{P_1, P_3, P_6\}, \quad \{P_1, P_4, P_7\}.$$

De ellas, sólo dos pasan por P_2 , luego tiene que haber una tercera, que contendrá necesariamente los puntos $\{P_2, P_6, P_7\}$. Del mismo modo, la tercera recta que tiene que pasar por P_3 no puede ser más que $\{P_3, P_5, P_7\}$, y la tercera recta que pasa por P_4 no puede ser más que $\{P_4, P_5, P_6\}$.

Así, si partimos de dos planos proyectivos de orden 2 y numeramos sus puntos con el criterio que acabamos de indicar, la biyección que hace corresponder los puntos con la misma numeración transforma rectas en rectas, luego es una colineación entre ambos planos, luego hay un único plano proyectivo de orden 2

salvo colineación, que es naturalmente isomorfo a $P^2(K)$, donde K es el cuerpo de dos elementos. El apéndice B muestra una aplicación del plano de Fano. ■

Los teoremas 7.44 y 7.45 que prueban la unicidad de los planos afines de orden 3 y 4 respectivamente implican a su vez la unicidad de los planos proyectivos de dichos órdenes, mientras que la no existencia de planos afines de orden 6 que hemos visto tras el teorema 7.46 implica que tampoco existen planos proyectivos de dicho orden. Se sabe que los únicos planos proyectivos de órdenes 2, 3, 5, 7, 8 son los arguesianos, pero hay cuatro planos proyectivos de orden 9 no isomorfos (colineales) entre sí, de los cuales, naturalmente, tres son no arguesianos. Vamos a mostrar uno de ellos.

Un plano proyectivo no arguesiano Sea K el cuerpo de 9 elementos. Como el polinomio $x^2 + 1$ es irreducible sobre el cuerpo de 3 elementos, los elementos de K pueden expresarse en la forma $a + bi$, donde a, b son elementos del cuerpo $\{0, 1, -1\}$ de tres elementos y $i^2 = -1$. Necesitaremos algunas propiedades elementales de K :

1. Los cuadrados en K son $0, \pm 1, \pm i$.
2. $x \in K$ cumple $-x^5 - x = x$ o $-x^5 - x = 0$ según si x es o no un cuadrado.

Pues todo elemento no nulo de K cumple $x^8 = -1$, luego $x^4 + 1$ vale $2 = -1$ en los cuadrados y se anula en otro caso. Similarmente:
3. $x^5 - x = 0$ o $x^5 - x = x$ según si x es o no un cuadrado.
4. $x^2 + x^4 + x^6 = -1$ salvo si $x = 0, 1, -1$.

Pues, sumando la progresión geométrica, el miembro izquierdo es

$$\frac{x^8 - x^2}{x^2 - 1} = \frac{1 - x^2}{x^2 - 1} = -1.$$

Llamaremos $V = K$, pero considerado con el producto dado por

$$x \cdot y = x(-y^5 - y) + x^3(y^5 - y),$$

(y con la misma suma de K). Por las propiedades anteriores, esto significa que

$$x \cdot y = \begin{cases} xy & \text{si } y \text{ es un cuadrado en } K, \\ x^3y & \text{si } y \text{ no es un cuadrado en } K. \end{cases}$$

Veamos las propiedades que tiene este producto:

1. $0 \cdot x = x \cdot 0 = 0$, $1 \cdot x = x \cdot 1 = x$.
2. Si $a, b \in V$ son no nulos, las ecuaciones $x \cdot a = b$ y $a \cdot y = b$ tienen solución única (no necesariamente la misma).

En efecto, si a es un cuadrado, la ecuación $x \cdot a = b$ equivale a $xa = b$, que claramente tiene solución única, y si no es un cuadrado, la ecuación

es $x^3a = b$, que también tiene solución única porque todo elemento de K tiene una única raíz cúbica en K (porque el automorfismo de Frobenius es un automorfismo).

Ahora basta probar que si $a \cdot y_1 = a \cdot y_2$ entonces $y_1 = y_2$ pues, como V es finito, esto significa que, cuando y recorre V , el producto $a \cdot y$ recorre cada elemento de V una única vez. La conclusión es obvia si y_1, y_2 son ambos cuadrados o ambos no cuadrados. Si y_1 es un cuadrado e y_2 no lo es, tenemos que $ay_1 = a^3y_2$, luego $y_1 = a^2y_2$ y así y_2 es un cuadrado, contradicción.

3. $(x + y) \cdot z = x \cdot z + y \cdot z$.

Si z es un cuadrado es trivial, y en caso contrario

$$(x + y) \cdot z = (x + y)^3 z = x^3 z + y^3 z = x \cdot z + y \cdot z.$$

4. Si $a, b, c \in V$ con $a \neq b$, entonces $x \cdot a = x \cdot b + c$ tiene solución única en x .

Si a, b son ambos cuadrados o ambos no cuadrados, la conclusión es inmediata. Supongamos que a es un cuadrado y b no. Entonces la ecuación es $xa = x^3b + c$, que equivale a $x^3b - xa + c = 0$. De nuevo por la finitud de V , basta probar que si

$$x_1^3 b - x_1 a + c = x_2^3 b - x_2 a + c,$$

entonces $x_1 = x_2$. En efecto, en tal caso

$$x_1^3 - x_1 ab^{-1} + cb^{-1} = x_2^3 - x_2 ab^{-1} + cb^{-1},$$

luego

$$(x_1 - x_2)^3 = (x_1 - x_2) ab^{-1}.$$

Si fuera $x_1 \neq x_2$, tendríamos $(x_1 - x_2)^2 = ab^{-1}$ y b sería un cuadrado, contradicción. El caso en que a no es un cuadrado y b sí se obtiene del anterior aplicado a $x \cdot b = x \cdot a - c$.

5. Si $a, b, c, d \in V$ cumplen $a \neq c$, el sistema de ecuaciones $a \cdot x + y = b$, $c \cdot x + y = d$ tiene solución única.

Si $a = 0$, la solución es $y = b$, $c \cdot x = d - b$ (hemos visto que la última ecuación tiene solución única). Supongamos que $a \neq 0$. Entonces, para cada valor de y , hay un único valor de x que cumple la primera ecuación y viceversa, por lo que tenemos una biyección $f : V \rightarrow V$ tal que los pares $(x, f(x))$ son los únicos que cumplen la primera ecuación.

Si $x_1 \neq x_2$, se cumple que $c \cdot x_1 + f(x_1) \neq c \cdot x_2 + f(x_2)$, pues en otro caso tendríamos que

$$c \cdot x_1 + f(x_1) = c \cdot x_2 + f(x_2), \quad a \cdot x_1 + f(x_1) = a \cdot x_2 + f(x_2),$$

con lo que a y c serían dos soluciones de la ecuación

$$x \cdot x_1 = x \cdot x_2 + f(x_2) - f(x_1),$$

en contra del apartado anterior. Por lo tanto, existe un único x tal que $c \cdot x + f(x) = d$, y así $(x, f(x))$ es la única solución del sistema de ecuaciones.

6. Si $x \neq 0, 1, -1$, entonces $x \cdot x = -1$.

Por la definición del producto en V y la última propiedad que hemos probado sobre V ,

$$x \cdot x = x(-x^5 - x) + x^3(x^5 - x) = -x^6 - x^2 - x^4 + 1 = 1 + 1 = -1.$$

7. $(-1) \cdot x = -x$.

$$-1 \cdot x = (1 + 1) \cdot x = x + x = 2x = -x.$$

Ahora definimos el plano $P = V^2 \cup V \cup \{\infty\}$, donde ∞ es cualquier objeto que no esté en V y en él consideramos las rectas siguientes:

1. La recta l_∞ que contiene a ∞ y los 9 elementos de V .
2. Las 9 rectas $x = a$ formadas por ∞ y los 9 pares (a, b) .
3. Las 81 rectas $y = x \cdot m + b$ que contienen al punto m y a los 9 pares (x, y) que cumplen la ecuación indicada.

Comprobemos que P es un plano proyectivo (obviamente de orden 9). En primer lugar probamos que por dos puntos distintos pasa una única recta:

- Si los dos puntos son de la forma (a, y_1) y (a, y_2) , por ellos pasa una única recta de tipo 2), y no pasa ninguna de tipo 3), ni la recta infinita 1).
- Si los dos puntos son de la forma (x_1, y_1) y (x_2, y_2) , con $x_1 \neq x_2$, entonces por ellos no pasa ni la recta infinita 1) ni las rectas verticales 2), y sólo pasa una recta de tipo 3), pues el sistema de ecuaciones

$$x_1 \cdot m + b = y_1, \quad x_2 \cdot m + b = y_2$$

tiene solución única (m, b) .

- Si los dos puntos son (x, y) y m , la única recta que pasa por ellos es de tipo 3), con $b = y - x \cdot m$.
- Si los puntos son (a, y) y ∞ , la única recta es de tipo 2).
- Si los puntos son m y m' o bien m y ∞ , la única recta es la recta infinita de tipo 1).

Ahora comprobamos que dos rectas siempre se cortan en un punto (no es necesario probar la unicidad).

- La recta infinita corta obviamente a todas las rectas finitas.

- Dos rectas de tipo 2) se cortan en ∞ .
- Una recta de tipo 2) y otra de tipo 1) se cortan en $(a, a \cdot m + b)$.
- Consideremos ahora dos rectas de tipo 3), digamos

$$y = x \cdot m_1 + b_1, \quad y = x \cdot m_2 + b_2.$$

Si $m_1 = m_2$, entonces se cortan en el punto infinito m_1 . Si $m_1 \neq m_2$, hemos visto que la ecuación

$$x \cdot m_1 = x \cdot m_2 + b_2 - b_1$$

tiene solución única x_0 , y entonces las rectas se cortan en (x_0, y_0) , donde $y_0 = x_0 \cdot m_1 + b_1 = x_0 \cdot m_2 + b_2$.

Los otros dos axiomas son inmediatos. Por último, vamos a demostrar que el plano P no cumple el teorema de Desargues. Para ello fijamos $b \in V$ tal que $b \neq 0, 1, -1$ y consideramos los triángulos de vértices

$$A = (1, 1), \quad B = (-b + 1, b - 1), \quad C = (1, b),$$

$$A' = (-b + 1, 1), \quad B' = (-1, -b + 1), \quad C' = (-1, -b - 1).$$

Vamos a probar que tienen un centro de perspectiva, pero no un eje de perspectiva. Para ello observamos en primer lugar que las rectas que unen los vértices tienen las ecuaciones siguientes:

AA'	$y = 1$
BB'	$y = x \cdot b + 1$
CC'	$y = x \cdot (b - 1) + 1$

La única que requiere algún cálculo es BB' . El punto B cumple la ecuación indicada:

$$(-b + 1) \cdot b + 1 = (b + b + 1) \cdot b + 1 = b \cdot b + b \cdot b + b + 1 = -1 - 1 + b + 1 = b - 1,$$

y también el punto B' , pues $(-1) \cdot b + 1 = -b + 1$. Y es claro que las tres rectas se cortan en el punto $(0, 1)$, que es el centro de perspectiva del triángulo. Ahora consideramos las rectas

AB	$y = x \cdot (-b + 1) + b$	$A'B'$	$y = x \cdot (b - 1)$	$(b + 1, -b)$
BC	$y = x \cdot (-b) - b$	$B'C'$	$x = -1$	$(-1, 0)$
AC	$x = 1$	$A'C'$	$y = x \cdot (-b) + b - 1$	$(1, -1)$

En efecto, claramente A cumple la ecuación de AB y en cuanto a B , como $b \neq 0, 1, -1$, lo mismo le sucede a $-b + 1$, luego $(-b + 1) \cdot (-b + 1) + b = -1 + b$.

El punto A' cumple la ecuación de $A'B'$:

$$(-b + 1) \cdot (b - 1) = (b - 1 + b - 1) \cdot (b - 1) = (b - 1) \cdot (b - 1) + (b - 1) \cdot (b - 1) = -2 = 1.$$

El punto B cumple la ecuación de BC :

$$(-b + 1) \cdot (-b) - b = (-b) \cdot (-b) - b - b = -1 + b.$$

El punto A' cumple la ecuación de $A'C'$:

$$(-b + 1) \cdot (-b) + b - 1 = (-b) \cdot (-b) - b + b - 1 = 1.$$

Ahora comprobamos que la intersección de cada par de rectas es el punto que se indica en la última columna de la tabla. El único que no es inmediato es el primero:

$$(b + 1) \cdot (-b + 1) + b = (-b + 1 - b + 1 - 1) \cdot (-b + 1) + b =$$

$$(-b + 1) \cdot (-b + 1) + (-b + 1) \cdot (-b + 1) + b - 1 + b = -b,$$

$$(b + 1) \cdot (b - 1) = (b - 1 - 1) \cdot (b - 1) = (b - 1) \cdot (b - 1) - b + 1 = -b.$$

Ahora tenemos que probar que los tres puntos no son colineales. Pero la recta que pasa por los dos últimos es claramente $y = x + 1$, y ésta no contiene al primero. ■

A su vez, si a un plano proyectivo no arguesiano le quitamos una recta, obtenemos un plano afín no arguesiano (pues si fuera arguesiano, sería un plano afín en el sentido algebraico, luego al añadirle la recta eliminada obtendríamos un plano proyectivo en sentido algebraico, luego arguesiano).

Capítulo IX

Cónicas

En [IGE 7.1] definimos la cónica con un foco en el punto F , una directriz en la recta d y excentricidad $e > 0$ como el lugar geométrico formado por los puntos P tales que el cociente de su distancia a F sobre su distancia a d vale e . Vimos que las cónicas se dividen en tres clases según su excentricidad: las hipérbolas son las cónicas de excentricidad $e > 1$, las parábolas tienen excentricidad $e = 1$ y los elipses son las cónicas de excentricidad $0 < e < 1$, aunque resulta conveniente considerar las circunferencias como elipses de excentricidad 0, a pesar de que no cumplen la definición general.

Entre muchos otros resultados, demostramos que, eligiendo adecuadamente el sistema de referencia (ortonormal) toda elipse, parábola o hipérbola está formada por los puntos que cumplen, respectivamente, una de las ecuaciones siguientes:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad y^2 = px, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \quad (9.1)$$

Por consiguiente, una característica común a todas las cónicas es que, respecto de un sistema de referencia adecuado, están formadas por los puntos que cumplen una ecuación de la forma

$$ax^2 + by^2 + dxy + ex + fy + c = 0, \quad (9.2)$$

con algún coeficiente no nulo. Notemos que no toda curva definida por una ecuación de este tipo es una cónica, pues si, por ejemplo, $a = b = d = 0$, nos queda la ecuación de una recta, pero vamos a ver que si definimos una cónica afín sobre un cuerpo K como el conjunto de los puntos de un plano afín cuyas coordenadas respecto de cierto sistema de referencia satisfacen una ecuación de tipo (9.2), en el caso en que $K = \mathbb{R}$ obtenemos las cónicas clásicas y unos pocos casos triviales adicionales que será fácil distinguir de las “cónicas propiamente dichas”.

El paso siguiente consiste en observar que la geometría proyectiva resulta más adecuada para exponer los resultados básicos sobre cónicas, lo cual supone completar cada cónica en un sentido análogo al hecho de que cada recta afín se

completa hasta una recta proyectiva añadiéndole un punto infinito. Para ello una cónica proyectiva puede definirse como la curva formada por los puntos de un plano proyectivo que, respecto de un cierto sistema de referencia, cumplen una ecuación de la forma

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz = 0 \quad (9.3)$$

con algún coeficiente no nulo.

Así, si fijamos como recta infinita la de ecuación $z = 0$, vemos que los puntos finitos (es decir, de coordenadas afines $(x, y, 1)$) que cumplen (9.3) son exactamente los que cumplen (9.2), de modo que una cónica proyectiva no es más que una cónica afín completada con unos pocos puntos infinitos, donde, como es fácil ver haciendo $z = 0$ en (9.3), “unos pocos” son a lo sumo dos puntos infinitos.

Y sucede que estos puntos infinitos que aparecen al considerar cónicas proyectivas en lugar de cónicas afines no son “una molestia”, sino todo lo contrario, pues tienen una interpretación muy natural: en estos términos, veremos que una elipse resulta ser una cónica sin puntos infinitos, una parábola es una cónica con un punto infinito y una hipérbola es una cónica con dos puntos infinitos.

9.1 Cuádricas proyectivas

Parte de la teoría sobre cónicas proyectivas puede desarrollarse en espacios proyectivos sobre cuerpos arbitrarios (descartando unos pocos casos triviales) de dimensión arbitraria. Aunque vamos a tratar principalmente el caso de las cónicas (curvas planas), presentamos en esta sección la parte de la teoría que puede desarrollarse en espacios de dimensión arbitraria sin apenas coste alguno.

Definición de cuádrice Teniendo en cuenta que una ecuación puede tener más o menos soluciones en función del cuerpo de escalares en el que las buscamos, para evitar que ecuaciones muy distintas definan la misma cuádrice (por falta de soluciones en un cuerpo dado), vamos a introducir un tecnicismo según el cual una cuádrice no será un conjunto de puntos, sino una cierta clase de equivalencia vinculada a las ecuaciones mismas que queremos que definan cuádricas y no a sus posibles soluciones, pues incluso puede no haber ninguna en un cuerpo “demasiado pequeño”.

Definición 9.1 Dado un espacio vectorial V de dimensión finita sobre un cuerpo K de característica distinta de 2, llamaremos \mathcal{F}_V al conjunto de todas las formas bilineales simétricas ([Al, 6.41]) $F : V \times V \rightarrow K$.

Es claro que \mathcal{F}_V es un K -espacio vectorial con las operaciones dadas por

$$(F + G)(u, v) = F(u, v) + G(u, v), \quad (\alpha F)(u, v) = \alpha F(u, v).$$

Fijada una base B de V , sabemos ([Al 6.42]) que existe una única matriz simétrica $M_B(F)$ tal que si dos vectores u y v tienen coordenadas $x, y \in K^n$,

respectivamente, entonces $F(u, v) = xM_B(F)y^t$. Claramente, la aplicación $\mathcal{F}_V \rightarrow \text{Mat}_n(K)$ dada por $F \mapsto M_B(F)$ es un monomorfismo de espacios vectoriales, cuya imagen es el subespacio formado por las matrices simétricas.

Si A es la matriz de una forma F en una base B , entonces su matriz en otra base B' es CAC^t , donde C es la matriz de cambio de base y, recíprocamente, toda matriz de esta forma (que siempre es una matriz simétrica) es la matriz de F en una cierta base. Por lo tanto, dos matrices simétricas corresponden a la misma forma en bases distintas si y sólo si son congruentes en el sentido de [Al 6.47].

Llamaremos *cuádricas* en un espacio proyectivo $X = P(V)$ a los elementos del espacio proyectivo $P(\mathcal{F}_V)$, de modo que cada cuádrica está determinada por una forma bilineal no nula F , pero entendiendo que dos formas determinan la misma cuádrica si una es un múltiplo de la otra por un escalar. Las cuádricas de los planos proyectivos se llaman *cónicas*.

Si \mathcal{C} es una cuádrica en un espacio proyectivo X determinada por una forma bilineal F , definimos el conjunto de los puntos de \mathcal{C} como el conjunto $\bar{\mathcal{C}} \subset X$ formado por los puntos $P = \langle v \rangle \in X$ tales que $F(v, v) = 0$.

Claramente, esto no depende ni de la elección de v como representante de P ni de la elección de F como representante de \mathcal{C} .

En la práctica escribiremos $P \in \mathcal{C}$ en lugar de $P \in \bar{\mathcal{C}}$, pero debemos recordar que una cónica \mathcal{C} no está definida como el conjunto $\bar{\mathcal{C}}$ de sus puntos, de modo que dos cónicas distintas pueden tener en principio los mismos puntos.

Expresiones en coordenadas Si fijamos un sistema de referencia proyectivo en X , éste determina una base de V salvo múltiplos por un mismo escalar, la cual nos permite asociar a cada punto $P \in X$ un vector de coordenadas homogéneas $x \in K^{n+1}$. Si la cónica \mathcal{C} está determinada por una forma F cuya matriz en una de las bases determinadas por el sistema de referencia es A , entonces la condición $P \in \mathcal{C}$ equivale a que las coordenadas homogéneas x de P cumplan la ecuación $xAx^t = 0$.

Una vez más observamos que esto no depende de la elección de la base de V asociada al sistema de referencia, ni de la elección de la forma F , ni de la elección de la matriz A , pues un cambio en cualquiera de estas elecciones se traduce al final en que la ecuación $xAx^t = 0$ quede multiplicada por un escalar no nulo que puede simplificarse.

Ejemplo La ecuación (9.3) puede expresarse matricialmente como

$$(x, y, z) \begin{pmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

(y aquí es esencial que hemos excluido los cuerpos de característica 2) y, recíprocamente, toda ecuación de la forma $(x, y, z)A(x, y, z)^t = 0$, donde A es una

matriz simétrica no nula, es de la forma (9.3). Por lo tanto, los conjuntos de puntos de las cónicas son los conjuntos cuyas coordenadas homogéneas en un sistema de referencia prefijado satisfacen una ecuación de tipo (9.3). ■

En general, cualquier conjunto de puntos en un espacio proyectivo X cuyas coordenadas homogéneas respecto de cierto sistema de referencia estén determinadas por una ecuación de la forma $xAx^t = 0$, para una cierta matriz simétrica no nula A , es el conjunto de puntos de la cónica asociada a la forma bilineal que en una base de V asociada al sistema de referencia dado tiene matriz A , y esto hace que si tomamos otro sistema de referencia, las coordenadas homogéneas del mismo conjunto de puntos estén determinadas por otra ecuación de la forma $xBx^t = 0$, donde B es una matriz simétrica congruente con A . Por lo tanto, también podemos decir que los conjuntos de puntos de las cuádricas son los conjuntos de puntos de un espacio proyectivo que en cualquier sistema de referencia están determinados por una ecuación en coordenadas de la forma $xAx^t = 0$.

Homografías y cuádricas Sea $f : X \rightarrow Y$ una homografía entre dos espacios proyectivos y \mathcal{C} una cónica en X . Si $X = P(V)$, $Y = P(W)$, $f = H(u)$ está determinada por el isomorfismo $u : V \rightarrow W$ y \mathcal{C} está determinada por la forma bilineal F en V , definimos $f[\mathcal{C}]$ como la cónica en Y determinada por la forma bilineal simétrica dada por $G(w_1, w_2) = F(u^{-1}(w_1), u^{-1}(w_2))$.

Como siempre, la definición de $f[\mathcal{C}]$ no depende de la elección de F ni de u . Es claro además que $\overline{f[\mathcal{C}]} = f[\overline{\mathcal{C}}]$, es decir, que los puntos de $f[\mathcal{C}]$ son las imágenes por f de los puntos de \mathcal{C} . Notemos también que si A es la matriz de \mathcal{C} en un cierto sistema de referencia y M es la matriz de f en ese mismo sistema, entonces la matriz de $f[\mathcal{C}]$ es MAM^t .

Esto hace que dos ecuaciones de la forma $xAx^t = 0$ y $xBx^t = 0$, donde las matrices simétricas A y B son congruentes, pueden relacionarse de dos formas: pueden verse como ecuaciones de la misma cuádrica en sistemas de referencia distintos, o bien como ecuaciones en el mismo sistema de referencia de una cónica y su imagen por una homografía del espacio en sí mismo.

Rango y vértice de una cuádrica Sea $X = P(V)$ un espacio proyectivo de dimensión n y \mathcal{C} una cuádrica en X determinada por una forma bilineal F . Consideremos la aplicación lineal $\iota_F : V \rightarrow V^*$ dada por $\iota_F(v)(w) = F(v, w)$. La dimensión de su imagen es el rango de la matriz de F en cualquier base de V , luego éste es independiente de la elección de la base, y tampoco se altera si cambiamos F por un múltiplo no nulo.

Por consiguiente, podemos definir el *rango* de \mathcal{C} como el rango de cualquiera de sus matrices respecto de cualquier sistema de referencia de X . Éste varía entre 1 y $n + 1$, y es un invariante de \mathcal{C} , en el sentido de que las homografías de un espacio en sí mismo conservan el rango de las cuádricas. En efecto, el hecho de que el rango no dependa de la matriz elegida para F se traduce en que dos matrices congruentes tienen el mismo rango, y esto implica a su vez

la propiedad de las homografías, porque las matrices de $f[\mathcal{C}]$ recorren la misma clase de congruencia que las matrices de \mathcal{C} .

A las cuádricas de rango máximo se las llama *cuádricas no degeneradas*, mientras que las restantes son *degeneradas*.

Por otra parte, el núcleo $N(\iota_F)$ es un subespacio vectorial de V que determina una variedad proyectiva $v(\mathcal{C}) = P(N(\iota_F)) \subset X$. Como siempre, es claro que $v(\mathcal{C})$ no depende de la elección de F entre las formas que determinan la cuádrlica.

Respecto de un sistema de referencia en el que F tiene matriz A , el vértice está formado por los puntos de coordenadas x que cumplen $xA = 0$, lo cual implica $xAx^t = 0$, luego $v(\mathcal{C}) \subset \mathcal{C}$. Como

$$n + 1 = \dim V = \dim N(\iota_F) + \dim \text{Im } \iota_F = \dim v(\mathcal{C}) + 1 + \text{rang}(\mathcal{C}),$$

vemos que $\dim v(\mathcal{C}) = n - \text{rang}(\mathcal{C})$. En particular, las cónicas no degeneradas son las que tienen el vértice de dimensión -1 , es decir, vacío.

Es fácil ver que si f es una homografía entonces $v(f[\mathcal{C}]) = f[v(\mathcal{C})]$.

Complexificación de una cuádrlica Vamos a estudiar las cuádrlicas sobre cuerpos arbitrarios (de característica distinta de 2), pero hay dos casos que tienen especial interés. Uno es el correspondiente a los espacios proyectivos sobre un cuerpo ordenado euclídeo R , y el segundo el correspondiente a un cuerpo C en el que todo elemento tiene raíz cuadrada. Esto incluye el caso en que C es algebraicamente cerrado, y también el caso en que $C = R[i]$, donde R es un cuerpo euclídeo e i es una raíz cuadrada de -1 . Nos referiremos a estos casos como el *caso real* y el *caso complejo*, respectivamente.

En este apartado vamos a ver que podemos relacionar ambos casos. Concretamente, si $C = R[i]$, tenemos que $R^{n+1} \subset C^{n+1}$, lo que nos permite definir una aplicación $j : P^n(R) \rightarrow P^n(C)$ inyectiva mediante $j(\langle v \rangle) = \langle v \rangle$, donde el primer espacio vectorial es sobre R y el segundo sobre C .

Observemos ahora que toda base de R^{n+1} como R -espacio vectorial es también una base de C^{n+1} como C -espacio vectorial, pues ambas afirmaciones equivalen a que el determinante de la matriz formada por los tres vectores sea no nula. Más en general, como todo conjunto libre se extiende a una base, es claro que un conjunto de vectores de R^{n+1} es libre en R^{n+1} si y sólo si lo es en C^{n+1} .

Como consecuencia, j transforma conjuntos de puntos proyectivamente independientes en conjuntos de puntos proyectivamente independientes, y sistemas de referencia en sistemas de referencia.

Si fijamos un sistema de referencia de $P^n(R)$, todo vector de coordenadas homogéneas de un punto $P \in P^n(R)$ respecto a dicho sistema de referencia es también un vector de coordenadas homogéneas de $j(P)$ respecto de la imagen del sistema de referencia (aunque $j(P)$ tendrá otros vectores de coordenadas homogéneas con valores imaginarios).

A partir de aquí usaremos j para considerar que $P^n(R) \subset P^n(C)$. A los puntos de $P^n(R)$ los llamaremos *puntos reales* de $P^n(C)$, mientras que los restantes serán *puntos imaginarios*. A los sistemas de referencia de $P^n(R)$, que

a través de la identificación son también sistemas de referencia de $P^n(C)$, los llamaremos *sistemas de referencia reales* de $P^n(C)$. Como hemos señalado, todo vector de coordenadas homogéneas de un punto real respecto de un sistema de referencia real sigue siendo un vector de coordenadas homogéneas de dicho sistema de referencia cuando se considera como sistema de referencia de $P^n(C)$, si bien entonces admite más vectores de coordenadas homogéneas.

Si $Y = P(\langle v_0, \dots, v_d \rangle)$ es una variedad proyectiva de $P^n(R)$, vista como subconjunto de $P^n(C)$ no es ya una variedad, pero está contenida en una única variedad $\bar{X} = P(\langle v_1, \dots, v_d \rangle)$ (donde ahora $\langle v_1, \dots, v_d \rangle$ es el subespacio vectorial generado en C^{n+1} y no en R^{n+1}) a la que llamaremos la *complexificación* de X . Es única porque dos variedades de $P^n(C)$ que contengan a X tendrán al menos $d+1$ puntos proyectivamente independientes en común. A las complexificaciones de las variedades de $P^n(R)$ las llamaremos *variedades reales* de $P^n(C)$. Equivalentemente, son las variedades d -dimensionales que pasan al menos por $d+1$ puntos proyectivamente independientes reales o, alternativamente, las que, respecto de un sistema de referencia real, están determinadas por ecuaciones con coeficientes reales.

Podemos definir un monomorfismo de grupos $LGP(n, R) \rightarrow LGP(n, C)$ que a cada homografía de $P^n(R)$ le hace corresponder la homografía de $P^n(C)$ que en un sistema de referencia real está determinada por la misma matriz. Es fácil ver que dicha homografía extiende a la dada, y es su única extensión (y en particular es independiente del sistema de referencia en el que hemos considerado la matriz), pues dos extensiones de una misma homografía coincidirán sobre un sistema de referencia real, luego serán iguales. A las homografías de $P^n(C)$ que están en la imagen de este monomorfismo las llamaremos *homografías reales* de $P^n(C)$.

Por último, cada cuádrica de $P^n(R)$ determina una única cuádrica real de $P^n(C)$ que en cada sistema de referencia real tiene la misma matriz que la cuádrica dada. A esta extensión la llamaremos *complexificación* de la cuádrica dada. La complexificación contiene los mismos puntos reales que la cuádrica dada, pero contendrá además otros puntos con coordenadas imaginarias. Además su rango será el mismo. A las complexificaciones de las cuádricas de $P^n(R)$ las llamaremos *cuádricas reales* de $P^n(R)$.

Veremos que, del mismo modo que resulta útil considerar que las rectas o las cónicas tienen puntos invisibles “en el infinito”, también es conveniente considerar que tienen “puntos imaginarios” más invisibles aún, en el sentido que acabamos de precisar.

Clasificación de cuádricas Dos cuádricas en un espacio proyectivo son *equivalentes* si una se transforma en la otra mediante una homografía. En general, en un espacio proyectivo dado puede haber muchas clases de congruencia de cónicas, de modo que no sea fácil describir todas las posibilidades, pero vamos a ver que en los casos real y complejo que hemos descrito en el apartado anterior sí que es posible dar una enumeración explícita de todas las clases de equivalencia de cónicas.

En el caso complejo podemos aplicar el teorema [Al 6.50], según el cual toda matriz simétrica de rango r sobre C es congruente con una matriz diagonal cuya diagonal está formada por r unos y $n + 1 - r$ ceros (donde n es la dimensión del espacio proyectivo). Por lo tanto, existen tantas clases de equivalencia de cuádricas como rangos posibles y cada cuádrica de rango r , en el sistema de referencia adecuado, admite una ecuación de la forma $x_1^2 + \cdots + x_r^2 = 0$. Dicha ecuación recibe el nombre de *forma canónica* de la cuádrica.

Clasificación de las cónicas proyectivas complejas:

Rango	dim $v(\mathcal{C})$	Ecuación	
3	-1	$x^2 + y^2 + z^2 = 0$	no degenerada
2	0	$x^2 + y^2 = 0$	par de rectas
1	1	$x^2 = 0$	recta

Tal y como indica la tabla, la ecuación canónica de la cónica de rango 2 factoriza como $(x + iy)(x - iy) = 0$, con lo que se ve que sus puntos forman la unión de dos rectas que se cortan en el punto $(0, 0, 1)$, que es el vértice de la cónica.

Los puntos de la cónica de rango 1 son claramente los de la recta $x = 0$, que coincide con el vértice de la cónica.

Clasificación de las superficies cuádricas proyectivas complejas

Rango	dim $v(\mathcal{C})$	Ecuación	
4	-1	$x^2 + y^2 + z^2 + w^2 = 0$	no degenerada
3	0	$x^2 + y^2 + z^2 = 0$	cono
2	1	$x^2 + y^2 = 0$	par de planos
1	2	$x^2 = 0$	plano

La ecuación canónica de la superficie cuádrica no degenerada se puede expresar como

$$(x + iy)(x - iy) = -(z + iw)(z - iw),$$

por lo que, para cada $\lambda \in C$, la recta r_λ de ecuaciones

$$x + iy = \lambda(z + iw), \quad \lambda(x - iy) = -(z - iw)$$

está contenida en \mathcal{C} , al igual que la recta r_∞ dada por $x - iy = 0$, $z + iw = 0$, y es fácil ver que cada punto de \mathcal{C} está contenido en una única de estas rectas.

La superficie cuádrica de rango 3, en su forma canónica, tiene por vértice el punto $(0, 0, 0, 1)$, y es la unión de todas las rectas que pasan por dicho vértice y por la cónica no degenerada que en el plano $w = 0$ tiene ecuación $x^2 + y^2 + z^2 = 0$. Por analogía con el caso real, en el que las rectas que unen un punto con una elipse forman un cono, se dice que se trata de un cono imaginario.

La superficie cuádrica de rango 2 es claramente la unión de los planos de ecuaciones $x + yi = 0$ y $x - yi = 0$, cuya intersección es la recta $x = y = 0$, que es el vértice de la cuádrica.

Por último, la superficie cuádrica de rango 1 es claramente el plano $x = 0$, que coincide con su vértice.

Pasamos ahora al caso real, en el cual aplicamos el teorema [A1 6.51] (la ley de inercia de Sylvester), según la cual, en el sistema de referencia adecuado, la ecuación de una cuádrica es de la forma $x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_r^2 = 0$, donde r es el rango de la cónica y s la signatura de cualquiera de sus matrices. Ahora debemos tener presente que si multiplicamos la ecuación por -1 obtenemos otra ecuación para la misma cuádrica, por lo que la clase de equivalencia de una cuádrica está determinada por su rango r y por su *índice*, definido como el mínimo de r y $r - s$.

Clasificación de las cónicas proyectivas reales

Rango	Índice	$\dim v(\mathcal{C})$	Ecuación	
3	0	-1	$x^2 + y^2 + z^2 = 0$	imaginaria
3	1	-1	$x^2 + y^2 - z^2 = 0$	real
2	0	0	$x^2 + y^2 = 0$	rectas imaginarias
2	1	0	$x^2 - y^2 = 0$	rectas reales
1	0	1	$x^2 = 0$	recta

Así pues, ahora hay dos clases de cónicas no degeneradas, una que no tiene puntos y otra que sí tiene. A las cónicas del primer tipo se las llama *cónicas imaginarias*, a las del segundo tipo se las llama *cónicas reales*.

La cónica de rango 2 e índice 0 consta de un único punto, que respecto de su ecuación canónica es el $(0, 0, 1)$, y es su vértice, pero observamos que su complejificación consta del par de rectas imaginarias $x + iy = 0$, $x - iy = 0$, que se cortan en el vértice de la cónica.

La cónica de rango 2 e índice 1 es claramente la unión de las rectas reales $x - y = 0$ y $x + y = 0$, que se cortan en el punto $(0, 0, 1)$, que es su vértice.

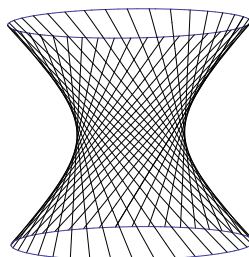
Por último, los puntos de la cónica de rango 1 son los de la recta $x = 0$, que es su vértice.

Clasificación de las superficies cuádricas reales

Rango	Índice	$\dim v(\mathcal{C})$	Ecuación	
4	0	-1	$x^2 + y^2 + z^2 + w^2 = 0$	imaginaria
4	1	-1	$x^2 + y^2 + z^2 - w^2 = 0$	elíptica
4	2	-1	$x^2 + y^2 - z^2 - w^2 = 0$	hiperbólica
3	0	0	$x^2 + y^2 + z^2 = 0$	cono imaginario
3	1	0	$x^2 + y^2 - z^2 = 0$	cono real
2	0	1	$x^2 + y^2 = 0$	planos imaginarios
2	1	1	$x^2 - y^2 = 0$	planos reales
1	0	2	$x^2 = 0$	plano

La tabla contiene la clasificación de las superficies cuádricas reales. Vemos que hay tres clases de cuádricas no degeneradas: las imaginarias, que son cuádricas complejas sin puntos reales, las elípticas y las hiperbólicas.

Si representamos gráficamente la parte finita de una cuádrica elíptica en forma canónica haciendo $w = 1$, obtenemos la esfera de radio 1. La representación de la cuádrica hiperbólica en forma canónica, es decir, de la ecuación $x^2 + y^2 - z^2 = 1$, es la que muestra la figura, en la que hay que tener presente que no estamos viendo sus puntos infinitos. En ella podemos observar que las cuádricas hiperbólicas se pueden expresar como unión de rectas. En efecto, la ecuación canónica se puede expresar como



$$(x + z)(x - z) = (w + y)(w - y),$$

luego es la unión de las rectas r_λ de ecuaciones

$$x + z = \lambda(w + y), \quad \lambda(x - z) = (w - y)$$

más la recta r_∞ de ecuaciones $w + y = 0, x - z = 0$.

La cuádrica de rango 3 e índice 1 es un cono de vértice en el punto $(0, 0, 1)$ (que es, de hecho, el vértice de la cónica en el sentido algebraico). Si el índice es 0 la ecuación sólo la cumple el vértice $(0, 0, 1)$, pero su complexificación es un cono imaginario con un único punto real.

La situación con las cuádricas de rango 2 es similar: la de índice 1 es la unión de los planos $x + y = 0$ y $x - y = 0$, que se cortan en la recta $x = y = 0$, que es precisamente el vértice de la cuádrica; la de índice 0 se reduce a dicho vértice, pero su complexificación está formada también por dos planos imaginarios, $x + iy = 0$ y $x - iy = 0$.

Por último, los puntos de la cuádrica de rango 1 son claramente los del plano $x = 0$, que es su vértice.

Intersección con una subvariedad A partir de aquí volvemos a considerar espacios proyectivos sobre un cuerpo arbitrario K de característica distinta de 2.

Como no hemos definido las cuádricas como conjuntos de puntos, el concepto de intersección de una cuádrica con una variedad proyectiva requiere una definición que va más allá de la mera intersección conjuntista.

Si $X = \mathbb{P}(V)$ es un espacio proyectivo, \mathcal{C} es una cuádrica en X determinada por una forma bilineal $F : V \times V \rightarrow K$ y $E = \mathbb{P}(W)$ es una variedad proyectiva, definimos $\mathcal{C} \cap E$ como la cuádrica en E determinada por la restricción $F|_{W \times W}$. Como siempre, es claro que la definición de $\mathcal{C} \cap E$ no depende de la elección de F , pues si dos formas son una un múltiplo de la otra, lo mismo sucede con sus restricciones.

También es inmediato comprobar que $\overline{\mathcal{C} \cap E} = \overline{\mathcal{C}} \cap E$. Por lo tanto, vemos que $\mathcal{C} \cap E$ es una cónica cuyo conjunto de puntos asociado está formado por los puntos que están en \mathcal{C} y en E .

Otra comprobación elemental muestra que si $f : X \rightarrow X$ es una homografía, entonces $f[\mathcal{C} \cap E] = f[\mathcal{C}] \cap f[E]$.

Hiperplanos tangentes Sea \mathcal{C} una cuádrica en un espacio proyectivo X y sea A su matriz respecto de un sistema de referencia dado. Sea $P \in \mathcal{C}$ un punto que, respecto de dicho sistema de referencia, tenga coordenadas homogéneas x . Sea Q un punto arbitrario con coordenadas homogéneas y . Entonces, la recta PQ está formada por los puntos de coordenadas $\lambda x + y$, con $\lambda \in K$, más el propio punto P de coordenadas x , que es el único punto de la recta que no recoge la expresión $\lambda x + y$. Que el punto correspondiente a λ esté en la cónica equivale a que se cumpla $(\lambda x + y)A(\lambda x + y)^t = 0$ o, equivalentemente,

$$2\lambda(xAy^t) + yAy^t = 0, \quad (9.4)$$

donde hemos usado que $xAx^t = 0$, porque $P \in \mathcal{C}$. Claramente, esta ecuación tiene solución única en λ salvo si $xAy^t = 0$. Esto significa que la recta PQ corta a \mathcal{C} exactamente en dos puntos (P y el correspondiente al único λ que cumple la ecuación (9.4)) salvo si $xAy^t = 0$, en cuyo caso hay dos posibilidades: si $yAy^t \neq 0$ entonces (9.4) no tiene solución, luego la recta PQ sólo corta a \mathcal{C} en el punto P , mientras que si $yAy^t = 0$ entonces todos los puntos de PQ están en la cónica \mathcal{C} .

Vamos a dar algunas definiciones que nos permitan describir la situación que acabamos de estudiar:

Definición 9.2 Una recta r en un espacio proyectivo es *exterior* a una cuádrica \mathcal{C} si $r \cap \mathcal{C} = \emptyset$; diremos que r es *secante* a \mathcal{C} si $r \cap \mathcal{C}$ consta de dos puntos, mientras que r es *tangente* a \mathcal{C} si $r \cap \mathcal{C}$ consta de un único punto (en cuyo caso diremos que la tangente es *propia*) o bien $r \subset \mathcal{C}$ (en cuyo caso diremos que la tangente es *impropia*).

Hemos probado que una recta sólo puede ser exterior, secante o tangente a una cónica \mathcal{C} y que las rectas que pasan por un punto $P \in \mathcal{C}$ de coordenadas homogéneas x respecto a un sistema de referencia dado son tangentes a \mathcal{C} si y sólo si todos sus puntos Q cumplen la ecuación $xAy^t = 0$, donde A es la matriz de \mathcal{C} e y es el vector de coordenadas de Q .

Ahora tenemos que distinguir dos casos: si $P \in v(\mathcal{C})$, entonces $xA = 0$, y resulta que todas las rectas que pasan por P son tangentes a \mathcal{C} . Por el contrario, si $P \notin v(\mathcal{C})$, entonces $xA \neq 0$, luego los puntos Q cuyas coordenadas y cumplen la ecuación $xAy^t = 0$ forman un hiperplano. Así pues:

Teorema 9.3 Sea \mathcal{C} una cuádrica en un espacio proyectivo X y sea $P \in X$.

1. Si $P \in v(\mathcal{C})$, entonces todas las rectas que pasan por P son tangentes a \mathcal{C} . En particular, si contienen otro punto de \mathcal{C} , están contenidas en \mathcal{C} .
2. Si $P \in \mathcal{C} \setminus v(\mathcal{C})$, entonces la unión de todas las tangentes a \mathcal{C} por P forma un hiperplano de X llamado hiperplano tangente a \mathcal{C} por P .
3. Si $P \in \mathcal{C} \setminus v(\mathcal{C})$ e Y es una variedad proyectiva en X tal que $P \in Y \subset \mathcal{C}$, entonces Y está contenida en el hiperplano tangente a \mathcal{C} por P .

4. $v(\mathcal{C})$ está contenido en todos los hiperplanos tangentes a \mathcal{C} .
5. Si $P, Q \in \mathcal{C} \setminus v(\mathcal{C})$ son dos puntos distintos con el mismo hiperplano tangente, entonces la recta PQ corta a $v(\mathcal{C})$, y en particular \mathcal{C} es degenerada.
6. Si $\mathcal{C} \neq \emptyset$ contiene o está contenida en un hiperplano, es degenerada.

DEMOSTRACIÓN: Ya hemos probado 1) y 2). La propiedad 3) es trivial, pues si $Q \in Y$ es distinto de P , entonces $PQ \subset Y \subset \mathcal{C}$, luego PQ es tangente a \mathcal{C} por P , luego PQ está en el hiperplano tangente, y en particular lo mismo vale para Q .

4) Sea $P \in \mathcal{C} \setminus v(\mathcal{C})$ y $Q \in v(\mathcal{C})$. Entonces $PQ \subset \mathcal{C}$ por el apartado anterior, luego es tangente a \mathcal{C} por P , luego está contenida en el hiperplano tangente en P , luego Q está en dicho hiperplano.

5) Sea A la matriz de \mathcal{C} en un sistema de referencia y sean x, y las coordenadas homogéneas de P y Q . Sabemos que las ecuaciones de los hiperplanos tangentes respectivos son $xAz^t = 0$ e $yAz^t = 0$. Si ambos hiperplanos son el mismo, existe $\lambda \in K$ tal que $xA = \lambda yA$, luego $(x - \lambda y)A = 0$, luego $x - \lambda y$ son las coordenadas de un punto de $v(\mathcal{C}) \cap PQ$.

6) Si \mathcal{C} contiene un hiperplano, éste es el hiperplano tangente a \mathcal{C} por todos sus puntos, luego \mathcal{C} es degenerada por el apartado anterior.

Supongamos que \mathcal{C} no es degenerada, sea $P \in \mathcal{C}$, sea H su hiperplano tangente en P y sea H' cualquier otro hiperplano. Podemos tomar un punto Q que no esté ni en H ni en H' . Entonces la recta PQ no es tangente y exterior, luego corta a \mathcal{C} en un punto que no está en H' , luego $\mathcal{C} \not\subset H'$. ■

Más aún, hemos probado que si $P \in \mathcal{C} \setminus v(\mathcal{C})$ y, en un sistema de referencia, P tiene coordenadas homogéneas x y \mathcal{C} tiene matriz A , el hiperplano tangente a \mathcal{C} por P está formado por los puntos cuyas coordenadas homogéneas y satisfacen la ecuación $xAy^t = 0$. Por consiguiente:

Teorema 9.4 *Si \mathcal{C} es una cuádrlica no degenerada en un espacio proyectivo X , existe una única polaridad $p_{\mathcal{C}}$ en X que a cada punto $P \in \mathcal{C}$ le hace corresponder el hiperplano tangente a \mathcal{C} por P . La llamaremos polaridad asociada a \mathcal{C} . Un punto P está en \mathcal{C} si y sólo si es autopolar, es decir, si y sólo si $P \in p_{\mathcal{C}}(P)$.*

DEMOSTRACIÓN: La polaridad $p_{\mathcal{C}}$ es simplemente la que en cada sistema de referencia tiene la misma matriz que \mathcal{C} . Todas las afirmaciones son inmediatas. ■

Vamos a usar igualmente la notación $p_{\mathcal{C}}(P)$ para referirnos al hiperplano tangente a \mathcal{C} por P cuando \mathcal{C} sea degenerada, pero P no esté en su vértice.

Teorema 9.5 *Sea \mathcal{C} una cuádrlica en un espacio proyectivo X y sea H un hiperplano.*

1. Si \mathcal{C} no es degenerada, $\mathcal{C} \cap H$ es degenerada si y sólo si H es un hiperplano tangente a \mathcal{C} .
2. Si \mathcal{C} es degenerada, $\mathcal{C} \cap H$ es degenerada si y sólo si $H \cap v(\mathcal{C}) \neq \emptyset$.

DEMOSTRACIÓN: Sea $X = P(V)$, $H = P(W)$, donde $V = W + \langle v_0 \rangle$ y supongamos que \mathcal{C} está definida por una forma bilineal F .

Basta probar que $\mathcal{C} \cap H$ es degenerada si y sólo si $H \cap v(\mathcal{C}) \neq \emptyset$ o bien H es un hiperplano tangente a \mathcal{C} , pues esto cubre los dos casos.

Si $\mathcal{C} \cap H$ es degenerada podemos tomar $P \in v(\mathcal{C} \cap H)$. Entonces $P = \langle w \rangle$, con $w \in W$ y $F(w, v) = 0$ para todo $v \in W$. Distinguiamos dos casos:

1) Si $F(w, v_0) = 0$, entonces $F(w, v) = 0$ para todo $v \in V$, luego concluimos que $P \in H \cap v(\mathcal{C})$.

2) Si $F(w, v_0) \neq 0$, entonces $P \in \mathcal{C} \setminus v(\mathcal{C})$ y

$$W \subset \{v \in V \mid F(w, v) = 0\},$$

luego $H \subset p(P)$ y, como ambos son hiperplanos, de hecho $H = p(P)$ es un hiperplano tangente a \mathcal{C} .

Si existe $P \in H \cap v(\mathcal{C})$ o bien $H = p(P)$, para cierto $P \in \mathcal{C} \setminus v(\mathcal{C})$, en ambos casos $F(w, v) = 0$ para todo $v \in W$, luego $P \in v(H \cap \mathcal{C})$, luego $H \cap \mathcal{C}$ es degenerada. ■

Esto demuestra que al cortar un cono (cuádrica degenerada) con un plano que no pase por su vértice obtenemos una cónica no degenerada.

Nota Para aplicar el teorema anterior al caso en que X es un plano proyectivo conviene observar que una cónica en una recta proyectiva es degenerada si y sólo si consta de un punto, mientras que si es no degenerada puede constar de dos puntos o bien ser vacía. Esto se sigue de [Al 6.49], que permite reducir la ecuación a $y^2 = ax^2$. ■

Cuádricas regladas Una cuádrica se dice *reglada* si contiene rectas. Por ejemplo, toda cuádrica cuyo vértice tenga al menos dimensión 1 es reglada, pues el vértice contiene rectas. Si el vértice es un punto, entonces la cuádrica será reglada si y sólo si contiene algún otro punto, ya que entonces la recta que lo une con el vértice es tangente, luego está en la cuádrica.

Vamos a probar que, cuando el cuerpo de escalares es euclídeo, una cuádrica no degenerada es reglada si y sólo si su índice es mayor que 1.

En efecto, si la cuádrica tiene índice 0 entonces no contiene puntos reales, luego no es reglada. Si tiene índice mayor que 1, entonces su forma canónica es de la forma

$$x_0^2 - x_1^2 + x_2^2 - x_3^2 + \cdots = 0,$$

donde los puntos suspensivos representan sumas o restas de cuadrados de otras variables. Es claro entonces que la cuádrica contiene la recta de ecuaciones

$$x_0 = x_1, \quad x_2 = x_3, \quad x_4 = \cdots = x_n = 0.$$

Consideremos por último el caso de una cuádrica no degenerada de índice 1, cuya ecuación canónica será de la forma

$$x_1^2 + \cdots + x_n^2 = x_0^2.$$

Si \mathcal{C} contuviera una recta PQ , sean x e y sus coordenadas homogéneas. Entonces

$$x_1^2 + \cdots + x_n^2 = x_0^2, \quad y_1^2 + \cdots + y_n^2 = y_0^2, \quad x_1 y_1 + \cdots + x_n y_n = x_0 y_0,$$

donde las dos primeras ecuaciones expresan que $P, Q \in \mathcal{C}$ y la tercera que $Q \in pe(P)$. Ahora bien, en términos de la estructura euclídea de R^n , esto significa que $|\bar{x}\bar{y}| = \|\bar{x}\|\|\bar{y}\|$ (donde la barra indica que eliminamos la coordenada de índice 0), luego por el teorema de Cauchy-Schwarz 3.22 sabemos que $\bar{y} = \lambda\bar{x}$, para cierto $\lambda \in R$ no nulo, luego $x_0 y_0 = \lambda x_0^2$, con $x_0 \neq 0$ (o sería $x = 0$), luego $y_0 = \lambda x_0$, luego $y = \lambda x$, pero esto es imposible, porque x e y son las coordenadas homogéneas de dos puntos distintos. ■

Así pues, una diferencia puramente geométrica entre las superficies cuádricas elípticas y las hiperbólicas es que las primeras no son regladas y las segundas sí.

9.2 Cónicas proyectivas

A partir de aquí particularizamos el estudio de las cuádricas al caso de las cónicas en un plano proyectivo (siempre sobre cuerpos de característica distinta de 2). Un resultado básico es el siguiente:

Teorema 9.6 *Por cinco puntos no alineados tres a tres pasa una única cónica.*

DEMOSTRACIÓN: Podemos tomar a cuatro de los puntos como sistema de referencia, de modo que sus coordenadas serán $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ y $(1, 1, 1)$. Sea M la matriz en este sistema de una cónica que pase por los cinco puntos. Para que pase por los tres primeros ha de ser de la forma

$$M = \begin{pmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{pmatrix}.$$

Para que pase por $(1, 1, 1)$ ha de cumplirse la ecuación $a + b + c = 0$. Ninguno de los tres coeficientes puede ser 0. Por ejemplo, si $b = 0$ la ecuación de la cónica sería $2axy + 2cyz = 0$, es decir, $(ax + cz)y = 0$, luego la cónica sería la unión de las rectas $ax + cz = 0$ e $y = 0$, y en una unión de dos rectas no puede haber cinco puntos no colineales tres a tres.

Como los múltiplos no nulos de A definen la misma cónica, no perdemos generalidad si suponemos que $b = 1$. Por lo tanto la matriz tiene que ser de la forma

$$M = \begin{pmatrix} 0 & a & 1 \\ a & 0 & -1 - a \\ 1 & -1 - a & 0 \end{pmatrix}, \quad \text{con } a \neq 0, -1.$$

El quinto punto tendrá coordenadas (x, y, z) . La cónica pasará por él si y sólo si

$$a = \frac{yz - xz}{xy - yz}.$$

No puede ocurrir que $xy - yz = 0$, pues esto implica que el quinto punto está en la recta $y = 0$ o en la recta $x - z = 0$, luego estaría alineado con $(0, 0, 1)$ y $(1, 0, 1)$ o bien con $(0, 1, 0)$ y $(1, 1, 1)$. Tampoco puede ocurrir que el valor de a así determinado sea 0 o -1 , pues sabemos que en tal caso la cónica no podría contener cinco puntos no colineales dos a dos, luego el punto (x, y, z) sería colineal con dos de los puntos restantes. ■

Observación El determinante de la matriz M considerada en la prueba del teorema anterior es $|M| = -2a(a - 1) \neq 0$. Así pues, la matriz de una cónica que pasa por cinco puntos no colineales tres a tres es siempre no degenerada. ■

Definición 9.7 Diremos que una cónica en un plano proyectivo es *no trivial* si pasa por cinco puntos no colineales tres a tres.

Si el cuerpo de escalares es euclídeo, sabemos que todas las cónicas no degeneradas reales son equivalentes, y como hemos probado que algunas de ellas pasan por cinco puntos no colineales tres a tres, lo mismo vale para todas. Por consiguiente, ser no trivial equivale a ser no degenerada y real (ya que las otras cónicas no degeneradas son la imaginarias, y son triviales porque no tienen puntos).

Similarmente, en el caso complejo ser no trivial equivale a ser no degenerada, por el mismo argumento. Esto hace que ser “no trivial” sea la generalización más conveniente al trabajar con cuerpos arbitrarios de la hipótesis de “ser no degenerada” en el caso complejo y la de “ser no degenerada y real” en el caso real.

Observación Hay un caso peculiar en el que no pueden existir cónicas no triviales, y es el del plano proyectivo sobre el cuerpo de tres elementos. En efecto, dados cuatro puntos no colineales tres a tres, al tomarlos como sistema de referencia determinan seis rectas, de ecuaciones

$$x = 0, \quad y = 0, \quad z = 0, \quad x - y = 0, \quad x - z = 0, \quad y - z = 0,$$

luego un quinto punto no colineal con dos cualesquiera de ellos tiene que tener coordenadas (x, y, z) distintas de 0 y distintas entre sí, lo cual es imposible si y sólo si el cuerpo consta únicamente de los elementos $0, 1, -1$. En este caso, la prueba del teorema 9.6 muestra que por cuatro puntos no colineales tres a tres pasa una única cónica no degenerada (pues el único valor posible para a es $a = 1$). ■

Los resultados de la sección anterior implican que una cónica no degenerada no puede contener rectas, pero se cumple algo más fuerte:

Teorema 9.8 *Una cónica no trivial no contiene tres puntos colineales.*

DEMOSTRACIÓN: Si una cónica contiene tres puntos colineales, la recta que los une tiene que ser tangente, luego tiene que estar contenida en la cónica, lo cual es imposible, pues las cuádricas no degeneradas no contienen hiperplanos. ■

En particular las cónicas no triviales no tienen tangentes impropias: toda recta es exterior a una cónica dada (es decir, no la corta), es secante (la corta en dos puntos) o bien es tangente (la corta en un único punto). Además, por cada punto de una cónica no trivial pasa una única recta tangente (ya que en este caso el hiperplano tangente es simplemente una recta). Las demás rectas son secantes.

Consideremos ahora una cónica \mathcal{C} y sea p su polaridad asociada. Por un punto P que no esté en \mathcal{C} pueden pasar dos tangentes a \mathcal{C} o bien ninguna, pero no hay más posibilidades. En efecto, por P no pueden pasar tres tangentes, porque si A, B, C son los puntos de tangencia, tendríamos que $P \in p(A), p(B), p(C)$, de donde $A, B, C \in p(P)$, luego \mathcal{C} contendría tres puntos colineales.

Por otra parte, si por P pasa una tangente y A es el punto de tangencia, entonces $P \in p(A)$, luego $A \in p(P)$, pero $P \notin p(P)$, porque $P \notin \mathcal{C}$, luego $p(P) \neq p(A)$ y $p(A)$ es la tangente a \mathcal{C} por A . Como $p(P)$ pasa por A y no es la tangente, tiene que ser secante, luego existe un punto $B \in \mathcal{C}$ distinto de A tal que $B \in p(P)$, luego $P \in p(B)$, luego P pasa también por la tangente a \mathcal{C} en B .

Por consiguiente, cada cónica clasifica los puntos del plano en tres tipos:

Definición 9.9 Si \mathcal{C} es una cónica en un plano proyectivo, diremos que un punto P es *exterior* a \mathcal{C} si por él pasan dos tangentes a \mathcal{C} , mientras que P es *interior* si por él no pasan tangentes a \mathcal{C} . Sabemos que los puntos por los que pasa una única tangente a \mathcal{C} (los que no son ni interiores ni exteriores) son precisamente los puntos de \mathcal{C} .

Teorema 9.10 *Sea \mathcal{C} una cónica en un plano proyectivo, sea p su polaridad asociada y sea P un punto que no esté en \mathcal{C} .*

1. *Si P es exterior, entonces por P pasan dos tangentes a \mathcal{C} y $p(P)$ es la recta que pasa por los puntos de tangencia.*
2. *El polo de una recta secante a \mathcal{C} es la intersección de las tangentes en los puntos de corte.*
3. *El punto P es exterior a \mathcal{C} si y sólo si su polar es secante a \mathcal{C} , y es interior si y sólo si su polar es exterior a \mathcal{C} .*
4. *Todos los puntos de una recta exterior a \mathcal{C} son exteriores a \mathcal{C} .*
5. *Todos los puntos de una recta tangente a \mathcal{C} excepto el de tangencia son exteriores a \mathcal{C} .*
6. *Toda recta contiene puntos exteriores a \mathcal{C} .*

DEMOSTRACIÓN: 1) Si A y B son los puntos de tangencia, entonces tenemos que $P \in p(A) \cap p(B)$, luego $A, B \in p(P)$, luego $p(P) = AB$.

2) Si r corta a \mathcal{C} en A y B , entonces $p(r) \in p(A) \cap p(B)$.

3) Si P es exterior, su polar es secante por 1). Si P es interior, entonces, para todo $Q \in \mathcal{C}$, tenemos que $P \notin p(Q)$, luego $Q \notin p(P)$, luego $p(P)$ es exterior a \mathcal{C} .

4) Si r es una recta exterior y $P \in r$, entonces $p(r) \in p(P)$ y $p(r)$ es un punto interior, luego la recta $p(P)$ es secante, luego P es un punto exterior.

5) Si r es una recta tangente y $P \in r$ no es el punto de tangencia $p(r)$, entonces $p(r) \in p(P)$ y $p(P) \neq r$, es decir, $p(P)$ es una recta que pasa por $p(r)$ distinta de la recta tangente r , luego $p(P)$ es secante, luego P es exterior.

6) Si todos los puntos de una recta r fueran interiores a \mathcal{C} o estuvieran en \mathcal{C} , entonces todas las rectas que pasan por $p(r)$ serían exteriores o tangentes a \mathcal{C} , pero esto es absurdo, porque tangentes sólo puede haber dos y, uniendo $p(r)$ con un punto de \mathcal{C} distinto de los dos posibles puntos de tangencia obtenemos una recta secante a \mathcal{C} que pasa por P . De hecho, como toda cónica contiene al menos cinco puntos, toda recta contiene al menos tres puntos exteriores. ■

El teorema anterior no nos da ninguna interpretación geométrica para la recta polar de un punto interior. Ésta la obtendremos del teorema siguiente:

Teorema 9.11 *Si un cuadrilátero completo está inscrito en una cónica (es decir, si sus vértices están sobre ella) entonces la recta polar de un punto diagonal es la recta que pasa por los otros dos puntos diagonales.*

DEMOSTRACIÓN: Sean A, B, C, D los cuatro puntos del cuadrilátero completo. Podemos tomarlos como sistema de referencia, de modo que sus coordenadas sean $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ y $(1, 1, 1)$, respectivamente. Según vimos tras la definición 8.57, las coordenadas de los puntos diagonales son $(0, 1, 1)$, $(1, 0, 1)$ y $(1, 1, 0)$.

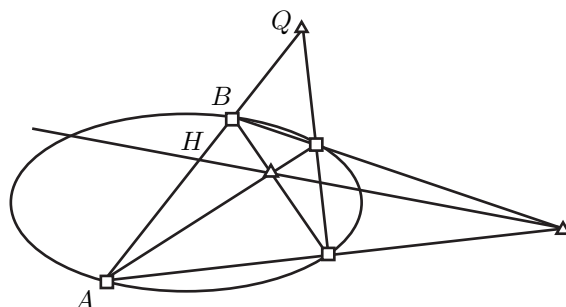
En la prueba del teorema 9.6 hemos visto que la matriz de la cónica tiene que ser de la forma

$$M = \begin{pmatrix} 0 & a & 1 \\ a & 0 & -1-a \\ 1 & -1-a & 0 \end{pmatrix}, \quad \text{con } a \neq 0, -1.$$

La recta polar del punto diagonal $(0, 1, 1)$ es entonces

$$(a+1)X - (a+1)Y - (a+1)Z = 0,$$

y vemos que, en efecto, pasa por los puntos $(1, 0, 1)$ y $(1, 1, 0)$. ■

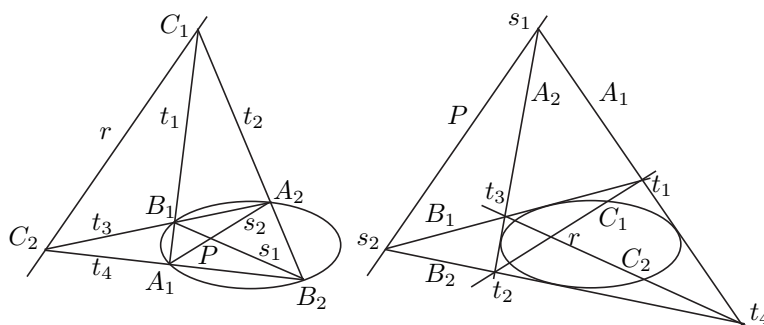


La figura ilustra el teorema, y si observamos el punto H , la propia definición de separación armónica nos da el teorema siguiente:

Teorema 9.12 *Sea \mathcal{C} una cónica, Q un punto y consideremos una recta que pase por Q y que corte a \mathcal{C} en dos puntos A y B . Entonces la recta polar de Q contiene al conjugado armónico de Q respecto de A y B .*

Nota Si la cónica es una circunferencia y AB es uno de sus diámetros, tomando coordenadas afines en AB de modo que el centro de la circunferencia sea 0 y A y B tengan coordenadas $\pm r$, un simple cálculo muestra que el conjugado armónico de un punto $Q = x$ es el punto $H = r^2/x$, que es el inverso de Q respecto de la circunferencia en el sentido de [IGE 5.3]. Teniendo en cuenta además que el polo del diámetro AB es el punto infinito común a todas sus perpendiculares (es la intersección de las tangentes a la circunferencia por A y B) concluimos que la recta polar de Q , que pasa por dicho polo, es perpendicular a AB , luego se trata de la misma recta polar definida en [IGE 5.25]. ■

Veamos cómo construir la recta polar de un punto arbitrario P que no esté sobre una cónica.



1. Tomamos dos rectas s_1 y s_2 que pasen por P y sean secantes a la cónica (si P es interior sirven dos rectas cualesquiera y si es exterior sólo hay que descartar las dos tangentes).
2. Los puntos de corte A_1, A_2, B_2, B_1 determinan un cuadrilátero completo inscrito en la cónica del cual P es un punto diagonal.
3. Calculamos los otros dos puntos diagonales C_1 y C_2 , y entonces $r = p(P)$ es la recta C_1C_2 .

La figura de la izquierda muestra el caso de un punto interior, que es el más interesante, porque si es exterior es más fácil considerar las dos tangentes a la cónica por P y unir los puntos de tangencia. No obstante, el método vale igualmente si el punto es exterior. La figura de la derecha muestra cómo calcular geoméricamente el polo de una recta. La notación está elegida para poner de manifiesto que la construcción es dual de la anterior:

1. Si antes partíamos de un punto P , ahora partimos de una recta P que no sea tangente a la cónica y cuyo polo $P' = p(P)$ queremos calcular.
2. Si antes tomábamos dos rectas s_1 y s_2 que pasaran por P y fueran secantes a la cónica, ahora tomamos dos puntos s_1 y s_2 que estén en P y sean exteriores a la cónica.
3. Llamamos A_1, A_2, B_1, B_2 a las rectas tangentes a la cónica que pasan por s_1 y s_2 . Sus polos A'_1, A'_2, B'_1, B'_2 son puntos de la cónica distintos dos a dos, pues en caso contrario la recta P sería tangente a la cónica.

Tenemos que $A'_1A'_2 = s'_1$ es la polar de s_1 y $B'_1B'_2 = s'_2$ es la polar de s_2 y $P' \in s'_1 \cap s'_2$. En otras palabras, A'_1, A'_2, B'_1, B'_2 forman un cuadrilátero completo y P' es uno de sus puntos diagonales. Por el teorema anterior, la recta P contiene a los otros dos.

4. Antes calculábamos C_1 como la intersección de las rectas $t_1 = A_1B_1$ y $t_2 = A_2B_2$. Ahora llamamos $t_1 = A_1 \cap B_1$, $t_2 = A_2 \cap B_2$ y calculamos $C_1 = t_1t_2$. Así las rectas polares $t'_1 = A'_1A'_2$ y $t'_2 = A'_2B'_2$, que son lados del cuadrilátero, se cortan en $C'_1 = p(C_1)$, que es un punto diagonal. Similarmenete obtenemos la recta C_2 cuyo polo C'_2 es el tercer punto diagonal. Sabemos que $P = C'_1C'_2$.
5. Antes calculábamos $r = C_1C_2$. Ahora calculamos $r = C_1 \cap C_2$, cuya recta polar $r' = C'_1C'_2$ es P , luego r es el polo de P .

Cónicas sobre cuerpos ordenados Los resultados que hemos obtenido sobre puntos interiores y exteriores tienen algunas lagunas, porque no hemos probado ninguno que implique la existencia de puntos interiores a una cónica. Es lógico que así sea pues, por ejemplo, en un plano proyectivo sobre un cuerpo algebraicamente cerrado no hay rectas exteriores a ninguna cónica, por lo que ninguna cónica tiene puntos interiores. Para obtener resultados positivos sobre puntos interiores necesitamos suponer que el cuerpo de escalares está ordenado:

Nota *Hasta el final de la sección trabajamos en un plano proyectivo sobre un cuerpo ordenado euclídeo R .*

Observemos que si r es una recta que no es tangente a una cónica \mathcal{C} y $P \in r$, entonces $p(P) \neq r$, pues si en caso contrario $P \in p(P)$, luego $P \in \mathcal{C}$, luego $r = p(P)$ sería tangente a \mathcal{C} . Esto nos permite aplicar el teorema 8.45, según el cual $I_r = p|_r$ es una polaridad simétrica en r , es decir, una involución. Concretamente, $I_r(P) = p(P) \cap r$, y así $I(P) = P$ si y sólo si $P \in p(P)$, si y sólo si $P \in \mathcal{C} \cap r$. En definitiva, I_r es elíptica si la recta r es exterior a \mathcal{C} y es hiperbólica si r es secante.

Ahora ya podemos probar que las rectas secantes pasan por puntos interiores:

Teorema 9.13 *Si una recta r corta a una cónica \mathcal{C} en dos puntos A y B , entonces, de los dos segmentos en que A y B dividen a r , uno está formado por puntos interiores y otro por puntos exteriores (sin contar los extremos).*

DEMOSTRACIÓN: Sea $P = p(A) \cap p(B)$, es decir, el punto donde se cortan las tangentes a \mathcal{C} por A y B (notemos que $P \notin r$ y que P es exterior a \mathcal{C}). Si identificamos cada punto con sus coordenadas homogéneas respecto a un sistema de referencia, entonces los puntos de r distintos de A son de la forma $Q = \lambda A + B$, con $\lambda \in R$, y un punto de la recta PQ distinto de P es de la forma $X = \mu P + \lambda A + B$. Dicho punto estará en \mathcal{C} si y sólo si $XM X^t = 0$, donde M es la matriz de \mathcal{C} en el sistema de referencia fijado. Teniendo en cuenta que $AMA^t = BMB^t = AMP^t = BMP^t = 0$, la condición $XM X^t = 0$ se reduce a

$$PMP^t \mu^2 + 2\lambda AMB^t = 0,$$

donde $PMP^t \neq 0$, pues $P \notin \mathcal{C}$ y $AMB^t \neq 0$ porque B no está en la tangente a \mathcal{C} por A . Así pues, $X \in \mathcal{C}$ si y sólo si

$$\mu^2 = -\frac{2\lambda AMB^t}{PMP^t}.$$

Esta ecuación tiene solución en μ en función del signo de λ (si y sólo si $\lambda \geq 0$ o bien si y sólo si $\lambda \leq 0$), y si $\lambda \neq 0$ tiene dos soluciones distintas. Esto significa que la recta PQ es secante o exterior según si $\lambda > 0$ o $\lambda < 0$.

Ahora bien, es claro que el signo de λ determina si Q pertenece a uno u otro de los dos segmentos S_+ y S_- en que los puntos A y B dividen a r . Podemos suponer, pues, que PQ es secante si y sólo si $Q \in S_+$ y que es exterior si y sólo si $Q \in S_-$ (para $Q \neq A, B$).

Consideremos ahora la involución I_r inducida por la cónica. Si $C \in r \setminus \{A, B\}$ y $Q = I_r(C)$, entonces $Q \in p(C)$, y también $P \in p(C)$, porque $C \in r = P(P)$, luego $p(C) = PQ$. Por lo tanto, $p(C)$ es secante o exterior según si $I_r(C) \in S_+$ o $I_r(C) \in S_-$. Equivalentemente, concluimos que C es exterior o interior según si $C \in I_r[S_+] = S_-$ o $C \in I_r[S_-] = S_+$. ■

Como consecuencia inmediata:

Teorema 9.14 *Por todo punto exterior a una cónica pasa al menos una recta exterior. Un punto es interior si y sólo si todas las rectas que pasan por él son secantes.*

DEMOSTRACIÓN: Si P es un punto exterior, entonces $p(P)$ es una recta secante, luego contiene puntos interiores Q , luego $p(Q)$ es una recta exterior que pasa por P .

Si P es un punto interior, por definición no pasan tangentes por P , ni tampoco rectas exteriores, ya que todos sus puntos son exteriores, luego todas las rectas que pasan por P son secantes. Recíprocamente, si todas las rectas que pasan por P son secantes, el punto es interior, ya que no puede ser exterior por la afirmación ya probada, ni tampoco estar en la cónica, ya que entonces por él pasaría una tangente. ■

9.3 Cónicas afines

Recordemos que una cónica afín puede definirse como un conjunto de puntos en un plano afín E que en un sistema de coordenadas satisface una ecuación de tipo (9.2), lo cual, si identificamos E con un plano proyectivo en el que hemos seleccionado una recta infinita E_∞ , equivale a que sea el conjunto de puntos finitos de una cónica proyectiva. Sin embargo, es más práctico definir las cónicas afines como las cónicas (proyectivas) de un espacio afín (proyectivo) E (en el sentido de la definición 8.32). En estos términos, una ecuación (9.2) es simplemente la ecuación que satisfacen las coordenadas afines de los puntos finitos de la cónica definida por la ecuación (9.3) respecto a un cierto sistema de referencia.

En la práctica, el paso de la ecuación afín a la ecuación proyectiva de una cónica se obtiene añadiendo una variable z para que todos los monomios de la ecuación tengan grado 2, pues es inmediato que los puntos finitos de la cónica determinada por la ecuación resultante son precisamente aquellos cuyas coordenadas afines cumplen la ecuación original.

Puesto que la recta infinita sólo puede cortar a la compleción proyectiva en 0, 1 o 2 puntos, obtenemos finalmente la división de las cónicas en elipses, parábolas e hipérbolas:

Definición 9.15 Diremos que una cónica afín no degenerada es una *elipse*, una *parábola* o una *hipérbola* según si tiene 0, 1 o 2 puntos infinitos.

En el caso real esto incluye a las cónicas imaginarias, que, como no tienen puntos, ni finitos ni infinitos, son técnicamente elipses. En el caso complejo una cónica afín puede ser una parábola o una hipérbola, pero nunca una elipse, pues los puntos infinitos de una cónica se obtienen resolviendo una ecuación de segundo grado, y todas las ecuaciones de segundo grado en C tienen al menos una raíz.

Ejemplo Consideremos la cónica de ecuación

$$x^2 + 3y^2 - 4xy + 2x + 4y + 4 = 0.$$

En coordenadas homogéneas es

$$x^2 + 3y^2 - 4xy + 2xz + 4yz + 4z^2 = 0.$$

Calculando el determinante de la matriz asociada vemos que no es degenerada. Haciendo $z = 0$ obtenemos la ecuación de sus puntos infinitos: $x^2 - 4xy + 3y^2 = 0$. Como $y = 0$ es claramente imposible podemos hacer $y = 1$, con lo que obtenemos $x^2 - 4x + 3 = 0$, cuyas soluciones son $x = 1, 3$. Así pues, la cónica pasa por los puntos infinitos $(1, 1, 0)$ y $(3, 1, 0)$ y es, por lo tanto, una hipérbola.

En cambio, si tomamos como recta infinita la de ecuación $x = 0$, los puntos finitos de la cónica pasan a ser los que cumplen la ecuación

$$1 + 3y^2 - 4y + 2z + 4yz + 4z^2 = 0,$$

y los puntos infinitos los que cumplen $3y^2 + 4yz + 4z^2 = 0$ y, haciendo $y = 1$ obtenemos la ecuación $3 + 4z + 4z^2 = 0$, que no tiene raíces reales, por lo que, desde este punto de vista alternativo, la cónica es una elipse. ■

Observemos que las elipses, parábolas e hipérbolas en el sentido euclídeo de [IGE] también lo son en este sentido. En efecto, las ecuaciones (9.1) se corresponden con las de las cónicas proyectivas

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - z^2 = 0, \quad y^2 = 4pxz, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} - z^2 = 0,$$

cuyos puntos infinitos cumplen, respectivamente, las ecuaciones

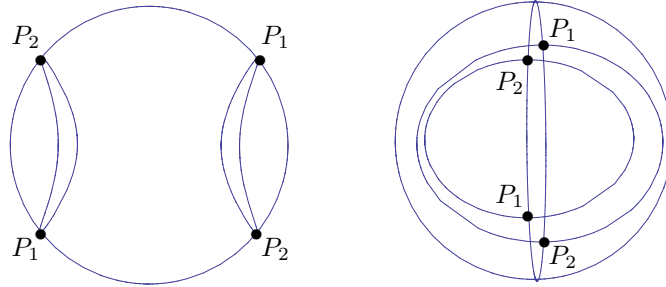
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0, \quad y^2 = 0, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 0.$$

La primera no tiene soluciones no nulas, luego las elipses euclídeas no tienen puntos infinitos. La segunda la cumple únicamente el punto $(1, 0, 0)$, luego, en efecto, las parábolas euclídeas tienen un único punto infinito. Por último, la tercera ecuación no tiene soluciones con $y = 0$, luego no perdemos generalidad si suponemos $y = b$, lo que nos da los dos puntos $(\pm a, b, 0)$ y así las hipérbolas euclídeas tienen dos puntos en el infinito.

Así pues, desde el punto de vista de la geometría proyectiva, la diferencia entre una elipse, una parábola y una hipérbola es simplemente una diferencia de punto de vista: cuando vemos una elipse, estamos viendo “todos sus puntos”, puesto que todos ellos son finitos; cuando vemos una parábola estamos viendo todos sus puntos menos uno, y por eso, en vez de verla como una curva cerrada, la vemos abierta, porque sus dos “brazos” se unen en el infinito; y cuando vemos una hipérbola la vemos “partida en dos ramas” porque dos de sus puntos quedan en el infinito, de modo que cada “brazo” de una rama se une en el infinito con un “brazo” de la rama opuesta.

Existe una forma de visualizar el plano proyectivo real completo (incluyendo los puntos infinitos) y consiste en observar que todo punto del plano proyectivo admite unas coordenadas homogéneas (x, y, z) tales que $x^2 + y^2 + z^2 = 1$, de modo que podemos identificarlo con un punto de la esfera de centro 0 y radio 1 de un espacio tridimensional euclídeo, con la única salvedad de que cada par de puntos antípodos (x, y, z) y $(-x, -y, -z)$ corresponden a un mismo punto del plano proyectivo.

Por ejemplo, la figura de la izquierda muestra una esfera vista desde arriba, de modo que el círculo que rodea la figura es el ecuador $z = 0$, es decir, la recta infinita. En ella vemos una cónica que aparece duplicada, porque, como hemos indicado, en esta representación cada punto aparece dos veces. Vemos que la cónica tiene dos puntos infinitos, P_1 y P_2 . Por lo tanto, respecto a dicha elección de la recta infinita, la cónica es una hipérbola, con dos ramas que unen P_1 y P_2 . La rama que está en el hemisferio Norte a la derecha es la misma que la del hemisferio Sur a la izquierda, y viceversa.



La figura de la derecha resulta de girar la esfera casi 90 grados, de modo que la circunferencia $z = 0$ se ve casi de perfil y ahora la circunferencia que rodea la figura es la de ecuación $x = 0$, que en la figura de la izquierda correspondería al par de meridianos verticales. Ahora una de las dos copias de la cónica queda completamente en el hemisferio norte, y la copia formada por sus puntos antípodas está en el hemisferio sur, de modo que si consideramos como recta infinita $x = 0$, la misma cónica se ve como una elipse, sin ningún punto en el infinito. ■

Ejemplo Vamos a calcular la recta tangente a la elipse

$$\frac{x^2}{3} + y^2 = 1$$

por el punto $(x, y) = (\sqrt{8/3}, 1)$.

Para ello consideramos sus coordenadas homogéneas $(\sqrt{8/3}, 1, 1)$ y las multiplicamos por la matriz de la elipse:

$$(\sqrt{8/3}, 1, 1) \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = (\sqrt{8/27}, 1/3, -1).$$

La ecuación de la recta tangente es

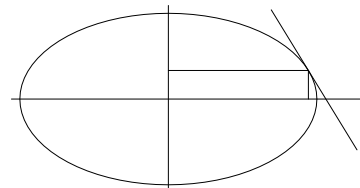
$$\sqrt{8/27}x + (1/3)y - z = 0,$$

que en coordenadas afines es

$$\sqrt{8/27}x + (1/3)y - 1 = 0$$

o, equivalentemente,

$$y = 3 - \sqrt{8/3}x. \quad \blacksquare$$



Definición 9.16 Las tangentes a una hipérbola por sus puntos infinitos se llaman *asíntotas* (gr. ‘que no coincide’). El *centro* de una cónica es el polo de la recta infinita.

Así, el centro de una cónica es un punto infinito si y sólo si pertenece a su propia recta polar, lo cual equivale a que esté en la cónica y a que la recta infinita sea tangente, es decir, a que la cónica sea una parábola. Las elipses y las hipérbolas, por el contrario tienen el centro finito. En el caso de una hipérbola, como la recta infinita es secante, su polo es el punto en el que se cortan las tangentes por los puntos de corte, es decir, el centro de una hipérbola es el punto donde se cortan sus asíntotas.

Ahora sabemos, por ejemplo, que cualquier biyección afín transforma el centro de una cónica en el centro de la imagen. También es inmediato que (en el caso real) el centro de una elipse es un punto interior, mientras que el de una hipérbola es exterior.

Ejemplo Consideremos de nuevo la hipérbola de ecuación

$$x^2 + 3y^2 - 4xy + 2x + 4y + 4 = 0.$$

Ya hemos visto que sus puntos infinitos son $(1, 1, 0)$ y $(3, 1, 0)$. Su centro es el punto $(x, y, 1)$ cuya recta polar tiene coordenadas $(0, 0, a)$, es decir

$$(x, y, z) \begin{pmatrix} 1 & -2 & 1 \\ -2 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix} = (0, 0, a),$$

lo que se traduce en las ecuaciones

$$\left. \begin{aligned} x - 2y + z &= 0 \\ -2x + 3y + 2z &= 0 \end{aligned} \right\}$$

La solución es $(7, 4, 1)$. En coordenadas cartesianas el centro es $(7, 4)$.

Las asíntotas son las polares de los puntos infinitos:

$$(1, 1, 0) \begin{pmatrix} 1 & -2 & 1 \\ -2 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

$$(3, 1, 0) \begin{pmatrix} 1 & -2 & 1 \\ -2 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

Al desarrollar obtenemos $-x + y + 3z = 0$ y $x - 3y + 5z = 0$. Las ecuaciones cartesianas son, pues,

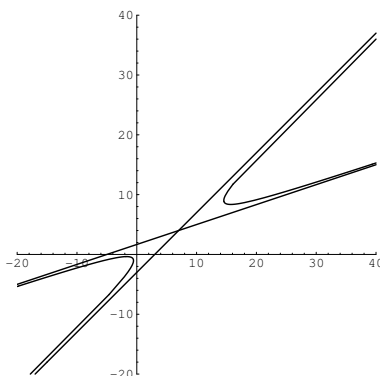
$$y = x - 3, \quad y = \frac{1}{3}x + \frac{5}{3}.$$

Para representar gráficamente la hipérbola conviene parametrizarla del modo siguiente: Tomamos uno cualquiera de sus puntos, por ejemplo $(1, 1, 0)$ y consideramos el haz de rectas que pasa por él. Se trata del haz de rectas paralelas a la asíntota por este punto, es decir, $y = x + m - 3$, donde el parámetro m varía en \mathbb{R} . La constante -3 hace que para $m = 0$ tengamos la asíntota, mientras que

para los demás valores la recta correspondiente corta a la hipérbola en un único punto finito (porque ha de ser secante y uno de los puntos de corte es infinito), que puede calcularse resolviendo el sistema formado por la ecuación de la recta y la de la cónica. El resultado es¹

$$\left(\frac{-3m^2 + 14m - 19}{2m}, \frac{-m^2 + 8m - 19}{2m} \right).$$

Cuando m recorre los intervalos $]-\infty, 0[$ y $]0, +\infty[$ la expresión anterior nos da cada una de las dos ramas de la hipérbola.



■

Nota Hasta el final de la sección nos restringimos a los casos que hemos llamado ‘real’ y ‘complejo’.

Hemos probado que dadas dos cónicas (no triviales) cualesquiera, existe una homografía que transforma una en la otra, lo que significa que, desde el punto de vista de la geometría proyectiva, todas las cónicas son iguales, indistinguibles. En particular, no hay nada en la geometría proyectiva que nos permita dividir a las cónicas en elipses, parábolas e hipérbolas. Esta distinción sólo tiene sentido en la geometría afín o, equivalentemente, cuando fijamos una recta infinita en un plano proyectivo.

Por el contrario, la geometría afín sí que distingue entre elipses, parábolas e hipérbolas, en el sentido de que una biyección afín es una homografía que fija a la recta infinita, luego transforma necesariamente elipses en elipses, parábolas en parábolas e hipérbolas en hipérbolas o, dicho de otro modo, que una misma cónica afín no puede verse de dos tipos distintos según el sistema de referencia afín que se elija.

Ahora vamos a demostrar que, salvo esta división de las cónicas en tres tipos, la geometría afín no hace más distinciones, en el sentido de que cualquier elipse puede transformarse en cualquier otra mediante una biyección afín adecuada, e igualmente con parábolas e hipérbolas.

¹Esta parametrización es un caso particular de un hecho general, véase la definición 9.29 más adelante.

Equivalentemente, vamos a ver que, eligiendo adecuadamente el sistema de referencia afín, toda cónica afín no trivial admite una de las ecuaciones siguientes:

$$x^2 + y^2 = -1, \quad x^2 + y^2 = 1, \quad x^2 - y^2 = 1, \quad x = y^2.$$

Además en el caso complejo la primera puede transformarse en la segunda con el cambio de coordenadas $x = ix'$, $y = iy'$.

En el caso real las cuatro ecuaciones que hemos obtenido corresponden, respectivamente a una elipse imaginaria, una elipse real, una hipérbola y una parábola, por lo que no es posible transformar una en otra mediante una biyección afín.

En efecto, en principio, en un sistema de referencia arbitrario, una cónica admite una ecuación (9.2). Podemos suponer que $ab \neq 0$, pues en caso contrario el cambio de coordenadas $x = x' - y'$, $y = y' - y'$ hace que se cumpla esta condición.

El cambio $x = x' - (d/2a)y'$, $y = y'$ transforma la ecuación en otra con $d = 0$, es decir:

$$ax^2 + by^2 + ex + fy + c = 0.$$

Tras el cambio puede que ya no se cumpla $ab \neq 0$, pero al menos tiene que cumplirse $a \neq 0$ o $b \neq 0$, pues en caso contrario la cónica sería degenerada. Distingamos dos casos:

1) Si $ab \neq 0$ hacemos el cambio $x = x' - e/2a$, $y = y' - f/2b$, con lo que la ecuación pasa a tener la forma

$$ax^2 + by^2 + c = 0.$$

No puede ser $c = 0$ o la cónica sería degenerada, luego dividiendo la ecuación entre c podemos suponer que $c = 1$. Ahora el cambio $x' = \sqrt{\pm a}x$, $y' = \sqrt{\pm b}$ (donde en el caso real los signos se eligen para que los radicandos sean positivos) nos da una ecuación $\pm x^2 \pm y^2 = 1$, donde las cuatro combinaciones de signos son posibles, y por lo tanto tenemos una de las ecuaciones

$$x^2 + y^2 = -1, \quad x^2 + y^2 = 1, \quad x^2 - y^2 = 1.$$

(El caso $-x^2 + y^2 = 1$ se reduce al tercero intercambiando las variables.)

2) Si $a = 0$ o $b = 0$, intercambiando las variables si es preciso podemos suponer que es $a = 0$. El cambio $x = x'$, $y = y' - f/2b$ nos da una ecuación

$$by^2 + ex + c = 0.$$

Como la cónica no es degenerada tiene que ser $b \neq 0$, y dividiendo entre b podemos suponer que $b = 1$. Por último, el cambio $x' = -ex - c$, $y' = y$ nos lleva a $x = y^2$. ■

Aunque hemos obtenido $x^2 - y^2 = 1$ como forma canónica para la hipérbola, aplicando la afinidad $(x', y') = (x + y, x - y)$ podemos transformarla en $xy = 1$, que a menudo resulta más manejable.

Con esto todavía no hemos justificado que las cónicas afines coincidan con las cónicas euclídeas definidas en [IGE]. Sabemos que toda cónica euclídea es una cónica afin, pero el recíproco no es evidente: dado un plano euclídeo y una curva dada por una ecuación (9.2) en un sistema de referencia ortonormal, ¿es una elipse, una parábola o una hipérbola en el sentido euclídeo de estar determinada por una directriz y un foco? Equivalentemente, ¿podemos encontrar otro sistema de referencia ortonormal en el que admita una ecuación de tipo (9.2)? Vamos a probar que así es:

Partimos de un plano afin euclídeo E , que completamos hasta un plano proyectivo Π . Consideremos una cónica afin en E cuya extensión a Π no sea degenerada (aunque admitimos que pueda ser imaginaria). Fijemos un sistema de referencia ortonormal en E respecto al cual la ecuación de la cónica sea (9.2). Vamos a reducirla a una forma canónica aplicando isometrías (o, equivalentemente, pasando a otro sistema de referencia ortonormal). La matriz asociada a la completación proyectiva es

$$A = \left(\begin{array}{cc|c} a & d/2 & e/2 \\ d/2 & b & f/2 \\ \hline e/2 & f/2 & c \end{array} \right).$$

Observemos que la submatriz A_0 formada por las dos primeras filas y columnas es no nula, pues de lo contrario la cónica sería degenerada. En virtud del teorema [Al 6.55] existe una matriz ortogonal² M tal que MA_0M^t es diagonal. El cambio de coordenadas

$$(x, y) = (x', y')M$$

corresponde a una isometría en E que transforma

$$ax^2 + by^2 + dxy = (x, y)A_0 \begin{pmatrix} x \\ y \end{pmatrix}$$

en

$$ux'^2 + vy'^2,$$

donde u y v son los valores propios de A_0 . Por lo tanto, transforma (9.2) en una ecuación de la forma

$$ax^2 + by^2 + ex + fy + c = 0,$$

donde los coeficientes a, b, e, f, c son distintos de los de la ecuación original. Si $a \neq 0$ la traslación $x' = x + e/2a$ transforma la ecuación en otra de la forma

$$ax^2 + by^2 + fy + c = 0.$$

²El teorema [Al 6.55] exige que el cuerpo al que pertenecen los coeficientes de la matriz sea realmente cerrado, pero esta hipótesis se usa únicamente para garantizar que su polinomio característico tenga sus raíces en $C = R[i]$. Ahora bien, como en este caso el polinomio es cuadrático, basta con exigir que R sea un cuerpo euclídeo.

Si $b \neq 0$ podemos hacer igualmente $f = 0$. Como a y b no pueden ser simultáneamente nulos, podemos acabar con una de estas dos ecuaciones:

$$ax^2 + by^2 + c = 0, \quad by^2 + ex + c = 0. \quad (9.5)$$

En el primer caso a , b y c han de ser no nulos, o la cónica sería degenerada. Distinguiamos a su vez varios casos:

- Si a , b y c tienen el mismo signo la cónica es imaginaria. Notemos que podemos suponer que los tres coeficientes son positivos y entonces la ecuación equivale a

$$-\left(\frac{x}{\sqrt{c/a}}\right)^2 - \left(\frac{y}{\sqrt{c/b}}\right)^2 = 1.$$

- En otro caso, si a y b son ambos negativos, podemos cambiar el signo para que sean ambos positivos, y entonces c es negativo. La ecuación equivale entonces a la de una elipse (incluyendo la posibilidad de que sea una circunferencia) en forma canónica:

$$\left(\frac{x}{\sqrt{-c/a}}\right)^2 + \left(\frac{y}{\sqrt{-c/b}}\right)^2 = 1.$$

- Si a y b tienen signos distintos (intercambiando x e y si es necesario, lo cual es una isometría), podemos suponer que b y c son negativos, con lo que tenemos la hipérbola

$$\left(\frac{x}{\sqrt{-c/a}}\right)^2 - \left(\frac{y}{\sqrt{c/b}}\right)^2 = 1.$$

Por último, en la segunda ecuación ha de ser $e \neq 0$ o la cónica sería degenerada. La traslación $x = x' - c/e$ la convierte en

$$by^2 + ex = 0,$$

que claramente es equivalente a $y^2 = -(e/b)x$, correspondiente a una parábola.

Con esto hemos demostrado:

Teorema 9.17 *Las cónicas en el sentido de la definición [IGE 7.1] coinciden con las cónicas afines reales.*

En efecto, hemos probado que, pasando de un sistema de referencia ortonormal a otro, la ecuación pasa a ser una de las ecuaciones canónicas que habíamos encontrado para cada tipo de cónica euclídea. Equivalentemente, hemos probado que toda cónica afín no degenerada en un plano euclídeo se puede transformar mediante una isometría en una dada por una de las cuatro ecuaciones

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad y^2 = 4px.$$

Observemos que las cónicas degeneradas pueden obtenerse también como secciones cónicas, al cortar un cono con un plano que pase por su vértice: si su inclinación es menor que la de las generatrices obtenemos un punto, si es igual obtenemos una recta, y si es mayor obtenemos dos rectas. Así pues, descartar las cónicas degeneradas en el caso afín o proyectivo equivale a descartar las secciones cónicas generadas por un plano que pase por el vértice del cono en la definición euclídea de sección cónica.

Nota Del mismo modo que hemos clasificado las cónicas afines en elipses, parábolas e hipérbolas, según cómo sea el corte con la recta infinita, igualmente es posible clasificar las superficies cuádricas afines según las formas en que el plano infinito puede cortar a cada cuádrica. Por ejemplo, existen cinco clases de superficies cuádricas afines reales no degeneradas, pero no vamos a entrar en detalles sobre la clasificación. ■

9.4 Homografías entre haces de rectas

El contenido de esta sección no está relacionado con las cónicas, pero son resultados que necesitaremos en la sección siguiente.

En un plano proyectivo X , consideremos el haz de rectas de centro un punto O . Lo representaremos por \mathcal{H}_O . Éste está determinado por dos cualesquiera de sus rectas, digamos r y r' . Fijemos un sistema de referencia, de modo que las rectas tengan ecuaciones

$$ax + by + cz = 0 \quad \text{y} \quad a'x + b'y + c'z = 0.$$

Consideremos la polaridad p asociada al sistema de referencia, esto es, la que en él tiene matriz identidad. Entonces $p(r)$ y $p(r')$ son los puntos A y B de coordenadas (a, b, c) y (a', b', c') , respectivamente. La polaridad biyecta \mathcal{H}_O con los puntos de la recta $p(O)$. Las coordenadas homogéneas de un punto cualquiera de $p(O) = AB$ son las dadas por $(a, b, c) + \lambda(a', b', c')$, donde $\lambda \in K \cup \{\infty\}$, entendiendo que $\lambda = \infty$ se corresponde con el punto (a', b', c') . Volviendo a aplicar p concluimos que los miembros de \mathcal{H}_O son las rectas de ecuación

$$(ax + by + cz) + \lambda(a'x + b'y + c'z) = 0, \quad \text{con } \lambda \in K \cup \{\infty\}.$$

Una homografía $H \in \text{LGP}(X)$ induce una biyección $\bar{H} : \mathcal{H}_O \rightarrow \mathcal{H}_{H(O)}$ dada por $r \mapsto H[r]$. A una biyección de este tipo la llamaremos una *homografía* entre los haces de rectas.

Por otra parte, si r una recta que no pasa por un punto O , podemos considerar la aplicación $\pi_O^r : \mathcal{H}_O \rightarrow r$ dada por $\pi(s) = s \cap r$, que es una biyección con inversa dada por $\pi_r^O(P) = OP$.

Teorema 9.18 Sean O y O' dos puntos cualesquiera y r, r' dos rectas tales que O no esté en r y O' no esté en r' . Existe una biyección entre las homografías $F : \mathcal{H}_O \rightarrow \mathcal{H}_{O'}$ y las homografías $G : r \rightarrow r'$ dada por $G = \pi_{r'}^{O'} \circ F \circ \pi_r^O$.

DEMOSTRACIÓN: Sea $F = \bar{H}$ una homografía entre los haces y llamemos $s = H[r]$. La proyección perspectiva de s en r' de centro O' se extiende a una perspectividad f de centro O' . Como fija a las rectas que pasan por O' , es claro que la homografía $H \circ f$ induce la misma homografía entre los haces de rectas, pero además transforma r en r' . Por lo tanto podemos suponer que $H[r] = r'$. En tal caso, la aplicación G que hemos definido entre r y r' es simplemente la restricción de H , luego es una homografía.

Recíprocamente, si $G : r \rightarrow r'$ es una homografía, tomamos un sistema de referencia formado por O , dos puntos A, B en r y un cuarto punto D . Por otra parte consideramos el sistema de referencia formado por $O', G(A), G(B), D'$, donde D' es cualquier punto de la recta $O'G(D \cap r)$ no contenido en r .

Sea H la homografía del plano que transforma un sistema de referencia en otro. Entonces H induce una homografía entre los haces de rectas y ésta a su vez induce una homografía de r en r' que coincide en tres puntos con la dada, luego es la dada. Es claro que $F = \bar{H}$ es la aplicación que le hemos asignado a G . ■

Esto nos permite definir el concepto de homografía entre una recta y un haz de rectas:

Definición 9.19 Una biyección $f : r \rightarrow \mathcal{H}_O$ entre una recta proyectiva y un haz de rectas proyectivas es una *homografía* si, dada una recta s que no pase por O , se cumple que $f \circ \pi_O^s : r \rightarrow s$ es una homografía. Similarmente, $g : \mathcal{H}_O \rightarrow r$ es una homografía si lo es $\pi_s^O \circ g$.

Observemos que la definición no depende de la elección de la recta s , pues si s' es otra recta que no pase por O y $f \circ \pi_O^{s'}$ es una homografía, entonces se cumple que $f \circ \pi_O^{s'} = (f \circ \pi_O^s) \circ (\pi_s^O \circ \pi_{O'}^{s'})$, y la última composición es la homografía correspondiente a la identidad en \mathcal{H}_O según el teorema anterior. Igualmente se razona con g .

También es claro que la composición de dos homografías $r \xrightarrow{f} \mathcal{H}_O \xrightarrow{g} \mathcal{H}_{O'}$ es una homografía, pues $f \circ g \circ \pi_{O'}^s = (f \circ \pi_{O'}^t) \circ (\pi_t^O \circ g \circ \pi_O^s)$ y los dos paréntesis son homografías, el segundo por el teorema anterior. Igualmente se comprueba que una composición de homografías $\mathcal{H}_O \xrightarrow{f} \mathcal{H}_{O'} \xrightarrow{g} r$ es una homografía.

En particular, las aplicaciones π_O^r y π_r^O son homografías. Otro ejemplo relevante es el siguiente:

Teorema 9.20 Si p es una polaridad simétrica en un plano proyectivo y O es un punto arbitrario, entonces p se restringe a una homografía $p : p(O) \rightarrow \mathcal{H}_O$.

DEMOSTRACIÓN: Sea $s = p(O)$, tomamos una recta r que no pase por O , y vamos a probar que $f = p|_s \circ \pi_O^r$ es una homografía.

Sea $P_0 = r \cap s, P_1$ un punto de r que no esté en s y P_2 un punto de s que no esté en r . Podemos completar estos tres puntos a un sistema de referencia proyectivo de modo que tengan coordenadas $(1, 0, 0), (0, 1, 0)$ y $(0, 0, 1)$, con lo que s tiene ecuación $y = 0$ y r tiene ecuación $z = 0$. Sea A la matriz de la

polaridad en este sistema de referencia. Si un punto $Q \in s$ tiene coordenadas $(p, 0, q)$, entonces $p(Q)$ tiene ecuación $(p, 0, q)A(x, y, z)^t = 0$, y el punto de corte con r debe cumplir

$$(p, 0, q) \begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = 0,$$

es decir, $(ap + eq)x + (dp + fq)z = 0$, luego la expresión en coordenadas de f es

$$(p, q) \mapsto (dp + fq, -ap - eq) = (p, q) \begin{pmatrix} d & -a \\ f & -e \end{pmatrix}.$$

Sólo hay que comprobar que el determinante de la matriz no es nulo, pero esto equivale a que r no pase por el polo de s . En efecto, dicho polo no está en la recta $z = 0$, luego admite coordenadas $(x, y, 1)$ tales que $(x, y, 1)A = (0, \lambda, 0)$ con $\lambda \neq 0$, y esto equivale a que el sistema

$$ax + dy + e = 0, \quad ex + fy + c = 0$$

tenga solución única en x e y , lo cual a su vez equivale a que $af - de \neq 0$. ■

Todas las propiedades que conocemos para homografías entre rectas se traducen ahora a propiedades análogas para homografías entre haces de rectas. Por ejemplo, dadas tres rectas en un haz \mathcal{H}_O , existe una única homografía que las transforma en tres rectas dadas de otro haz $\mathcal{H}_{O'}$.

Una homografía que transforme las rectas $ax + by + cz = 0$ y $a'x + b'y + c'z = 0$ en las rectas $ux + vy + wz = 0$ y $u'x + v'y + w'z = 0$ se corresponde a través de la polaridad p asociada al sistema de referencia con una homografía que transforma los puntos (a, b, c) y (a', b', c') en los puntos (u, v, w) y (u', v', w') . Una tercera recta del haz será de la forma

$$(ax + by + cz) + \lambda_0(a'x + b'y + c'z) = 0, \quad \text{con } \lambda_0 \in K \setminus \{0\},$$

que se transformará en una recta

$$(ux + vy + wz) + \mu_0(u'x + v'y + w'z) = 0, \quad \text{con } \mu_0 \in K \setminus \{0\}.$$

La homografía asociada a través de p transformará $(a, b, c) + \lambda_0(a', b', c')$ en $(u, v, w) + \mu_0(u', v', w')$. Es claro que una (y por tanto la única) homografía que cumple estas condiciones es la dada por

$$(a, b, c) + \lambda(a', b', c') \mapsto (u, v, w) + \lambda \frac{\mu_0}{\lambda_0}(a', b', c'),$$

luego la homografía entre los haces de rectas es la dada por

$$ax + by + cx + \lambda(a'x + b'y + c'z) = 0 \mapsto ux + vy + wz + \lambda \frac{\mu_0}{\lambda_0}(u'x + v'y + w'z) = 0.$$

Notemos que podemos multiplicar por μ_0/λ_0 los coeficientes (u', v', w') sin alterar por ello la ecuación de la recta, y entonces la ecuación de la homografía entre haces se expresa simplemente como

$$ax + by + cx + \lambda(a'x + b'y + c'z) = 0 \mapsto ux + vy + wz + \lambda(u'x + v'y + w'z) = 0.$$

Es claro que toda correspondencia de este tipo es una homografía entre haces.

Diremos que una homografía $\bar{H} : \mathcal{H}_O \rightarrow \mathcal{H}_{O'}$ entre dos haces con centros distintos $O \neq O'$ es una *proyección perspectiva* si H es una perspectiva (o, más exactamente, si existe una perspectiva H' , no necesariamente H , tal que $\bar{H} = \bar{H}'$).

Una homografía \bar{H} es una perspectiva si y sólo si fija la recta OO' . En efecto, si H es una perspectiva de centro P , como ha de cumplir $H(O) = O'$, los puntos P , O y O' han de ser colineales, luego H fija a OO' . Recíprocamente, si \bar{H} fija a OO' , tomamos dos rectas r y s en \mathcal{H}_O distintas de OO' y consideramos sus imágenes $r' = H[r]$ y $s' = H[s]$. Sean $Q = r \cap r'$, $Q' = s \cap s'$. Sea $P = OO' \cap QQ'$. La proyección perspectiva de centro P entre r y s' se extiende a una perspectiva de centro P que transforma O en O' y Q en Q' , luego induce una homografía entre los haces que coincide con la dada en OO' , r y s , luego es la dada.

Notemos que una homografía $\bar{H} : \mathcal{H}_O \rightarrow \mathcal{H}_{O'}$ con $O \neq O'$ no puede fijar a ninguna recta salvo a lo sumo a OO' , luego decir que no es perspectiva equivale a decir que no fija a ninguna recta.

9.5 El teorema de Steiner

Con las consideraciones de la sección anterior estamos en condiciones de probar una caracterización muy simple de las cónicas proyectivas de la que extraeremos varias consecuencias de interés:

Teorema 9.21 (Steiner) *Sean O y O' puntos distintos en un plano proyectivo. Sea $\bar{H} : \mathcal{H}_O \rightarrow \mathcal{H}_{O'}$ una homografía no perspectiva. Entonces cada recta $r \in \mathcal{H}_O$ corta a $\bar{H}(r)$ en un único punto P_r . El conjunto \mathcal{C} formado por todos los puntos P_r es una cónica no trivial, y toda cónica no trivial puede obtenerse de esta forma.*

DEMOSTRACIÓN: Fijemos un sistema de referencia en el plano X . La ecuación de la homografía entre las rectas es

$$ax + by + cx + \lambda(a'x + b'y + c'z) = 0 \mapsto ux + vy + wz + \lambda(u'x + v'y + w'z) = 0.$$

Un punto de coordenadas (x, y, z) estará en \mathcal{C} si y sólo si satisface las ecuaciones de ambos haces de rectas con el mismo valor de λ , es decir, si y sólo si

$$-\frac{ax + by + cz}{a'x + b'y + c'z} = -\frac{ux + vy + wz}{u'x + v'y + w'z},$$

o equivalentemente, si y sólo si

$$(ax + by + cz)(u'x + v'y + w'z) - (a'x + b'y + c'z)(ux + vy + wz) = 0.$$

(Hay que tratar aparte el caso $\lambda = \infty$, donde los denominadores se anulan, pero la conclusión es la misma.)

Al operar la última ecuación obtenemos ciertamente la ecuación de una cónica. Notemos que no puede ser idénticamente nula, pues no todos los puntos de X están en \mathcal{C} . Más exactamente, cada recta que pasa por O distinta de OO' contiene un único punto adicional de \mathcal{C} (su intersección con su imagen). Por último, veamos que la cónica contiene cinco puntos no colineales tres a tres. Para ello tomamos tres rectas en \mathcal{H}_O distintas³ de OO' y sean A, B, C los tres puntos correspondientes en la cónica. Es claro que de los cinco puntos O, O', A, B, C a lo sumo podrían ser colineales A, B, C , pero vamos a ver que no es el caso. Si estuvieran sobre la misma recta r , consideramos la homografía H' del plano X que fija a A, B y C y además cumple $H(O) = O'$. Entonces \bar{H}' coincide con \bar{H} en OA, OB, OC , luego $\bar{H}' = \bar{H}$. Pero H' fija a tres puntos de r , luego fija a todos los puntos de r , en particular a la intersección $r \cap OO'$, de donde se sigue que $H'[OO'] = OO'$, luego $\bar{H}[OO'] = OO'$, contradicción.

Recíprocamente, si \mathcal{C}' es una cónica no trivial, tomamos en ella cinco puntos A, B, C, D, E no colineales tres a tres. Existe una única homografía de \mathcal{H}_A en \mathcal{H}_B que transforma AC, AD y AE en BC, BD y BE respectivamente. Si fuera perspectiva, es decir, de la forma \bar{H} , donde H es una perspectividad, entonces el eje de H no pasaría por O ni O' , pues $H(O) = O'$, y los puntos C, D, E tendrían que estar sobre dicho eje, pero no son colineales. Así pues, la homografía no es perspectiva, luego por la parte ya probada determina una cónica \mathcal{C}' que claramente contiene a los cinco puntos que hemos tomado, luego $\mathcal{C} = \mathcal{C}'$. ■

Observemos que, en las condiciones del teorema anterior, si r es la tangente a \mathcal{C} por O , el único punto de \mathcal{C} en el que puede cortar a $\bar{H}(r)$ es O , luego $\bar{H}(r) = O' =$. Recíprocamente, $\bar{H}(O'O)$ tiene que ser la tangente a \mathcal{C} por O' .

En la segunda parte de la prueba del teorema de Steiner hemos demostrado en realidad algo ligeramente más fuerte:

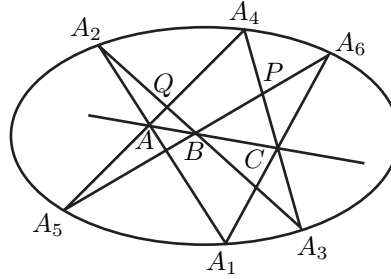
Teorema 9.22 Sean O y O' dos puntos en una cónica \mathcal{C} . Entonces, la aplicación $H : \mathcal{H}_O \rightarrow \mathcal{H}_{O'}$ que a cada recta OP con $P \in \mathcal{C}$ le asigna la recta $H(OP) = O'P$, (con el convenio de que OO representa a la tangente a \mathcal{C} por O) es una homografía de haces de rectas.

Como aplicación demostramos uno de los teoremas más famosos de la geometría proyectiva (compárese con [IGE 4.56]):

Teorema 9.23 (Pascal) Seis puntos no colineales tres a tres están contenidos en una misma cónica si y sólo si los lados opuestos de un hexágono cualquiera que los tenga como vértices se cortan en tres puntos colineales.

³Esto supone que el cuerpo de escalares no sea el cuerpo de tres elementos, condición necesaria, pues en caso contrario no existen cónicas no triviales.

DEMOSTRACIÓN: Consideremos los puntos definidos como indica la figura. Hay que probar que A, B, C son colineales. Sea $H : \mathcal{H}_{A_2} \rightarrow \mathcal{H}_{A_6}$ la homografía dada por el teorema anterior. En particular, H transforma $A_2A_5, A_2A_4, A_2A_3, A_2A_1$ en las rectas $A_6A_5, A_6A_4, A_6A_3, A_6A_1$, respectivamente.



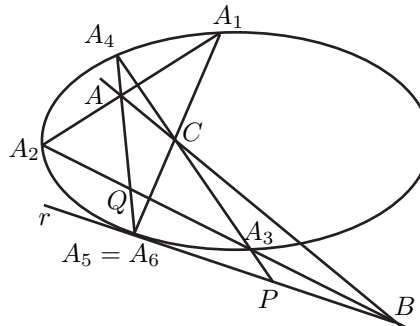
Ahora aplicamos el teorema 9.18, según el cual, H induce una homografía $h : A_4A_5 \rightarrow A_4A_3$ que transforma los puntos A_5, A_4, Q, A en P, A_4, A_3, C , respectivamente. Ahora bien, esta homografía coincide en los tres primeros puntos con la proyección perspectiva de centro B entre ambas rectas, luego ambas coinciden, y en particular la proyección perspectiva transforma también C en A , lo que implica que A, B, C son colineales.

Recíprocamente, si los seis puntos cumplen la condición, consideremos la cónica que pasa por A_2, A_3, A_4, A_5, A_6 , sea H la misma homografía considerada antes y sea A'_1 el punto $A_1A_2 \cap H(A_1A_2)$, que está en la cónica. Por la parte ya probada, el punto $C' = A'_1A_6 \cap A_3A_4$ tiene que estar en la recta AB , luego tiene que ser $C' = C$, de donde $A'_1A_6 = A_1A_6$, luego $A'_1 = A_1$. ■

El teorema siguiente nos da un procedimiento para trazar la tangente a una cónica por uno de sus puntos conociendo otros cuatro puntos:

Teorema 9.24 Sean A_1, A_2, A_3, A_4, A_5 cinco puntos en una cónica \mathcal{C} . Sea r la tangente a \mathcal{C} por A_5 . Entonces los puntos $r \cap A_2A_3, A_1A_2 \cap A_4A_5$ y $A_1A_5 \cap A_3A_4$ son colineales.

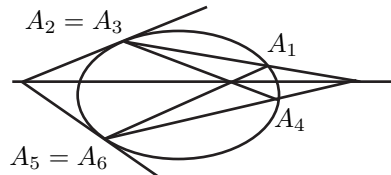
DEMOSTRACIÓN: Vale literalmente la prueba del teorema anterior sin más que considerar que $A_5 = A_6$ y que la recta A_5A_6 es la tangente a la cónica por A_5 . ■



Todavía podemos identificar dos puntos más y obtenemos una variante para cuadriláteros completos:

Teorema 9.25 Si un cuadrilátero completo está inscrito en una cónica, las tangentes por dos de sus vértices se cortan en un punto colineal con los dos puntos diagonales que no están sobre el lado determinado por dichos vértices.

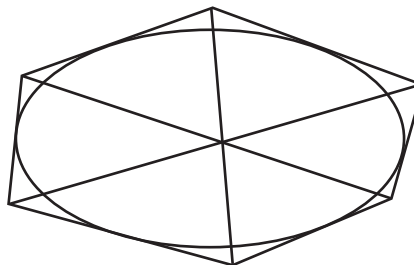
De nuevo, el argumento es literalmente el mismo empleado en la prueba del teorema de Pascal, ahora identificando $A_2 = A_3, A_5 = A_6$ y cambiando los lados correspondientes por tangente.



El teorema de Pascal admite una forma dual que se demuestra sin más que aplicar la polaridad de la cónica:

Teorema 9.26 (Brianchon) *Un hexágono está circunscrito a una cónica (es decir, sus lados son tangentes) si y sólo si las tres rectas que pasan por vértices opuestos son concurrentes.*

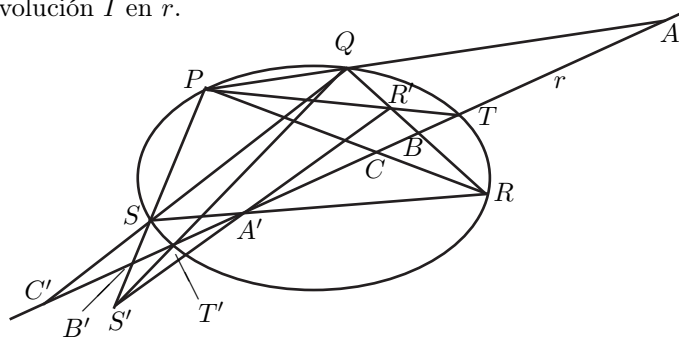
DEMOSTRACIÓN: Los polos de los lados del hexágono son seis puntos distintos situados sobre la cónica, y sus lados son las rectas polares de los vértices del hexágono dado. Las intersecciones de lados opuestos del nuevo hexágono son los polos de las rectas que unen vértices opuestos del hexágono dado, y como éstas son colineales por el teorema de Pascal, sus polos son concurrentes. El recíproco es análogo. ■



Veamos más aplicaciones del teorema de Pascal:

Teorema 9.27 (Desargues) *Consideremos un cuadrilátero completo inscrito en una cónica \mathcal{C} y una recta r que no pase por ninguno de sus vértices, pero que corte a \mathcal{C} en uno o dos puntos. Entonces éstos son conjugados respecto a la involución en r que determinan los pares de puntos donde los lados opuestos del cuadrilátero cortan a r .*

DEMOSTRACIÓN: El teorema 8.65 afirma que, realmente, los puntos determinados por pares de lados opuestos del cuadrilátero son conjugados respecto de una involución I en r .



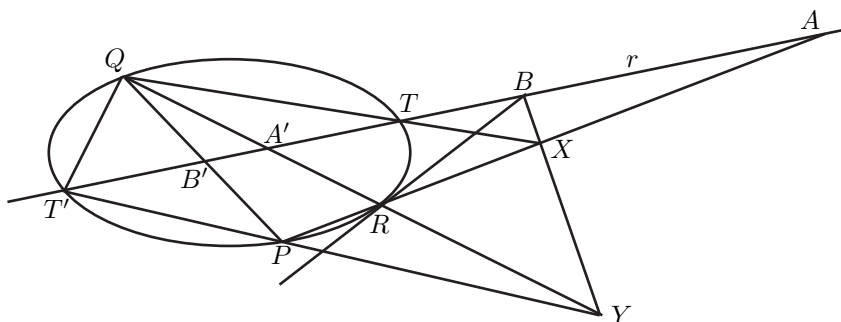
La figura muestra el cuadrilátero $PQRS$ y los pares de puntos conjugados que determina (A, A') , (B, B') , (C, C') . Falta probar que la involución que los conjuga también empareja los puntos de corte T y T' entre r y \mathcal{C} . Para ello llamamos $R' = PT \cap QR$ y $S' = QT' \cap PS$, con lo que formamos un nuevo cuadrilátero completo $PQR'S'$. El teorema de Pascal aplicado al hexágono $PSRQT'T$ nos da que R' , S' y A' son colineales, lo cual implica que (A, A') es uno de los pares en r determinados por el cuadrilátero, y los otros dos son (B, B') y (T, T') , por lo que son conjugados por una misma involución. Como tiene dos pares en común con I , de hecho es I , luego $I(T) = T'$.

Si r es tangente a \mathcal{C} , usamos la versión para cinco puntos del teorema de Pascal. ■

Una variante del argumento anterior prueba la versión siguiente para un triángulo en lugar de un cuadrilátero:

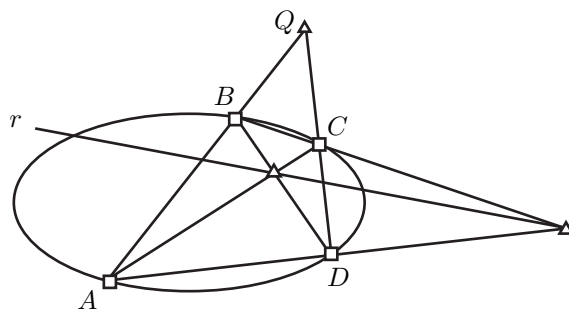
Teorema 9.28 *Consideremos un triángulo inscrito en una cónica \mathcal{C} y una recta r que no pase por sus vértices, pero que corte a \mathcal{C} en uno o dos puntos. Entonces éstos son conjugados respecto a la involución en r que determinan el par de puntos donde dos lados del triángulo cortan a r y el par donde lo hacen el tercer lado y la tangente a \mathcal{C} por el vértice opuesto.*

DEMOSTRACIÓN: La figura muestra un triángulo PQR y los pares (A, A') , (B, B') que determina en la recta r . Éstos determinan una involución I en r y se trata de probar que conjuga los puntos de corte T y T' .



Para ello formamos $X = QT \cap PR$, $Y = QR \cap PT'$ y consideramos el cuadrilátero completo $PQXY$. El teorema de Pascal aplicado al pentágono $TT'PRQ$ nos da que X, Y, B son colineales, lo que nos da que uno de los pares determinados en la recta es (B, B') , y los otros dos son claramente (T, T') y (A, A') , luego determinan la involución I y además vemos que ésta conjuga a T y T' como se requería. Si $T = T'$ el razonamiento vale igualmente usando 9.25 en lugar del teorema de Pascal para cinco puntos. ■

Por último veamos una prueba alternativa del teorema 9.11:



DEMOSTRACIÓN: Por el teorema 9.25, las tangentes a la cónica por C y D se cortan en un punto de la recta r , pero dicho punto es el polo de CD . Igualmente, las tangentes a la cónica por A y B se cortan en un punto (el polo de AB) de la recta r . Así pues, el polo de r está en las rectas AB y CD , luego es el punto diagonal Q . ■

Homografías entre cónicas y rectas Veamos otra consecuencia destacada del teorema 9.22. Para ello observamos en primer lugar que existen biyecciones naturales entre las rectas y las cónicas:

Definición 9.29 Sea \mathcal{C} una cónica y $O \in \mathcal{C}$. Sea r una recta que no pase por O . Definimos $h_O : \mathcal{C} \rightarrow r$ como la aplicación dada por $h_O(P) = OP \cap r$, con el convenio de que OO representa a la tangente a \mathcal{C} por O . Es claro que h_O es una aplicación biyectiva.

Vemos así que existe una correspondencia natural entre los puntos de cualquier cónica y los de cualquier recta. En particular, si el cuerpo K que determina el plano proyectivo es infinito, entonces las cónicas tienen infinitos puntos, y si es finito, entonces cada cónica contiene tantos puntos como cada recta, que son uno más que el número de elementos de K .

Lo relevante es que estas biyecciones pueden considerarse homografías:

Diremos que una biyección $h : s \rightarrow \mathcal{C}$ entre una recta y una cónica es una *homografía* si, dado un punto $O \in \mathcal{C}$ y una recta r que no pase por O , la composición $h \circ h_O : s \rightarrow r$ es una homografía.

Como en el caso de los haces de rectas, ahora podemos probar que esta definición no depende de la elección del punto O ni de la recta r . En efecto, si $h \circ h_O^r$ es una homografía y tomamos otro punto O' y otra recta r' , entonces $h \circ h_{O'}^{r'}$ se descompone como

$$s \xrightarrow{h \circ h_O^r} r \xrightarrow{\pi_r^O} \mathcal{H}_O \xrightarrow{H_O^{O'}} \mathcal{H}_{O'} \xrightarrow{\pi_{O'}^{r'}} r',$$

donde $H_O^{O'}$ es la homografía dada por el teorema 9.22. Por lo tanto, es también una homografía.

Una homografía $h : \mathcal{C} \rightarrow s$ de una cónica en una recta se define como la inversa de una homografía de una recta en una cónica o, equivalentemente como una aplicación tal que $h_O^{-1} \circ h : r \rightarrow \mathcal{C}$ sea una homografía entre rectas, para cualquier elección de un punto $O \in \mathcal{C}$ y una recta r que no pase por O .

Es inmediato que la composición de una homografía entre una cónica y una recta y una homografía entre rectas da de nuevo una homografía, así como que las aplicaciones $h_O : \mathcal{C} \rightarrow r$ son homografías.

Observemos también que existe una única homografía que transforma tres puntos dados de una cónica en tres puntos dados de una recta. En efecto, dados tres puntos A, B, C en la cónica, consideramos cualquier aplicación h_O en la recta y la componemos con la homografía de la recta que lleva los puntos $h_O(A), h_O(B), h_O(C)$ en los tres puntos deseados. Si $h_1, h_2 : \mathcal{C} \rightarrow r$ coinciden en tres puntos, entonces $h_1^{-1} \circ h_2$ deja fijos tres puntos, luego es la identidad, luego $h_1 = h_2$.

Podríamos definir las homografías entre cónicas de forma análoga a como hemos definido las homografías entre cónica y recta, pero hay una definición equivalente más simple:

Definición 9.30 Sean \mathcal{C} y \mathcal{C}' dos cónicas en un plano proyectivo X . Una homografía $h : \mathcal{C} \rightarrow \mathcal{C}'$ es la restricción de una homografía $H : X \rightarrow X$ tal que $H[\mathcal{C}] = \mathcal{C}'$.

Como una cónica contiene un sistema de referencia proyectivo, es claro que cada homografía de una cónica se extiende a una única homografía del plano.

Teorema 9.31 Dadas dos cónicas en un plano proyectivo, existe una única homografía entre ellas que transforma tres puntos dados en tres puntos dados.

DEMOSTRACIÓN: Sean $\mathcal{C}, \mathcal{C}'$ las dos cónicas y consideremos dos ternas de puntos distintos $A, B, C \in \mathcal{C}$ y $A', B', C' \in \mathcal{C}'$. Sea P el punto de corte de las tangentes a \mathcal{C} por A y B y sea P' el punto de corte de las tangentes a \mathcal{C}' por A' y B' . Es claro que A, B, C, P y A', B', C', P' son dos sistemas de referencia proyectivos, por lo que existe una única homografía H del plano que transforma uno en otro. Veamos que $H[\mathcal{C}] = \mathcal{C}'$. Ambas cónicas pasan por A', B', C' y sus tangentes por A' y B' pasan por P' . Por lo tanto, basta probar que existe una única cónica que pasa por tres puntos dados A', B', C' y cuyas tangentes por A' y B' pasan por un mismo punto P' .

Para ello tomamos los cuatro puntos como sistema de referencia, de modo que sus coordenadas serán $A' = (1, 0, 0)$, $B' = (0, 1, 0)$, $C' = (0, 1, 0)$ y $P' = (1, 1, 1)$. Según hemos visto en la prueba del teorema 9.6, el hecho de que la cónica pase por los tres primeros se traduce en que su matriz es de la forma

$$M = \begin{pmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{pmatrix}.$$

Las ecuaciones de las tangentes por A' y B' son $aY + bZ = 0$ y $aX + cZ = 0$, respectivamente, luego el hecho de que P' esté en ambas se traduce en que $b = c = -a$. Como podemos elegir arbitrariamente uno de los parámetros, no perdemos generalidad si suponemos que $a = 1$, $b = c = -1$, luego la cónica está completamente determinada.

Esto prueba la existencia. La unicidad se debe a que una homografía que transforme una cónica en la otra y lleve los tres puntos dados en los tres puntos dados transforma necesariamente P en P' , luego es la homografía que hemos construido. ■

Es claro que la composición de homografías entre cónicas es una homografía, y también lo es la composición de una homografía entre rectas con una homografía de una recta en una cónica. En efecto, tenemos la situación siguiente:

$$\begin{array}{ccc} r & \xrightarrow{h} & \mathcal{C} & \xrightarrow{h_O} & s \\ & & \downarrow H & & \downarrow H \\ & & \mathcal{C}' & \xrightarrow{h_{O'}} & s' \end{array}$$

donde s es una recta cualquiera y $O \in \mathcal{C}$ un punto cualquiera que no pase por s , $O' = H(O)$ y $s' = H[s]$. La construcción geométrica de h_O garantiza que el diagrama es conmutativo, luego el hecho de que $h \circ h_O$ sea una homografía (porque h lo es) implica que $r \circ H \circ h_{O'}$ también lo es, luego $r \circ H$ también es una homografía.

A partir de aquí es inmediato que podemos tratar a las cónicas proyectivas exactamente igual que si fueran rectas. Por ejemplo, dados tres puntos en una cónica \mathcal{C} , podemos considerar las coordenadas homogéneas de cualquier punto $P \in \mathcal{C}$ respecto a dicho sistema, definidas tomando cualquier homografía $h : \mathcal{C} \rightarrow r$ y considerando las coordenadas homogéneas de $h(P)$ respecto del sistema de referencia formado por las imágenes de los tres puntos dados. Es inmediato que la definición no depende de la elección de h .

Similarmente podemos hablar de la razón doble entre cuatro puntos de una cónica y, en general, de cualquier concepto definido para rectas y que se conserve por homografías. Por ejemplo, si consideramos un plano proyectivo sobre un cuerpo ordenado, podemos hablar de los segmentos determinados por dos puntos en una cónica, y de la separación de un par de puntos por otro par. El teorema siguiente nos da una caracterización de la separación:

Teorema 9.32 Sean A, B, C, D cuatro puntos en una cónica \mathcal{C} . Entonces $AB \parallel CD$ si y sólo si $AB \cap CD$ es un punto interior.

DEMOSTRACIÓN: Consideremos la proyección $h_C : \mathcal{C} \rightarrow AB$. Notemos que $h_C(A) = A$, $h_C(B) = B$, $h_C(D) = AB \cap CD$, mientras que $h_C(C)$ es la intersección de AB con la tangente a \mathcal{C} por C . En particular, $h_C(C)$ es un punto exterior de \mathcal{C} , pues todos los puntos de una tangente son exteriores, salvo el punto de tangencia. Entonces $h_C(D)$ será interior si y sólo si está en el segmento de extremos A y B opuesto al que contiene a $h_C(C)$ (pues uno consta de puntos interiores y el otro de puntos exteriores), es decir, si y sólo si $h_C(A)h_C(B) \parallel h_C(C)h_C(D)$, lo cual equivale por definición de separación sobre una cónica a $AB \parallel CD$. ■

Nota Observemos que podemos definir también una homografía $f : \mathcal{C} \rightarrow \mathcal{H}_O$ entre una cónica y un haz de rectas como cualquier biyección que compuesta con una homografía $\mathcal{H}_O \rightarrow r$ sea una homografía, y análogamente se definen las homografías $\mathcal{H}_O \rightarrow \mathcal{C}$. Los resultados que tenemos demostrados hacen inmediato que estas definiciones no dependen de la elección de la homografía auxiliar, así como que la composición de dos homografías de cualquier tipo (entre rectas, cónicas o haces) es también una homografía. ■

El centro y el eje de una homografía de una cónica Los teoremas de Pascal y Desargues nos permiten describir con detalle las homografías de las cónicas. El resultado básico es el siguiente:

Teorema 9.33 Si $A \neq A'$ y B, B' son dos pares de puntos homólogos por una homografía en una cónica, entonces $AB' \cap A'B$ está sobre una recta r independiente de la elección de los puntos A y B .

DEMOSTRACIÓN: Sea $f : \mathcal{C} \rightarrow \mathcal{C}$ una homografía en una cónica \mathcal{C} , que se extiende a una homografía del plano proyectivo y consideremos la composición de las homografías siguientes entre haces de rectas:

- La homografía $\mathcal{H}_{A'} \rightarrow \mathcal{H}_A$ dada por $A'P \mapsto AP$ según 9.22.
- La homografía $\mathcal{H}_A \rightarrow \mathcal{H}_{A'}$ dada por $r \mapsto f[r]$.
- La homografía $\mathcal{H}_{A'} \rightarrow \mathcal{H}_A$ dada por $A'P \mapsto AP$ según 9.22.

Al componerlas obtenemos una homografía $\mathcal{H}_{A'} \rightarrow \mathcal{H}_A$ dada por

$$A'P \mapsto AP \mapsto A'f(P) \mapsto Af(P).$$

Esta homografía fija la recta AA' , luego es una proyección perspectiva (por la última observación de la sección precedente), es decir, está inducida por una perspectividad H . Sea r su eje de perspectiva. Entonces cada par de rectas homólogas AB' y $B'A$ se cortan en r (para todo punto B). Falta probar que r es independiente de A . Ahora bien, si tomamos tres puntos distintos A, B, C con imágenes distintas de ellos tres A', B', C' , los puntos $AB' \cap A'B, A'C \cap C'A, BC' \cap C'B$ son colineales, por el teorema de Pascal aplicado al hexágono $AB'CA'BC'$. Si llamamos r_0 a la recta que los contiene, vemos que r_0 es la recta r determinada por A y también la determinada por B . Esto prueba que si A y B tienen imágenes distintas de ellos mismos (es decir, salvo si $A' = B$ o $B' = A$) ambos determinan la misma recta. Si $A' = B$ tomamos un tercer punto C tal que C y C' sean distintos de A, A', B, B' y concluimos que A, B determinan ambos la misma recta que C . ■

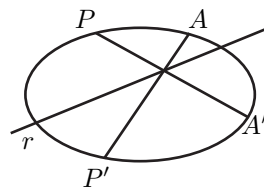
Por dualidad tenemos también el teorema siguiente:

Teorema 9.34 *Si $r \neq r', s \neq s'$ son dos pares de tangentes homólogas por una homografía en una cónica, entonces la recta que pasa por $r \cap s'$ y $r' \cap s$ pasa por un punto O independiente de la elección de las rectas r y s .*

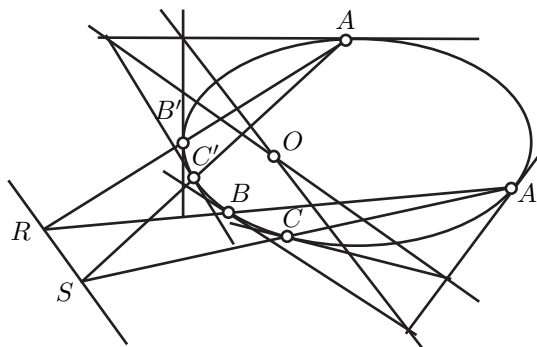
Definición 9.35 La recta r y el punto O determinados por los dos teoremas anteriores reciben el nombre de *eje* y *centro* de la homografía considerada.

Observemos que, por la propia definición, el eje de una homografía de una cónica pasa por sus puntos fijos, si es que los tiene.

Una homografía en una cónica distinta de la identidad queda completamente determinada en cuanto conocemos su eje y un par de puntos homólogos $A \neq A'$. En efecto, para encontrar el homólogo de otro punto P no tenemos más que unir A con $PA' \cap r$. La imagen P' será el punto de intersección de esta recta con la cónica.



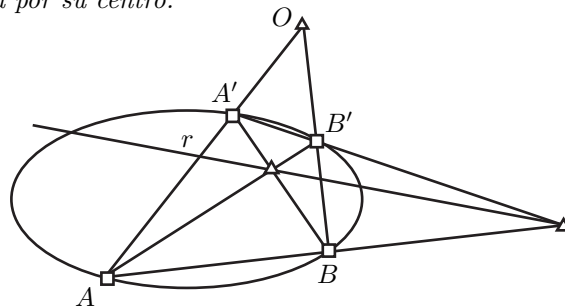
Teorema 9.36 *El centro de una homografía en una cónica es el polo de su eje.*



DEMOSTRACIÓN: Consideremos tres pares de puntos homólogos $A \neq A'$, $B \neq B'$, $C \neq C'$. Sean $R = A'B \cap B'A$ y $S = A'C \cap C'A$. De este modo RS es el eje de la homografía. Las tangentes en A y B' se cortan en el polo de AR , y las tangentes en A' y B se cortan en el polo de $A'R$, luego la recta polar de R es la que une ambos puntos de intersección. Similarmente se calcula la polar de S y la intersección O entre ambas es el polo de RS . Pero, por definición, el punto O así calculado es el centro de la homografía. ■

Veamos ahora que una involución queda determinada por su centro de una forma muy simple:

Teorema 9.37 Si A y A' son conjugados por una involución en una cónica, la recta AA' pasa por su centro.



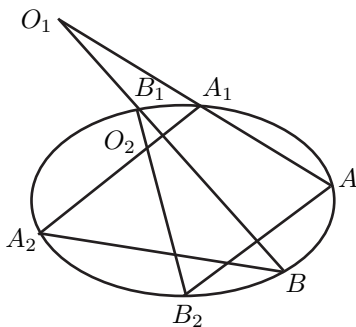
DEMOSTRACIÓN: Sean $A \neq A'$ y B, B' dos pares de puntos conjugados por la involución. Entonces $AB' \cap BA'$ está en el eje, y es claro que su imagen por la homografía que genera la involución es él mismo. Todo punto del eje puede expresarse de este modo para un B adecuado, luego la homografía fija a cada punto de su eje y es, por consiguiente, una perspectividad. Basta probar que su centro de perspectiva coincide con el centro de la involución. Por el teorema anterior basta probar que el centro de perspectiva O es el polo del eje de perspectiva r . Si tomamos dos pares de conjugados $A \neq A'$, $B \neq B'$, entonces los puntos $AB' \cap A'B$ y $AB \cap A'B'$ están en r (notemos que el segundo es $A''B \cap A'B'$). Ahora basta aplicar el teorema 9.11. ■

Es claro que todo punto O que no esté contenido en una cónica \mathcal{C} es el centro de una única involución en \mathcal{C} . Basta considerar dos rectas secantes a \mathcal{C} que pasen por O . Si éstas cortan a la cónica en A, A' y B, B' , la involución para la que estos puntos forman dos pares de conjugados tiene claramente centro O .

En el capítulo anterior vimos que toda homografía en una recta (y por lo tanto en una cónica) es producto de dos involuciones. El teorema siguiente nos da la relación entre los centros y los ejes.

Teorema 9.38 *El eje de la composición de dos involuciones de una cónica es la recta que une sus centros y el centro de la intersección de sus ejes.*

DEMOSTRACIÓN: Tomemos dos puntos A y B de la cónica que no sean puntos fijos o conjugados de ninguna de las involuciones ni de su composición. Entonces la primera involución los transforma en otros dos puntos A_1, B_1 , y la segunda transforma éstos en otros dos puntos A_2, B_2 , de modo que los seis puntos son distintos. Si O_1 y O_2 son los centros de las involuciones, por el teorema 9.37 sabemos que $O_1 = AA_1 \cap BB_1$ y $O_2 = A_1A_2 \cap B_1B_2$. Al aplicar el teorema de Pascal al hexágono $AA_1A_2BB_1B_2$ concluimos que O_1O_2 contiene a $AB_2 \cap BA_2$, que por definición es un punto del eje de la composición. Por lo tanto dicho eje ha de ser O_1O_2 . Aplicando la polaridad de la cónica obtenemos que el centro es la intersección de los ejes. ■



Capítulo X

La geometría parabólica

Vamos a reformular la geometría euclídea en términos de la geometría proyectiva. Sabemos que la forma natural de introducir una estructura euclídea en un espacio afín E (sobre un cuerpo ordenado pitagórico) es a través de un producto escalar en el espacio vectorial asociado \vec{E} . Por otra parte sabemos (véase la definición 8.32 y las observaciones posteriores) que un espacio afín puede identificarse con un par (E, E_∞) , donde E es un espacio proyectivo y E_∞ es un hiperplano seleccionado arbitrariamente como hiperplano infinito, y ahora vamos a ver que, en estos términos, una forma alternativa de introducir una estructura euclídea en E es a través de una polaridad simétrica en E_∞ .

Observemos en primer lugar que un producto escalar está determinado salvo un factor de escala por la relación de ortogonalidad que determina. Más en general, tenemos el resultado siguiente, válido para formas bilineales simétricas arbitrarias:

Teorema 10.1 Sean $F, G : V \times V \rightarrow K$ dos formas bilineales simétricas en un espacio vectorial V de dimensión finita sobre un cuerpo K de característica distinta de 2. Si $F(v, w) = 0$ equivale a $G(v, w) = 0$, para todo par de vectores $v, w \in V$, entonces existe $\alpha \in K$ no nulo tal que $G = \alpha F$.

DEMOSTRACIÓN: Por el teorema [Al 6.48] existe una base e_1, \dots, e_n de V ortogonal para F , pero por la hipótesis es también ortogonal para G . Esto significa que las matrices de F y G en esta base son diagonales, digamos $[a_1, \dots, a_n]$ y $[b_1, \dots, b_n]$. Claramente, como $a_i = F(e_i, e_i)$, tenemos que $a_i = 0$ si y sólo si $b_i = 0$. En particular, una forma es idénticamente nula si y sólo si lo es la otra, y en tal caso el teorema se cumple con $\alpha = 1$. Supongamos, pues, que las formas no son nulas. No perdemos generalidad si suponemos que $a_1 \neq 0$, luego también $b_1 \neq 0$. Ahora observamos que

$$F(a_i e_1 - a_1 e_i, e_1 + e_i) = a_i F(e_1, e_1) - a_1 F(e_i, e_i) = 0,$$

luego

$$G(a_i e_1 - a_1 e_i, e_1 + e_i) = a_i G(e_1, e_1) - a_1 G(e_i, e_i) = a_i b_1 - a_1 b_i = 0,$$

luego, o bien $a_i = b_i = 0$, o bien $b_i/a_i = b_1/a_1$. Llamando $\alpha = b_1/a_1 \neq 0$, en ambos casos $b_i = \alpha a_i$, para todo i , luego $G = \alpha F$. ■

Si en un espacio euclídeo cambiamos el producto escalar por un múltiplo positivo, obtenemos otro espacio euclídeo que se diferencia del dado únicamente en la elección de la unidad de medida. Así pues, si en un espacio afín especificamos una clase de equivalencia de productos escalares salvo un factor de escala, con ello estamos determinando lo que es “absoluto” en una estructura euclídea (la ortogonalidad, la congruencia entre segmentos y entre ángulos, etc.), pero no una unidad de medida, que es una elección arbitraria ajena a la geometría del espacio. Ahora vamos a ver que a partir, no de un producto escalar, sino meramente de la relación de ortogonalidad que determina, es posible definir una polaridad simétrica en el hiperplano infinito que determina dicha relación de ortogonalidad:

Teorema 10.2 *Sea E un espacio afín (proyectivo) de dimensión ≥ 2 sobre un cuerpo ordenado pitagórico y supongamos que en el espacio vectorial \vec{E} hay definido un producto escalar que dota a E_{fin} de estructura de espacio euclídeo. Fijado $O \in E_{\text{fin}}$, sea $p : E_{\infty} \rightarrow \mathcal{H}(E_{\infty})$ la aplicación que a cada $P \in E_{\infty}$ le asigna la intersección con E_{∞} de cualquier hiperplano perpendicular a OP . Entonces p es una polaridad simétrica en E_{∞} que no depende de la elección de O .*

DEMOSTRACIÓN: Si O' es cualquier otro punto finito de E , las rectas OP y $O'P$ son paralelas, luego tienen los mismos hiperplanos perpendiculares, y todos ellos son paralelos entre sí, luego cortan a E_{∞} en un mismo hiperplano $p(P)$. Esto prueba que p está bien definida y no depende de la elección de O .

Fijemos un sistema de referencia ortonormal en E_{fin} , que a su vez determina un sistema de referencia (proyectivo) en E en el que E_{∞} tiene ecuación $x_0 = 0$, de modo que las coordenadas homogéneas de un punto de coordenadas afines (x_1, \dots, x_n) son $(1, x_1, \dots, x_n)$. Tomemos como O el punto de coordenadas $(1, 0, \dots, 0)$. Entonces, si un punto $P \in E_{\infty}$ tiene coordenadas $(0, a_1, \dots, a_n)$, la recta OP está formada por los puntos de coordenadas $(1, 0, \dots, 0) + \lambda(0, a_1, \dots, a_n)$ (donde hemos perdido el punto infinito P) que corresponden a las coordenadas afines $\lambda(a_1, \dots, a_n)$. Como el sistema de referencia es ortonormal, los puntos del plano perpendicular que pasa por O cumplen la ecuación $a_1x_1 + \dots + a_nx_n = 0$, y la intersección con E_{∞} del plano está formada por los puntos $(0, x_1, \dots, x_n)$ que cumplen esta misma ecuación.

Al eliminar la primera coordenada de las coordenadas homogéneas de los puntos de E_{∞} obtenemos sus coordenadas homogéneas respecto de un sistema de referencia de E_{∞} , respecto del cual, el hiperplano $p(P)$ asociado al punto de coordenadas (a_1, \dots, a_n) tiene ecuación $a_1x_1 + \dots + a_nx_n = 0$, lo que significa que p es la polaridad asociada a la matriz identidad. En particular es simétrica. ■

Definición 10.3 Si E es un espacio euclídeo, la polaridad en E_{∞} dada por el teorema anterior recibe el nombre de *polaridad ortogonal* de E .

Así, una variedad afín A (es decir, una variedad proyectiva de X no contenida en E_∞) es perpendicular a una recta r si y sólo si $A_\infty = A \cap E_\infty$ está contenida en el hiperplano polar del punto infinito de r .

Dos puntos P y Q de un espacio proyectivo se dicen *conjugados* respecto de una polaridad p si $P \in p(Q)$ (o, equivalentemente, $Q \in p(P)$). En estos términos, dos rectas secantes r y s son perpendiculares si y sólo si sus puntos infinitos son conjugados respecto a la polaridad ortogonal.

En efecto, si los puntos infinitos de r y s son r_∞ y s_∞ , respectivamente, entonces r es perpendicular a s si y sólo si s está contenida en el hiperplano H perpendicular a r por el punto de corte, si y sólo si $s_\infty \in H_\infty = p(r_\infty)$.

En particular, un punto infinito no puede ser *autoconjugado*, es decir, no puede pertenecer a su hiperplano polar, pues eso significaría que las rectas afines que pasan por dicho punto infinito serían perpendiculares a sí mismas.

Esto nos lleva al concepto de espacio parabólico, que introducimos a continuación:

10.1 Espacios parabólicos

Definición 10.4 Un *espacio parabólico*¹ es un espacio afín (proyectivo) E de dimensión ≥ 2 en el que se ha seleccionado una polaridad simétrica p en E_∞ , a la que llamaremos *polaridad ortogonal*. Un espacio *euclídeo* es un espacio parabólico en el que ningún punto infinito P es *autoconjugado*, es decir, que ningún punto cumple $P \in p(P)$.

Diremos que una variedad afín A en un espacio parabólico es *perpendicular* a una recta r si se cumple que $A \cap r \neq \emptyset$ (admitiendo que la intersección pueda ser un punto infinito) y que la variedad $A_\infty = A \cap E_\infty$ está contenida en el hiperplano polar del punto infinito de r . En particular, una recta s es perpendicular a r si y sólo si sus puntos infinitos son conjugados respecto de la polaridad ortogonal.

Si E es un plano, el teorema 8.62 afirma que las polaridades simétricas de E_∞ coinciden con las involuciones, por lo que en este caso es más frecuente hablar de la *involución ortogonal*.

Así, si I es la involución ortogonal de E , dos rectas r y s son perpendiculares si y sólo si sus puntos infinitos cumplen $I(r_\infty) = s_\infty$.

Una recta afín es *isótropa* (gr. *igual dirección*) si es perpendicular a sí misma o, equivalentemente, si su punto infinito es autoconjugado. En estos términos, un espacio es euclídeo si y sólo si no tiene rectas isótropas.

¹No estamos en condiciones de entender ahora el por qué de este nombre, pero la idea de fondo (aunque ahora no pueda entenderse del todo) es que en una geometría parabólica cada recta tiene un punto infinito, en una geometría elíptica las rectas no tienen puntos infinitos, y en una geometría hiperbólica cada recta tiene dos puntos infinitos.

Obviamente, la existencia de rectas isotropas contradice el concepto intuitivo de perpendicularidad, pero veremos que parte de la geometría parabólica puede desarrollarse sin necesidad de exigir que no haya rectas isotropas, y así obtendremos, por ejemplo, una geometría parabólica sobre cuerpos algebraicamente cerrados, donde veremos que necesariamente existen tales rectas.

Veamos ahora que la traducción a la geometría afín de una polaridad ortogonal es la esperada:

Teorema 10.5 *Si E es un espacio parabólico sobre un cuerpo K de característica $\neq 2$, existe una forma bilineal simétrica regular $F : \vec{E} \times \vec{E} \rightarrow K$ tal que dos rectas con vectores directores \vec{u} y \vec{v} son perpendiculares si y sólo si $F(\vec{u}, \vec{v}) = 0$.*

DEMOSTRACIÓN: Fijemos un sistema de referencia en E , de modo que E_∞ esté formado por los puntos con $x_0 = 0$. La recta que pasa por un punto $(1, a_1, \dots, a_n)$ con vector director \vec{u} de coordenadas (u_1, \dots, u_n) está formada por los puntos finitos de coordenadas

$$(1, a_1, \dots, a_n) + \lambda(0, u_1, \dots, u_n), \quad \lambda \in K,$$

que son los puntos finitos de la recta proyectiva r formada por los puntos de coordenadas homogéneas

$$\mu(1, a_1, \dots, a_n) + \lambda(0, u_1, \dots, u_n), \quad \lambda, \mu \in K.$$

El punto infinito r_∞ es, pues, el de coordenadas homogéneas $(0, u_1, \dots, u_n)$. Al eliminar el 0 inicial obtenemos las coordenadas homogéneas de r_∞ respecto de un sistema de referencia de E_∞ . Sea A la matriz (simétrica, regular) de la polaridad ortogonal en dicho sistema de referencia. Dados dos vectores \vec{u} y \vec{v} de coordenadas (u_1, \dots, u_n) y (v_1, \dots, v_n) , respectivamente, podemos definir

$$F(\vec{u}, \vec{v}) = (u_1, \dots, u_n)A(v_1, \dots, v_n)^t.$$

Claramente, F es una forma bilineal simétrica regular en \vec{E} , y las rectas r y s de vectores directores \vec{u} y \vec{v} son perpendiculares si y sólo si sus puntos infinitos cumplen $s_\infty \in p(r_\infty)$, lo cual equivale a que $F(\vec{u}, \vec{v}) = 0$. ■

Observaciones Aunque en la prueba del teorema anterior hemos definido F en términos de un sistema de referencia elegido arbitrariamente, el teorema 10.1 implica que F está unívocamente determinado por la polaridad ortogonal salvo un factor de escala. La no existencia de puntos infinitos autoconjugados equivale a que $F(\vec{u}, \vec{u}) \neq 0$ para todo vector no nulo. Si K es un cuerpo ordenado euclídeo, entonces, en un sistema de referencia adecuado, F admite una matriz diagonal cuya diagonal está formada por valores 1 y -1 , y es fácil ver entonces que la no existencia de puntos autopolares equivale a que todas las entradas tengan el mismo signo, lo cual equivale a que $\pm F$ (con el signo adecuado) sea un producto escalar.

En suma, los espacios euclídeos en el sentido de 10.4 sobre un cuerpo ordenado euclídeo coinciden con los espacios euclídeos en el sentido del capítulo III salvo por el hecho de que el producto escalar está determinado salvo un factor de escala.

Más aún, es inmediato que el producto escalar del espacio del teorema 10.2 es uno de los que determina la polaridad ortogonal definida en dicho teorema. ■

Vamos a encontrar caracterizaciones de algunos conceptos de la geometría euclídea en términos de la geometría parabólica, algunos de los cuales darán lugar a conceptos análogos en espacios parabólicos arbitrarios, no necesariamente euclídeos. Puesto que vamos a admitir la posibilidad de que un espacio parabólico E tenga puntos infinitos autoconjugados, conviene hacer algunas observaciones generales sobre éstos:

Fijado un sistema de referencia en E_∞ , la polaridad ortogonal vendrá determinada por una matriz regular simétrica A , de modo que el hiperplano polar de un punto P de coordenadas homogéneas $c \in K^n$ es el hiperplano de ecuación $cAx^t = 0$, y los puntos autoconjugados son los que cumplen $cAc^t = 0$, pero esto es la ecuación de una cuádrica no degenerada. Esto nos lleva a la definición siguiente:

Definición 10.6 Si E es un espacio parabólico de dimensión ≥ 2 , llamaremos *cuádrica esférica* a la cuádrica en E_∞ cuya matriz en un sistema de referencia dado es la de la polaridad ortogonal.

Es inmediato que esta definición no depende de la elección del sistema de referencia, pues si pasamos a otro la matriz de una cuádrica y la de una polaridad se transforman igual. El teorema siguiente es inmediato:

Teorema 10.7 Si E es un espacio parabólico de dimensión ≥ 2 , la cuádrica esférica es una cuádrica en E_∞ no degenerada (tal vez vacía) cuyos puntos son los puntos autoconjugados de E , y de modo que la polaridad que induce en E_∞ es la polaridad ortogonal.²

En el caso de un plano parabólico los puntos infinitos autoconjugados son los puntos fijos de la involución ortogonal, de modo que el plano será euclídeo si y sólo si ésta es una involución elíptica.

Ahora es claro que, tal y como habíamos señalado, si el cuerpo de escalares es algebraicamente cerrado (o simplemente si todos sus elementos tienen raíz cuadrada) entonces todo espacio parabólico tiene puntos infinitos autoconjugados, pues las cuádricas son todas no vacías.

Veamos ahora algunas consecuencias elementales de la definición de espacio parabólico:

²El hecho de que hayamos definido las cuádricas como clases de equivalencia de formas bilineales simétricas hace que tenga sentido hablar de la polaridad de una cuádrica aunque ésta pueda no contener puntos.

- Las rectas perpendiculares a un hiperplano afín H son las rectas que pasan por el polo de H_∞ . En particular, por cada punto finito de E pasa una única perpendicular a H .

En efecto, una recta r es perpendicular a H si y sólo si $H_\infty = p(r_\infty)$. En efecto, en principio la definición sólo requiere la inclusión $H_\infty \subset p(r_\infty)$, pero como son dos hiperplanos de E_∞ , la inclusión implica la igualdad. Además, la condición $r \cap H \neq \emptyset$ se cumple siempre. Así pues, r_∞ es necesariamente el polo de H .

- Por cada punto finito P pasa un único hiperplano H perpendicular a una recta afín dada r . Una variedad que pasa por P es perpendicular a r si y sólo si corta a r y está contenida en H .

En efecto, sea r_∞ el punto infinito de r y sea H el hiperplano que contiene a P y a $p(r_\infty)$. Así $H_\infty = p(r_\infty)$, luego H es perpendicular a r por el punto anterior. Una variedad A que pasa por P y corta a r es perpendicular a r si y sólo si $A_\infty \subset p(r_\infty) = H_\infty$, lo cual equivale a que $A = \langle A_\infty, P \rangle \subset H$.

- Por un punto exterior a una recta no isótropa pasa una única perpendicular, que corta a ésta en un punto finito.

Si r es una recta y P un punto exterior, el hiperplano H perpendicular a r por P no puede cortar a r en su punto infinito, pues en tal caso $r_\infty \in H_\infty \subset p(r_\infty)$ sería autoconjugado, luego la intersección es un punto finito $Q \in r$, y entonces PQ es la única perpendicular a r por P .

- Dos rectas paralelas son perpendiculares si y sólo si son isótropas.

Si R es el punto infinito común de las rectas, éstas son perpendiculares si y sólo si $R \in p(R)$, es decir, si y sólo si R es autoconjugado, lo cual equivale a que las rectas sean isótropas.

Veamos ahora que el concepto de “esfera” puede generalizarse a espacios parabólicos arbitrarios:

Teorema 10.8 *Si E es un espacio parabólico y O y P son puntos finitos tales que la recta OP no es isótropa, entonces existe una única cuádrica no degenerada en E de centro³ O que pasa por P y que cuya polaridad induce en E_∞ la polaridad ortogonal.*

DEMOSTRACIÓN: Fijemos un sistema de referencia afín en el que O tenga coordenadas $(0, \dots, 0)$ y P tenga coordenadas $(1, 0, \dots, 0)$. Así, sus coordenadas homogéneas son $(1, 0, \dots, 0)$ y $(1, 1, 0, \dots, 0)$, respectivamente. Además, el punto infinito de OP tiene coordenadas $(0, 1, 0, \dots, 0)$. Vamos a probar que sólo existe una matriz simétrica A tal que la cuádrica que define en este sistema de coordenadas cumple lo requerido. En efecto, tiene que ser de la forma

$$A = \left(\begin{array}{c|c} \alpha & 0 \\ \hline 0 & A' \end{array} \right),$$

³El centro de una cuádrica afín se define igual que el de una cónica, es decir, como el polo del hiperplano infinito.

donde A' es la matriz de la polaridad ortogonal en el sistema de referencia de E_∞ respecto al que las coordenadas homogéneas de un punto infinito son las que resultan de eliminar la primera coordenada homogénea (nula) respecto del sistema que hemos fijado. En efecto, la presencia de la matriz A' en A se ve en la demostración de 8.45, y los ceros son la condición necesaria y suficiente para que O sea el centro de la cónica (al multiplicar la matriz por sus coordenadas $(1, 0, \dots, 0)$ debemos obtener un vector $(\alpha, 0, \dots, 0)$, para que la ecuación del hiperplano polar sea $x_0 = 0$).

Falta imponer que P esté en la cuádrlica determinada por la matriz, lo cual equivale a que $(1, 1, 0, \dots, 0)A(1, 1, 0, \dots, 0)^t = 0$. A su vez, esto equivale a que $\alpha + \beta = 0$, donde β es el primer coeficiente de A' . Esto nos da un único valor posible para α , pero que sólo será válido si $\alpha \neq 0$, es decir, si $\beta \neq 0$, porque en caso contrario la matriz A sería singular y la cuádrlica sería degenerada. Ahora bien, $\beta = (1, 0, \dots, 0)A'(1, 0, \dots, 0)^t \neq 0$, ya que lo contrario significaría que el punto infinito de OP sería autoconjugado, en contra de la hipótesis de que la recta no es isótropa. ■

Definición 10.9 Si E es un espacio parabólico, llamaremos *esferas*⁴ en E a las cuádrlicas no degeneradas de centro finito que inducen en E_∞ la polaridad ortogonal. En los planos parabólicos las esferas se llaman *circunferencias*.

He aquí una caracterización más simple de las esferas:

Teorema 10.10 *En un espacio parabólico E , una cuádrlica no degenerada \mathcal{C} es una esfera si y sólo si $\mathcal{C} \cap E_\infty$ es la cuádrlica esférica.*

DEMOSTRACIÓN: Si \mathcal{C} es una esfera, entonces tiene centro finito, y si $\mathcal{C} \cap E_\infty$ es la cuádrlica esférica (que es no degenerada) entonces E_∞ no puede ser tangente a \mathcal{C} (por el teorema 9.5), luego el centro $p_{\mathcal{C}}(E_\infty)$ tiene que ser finito. Así pues, podemos probar la equivalencia bajo el supuesto de que \mathcal{C} es una cuádrlica de centro finito O . Tomamos un sistema de referencia en el que E_∞ tenga ecuación $x_0 = 0$ y cuyo origen de coordenadas afines sea O . Esto se traduce en que la matriz de \mathcal{C} será de la forma

$$A = \left(\begin{array}{c|c} \alpha & 0 \\ \hline 0 & A' \end{array} \right).$$

Así \mathcal{C} es una esfera si y sólo si A' es la matriz de la polaridad ortogonal, lo cual equivale también a que $\mathcal{C} \cap E_\infty$ sea la cuádrlica de matriz A' , es decir, la cuádrlica esférica. ■

⁴Es fácil ver que si E es un espacio euclídeo y p es la polaridad dada por el teorema 10.2, las esferas que acabamos de definir coinciden con el concepto usual de esfera. En efecto, fijado un sistema de referencia ortonormal con origen en un punto O , la ecuación de una esfera de centro O es $x_1^2 + \dots + x_n^2 = r^2$, luego su matriz es la matriz diagonal $[-r^2, 1, \dots, 1]$, luego tiene centro O en el sentido proyectivo y la polaridad que induce en E_∞ tiene matriz identidad, que es precisamente la matriz de la polaridad ortogonal, tal y como se ve en la prueba de 10.2. Por lo tanto, toda esfera usual es una esfera en el sentido de la geometría parabólica. Como existe una única esfera usual y una única esfera parabólica con un centro dado y que pasa por un punto dado, ambos conceptos son equivalentes.

En particular, en un plano parabólico con involución ortogonal hiperbólica, las circunferencias son exactamente las cónicas no degeneradas que pasan por los dos puntos infinitos autoconjugados. Por esta razón se conocen también como *puntos circulares*.

Observemos que las circunferencias en un plano euclídeo son elipses (porque no contienen puntos infinitos), mientras que en un plano con involución ortogonal hiperbólica, las circunferencias son hipérbolas.

Hemos probado que existe una única esfera con un centro finito dado O que pasa por un punto finito dado P , con la única restricción de que la recta OP no sea isótropa. Esto no es ninguna restricción en un espacio euclídeo, pero si hay rectas isótropas la condición es necesaria:

Dada una esfera \mathcal{C} de centro (finito) O en un espacio parabólico y un punto infinito autoconjugado P_∞ , tenemos que $P_\infty \in E_\infty = p_{\mathcal{C}}(O)$, y por consiguiente $O \in p_{\mathcal{C}}(P_\infty)$, luego $OP_\infty \subset p_{\mathcal{C}}(P_\infty)$, luego la recta OP_∞ es tangente a \mathcal{C} por P_∞ . Si contuviera un punto finito de \mathcal{C} , entonces estaría contenida en \mathcal{C} , luego $O \in \mathcal{C}$, luego $O \in p_{\mathcal{C}}(O) = E_\infty$, en contradicción con que O era un punto finito.

En otras palabras: toda recta isótropa r que pasa por el centro de una esfera es tangente a ésta por r_∞ , y es una tangente propia, porque en caso contrario el centro sería un punto infinito, luego una esfera corta a una recta isótropa que pase por su centro únicamente en su punto infinito.

En particular, en un plano parabólico no euclídeo, las circunferencias son hipérbolas y las rectas isótropas que pasan por su centro son sus asíntotas.

También es fácil caracterizar las semejanzas euclídeas (a la vez que extendemos el concepto a espacios parabólicos arbitrarios):

Definición 10.11 Llamaremos *grupo parabólico* de un espacio parabólico E al grupo de todas las biyecciones afines $f \in \text{GA}(E)$ que conservan la polaridad ortogonal, es decir, que cumplen $p(f(P)) = f[p(P)]$ para todo $P \in E_\infty$. Lo representaremos por $\text{Sem}(E)$ y a sus elementos los llamaremos *semejanzas* de E .

Equivalentemente, las semejanzas de E son las biyecciones afines que transforman rectas perpendiculares en rectas perpendiculares.⁵

En efecto, si f cumple esta condición, $P \in E_\infty$ y $Q \in p(P)$, tomamos cualquier punto finito R y consideramos las rectas perpendiculares RP y RQ . Entonces $f[RP]$ y $f[RQ]$ también son perpendiculares, lo que significa que $f(Q) \in p(f(P))$, luego $f[p(P)] \subset p(f(P))$. Como son dos hiperplanos de E_∞ , se da la igualdad. El recíproco es claro. ■

Observemos que las homotecias y las traslaciones de E son semejanzas, pues son las homografías que se restringen a la identidad en E_∞ , luego trivialmente conservan la polaridad ortogonal.

⁵Teniendo en cuenta el teorema 5.24, de aquí se sigue que las semejanzas en este sentido de un espacio euclídeo respecto de la estructura euclídea dada por el teorema 10.2 son las semejanzas en el sentido usual.

Para caracterizar las isometrías nos valdremos del teorema 5.36, según el cual las isometrías están generadas por las reflexiones, las cuales admiten una caracterización sencilla:

Definición 10.12 Una *reflexión* en un espacio parabólico E es una homología de orden 2 que fija al hiperplano infinito E_∞ , cuyo centro C es un punto infinito y cuyo eje H cumple $H_\infty = H \cap E_\infty = p(C)$.

Recordemos que, por definición de homología, el centro C no está en el eje, luego éste no puede ser el hiperplano infinito. Más aún, como $C \in E_\infty$, $C \notin H$ y $p(C) = H_\infty$, necesariamente $C \notin p(C)$, es decir, que el centro no es autoconjugado. Equivalentemente, el eje H no es un *hiperplano isótropo*, en el sentido de que no contiene al polo de H_∞ (que es C). Recíprocamente:

Teorema 10.13 Sea E un espacio parabólico y sea H un hiperplano afín en E que no sea isótropo (es decir, que no contenga al polo de H_∞). Entonces existe una única reflexión⁶ en E de eje H .

DEMOSTRACIÓN: Veamos primero la unicidad. Si f es una reflexión de centro C y eje H , necesariamente $C = p(H_\infty)$. Dado un punto finito $P \in E \setminus H$, la recta $r = CP$ corta a H en un punto P que no puede ser C (porque no está en H), luego es finito. Notemos que r es la única perpendicular a H que pasa por P . Por definición de homología sabemos que f fija a r , es decir, que se restringe a una involución en r con dos puntos fijos P y C (no puede ser la identidad, pues entonces f fijaría a todos los puntos de r y de H , luego sería la identidad). Por el teorema 8.63, la imagen de cualquier otro punto $A \in r$ es necesariamente su conjugado armónico respecto de (P, C) , es decir, el único punto B tal que P es el punto medio de A y B .

Esto prueba que f está unívocamente determinada por H : fija a los puntos de H y la imagen de cualquier punto finito P que no esté en H es su simétrico respecto al pie de la perpendicular a H por P .

Sólo falta probar que realmente existe una homografía con estas características: tomamos $n - 1$ puntos $P_1, \dots, P_{n-1} \in H \cap E_\infty$ proyectivamente independientes, añadimos $P_n = C$ y completamos con un punto $O = P_0 \in H \setminus E_\infty$. Así tenemos $n + 1$ puntos proyectivamente independientes en E , que completamos a un sistema de referencia proyectivo con cualquier punto $P_{n+1} \in E$ que cumpla lo necesario. Respecto a este sistema, estos puntos tienen coordenadas

$$O = (1, 0, \dots, 0), P_1 = (0, 1, \dots, 0), \dots, P_n = (0, \dots, 0, 1).$$

Por lo tanto, el hiperplano infinito tiene ecuación $x_0 = 0$ y el eje H tiene ecuación $x_n = 0$. En coordenadas afines, O es el origen de coordenadas, H el hiperplano $x_n = 0$ y la recta OC perpendicular a H por O es el eje x_n . Basta considerar la biyección afín f dada por $(x_1, \dots, x_n) \mapsto (x_1, \dots, x_{n-1}, -x_n)$. Es

⁶La prueba muestra que, en el caso de un espacio euclídeo dado por el teorema 10.2, las reflexiones en el sentido de la definición que acabamos de dar coinciden con las usuales, es decir, con las simetrías respecto de hiperplanos.

claro que fija a todos los puntos de H y a todos los hiperplanos que pasan por C (que son los que tienen ecuaciones de la forma $a_1x_1 + \dots + a_{n-1}x_{n-1} = b$), y además tiene orden 2, luego es una homología de eje H . ■

Teorema 10.14 *Las reflexiones son semejanzas.*

DEMOSTRACIÓN: Toda reflexión admite una expresión en coordenadas como la considerada en la demostración del teorema anterior, en la que el hiperplano infinito tiene ecuación $x_0 = 0$, el eje H tiene ecuación $x_n = 0$ y el centro C tiene coordenadas $(0, \dots, 0, 1)$. Al eliminar la primera coordenada homogénea (nula) de un punto infinito obtenemos sus coordenadas homogéneas respecto de un sistema de referencia de E_∞ . Sea A la matriz (simétrica, regular, de dimensión $n \times n$) de la polaridad ortogonal en dicho sistema de referencia. Como $p(C) = H_\infty$, tiene que cumplirse que $(0, \dots, 0, 1)A = (0, \dots, 0, 1)$, luego

$$A = \left(\begin{array}{c|c} A^* & 0 \\ \hline 0 & 1 \end{array} \right),$$

para cierta matriz A^* . Por lo tanto, $xAy^t = \bar{x}A^*\bar{y}^t + x_ny_n$, donde \bar{x}, \bar{y} son los vectores que resultan de eliminar la última coordenada a x, y , respectivamente.

Por otra parte, la expresión de la reflexión f en coordenadas homogéneas es $f(x_0, \dots, x_n) = (x_0, \dots, x_{n-1}, x_n)$, luego su restricción a E_∞ viene dada por $f(x) = (\bar{x}, -x_n)$, y así es claro que si dos puntos infinitos P y Q de coordenadas x, y cumplen $Q \in p(P)$, es decir, $xAx^t = 0$, entonces también $(\bar{x}, -x_n)A(\bar{y}, -y_n)^t = 0$, luego $f(Q) \in p(f(P))$. Por lo tanto $f[p(P)] \subset p(f(P))$, y como son hiperplanos, se tiene la igualdad. ■

Definición 10.15 Definimos el *grupo de isometrías* de un espacio parabólico E como el grupo generado por las reflexiones.⁷ Lo representaremos por $\text{Is}(E)$.

Por el teorema anterior sabemos que $\text{Is}(E)$ es un subgrupo de $\text{Sem}(E)$.

Ya hemos señalado que las homotecias y las traslaciones son semejanzas. Veamos ahora que las traslaciones son también isometrías:

Teorema 10.16 *El producto de dos reflexiones con ejes paralelos es una traslación. Toda traslación cuyo centro no sea un punto autoconjugado es producto de dos reflexiones de ejes paralelos. Si el centro es autoconjugado entonces la traslación es producto de cuatro reflexiones. En particular todas las traslaciones son isometrías.*

DEMOSTRACIÓN: Dada una reflexión f , en la prueba del teorema 10.13 hemos visto que, en un sistema de referencia afín adecuado, su expresión en coordenadas es $f(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}, -x_n)$, de modo que su eje es el

⁷De acuerdo con el teorema 5.36, las isometrías en este sentido son las mismas que las definidas en términos de un producto escalar para espacios euclídeos sobre cuerpos realmente cerrados.

hiperplano $x_n = 0$. Si otra reflexión g tiene un eje paralelo, entonces, en ese mismo sistema de referencia el eje tendrá ecuación $x_n = c$, para cierto escalar c , y es fácil ver entonces que la expresión en coordenadas es $g(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}, 2c - x_n)$ (se comprueba fácilmente que g cumple la definición de reflexión con el hiperplano $x_n = c$ como eje, y por la unicidad tiene que ser la reflexión dada). Entonces, $(fg)(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}, x_n + 2c)$ es claramente una traslación.

Recíprocamente, si T es una traslación de centro C no autoconjugado, tomamos cualquier punto finito O , de modo que la recta OC no es isótropa. Si H es el hiperplano perpendicular a OC que pasa por O , tenemos que $H_\infty = p(C)$, luego $C \notin H$, luego OC no está contenida en H , sino que $OC \cap H = O$. Es fácil construir entonces un sistema de referencia afín respecto del cual H tenga ecuación $x_n = 0$ y OC sea el eje x_n . Esto implica que la expresión coordenada de T es $T(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}, x_n + d)$. Llamando⁸ $c = d/2$ las reflexiones f y g con las expresiones coordenadas de la parte precedente cumplen que $fg = T$.

Si T es una traslación, pero su centro C es autoconjugado, tomamos una recta r que pase por C y sea secante a la cuádrlica esférica. Entonces r sólo contiene dos puntos autoconjugados, uno de los cuales es C . Tomemos dos puntos A y B en r que no sean autoconjugados y sea O un punto finito cualquiera, sea $P = T(O) \in OC$ y sea $Q = AP \cap OB$ (las rectas se cortan porque están en el mismo plano OAB , y la intersección es finita porque $A \neq B$). Entonces $\vec{OP} = \vec{OQ} + \vec{QP}$, luego T es la composición de las traslaciones definidas por los otros dos vectores, cuyos centros son A y B , no autoconjugados. Por el caso precedente ambas traslaciones se descomponen en dos reflexiones cada una, luego T es composición de cuatro reflexiones. ■

Con esto podemos caracterizar la congruencia de segmentos y de vectores:

Dados cuatro puntos finitos A, B, C, D en un espacio parabólico E , podemos definir $\overline{AB} \equiv \overline{CD}$ como que existe una isometría $f: E \rightarrow E$ tal que $f(A) = C$ y $f(B) = D$.

La misma definición vale para $\vec{AB} \equiv \vec{CD}$, sólo que ahora hay que justificar que la definición no depende de la elección de los pares de puntos que representan a los vectores, pero esto se debe a que si tenemos $\vec{AB} = \vec{A'B'}$, entonces la traslación de vector $\vec{AA'}$ es una congruencia entre (A, B) y (A', B') , luego no importa qué par tomamos para compararlo con el de otro vector.

10.2 Espacios euclídeos

Los resultados de la sección precedente muestran que en el estudio de la geometría euclídea el producto escalar puede sustituirse por una polaridad ortogonal sin puntos autoconjugados, y que algunos conceptos y resultados son

⁸Recordemos que sólo consideramos espacios proyectivos sobre cuerpos de característica distinta de 2.

generalizables (aunque sea parcialmente) a espacios parabólicos arbitrarios. La razón por la que hemos estudiado espacios parabólicos en general, sin excluir la presencia de rectas isotropas, es porque éstas aparecen cuando se considera la complexificación de un espacio euclídeo:

La complexificación de un espacio euclídeo Sabemos que todo espacio euclídeo sobre un cuerpo ordenado euclídeo R puede identificarse con el conjunto de los puntos finitos del espacio proyectivo $E = P^n(R)$ respecto a un hiperplano infinito E_∞ seleccionado arbitrariamente. La estructura euclídea de E determina una polaridad p en E_∞ sin puntos autoconjugados con la que E se convierte en un espacio parabólico.

A su vez, en las páginas 351 y siguientes hemos visto que E puede identificarse con el conjunto de los puntos reales del espacio complejo $E_C = P^n(C)$, y que el hiperplano infinito E_∞ se identifica con el conjunto de los puntos reales de un único hiperplano de E_C , que seguiremos llamando E_∞ . A su vez, la polaridad ortogonal p se extiende a una única polaridad del hiperplano infinito extendido (la dada por la misma matriz respecto a cualquier sistema de referencia real) que seguiremos llamando p , y que dota a E_C de estructura de espacio parabólico, sólo que en E_C la cuádrica esférica ya no es vacía, por lo que hay puntos autoconjugados.

De este modo, al espacio euclídeo de partida, no sólo le hemos “añadido” puntos infinitos, sino también puntos imaginarios, pero es importante tener presente que con esto no hemos perdido nada de generalidad. Todo espacio euclídeo (entendido como espacio afín) se puede sumergir en un espacio parabólico con puntos infinitos e imaginarios, de modo que al estudiar un espacio euclídeo sumergido en un espacio parabólico complejo estamos estudiando realmente un espacio euclídeo arbitrario.

Si fijamos en E un sistema de referencia ortonormal de modo que E_∞ sea el hiperplano $x_0 = 0$, una esfera de centro el origen de coordenadas está formado por los puntos cuyas coordenadas afines cumplen una ecuación de la forma

$$x_1^2 + \cdots + x_n^2 = r^2,$$

luego las coordenadas homogéneas de sus puntos cumplen la ecuación

$$-r^2x_0^2 + x_1^2 + \cdots + x_n^2 = 0,$$

luego la cuádrica esférica, que es la intersección de la esfera con el hiperplano infinito, tiene ecuación

$$x_1^2 + \cdots + x_n^2 = 0,$$

que es, pues, independiente del sistema de referencia ortonormal elegido.

Por ejemplo, en el caso de un plano euclídeo, considerando que la recta infinita es $z = 0$, la ecuación es $x^2 + y^2 = 0$, luego los puntos circulares son los de coordenadas homogéneas $(1, \pm i, 0)$.

Un concepto útil a la hora de tratar con puntos imaginarios es el siguiente:

Fijado un sistema de referencia proyectivo (en E , es decir, formado por puntos reales), si un punto P tiene coordenadas $x = (x_0, \dots, x_n)$, definimos su *conjugado complejo* como el punto \bar{P} de coordenadas $\bar{x} = (\bar{x}_0, \dots, \bar{x}_n)$, donde la barra indica el R -automorfismo de C que intercambia i con $-i$.

Es claro que la definición no depende de la elección de las coordenadas homogéneas de P , y tampoco de la elección del sistema de referencia, pues si fijamos otro (siempre formado por puntos reales), existirá una matriz real M tal que las coordenadas de P en el nuevo sistema serán xM , y al conjugarlas obtenemos $\bar{x}M$, que son las coordenadas de \bar{P} en el nuevo sistema de referencia.

Por ejemplo, hemos visto que los puntos circulares de un plano euclídeo son conjugados complejos, pues sus coordenadas en un sistema de referencia ortonormal son $(1, \pm i, 0)$.

Obviamente, $\overline{\bar{P}} = P$ y un punto P es real si y sólo si $P = \bar{P}$. Esto último no es difícil de ver, pero no es trivial. En efecto, en principio, la igualdad $P = \bar{P}$ significa que las coordenadas de P cumplen $\bar{x} = \lambda x$, para cierto $\lambda \in C$ no nulo. Tomemos una coordenada no nula x_i , de modo que $\bar{x}_i = \lambda x_i$, luego $x_i = \lambda \bar{x}_i = \lambda \lambda x_i$, luego $\lambda \lambda = 1$. Tomemos ahora un $\mu \in C$ tal que $\mu^2 = \lambda$, con lo que $(\mu \bar{\mu})^2 = 1$, luego $\mu \bar{\mu} = 1$ (pues $\mu \bar{\mu} \geq 0$). Entonces $\bar{\mu} \bar{x} = \bar{\mu} \lambda x = \mu x$ es un vector de coordenadas homogéneas reales para P .

Similarmente podemos definir la conjugación de una variedad proyectiva $V = \langle P_1, \dots, P_d \rangle$ como la variedad $\bar{V} = \langle \bar{P}_1, \dots, \bar{P}_d \rangle$. Es fácil ver que $P \in V$ si y sólo si $\bar{P} \in \bar{V}$, lo que prueba que la definición no depende del generador elegido.

En particular, el conjugado de un hiperplano es el hiperplano que resulta de conjugar los coeficientes de su ecuación respecto de cualquier sistema de referencia real.

Similarmente, definimos la conjugación $\bar{\mathcal{C}}$ de una cuádrica \mathcal{C} como la cuádrica que resulta de conjugar los coeficientes de su ecuación respecto de un sistema de referencia (real) prefijado. Una vez más, la elección del sistema de referencia es irrelevante, pues si la ecuación de \mathcal{C} en un sistema de referencia es $xAx^t = 0$, su ecuación respecto de otro sistema de referencia es $xMAM^t x^t = 0$, donde M es una matriz real, y al conjugar esta ecuación obtenemos $xM\bar{A}M^t x^t = 0$, donde \bar{A} es la matriz que resulta de conjugar los coeficientes de A , y ésta es la ecuación correspondiente a la cuádrica $x\bar{A}x^t = 0$ en el nuevo sistema de referencia.

Así, se cumple también que $P \in \mathcal{C}$ si y sólo si $\bar{P} \in \bar{\mathcal{C}}$. Por el mismo razonamiento que en el caso de los puntos, un hiperplano o una cuádrica son reales (es decir, admiten una ecuación con coeficientes reales) si y sólo si coinciden con su conjugado.

Dualización de teoremas euclídeos Una de las razones por las que tiene interés plantear la geometría euclídea en términos de la geometría parabólica es que veremos que este enfoque puede adaptarse para construir geometrías no

euclídeas (la geometría hiperbólica y la geometría elíptica). De momento vamos a ver una aplicación más elemental e inmediata, y es que el punto de vista proyectivo permite en ocasiones aplicar el principio de dualidad a la geometría euclídea. La idea es tomar un teorema de la geometría euclídea, generalizarlo a un teorema de la geometría proyectiva, considerar el enunciado dual y particularizarlo de nuevo a la geometría euclídea. Consideremos por ejemplo:

Las alturas de un triángulo se cortan en un punto.

El siguiente enunciado proyectivo es equivalente al teorema clásico que acabamos de citar:

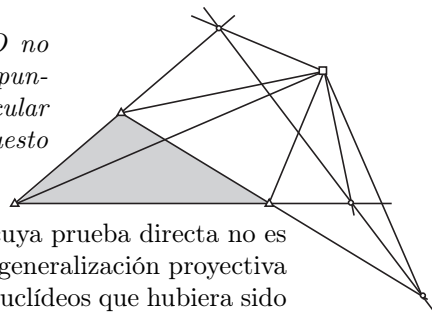
Si T es un triángulo en un plano sobre un cuerpo ordenado euclídeo, r es una recta que no contiene a ninguno de sus vértices e I es una involución elíptica en r , entonces las rectas que unen cada vértice de T con el conjugado de la intersección con r del lado opuesto son concurrentes.

Para probarlo basta tomar a r como recta infinita y a I como involución ortogonal. Entonces las rectas descritas son las alturas del triángulo. Ahora consideramos el enunciado dual. Notemos que una polaridad transforma una involución en una recta en una involución en el haz de rectas de su polo. Así pues, queda:

Si T es un triángulo en un plano proyectivo sobre un cuerpo ordenado euclídeo, O es un punto exterior a sus lados e I es una involución elíptica en el haz de rectas de O , entonces los puntos de corte entre cada lado de T y la recta conjugada a la que une O y el vértice opuesto son colineales.

Ahora particularizamos este enunciado al caso en que I es la involución en O que hace corresponder cada recta con su perpendicular (la inducida en el haz de O desde la recta infinita). El resultado es relativamente simple:

Dado un triángulo T y un punto O no contenido en ninguno de sus lados, los puntos donde cada lado corta a la perpendicular a la recta que une O con el vértice opuesto son colineales.



Obtenemos así un enunciado “dual” cuya prueba directa no es evidente. De hecho, por este método de generalización proyectiva y dualización se han obtenido teoremas euclídeos que hubiera sido difícil descubrir de otro modo.

La polaridad de una esfera Vamos a describir las polaridades de las esferas en un espacio euclídeo, es decir, las polaridades en la parte finita del espacio de determinan la polaridad ortogonal en el hiperplano infinito.

Consideremos una esfera \mathcal{C} de centro O y radio r y tomemos un punto finito Q distinto de O (pues ya sabemos que el hiperplano polar de O es el hiperplano infinito). Sea s_∞ el punto infinito de la recta $s = OQ$ y sea $H = p_e(Q)$. El hecho de que la polaridad de la esfera induce la polaridad ortogonal significa que $H_\infty = H \cap E_\infty = p(s_\infty)$, de modo que H es perpendicular a s . Para determinarlo completamente sólo hay que determinar qué punto es $H \cap s$, es decir, el conjugado de Q respecto de la involución que p_e induce en s . Esta involución está determinada por que fija a los dos puntos en los que s corta a \mathcal{C} .

Sea $\vec{v} = \vec{OQ} / \|\vec{OQ}\|$, de modo que (O, \vec{v}) es un sistema de referencia afín de s en el que la coordenada de Q es $\|\vec{OQ}\|$. Respecto de este sistema de referencia, los puntos de $\mathcal{C} \cap s$ tienen coordenadas $\pm r$, y la única involución que los fija es la que tiene ecuación

$$y = \frac{r^2}{x},$$

pues, efectivamente, ésta cumple las condiciones y es única. Vemos, por tanto, que si $p(Q) \cap OQ = R$, entonces, como la coordenada de R es $\pm \|\vec{OR}\|$, se cumple que $\pm \|\vec{OQ}\| \|\vec{OR}\| = r^2$, luego el signo tiene que ser positivo. Así pues, Q y R están en la misma semirrecta respecto de O y satisfacen la relación $\|\vec{OQ}\| \|\vec{OR}\| = r^2$. Esto caracteriza completamente al hiperplano polar.

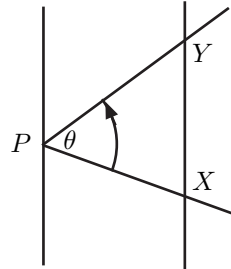
Definición 10.17 Diremos que dos puntos Q y R son mutuamente *inversos* respecto a la esfera de centro O y radio r si ambos se encuentran sobre la misma semirrecta de origen O y además $\|\vec{OQ}\| \|\vec{OR}\| = r^2$. Convenimos en que O es el inverso de todos los puntos infinitos.

En estos términos, el hiperplano polar de un punto finito Q respecto a una esfera de centro O es el hiperplano perpendicular a la recta OP que pasa por el inverso de P .

Una fórmula para la razón doble Vamos a probar una conocida fórmula para la razón doble de cuatro puntos A, B, C, D sobre una recta r . Si P es cualquier punto exterior a r entonces

$$\mathcal{R}(A, B, C, D) = \frac{\widehat{\text{sen APC}}}{\widehat{\text{sen APD}}} : \frac{\widehat{\text{sen BPC}}}{\widehat{\text{sen BPD}}}.$$

Es importante aclarar que los senos deben tomarse con el signo apropiado: Dados tres puntos no colineales P, X, Y y fijado un sistema de referencia ortonormal de origen P , existe un único ángulo $\theta \in]-\pi, \pi[$ tal que el giro de ángulo θ transforma la semirrecta \vec{PX} en \vec{PY} . Definimos $\widehat{\text{sen XPY}} = \text{sen } \theta$. El signo de θ depende del sistema de referencia, pero el miembro derecho de la fórmula es independiente de éste pues, si lo cambiamos, todos los signos se conservan o todos cambian, y en ambos casos la expresión queda inalterada.



Si en particular tomamos el sistema de referencia con el segundo eje paralelo a una recta r que no contenga a P , es fácil ver que podemos escoger una ordenación en r de modo que para todo par de puntos X, Y en r , el signo de \widehat{XPY} es positivo o negativo según si $X < Y$ o $Y < X$. Teniendo esto en cuenta así como la expresión

$$\mathcal{R}(A, B, C, D) = \frac{\overrightarrow{AC}}{\overrightarrow{AD}} : \frac{\overrightarrow{BC}}{\overrightarrow{BD}},$$

es claro que los signos de los dos miembros de la fórmula que queremos probar son iguales, y basta probar que los valores absolutos también lo son.

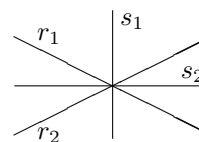
Recordemos que (ABC) es el área del triángulo \widehat{ABC} . Claramente tenemos

$$\begin{aligned} |\mathcal{R}(A, B, C, D)| &= \frac{\overline{AC}}{\overline{AD}} : \frac{\overline{BC}}{\overline{BD}} = \frac{(APC)(BPD)}{(APD)(BPC)} \\ &= \left| \frac{\overline{AP} \overline{CP} \widehat{APC} \overline{BP} \overline{DP} \widehat{BPD}}{\overline{AP} \overline{DP} \widehat{APD} \overline{BP} \overline{CP} \widehat{BPC}} \right| \\ &= \left| \frac{\widehat{APC}}{\widehat{APD}} : \frac{\widehat{BPC}}{\widehat{BPD}} \right|. \end{aligned}$$

■

Bisectrices Ahora caracterizaremos la bisectriz de un ángulo en términos proyectivos. Ante todo conviene definir *la razón doble* de cuatro rectas concurrentes como la de los puntos donde éstas cortan a una recta cualquiera. La elección de la recta es irrelevante, pues dos cuádruplas correspondientes a dos elecciones distintas están en perspectiva. Es claro que la razón doble es un invariante respecto a homografías entre haces. A su vez diremos que dos pares de rectas en un haz están separadas armónicamente si su razón doble es -1 .

Dadas dos rectas cualesquiera r_1 y r_2 concurrentes en un punto P , por el teorema 8.76 (que se traduce inmediatamente a haces de rectas) existen dos únicas rectas perpendiculares s_1 y s_2 en el haz \mathcal{H}_P que separan armónicamente a r_1 y r_2 . La reflexión de eje s_1 induce una involución en el



haz que fija a s_1 y s_2 , luego ha de hacer corresponder r_1 con r_2 , y lo mismo vale para la reflexión de eje s_2 . Esto implica que cada una de las rectas s_1 y s_2 es la bisectriz de dos de los ángulos opuestos por el vértice que determinan r_1 y r_2 .

Así pues, las bisectrices de un par de rectas concurrentes son el único par de rectas perpendiculares del haz formado en el plano que las contiene por su punto de intersección que separan armónicamente a las rectas dadas. ■

La fórmula de Cayley Veamos una importante fórmula que relaciona el ángulo entre dos rectas con cierta razón doble imaginaria. Dadas dos rectas r_1 y r_2 concurrentes en un punto O , podemos razonar en el plano que las contiene. Existen dos giros de centro O que transforman una en la otra. Fijado

un sistema de referencia ortonormal de centro O , cada giro tiene asociado un ángulo módulo 2π (que en realidad es independiente del sistema de referencia salvo por el signo). Si uno de los giros que transforman r_1 en r_2 es el de ángulo α , el otro es el de ángulo $\alpha + \pi$. Definimos el ángulo entre r_1 y r_2 como éste valor de α , que está determinado módulo π , y lo representaremos por $\widehat{r_1 r_2}$. Si tomamos $-\pi/2 \leq \widehat{r_1 r_2} \leq \pi/2$ entonces el valor absoluto es el menor ángulo que forman las dos rectas, mientras que el signo indica si para girar r_1 hasta superponerla a r_2 —recorriendo el menor de los dos ángulos posibles—hemos de girar en sentido horario (argumentos decrecientes) o antihorario (argumentos crecientes). Este signo depende del sistema de referencia respecto al cual definimos los argumentos.

Una caracterización más manejable del ángulo entre dos rectas es la siguiente: tomamos un argumento θ_i de un punto de r_i distinto de O . Esto significa que el vector $(\cos \theta_i, \text{sen } \theta_i)$ tiene la dirección de r_i . Es claro que θ_i está determinado módulo π . Su interpretación es que un giro de ángulo θ_i transforma el primer eje coordenado en r_i , luego un giro de ángulo $-\theta_i$ seguido de un giro de ángulo θ_2 transformará r_1 en el eje coordenado y luego en r_2 . Así pues, $\widehat{r_1 r_2} = \theta_2 - \theta_1$.

Considerando las coordenadas homogéneas asociadas al sistema de referencia elegido, es claro que los puntos infinitos de r_i son $(\cos \theta_i, \text{sen } \theta_i, 0)$. Por otra parte, hemos visto que en todo sistema de referencia ortonormal los puntos circulares tienen coordenadas $(1, \pm i, 0)$. Sean I_1 e I_2 las rectas imaginarias que pasan por O y por los puntos circulares.

Calculamos la razón doble $\mathcal{R}(r_1, r_2, I_1, I_2)$ usando los puntos donde las cuatro rectas cortan a la recta infinita, que en coordenadas cartesianas son:

$$\frac{\text{sen } \theta_1}{\cos \theta_1}, \quad \frac{\text{sen } \theta_2}{\cos \theta_2}, \quad -i, \quad i.$$

Si llamamos $\theta = \theta_2 - \theta_1 = \widehat{r_1 r_2}$ tenemos que

$$\begin{aligned} \mathcal{R}(r_1, r_2, I_1, I_2) &= \frac{-i - \frac{\text{sen } \theta_1}{\cos \theta_1}}{i - \frac{\text{sen } \theta_1}{\cos \theta_1}} : \frac{-i - \frac{\text{sen } \theta_2}{\cos \theta_2}}{i - \frac{\text{sen } \theta_2}{\cos \theta_2}} \\ &= \frac{\cos \theta_1 \cos \theta_2 + \text{sen } \theta_1 \text{sen } \theta_2 + i(\cos \theta_1 \text{sen } \theta_2 - \text{sen } \theta_1 \cos \theta_2)}{\cos \theta_1 \cos \theta_2 + \text{sen } \theta_1 \text{sen } \theta_2 - i(\cos \theta_1 \text{sen } \theta_2 - \text{sen } \theta_1 \cos \theta_2)} \\ &= \frac{\cos(\theta_2 - \theta_1) + i \text{sen}(\theta_2 - \theta_1)}{\cos(\theta_2 - \theta_1) - i \text{sen}(\theta_2 - \theta_1)} = \cos^2 \theta - \text{sen}^2 \theta + 2i \cos \theta \text{sen } \theta \\ &= \cos 2\theta + i \text{sen } 2\theta. \end{aligned}$$

En definitiva:

$$\widehat{r_1 r_2} = \frac{1}{2} \arg \mathcal{R}(r_1, r_2, I_1, I_2),$$

que está determinado módulo π , como ya sabíamos.⁹ Es claro que el segundo miembro de esta fórmula es independiente del sistema de referencia salvo por un

⁹La expresión habitual de esta fórmula es $\widehat{r_1 r_2} = -\frac{i}{2} \log \mathcal{R}(r_1, r_2, I_1, I_2)$, donde \log representa al logaritmo complejo.

detalle: si cambiamos de orden I_1, I_2 la razón doble se invierte y su argumento cambia de signo. Para que la expresión de la derecha nos dé el ángulo entre las dos rectas con el signo adecuado a un sistema de referencia dado hemos de llamar I_1 al punto circular cuyas coordenadas son $(-i, 1, 0)$ y no $(i, 1, 0)$. Notemos además que el segundo miembro no varía si sustituimos una recta r_i por una paralela.

Esta fórmula puede usarse como definición del ángulo entre dos rectas (o dos vectores) en un espacio euclídeo, alternativamente a la que proporciona el producto escalar.

A partir de la fórmula de la razón doble es inmediato que

$$\mathcal{R}(r_1, r_2, I_1, I_2)\mathcal{R}(r_2, r_3, I_1, I_2) = \mathcal{R}(r_1, r_3, I_1, I_2),$$

de donde $\widehat{r_1 r_3} = \widehat{r_1 r_2} + \widehat{r_2 r_3}$, o también $\widehat{r_1 r_2} + \widehat{r_2 r_3} + \widehat{r_3 r_1} = 0$, lo que significa que los tres ángulos suman π . De ahí es fácil deducir que los ángulos de un triángulo suman π .

10.3 Cónicas euclídeas

Estudiamos ahora las cónicas en un plano euclídeo en términos de la geometría parabólica. Vamos a ver que todas las propiedades de las elipses, parábolas e hipérbolas pueden deducirse de su mera definición como cónicas no degeneradas que cortan a la recta infinita en 0, 1 o 2 puntos. Fijamos un plano euclídeo E y llamaremos C_1 y C_2 a sus puntos circulares.

Empezamos dando definiciones adecuadas de los elementos de una cónica:

Diremos que dos rectas de un haz de rectas son *conjugadas* respecto de una polaridad si una pasa por el polo de la otra. Diremos que son conjugadas respecto de una cónica si lo son respecto de la polaridad asociada a la cónica.

Definición 10.18 Sea \mathcal{C} una cónica. Un *eje* de \mathcal{C} es una recta tal que la reflexión respecto a ella deja invariante a \mathcal{C} . Un *foco* F de \mathcal{C} es un punto tal que los pares de rectas conjugadas respecto a \mathcal{C} que pasan por F son perpendiculares. Una *directriz* es la recta polar de un foco. Un *diámetro* de \mathcal{C} es una recta que pasa por su centro. Un *vértice* de \mathcal{C} es un punto de \mathcal{C} tal que su tangente es perpendicular al diámetro que pasa por él.

Todos estos conceptos ya los habíamos definido ya de otro modo en [IGE]. Veremos que las definiciones son equivalentes.

Empezamos estudiando estos elementos sobre las parábolas:

Teorema 10.19 *Una parábola tiene un único eje, un único vértice y un único foco. Además el eje es el diámetro que pasa por el vértice y por el foco y la directriz correspondiente es perpendicular al eje.*

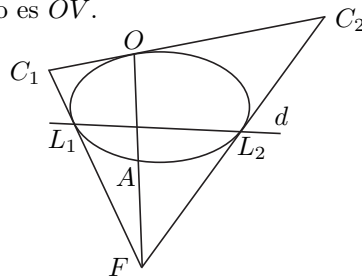
DEMOSTRACIÓN: Recordemos que una parábola es una cónica \mathcal{C} tal que la recta infinita es tangente por uno de sus puntos O , que será, pues, el polo de la recta infinita, luego el centro de la parábola, por definición de centro.

Que un punto V sea un vértice significa que la recta OV es perpendicular a la tangente por V , es decir, que ésta pasa por el conjugado ortogonal O' de O . Pero por O' pasan sólo dos tangentes a la parábola, una es la recta infinita, y la otra toca a la cónica en un punto V que resulta ser, pues, su único vértice.

Además, se cumple que OV es un eje de la parábola. En efecto, dado un punto P en la parábola que no sea V , la perpendicular a OV por P pasa por O' , luego no es tangente a la parábola, luego la corta en un segundo punto P' . Como OV es la polar de O' , el teorema 9.12 nos permite concluir que P y P' se corresponden por la reflexión respecto a OV .

Veamos ahora que OV es el único eje. Si una recta r es un eje, entonces la reflexión de eje r ha de fijar a la parábola. Como es una isometría, es claro que ha de fijar a su único vértice, luego V está en r . La reflexión ha de fijar también a la tangente por V , lo que sólo es posible si dicha tangente t es igual a r o perpendicular a r . No puede ser $t = r$, pues, dado un punto P en la parábola distinto de V , la perpendicular a t por P pasa por O , luego no corta a la parábola en otro punto finito, luego la imagen de P por la reflexión respecto a t , que es otro punto en dicha perpendicular, queda fuera de la cónica. Así pues, r es perpendicular a t y pasa por V , luego es OV .

Como la recta infinita es tangente a \mathcal{C} , todos sus puntos distintos de O son exteriores, en particular los puntos circulares (ninguno de los cuales es O , porque O es real). Por consiguiente, por C_i pasan dos tangentes a \mathcal{C} . Una de ellas es la recta infinita y la otra una recta l_i . Sea L_i el punto de tangencia y sea $F = l_1 \cap l_2$. Así la polar de F es la recta $d = L_1L_2$. Es fácil ver que la imagen por una homografía real de dos puntos imaginarios conjugados son dos puntos imaginarios conjugados, luego la recta d que los une es real,¹⁰ y su polo F también, así como la recta OF y el punto A donde ésta corta a la parábola.



Vamos a probar que F es el único foco de la parábola. En efecto, consideramos en el haz \mathcal{H}_F la involución inducida por la involución ortogonal (la que lleva cada recta a su perpendicular) y la inducida por la cónica. Concretamente, son las composiciones

$$\mathcal{H}_F \longrightarrow E_\infty \xrightarrow{I} E_\infty \longrightarrow \mathcal{H}_F, \quad \mathcal{H}_F \xrightarrow{p_c} d \longrightarrow \mathcal{H}_F.$$

Es inmediato comprobar que ambas involuciones fijan a las rectas l_i , por lo que son la misma involución hiperbólica (todas las propiedades sobre involuciones en rectas se traducen inmediatamente a involuciones sobre haces de rectas). Esto significa que los pares de rectas conjugadas respecto de la cónica son precisamente los pares de rectas perpendiculares, luego F es un foco.

¹⁰En general, si una recta pasa por dos puntos imaginarios conjugados, es real, ya que su conjugada pasa por esos mismos puntos.

Recíprocamente, si F' fuera otro foco, las rectas isotropas $F'C_i$ deberían ser autoconjugadas respecto de la cónica, porque lo son para la involución ortogonal, pero eso obliga a que sean las tangentes a la cónica por F' , o también las tangentes finitas a la cónica por los puntos C_i , es decir, las rectas l_i , luego $F' = F$.

Veamos que A es el vértice de la parábola, con lo que OA será su eje. Sea B_∞ el polo de OA , es decir, el punto infinito de la tangente por A . Entonces las rectas FB_∞ y OA son conjugadas y pasan por F , luego son perpendiculares. Por consiguiente B_∞ es el conjugado ortogonal de O . En particular la tangente por A es perpendicular a OA , lo que prueba que A es el vértice. Puesto que F está en la polar de B_∞ , tenemos además que B_∞ está en d , luego la directriz es perpendicular al eje. ■

Notemos que hemos probado que el eje de una parábola es la polar del conjugado ortogonal de su centro, y que su foco es la intersección de las tangentes isotropas finitas.

Ejemplo Consideremos la cónica de ecuación $x^2 - 2xy + y^2 - 4x - 3 = 0$. Su compleción proyectiva es

$$x^2 - 2xy + y^2 - 4xz - 3z^2 = 0,$$

cuya matriz asociada es

$$A = \begin{pmatrix} 1 & -1 & -2 \\ -1 & 1 & 0 \\ -2 & 0 & -3 \end{pmatrix}.$$

Haciendo $z = 0$ obtenemos $x^2 - 2xy + y^2 = 0$, luego tiene un único punto infinito $O = (1, 1, 0)$. Se trata, pues, de una parábola de centro O . El conjugado ortogonal de O es $O' = (-1, 1, 0)$ y el eje de la parábola es la polar de O' , cuya ecuación es $O'A(x, y, 1)^t = 0$, es decir, $x - y = 1$. Calculando la intersección de esta recta con la parábola obtenemos el vértice $V = (-1/2, -3/2)$. Las polares de los puntos circulares tienen ecuaciones $(1, \pm i, 0)A(x, y, 1)^t = 0$, es decir,

$$(1 - i)x - (1 - i)y - 2 = 0, \quad (1 + i)x - (1 + i)y - 2 = 0.$$

Los puntos de corte finitos con la parábola son

$$\left(-\frac{3}{4} + \frac{i}{2}, -\frac{7}{4} - \frac{i}{2}\right), \quad \left(-\frac{3}{4} - \frac{i}{2}, -\frac{7}{4} + \frac{i}{2}\right)$$

y las tangentes isotropas de la parábola son:

$$\begin{vmatrix} x & y & 1 \\ -\frac{3}{4} + \frac{i}{2} & -\frac{7}{4} - \frac{i}{2} & 1 \\ 1 & i & 0 \end{vmatrix} = ix - y = \frac{5}{4} - \frac{i}{4} = 0, \quad -ix - y = \frac{5}{4} + \frac{i}{4} = 0.$$

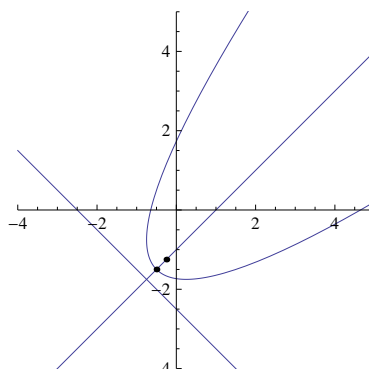
El foco es la intersección de estas tangentes: $F = (-1/4, -5/4)$, y la directriz es la polar del foco, cuya ecuación es

$$(-1/4, -5/4, 1)A(x, y, 1)^t = 0,$$

es decir, $x + y = -5/2$.

El haz de rectas de O está formado por las rectas de ecuaciones $y = x + m$. Calculando sus intersecciones con la parábola obtenemos la parametrización

$$(x, y) = \left(\frac{m^2 - 3}{4}, \frac{m^2 + 4m - 3}{4} \right). \quad \blacksquare$$



Para los otros tipos de cónicas probamos en primer lugar lo siguiente:

Teorema 10.20 *Sea \mathcal{C} una cónica que no sea una parábola. Entonces:*

1. *Una recta es un eje de \mathcal{C} si y sólo si es un diámetro perpendicular a su diámetro conjugado.*
2. *Un punto de \mathcal{C} es un vértice si y sólo si está sobre un eje.*

DEMOSTRACIÓN: Si r es un eje de \mathcal{C} , la reflexión de eje r es una isometría que deja fija a \mathcal{C} , luego fija a su centro O , luego O está en r , es decir, r es un diámetro. Al fijar a \mathcal{C} , la reflexión respecto a r transforma pares de rectas conjugadas de \mathcal{H}_O en pares de rectas conjugadas. Como fija a r , también fija a su diámetro conjugado, pero esto sólo es posible si éste es perpendicular a r .

Recíprocamente, si r es un diámetro perpendicular a su diámetro conjugado r' , entonces el polo de r es el punto infinito de r' , digamos P_∞ . Dado un punto Q en \mathcal{C} , la perpendicular a r por Q es QP_∞ . Si es la tangente por Q entonces Q está en la polar de P_∞ , o sea, en r , luego es fijado por la reflexión respecto de r . Si por el contrario QP_∞ corta a la cónica en otro punto Q' , entonces el teorema 9.12 garantiza que $r = p(P_\infty)$ contiene al punto medio de Q y Q' , luego éstos se corresponden por la reflexión respecto a r . En cualquier caso esta reflexión fija a \mathcal{C} .

Si un punto V de \mathcal{C} está en un eje r , entonces el polo de r es el punto infinito de su diámetro conjugado r' , por donde pasa también la tangente por V a \mathcal{C} , luego ésta es perpendicular a r . Recíprocamente, si V es un vértice y r es el diámetro que pasa por V , entonces su polo es el punto infinito de la tangente por V y también el del diámetro conjugado, luego éstas son rectas paralelas y la tangente es perpendicular a r , luego el diámetro conjugado también. \blacksquare

El caso más simple es el de las circunferencias:

Teorema 10.21 *Una circunferencia tiene un único foco, que coincide con su centro y es un punto interior. Todos sus diámetros son ejes y todos sus puntos son vértices. La directriz es la recta infinita.*

DEMOSTRACIÓN: Sea \mathcal{C} una circunferencia (real o imaginaria), lo cual significa que pasa por los puntos circulares C_1 y C_2 . Si F es un foco de \mathcal{C} , esto significa que la involución que la conjugación respecto de \mathcal{C} induce en \mathcal{H}_F coincide con la involución inducida por la involución ortogonal. Como ésta fija a las rectas FC_i , lo mismo le sucede a aquélla, lo cual equivale a que las rectas FC_i sean las tangentes a \mathcal{C} por los puntos C_i . Así pues, F es necesariamente la intersección de dichas tangentes. Por lo tanto F es el polo de la recta C_1C_2 , es decir, de la recta infinita y, por consiguiente, es el centro de la circunferencia.

Recíprocamente, el centro O de \mathcal{C} es ciertamente un polo, pues las involuciones en \mathcal{H}_O inducidas tanto por \mathcal{C} como por la involución ortogonal fijan ambas a las rectas OC_i , luego son la misma.

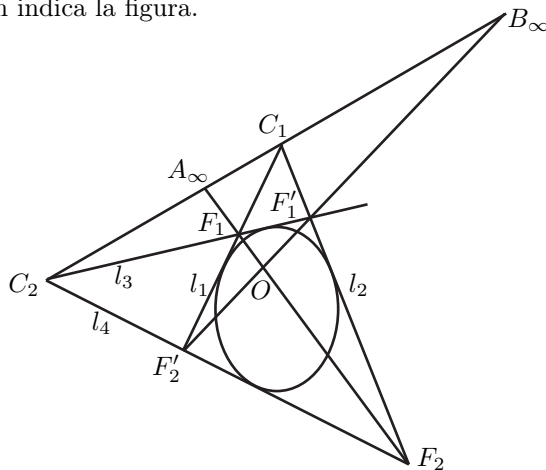
En particular esto quiere decir que los pares de diámetros conjugados son también pares de rectas perpendiculares, lo que significa que todos los diámetros de \mathcal{C} son ejes, por el teorema anterior, del cual se sigue a su vez que todos los puntos de \mathcal{C} son vértices.

Como la recta infinita es exterior a la circunferencia (porque la corta en puntos imaginarios), concluimos que el centro es un punto interior, de modo que cada diámetro es secante y corta a la circunferencia en dos vértices. Por último, la directriz es obviamente la recta infinita. ■

El teorema siguiente recoge todos los casos que faltan:

Teorema 10.22 *Una hipérbola o una elipse real o imaginaria que no sea una circunferencia tiene exactamente dos ejes y dos focos. Ambos están sobre el mismo eje, son interiores y están simétricamente situados respecto del centro. Las directrices son perpendiculares a dicho eje.*

DEMOSTRACIÓN: Sea \mathcal{C} una cónica en las condiciones del enunciado. Hay que entender que es una cónica real (es decir, que viene dada por una ecuación real, aunque pueda ser una elipse imaginaria). Que no sea una circunferencia equivale a que no pasa por los puntos circulares C_1 y C_2 (si pasa por uno, tiene que pasar por el otro, por ser real). Sean l_1 y l_2 las tangentes a \mathcal{C} por C_1 y l_3 y l_4 las tangentes por C_2 . Sean F_1, F_2, F'_1 y F'_2 las intersecciones entre estas tangentes, según indica la figura.



Observemos que F_1 está en la intersección de una tangente por C_1 y otra por C_2 . Como la cónica es real, la conjugada de una tangente es otra tangente, luego \bar{F}_1 es la intersección de una tangente por $\bar{C}_1 = C_2$ y otra por $\bar{C}_2 = C_1$, luego \bar{F}_1 tiene que ser F_1 o F_2 , y lo mismo sucede con F'_1 y F'_2 : ambos pares son reales o imaginarios conjugados.

Más aún, si un par es imaginario, por ejemplo F'_1, F'_2 , entonces el otro par es real, como se deduce de conjugar la relación $F_1 = C_1 F'_2 \cap C_2 F'_1$ y, recíprocamente, si F'_1, F'_2 son reales, entonces, conjugando la relación anterior resulta que $\bar{F}_1 = C_2 F'_2 \cap C_1 F'_1 = F_2$.

Por consiguiente, uno de los pares es real y el otro imaginario. No perdemos generalidad si suponemos que F_1 y F_2 son el par real. Esto implica además que ambos puntos son interiores, pues la conjugación compleja intercambia las tangentes $F_1 C_i$, luego intercambia los puntos de tangencia, luego éstos no son reales, luego por los focos no pasan tangentes.

Veamos que F_1 y F_2 son los dos únicos focos de la cónica. En efecto, si F es cualquiera de ellos, como en el caso de la parábola, la involución inducida en \mathcal{H}_F por la involución ortogonal coincide con la inducida por la polaridad de la cónica, pues ambas fijan a las rectas FC_i , luego los pares de rectas conjugadas respecto de \mathcal{C} que pasan por F son perpendiculares. Recíprocamente, si F es un foco de \mathcal{C} , las rectas FC_i deben quedar fijas por la involución inducida por la cónica, luego deben ser tangentes a la cónica, luego F tiene que ser uno de los F_i o F'_i , pero como tiene que ser real, tiene que ser un F_i . Con esto hemos probado que la cónica tiene únicamente dos focos, F_1 y F_2 .

Vamos a aplicar el teorema dual del teorema 9.11 al cuadrilátero formado por las rectas l_1, l_2, l_3, l_4 o, equivalentemente, vamos a aplicar dicho teorema al cuadrilátero inscrito formado por los polos de dichas rectas. Si damos el mismo nombre a una recta que a su polo, tenemos que los puntos l_1, l_2, l_3, l_4 son los vértices de un cuadrilátero cuyos lados son $l_1 l_2 = C_1$, $l_1 l_3 = F_1$, $l_1 l_4 = F'_2$, $l_2 l_3 = F'_1$, $l_2 l_4 = F_2$, $l_3 l_4 = C_2$. Los puntos diagonales son las rectas $C_1 C_2$, $F_1 F_2$, $F'_1 F'_2$, y lo que afirma el teorema es que la recta polar (o sea, el polo) de cada una de estas rectas es la intersección de las otras dos. Así, el polo de $C_1 C_2$ (o sea, el centro de \mathcal{C}) es $O = F_1 F_2 \cap F'_1 F'_2$, el polo de $F'_1 F'_2$ es el punto infinito A_∞ de $F_1 F_2$ y el polo de $F_1 F_2$ es el punto infinito B_∞ de $F'_1 F'_2$.

Observemos ahora que las rectas $F_2 A_\infty$ y $F_2 B_\infty$ son conjugadas respecto de la cónica \mathcal{C} en \mathcal{H}_{F_2} , por lo que también son conjugadas respecto de la involución ortogonal, lo que significa que A_∞ y B_∞ son conjugados ortogonales. En particular, OB_∞ es la perpendicular a OA_∞ por O , y como O es real (por ser el centro de una cónica real) y OA_∞ es real (porque pasa por los puntos reales F_1 y F_2), concluimos que también OB_∞ es real.

Ahora consideremos el haz \mathcal{H}_O , en el cual tenemos definidas la involución ortogonal y la de la cónica. No son la misma, porque OC_1 es fijado por la involución ortogonal, pero no por la de \mathcal{C} , ya que no es una recta tangente. Sin embargo, las rectas OA_∞ y OB_∞ son conjugadas para ambas, para la involución ortogonal porque hemos visto que A_∞ y B_∞ son conjugados ortogonales, y para la de la cónica porque hemos visto que B_∞ es el polo de $F_1 F_2 = OQ_\infty$.

El teorema 8.76, como cualquier enunciado sobre homografías, se traduce trivialmente al caso de homografías de un haz de rectas, y como la involución ortogonal en el haz real \mathcal{H}_O es elíptica, sabemos que hay un único par de rectas reales en \mathcal{H}_O que son a la vez perpendiculares y conjugadas respecto de la cónica. Esto prueba que OA_∞ y OB_∞ son los únicos ejes de la cónica.

Considerando el cuadrilátero formado por los puntos F_i y F'_i , cuyos puntos diagonales son O , C_1 , C_2 , la propia definición de separación armónica nos da $\mathcal{H}(F_1, F_2; O, A_\infty)$. Por lo tanto, O es el punto medio de F_1 y F_2 . Como F_1 y F_2 están en la polar de B_∞ , las polares de F_1 y F_2 , es decir, las directrices, pasan por B_∞ , luego son perpendiculares al eje OA_∞ . ■

El eje de una elipse o de una hipérbola que contiene a los focos recibe el nombre de *eje mayor*, mientras que el otro eje es el *eje menor*.

En una elipse real, el centro es interior, luego los ejes son secantes, luego la elipse tiene exactamente cuatro vértices. En cambio, una hipérbola sólo tiene dos vértices sobre su eje mayor, pues éste pasa por los focos, que son interiores, y todas las rectas que pasan por un punto interior son secantes. En cambio, el semieje menor es exterior. Para probarlo basta ver que su polo, que es el punto infinito A_∞ del eje mayor, es interior, y ello se debe a que los vértices dividen al eje mayor en dos segmentos, uno formado por puntos interiores y otro por puntos exteriores. Como O es exterior, basta probar que O y A_∞ están en segmentos opuestos, pero ello se debe a que son conjugados por la involución inducida por la cónica, luego según 8.63 están separados armónicamente por los vértices (que son los puntos fijos).

Ejemplo Consideremos la cónica determinada por la matriz

$$A = \begin{pmatrix} 28 & 64 & -212 \\ 64 & 28 & -248 \\ -212 & -248 & 925 \end{pmatrix}.$$

Resolviendo $(x, y, 0)A(x, y, 0)^t = 0$ vemos que se trata de una hipérbola que pasa por los puntos infinitos $(7, -16 - 3\sqrt{23}, 0)$ y $(7, -16 + 3\sqrt{23}, 0)$. Resolviendo $(x, y, 1)A = (0, 0, a)$ obtenemos que el centro es $(3, 2)$. Las asíntotas son las rectas polares de los puntos infinitos, y resultan ser:

$$249.829x + 57.55y = 864.586, \quad 13.2571x + 57.55y = 154.871.$$

Las polares de los puntos circulares son

$$(7 + 16i)x + (16 + 7i)y = 53 + 62i, \quad (7 - 16i)x + (16 - 7i)y = 53 - 62i.$$

Resolviendo el sistema formado por una de estas ecuaciones y la ecuación de la hipérbola obtenemos los puntos de tangencia:

$$\begin{aligned} \frac{1}{16}(39 - 23i, 23 + 23i), & \quad \frac{1}{16}(57 + 23i, 41 - 23i), \\ \frac{1}{16}(39 + 23i, 23 - 23i), & \quad \frac{1}{16}(57 - 23i, 41 + 23i). \end{aligned}$$

A partir de ellos calculamos las cuatro tangentes isotropas:

$$ix - y = i, \quad -ix - y = -i, \quad ix - y = -4 + 5i, \quad -ix - y = -4 - 5i.$$

Calculamos las intersecciones de las que no son paralelas:

$$F_1 = (1, 0), \quad F_2 = (5, 4), \quad F'_1 = (3 - 2i, 2 + 2i), \quad F'_2 = (3 + 2i, 2 - 2i).$$

Por lo tanto, los focos son los dos primeros puntos. La recta que los une es el eje mayor: $x + y = 5$, y su perpendicular por el centro es el eje menor: $x - y = 1$.

La intersección del eje mayor con la hipérbola nos da los vértices:

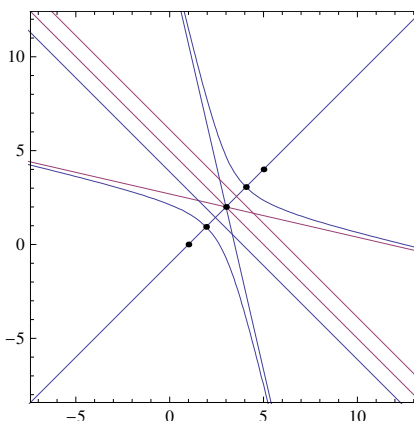
$$\frac{1}{4}(12 - 3\sqrt{2}, 8 - 3\sqrt{2}),$$

$$\frac{1}{4}(12 + 3\sqrt{2}, 8 + 3\sqrt{2}).$$

Y las polares de los focos son las directrices:

$$184x + 184y = 713, \quad 184x + 184y = 1127.$$

■



Ejemplo Consideremos ahora la elipse imaginaria $x^2 + 4y^2 = -1$. Los mismos cálculos que en el ejemplo anterior muestran que sus puntos infinitos son $(2, \pm i, 0)$ (por lo que no tiene asíntotas reales), su centro es $(0, 0)$, las polares de los puntos circulares son $x \pm 4i = 0$, los puntos de tangencia son $(\pm \frac{2i}{\sqrt{3}}, \pm \frac{1}{2\sqrt{3}})$, las tangentes isotropas son $\pm ix + y = \pm \sqrt{3}/2$, y sus intersecciones son

$$(0, \pm \sqrt{3}/2), \quad (\pm i\sqrt{3}/2, 0),$$

luego los focos son los dos primeros puntos. El eje mayor es el eje $y = 0$ y el eje menor es $x = 0$. Las directrices son $y = \pm \sqrt{3}/2$. ■

Ahora demostramos que los focos y las directrices que hemos definido en esta sección coinciden con los que habíamos definido en [IGE 7.1]:

Teorema 10.23 Si \mathcal{C} es una cónica real que no sea una circunferencia, la razón entre las distancias de un punto de \mathcal{C} a uno de sus focos y a su directriz correspondiente es constante, y recibe el nombre de excentricidad de \mathcal{C} .

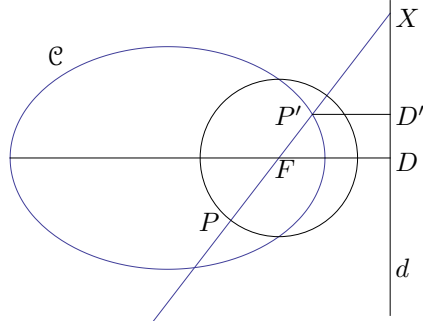
DEMOSTRACIÓN: Sea F un foco de \mathcal{C} y sea d la directriz correspondiente. Sea D el punto en que d corta al eje mayor a , sea H el punto tal que D es el punto medio de F y H y sea h la paralela a d que pasa por h . Consideremos la homología f de centro F y eje h que transforma D en el punto infinito de DH . Considerando razones dobles se ve enseguida que la imagen del punto infinito

ha de ser D (por ser el conjugado armónico de D respecto a F y H). Por otra parte f deja invariante el punto infinito común a d y h , luego f intercambia d con la recta infinita. Más aún, f tiene orden 2, luego induce involuciones en todas las rectas que pasan por F .

Como F es el polo de d respecto a \mathcal{C} , es claro que F será el polo de la recta infinita respecto de $f[\mathcal{C}]$, es decir, F es el centro de $f[\mathcal{C}]$. Las tangentes a \mathcal{C} que pasan por F cortan a \mathcal{C} en dos puntos (imaginarios) de d , y f los transforma en los puntos infinitos de dichas tangentes, que son los puntos circulares. Así pues, $f[\mathcal{C}]$ contiene a los puntos circulares, luego es una circunferencia.

Sea P' cualquier punto de \mathcal{C} y sea $P = f(P')$. Sea $X = FP' \cap d$ y supongamos que es un punto finito. Si P_∞ el punto infinito de FP' , entonces $f|_{FP'}$ intercambia X con P_∞ y fija a F , luego

$$\begin{aligned} \mathcal{R}(X, P_\infty, P, F) &= \mathcal{R}(P_\infty, X, P', F) \\ &= \mathcal{R}(X, P_\infty, F, P'), \end{aligned}$$



lo cual equivale a

$$\frac{\overrightarrow{XP}}{\overrightarrow{XF}} = \frac{\overrightarrow{XP'}}{\overrightarrow{XF}},$$

de donde

$$\frac{\overrightarrow{XF} + \overrightarrow{FP}}{\overrightarrow{XF}} = \frac{\overrightarrow{XP'} + \overrightarrow{P'F}}{\overrightarrow{XP'}} \Rightarrow \frac{\overrightarrow{FP}}{\overrightarrow{XF}} = \frac{\overrightarrow{P'F}}{\overrightarrow{XP'}}.$$

Ahora tomamos módulos y aplicamos el teorema de Tales

$$\frac{\overline{FP}}{\overline{FP'}} = \frac{\overline{XF}}{\overline{XP'}} = \frac{\overline{FD}}{\overline{P'D'}} \Rightarrow \frac{\overline{P'F}}{\overline{P'D'}} = \frac{\overline{PF}}{\overline{FD}},$$

donde D' es el punto donde la perpendicular a d desde P' corta a d . Por último notamos que $\overline{P'D'}$ es la distancia de P' a d , así como que \overline{PF} no depende de P , pues es el radio de $f[\mathcal{C}]$. Así pues, el miembro izquierdo de la última igualdad es constante.

En el caso en que X es infinito, eso significa que FP' es paralela a d , luego $\overline{P'D'} = \overline{FD}$. Si probamos que $\overline{P'D'} = \overline{FD}$ tendremos que P' también cumple el teorema en este caso. Ahora bien, la involución que f induce en la recta FP' fija a F y a X , luego envía cada punto a su conjugado armónico respecto de ambos puntos. En particular, P y P' son conjugados armónicos respecto de F y X , lo cual equivale a que F es su punto medio. Esto implica a su vez que $P' \in f[\mathcal{C}]$, luego $\overline{P'F} = \overline{PF}$. ■

Puesto que la reflexión respecto al eje menor transforma la cónica en sí misma, es fácil ver que la excentricidad no depende del foco con que se calcula. En particular, manteniendo la notación de la prueba del teorema anterior, si A

es el vértice que se encuentra en la misma semirrecta que F respecto al centro, la excentricidad puede calcularse como

$$e = \frac{\overline{AF}}{\overline{FD}}.$$

Puesto que la homología f transforma la recta infinita en la directriz d , es claro que \mathcal{C} será una elipse, una parábola o una hipérbola según si d es exterior, tangente o secante a $f[\mathcal{C}]$. Según el caso, el cociente anterior será menor, igual o mayor que 1. Por consiguiente:

Teorema 10.24 *Una cónica \mathcal{C} es una elipse, una parábola o una hipérbola según si su excentricidad es menor, igual o mayor que 1.*

Esto vale para las circunferencias si convenimos en que su excentricidad es 0.

Por otra parte, dado un punto cualquiera F , una recta cualquiera d que no pase por F y un número $e > 0$, la figura \mathcal{C} de los puntos tales que la razón de las distancias a F y d es e es una cónica real no degenerada. En efecto, tomando un sistema de referencia ortonormal adecuado podemos suponer que $F = (0, 0)$ y que d tiene ecuación $x = D$, con $D > 0$. Entonces la condición equivale a

$$x^2 + y^2 = e^2(D - x)^2.$$

Esta ecuación corresponde a la cónica de matriz

$$A = \begin{pmatrix} 1 - e^2 & 0 & e^2 D \\ 0 & 1 & 0 \\ e^2 D & 0 & -e^2 D^2 \end{pmatrix}. \quad (10.1)$$

Su determinante es $-e^2 D^2$, luego no es degenerada. Además haciendo $y = 0$ obtenemos que la ecuación se cumple para $x = \frac{e}{e+1}D$, luego \mathcal{C} no es imaginaria.

Más aún, F es un foco de \mathcal{C} . Para probarlo notamos primero que $d = p(F)$, luego la intersección de D con \mathcal{C} está formada por los puntos de contacto de las tangentes que pasan por F . Haciendo $x = D$ en la ecuación obtenemos que éstos son $(D, \pm Di, 1)$, luego las tangentes son las rectas que pasan por estos puntos y $(0, 0, 1)$. Es claro que contienen a los puntos $(D, \pm Di, 0) = (1, \pm i, 0)$, que son los puntos circulares, luego F es ciertamente un foco y d la directriz correspondiente. Así pues:

Teorema 10.25 *Dada una recta d , un punto F que no esté en d y un número $e > 0$, la figura formada por todos los puntos tales que la razón de las distancias a F y d sea igual a e es una cónica real con un foco igual a F , directriz d y excentricidad e . Toda cónica real que no sea una circunferencia es de esta forma.*

Esto implica que toda cónica que no sea una circunferencia está completamente determinada por su excentricidad, uno de sus focos y su directriz correspondiente. Las semejanzas conservan las proporciones, por lo que también

conservan la excentricidad de las cónicas. Como fijan o invierten los puntos circulares, es claro que conservan los focos (luego también las directrices). Por otra parte, dadas dos cónicas, siempre es posible transformar la primera mediante una semejanza para obtener otra con un foco y una directriz común con la segunda. Si ambas tenían la misma excentricidad, entonces la cónica que obtenemos coincide con la segunda. En resumen:

Teorema 10.26 *Dos cónicas reales son semejantes si y sólo si tienen la misma excentricidad.*

Como siempre, el argumento no incluye las circunferencias, pero el resultado es válido igualmente para ellas. En particular dos parábolas cualesquiera son semejantes.

Algunas relaciones Vamos a probar varias relaciones importantes entre los elementos de una cónica, la mayoría de las cuales las conocemos ya de [IGE]. Conviene introducir algunas definiciones:

Llamaremos *semiejes mayores* de una elipse a los segmentos que unen su centro con sus vértices en el eje mayor. También llamaremos semieje mayor a la longitud de estos segmentos, y la representaremos por a . Análogamente se define el semieje menor, que representaremos por b . En una hipérbola sólo tenemos definido el semieje mayor. La *distancia focal* de una elipse o una hipérbola es la distancia de su centro O a cualquiera de sus focos, y la representaremos por c . Para una hipérbola, definimos formalmente $b = \sqrt{c^2 - a^2}$.

Teorema 10.27 *Sea \mathcal{C} una cónica real. Entonces:*

1. Si \mathcal{C} es una elipse, sus puntos son exactamente los que cumplen

$$\overline{PF_1} + \overline{PF_2} = 2a.$$

2. Si \mathcal{C} es una hipérbola, sus puntos son exactamente los que cumplen

$$\overline{PF_1} - \overline{PF_2} = \pm 2a.$$

3. Si \mathcal{C} es una elipse, la distancia de un vértice menor a uno de los focos es a , y se cumple $a^2 = b^2 + c^2$.

4. Si \mathcal{C} es una elipse o una hipérbola, $e = c/a$.

5. Si \mathcal{C} es una hipérbola, el ángulo que forman sus asíntotas es el dado por

$$\cos \theta = 1 - \frac{2}{e^2}.$$

DEMOSTRACIÓN: 1) Dado un punto P de \mathcal{C} , sean D_1 y D_2 los puntos de las directrices donde se alcanza la distancia de éstas a P . Entonces

$$\overline{PF_1} + \overline{PF_2} = e \overline{PD_1} + e \overline{PD_2} = e \overline{D_1 D_2}.$$

Por lo tanto la suma de las distancias a los focos es constante. Evaluándola en un vértice es fácil ver que es exactamente $2a$. Se comprueba sin dificultad que los puntos que cumplen esta ecuación forman una cónica que contiene a \mathcal{C} , luego ha de ser \mathcal{C} .

2) La prueba es similar a la anterior, salvo que ahora $\overline{PD_1} - \overline{PD_2}$ es la distancia entre las dos directrices con signo positivo o negativo según la rama en que esté P , luego la diferencia de distancias sólo es constante en módulo.

3) Si B es uno de los vértices menores, tenemos que $\overline{BF_1} + \overline{BF_2} = 2a$, pero ambos sumandos son iguales. Si A es un vértice mayor, el triángulo $\widehat{BAF_1}$ es rectángulo, lo que nos da la relación $a^2 = b^2 + c^2$. En particular vemos que el semieje mayor es ciertamente mayor que el semieje menor.

4) La matriz (10.1) corresponde a una cónica arbitraria con un foco en $(0, 0)$ y directriz $x = D$. Es fácil calcular su centro y su vértice:

$$O\left(-\frac{e^2 D}{1 - e^2}\right), \quad A\left(\frac{eD}{1 + e}, 0\right),$$

luego

$$a = \frac{ed}{|1 - e^2|}, \quad c = \frac{e^2 D}{|1 - e^2|},$$

luego ciertamente $e = c/a$.

5) La cónica de matriz A según 10.1 corta a la recta infinita en los puntos $(1, \pm\sqrt{e^2 - 1}, 0)$, luego sus asíntotas son las polares de estos puntos, cuyas ecuaciones son $(1 - e^2)x \pm \sqrt{e^2 - 1}y + e^2 D = 0$. Los vectores directores son $(\pm\sqrt{e^2 - 1}, 1 - e^2)$ y al calcular el ángulo que forman obtenemos la fórmula pedida. ■

Las hipérbolas cuyas asíntotas forman ángulos rectos se llaman *equiláteras*. Según el teorema anterior son exactamente las de excentricidad $e = \sqrt{2}$.

Ejercicio: Si A_1 y A_2 son los vértices mayores de una elipse o una hipérbola y F es uno de sus focos, $\overline{FA_1} \overline{FA_2} = b^2$.

Los focos de las cónicas tienen otra propiedad destacada. En el caso de las elipses e hipérbolas el enunciado es el siguiente:

Teorema 10.28 *Si \mathcal{C} es una elipse real que no sea una circunferencia (o una hipérbola), entonces la normal¹¹ (tangente) a \mathcal{C} por uno cualquiera de sus puntos X es la bisectriz del ángulo $\widehat{F_1 X F_2}$, donde F_1 y F_2 son sus focos.*

DEMOSTRACIÓN: Si X está en uno de los ejes de la cónica la conclusión se cumple trivialmente. Supongamos que no es así. Vamos a aplicar el dual del teorema 9.27 al haz de rectas \mathcal{H}_X y a las cuatro tangentes a \mathcal{C} por los puntos circulares C_1 y C_2 . Equivalentemente, aplicamos dicho teorema a la recta t tangente a \mathcal{C} por X (cuyos puntos son los polos de las rectas de \mathcal{H}_X) y al

¹¹Se llama *recta normal* a una cónica por un punto a la perpendicular a su tangente.

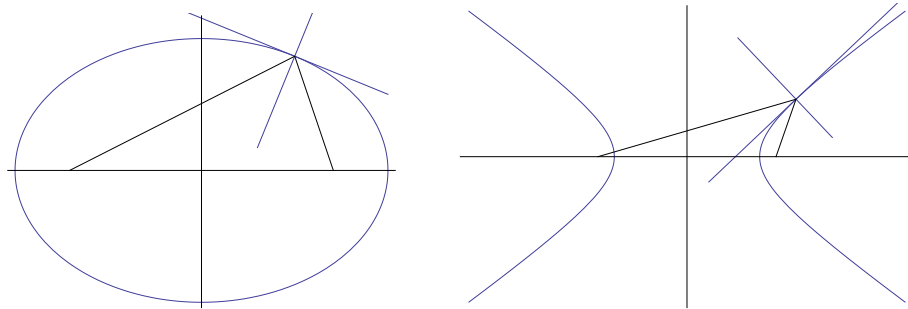
cuadrilátero formado por los cuatro puntos de tangencia. (Véase la figura en la demostración del teorema 10.22). Si los llamamos igual que a las tangentes, dichos vértices son l_1, l_2, l_3, l_4 . Los lados opuestos l_1l_2 y l_3l_4 tienen por polos a C_1 y C_2 y, análogamente, los otros dos pares de lados opuestos se corresponden con los polos F_1, F_2 y F'_1, F'_2 . La intersección con t de l_1l_2 tiene por polar a la recta XC_1 , de modo que los pares de rectas en \mathcal{H}_X que son polares de intersecciones con t de pares de lados opuestos son (XC_1, XC_2) , (XF_1, XF_2) y (XF'_1, XF'_2) . A su vez, la intersección X de t con la cónica tiene por polar a la propia t .

Lo que dice el teorema es que existe una involución en t que fija a X y que conjuga los pares de intersecciones con lados opuestos. A través de la polaridad de \mathcal{C} , dicha involución se transforma en una involución de \mathcal{H}_X que fija a t y que conjuga los pares (XC_1, XC_2) , (XF_1, XF_2) y (XF'_1, XF'_2) .

Pero una involución con un punto fijo tiene, de hecho, dos puntos fijos. La otra recta fijada n tiene que ser la conjugada harmónica de t respecto de XC_1 y XC_2 , lo cual significa que es la imagen de t por la involución inducida en \mathcal{H}_X por la involución ortogonal (ya que ésta tiene a XC_1 y XC_2 como puntos fijos). En otras palabras, n es la perpendicular a t por X , es decir, la recta normal a \mathcal{C} por X .

Por otra parte, XF_1 y XF_2 están también separadas harmónicamente por t y n , y hemos visto que esto significa que t y n son las bisectrices de los ángulos formados por XF_1 y XF_2 . ■

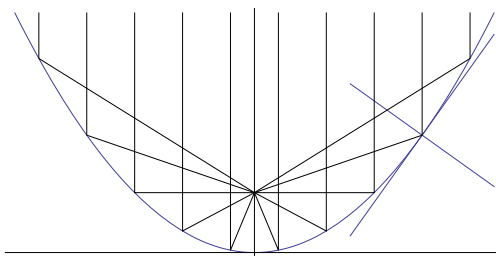
Esto se interpreta como que un rayo de luz emitido desde un foco de una elipse que se refleje en ésta pasará por el otro foco, o que en una mesa de billar elíptica una bola lanzada desde un foco sigue una trayectoria que pasa siempre por el otro foco. En el caso de una hipérbola la situación es un poco más artificial: un rayo de luz emitido desde un foco sigue al reflejarse la recta que pasa por el otro foco, pero se aleja de él.



La propiedad análoga para las parábolas es la siguiente:

Teorema 10.29 *Si \mathcal{C} es una parábola y F es su foco, la tangente y la normal a \mathcal{C} por uno cualquiera de sus puntos X son las bisectrices de los ángulos determinados por XF y la paralela al eje por X .*

Esto significa que todos los rayos de luz emitidos desde el foco de una parábola se reflejan en ésta de modo que sus trayectorias pasan a ser paralelas a su eje o, recíprocamente, que todos los rayos de luz que llegan a una parábola paralelos a su eje confluyen en su foco. Por eso las antenas parabólicas tienen forma de paraboloides, que es la superficie de revolución generada por una parábola (una cuádrica). Lo mismo sucede con los espejos que rodean las fuentes luminosas de los focos, faros, linternas, etc.



DEMOSTRACIÓN: El razonamiento es el mismo que el empleado en la prueba de 10.28, sólo que ahora aplicamos el dual del teorema 9.28 al triángulo formado por las tres tangentes a \mathcal{C} por los puntos circulares (una de las cuales es la recta infinita). Ahora concluimos que existe una involución del haz \mathcal{H}_X que fija a la tangente t por X y que conjugua los pares de rectas XC_1 , XC_2 y XF , XO (véase la figura de la prueba del teorema 10.19). Igual que en 10.28 se razona que la otra recta fijada por la involución es la normal n a \mathcal{C} por X , luego XF y XO están separados armónicamente por t y n , lo cual significa que éstas son las bisectrices de los ángulos determinados por aquéllas. ■

10.4 Homografías de la esfera

En esta sección, cuando hablemos de una esfera S se entenderá que es una esfera en un espacio parabólico euclídeo tridimensional E . Por simplicidad supondremos que el cuerpo de escalares es \mathbb{R} .

Al decir que E es un espacio parabólico euclídeo y no simplemente un espacio euclídeo estamos indicando que no fijamos una unidad de longitud en E o, equivalentemente, que la estructura de E no determina un producto escalar en su parte finita, sino sólo una clase de equivalencia de productos escalares, determinados salvo producto por una constante. En la práctica esto significa que siempre podremos elegir un producto escalar sin pérdida de generalidad para que S tenga radio unitario. En otros términos, respecto de cualquier sistema de referencia ortonormal con origen O en el centro de S , la esfera está formada por los puntos cuyas coordenadas afines cumplen una ecuación $x^2 + y^2 + z^2 = r^2$ y, ajustando el producto escalar, podemos convertirla en $x^2 + y^2 + z^2 = 1$.

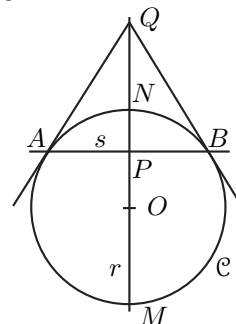
Las homografías de una esfera se definen igual que las de una cónica (de hecho, podríamos definir en general las homografías de una cuádrica):

Definición 10.30 Una *homografía* en una esfera S en un espacio E es una biyección $f : S \rightarrow S$ inducida por restricción a partir de una homografía en E que fije a S . Llamaremos $LGP(S)$ al grupo formado por todas las homografías de S .

Necesitaremos el siguiente resultado técnico:

Teorema 10.31 *Si S es una esfera y C es una circunferencia en S , existe una homografía no trivial en S que fija a todos los puntos de C .*

DEMOSTRACIÓN: Sea Π el plano de C , sea r la perpendicular a Π que pasa por el centro O de S , sean N y M los puntos donde r corta a S , sea Π' cualquier plano que contenga a r , sea \mathcal{C} la circunferencia en que Π' corta a S , sea $s = \Pi \cap \Pi'$, que es una recta perpendicular a r . Sea $P = r \cap s = \Pi \cap r$. Sean A y B los puntos donde s corta a r y sea Q el polo de s . El punto Q es el conjugado harmónico de P respecto a M y N , luego no depende de la elección de Π' , al igual que la homología f de centro Q y eje Π que envía M a N . Así f fija a todos los puntos de $C \subset \Pi$ y no es la identidad. Basta probar que fija a S .



Ahora bien, $f[\mathcal{C}]$ es una cónica en Π' que pasa por A, B, N , al igual que \mathcal{C} , y sus tangentes por A y B pasan por $f(Q) = Q$. En la prueba de 9.31 vimos que esto determina una cónica, luego $f[\mathcal{C}] = \mathcal{C}$. Como todo punto de S está en una de las circunferencias \mathcal{C} , para cierta elección de Π' , concluimos que $f[S] = S$. ■

Vamos a estudiar la geometría de S y sus circunferencias usando las proyecciones estereográficas como sistemas de coordenadas (véase el apéndice A). Si Π es cualquier plano en E que pase por el centro de O y N es uno de los dos puntos en los que la perpendicular a Π por O corta a S , está definida la proyección estereográfica $\pi : S \rightarrow \mathbb{C}^\infty$ en la esfera de Riemann, de modo que $\pi(N) = \infty$.

Según el teorema A.2, esta proyección biyecta las circunferencias de S con las rectas y las circunferencias de \mathbb{C}^∞ (donde en $\mathbb{C} = \mathbb{R}^2$ consideramos la estructura euclídea usual). Concretamente, las circunferencias que pasan por el polo norte N son las que se transforman en rectas, y las restantes se transforman en circunferencias. Recordemos también que llamamos rectas de \mathbb{C}^∞ a las rectas de \mathbb{R}^2 con el punto ∞ añadido.

El teorema fundamental es el siguiente:

Teorema 10.32 *Si $\pi : S \rightarrow \mathbb{C}^\infty$ es una proyección estereográfica, entonces $f \mapsto \pi^{-1}f\pi$ determina un isomorfismo entre el grupo de las homografías de S y el grupo de las transformaciones de Möbius de \mathbb{C}^∞ .*

DEMOSTRACIÓN: Sea E el espacio euclídeo que contiene a S . Las homografías de E transforman planos en planos, luego las homografías de S transforman circunferencias en circunferencias. Si f es una homografía de S , entonces $\pi^{-1}f\pi$ es una biyección en \mathbb{C}^∞ que transforma circunferencias en circunferencias, luego es una transformación de Möbius por el teorema A.8.

Es evidente que $f \mapsto \pi^{-1}f\pi$ es un monomorfismo de grupos, y sólo falta probar que toda transformación circular puede obtenerse de este modo. De

hecho basta probar que todas las inversiones de la recta compleja provienen de homografías en la esfera. Ahora bien, dada una inversión h respecto de una circunferencia C , se cumple que $\pi^{-1}[C]$ es una circunferencia en S , y el teorema 10.31 nos da una homografía f en S que no es la identidad y fija a los puntos de $\pi^{-1}[C]$. Entonces $\pi^{-1}f\pi$ es una transformación circular distinta de la identidad que fija a C , luego $\pi^{-1}f\pi = h$. ■

Así pues, todos los resultados que probamos en el apéndice A sobre transformaciones de Möbius se traducen ahora a resultados sobre homografías de esferas. Por ejemplo, dada una circunferencia C en una esfera S , existe una única homografía en S distinta de la identidad que fija a todos sus puntos, a la cual llamaremos *inversión* respecto a C . Toda homografía en una esfera es producto de inversiones. Si una homografía fija a tres puntos es la identidad o la inversión respecto a la circunferencia que los contiene. Las homografías de una esfera son las biyecciones que transforman circunferencias en circunferencias.

A su vez, el teorema anterior nos aporta un dato importante sobre las transformaciones de Möbius:

Teorema 10.33 *La restricción de una transformación de Möbius a una circunferencia es una homografía. Más aún, cada homografía entre dos circunferencias se extiende a una única transformación de Möbius directa.*

DEMOSTRACIÓN: Tomemos una transformación de Möbius $f : \mathbb{C}^\infty \rightarrow \mathbb{C}^\infty$ y una circunferencia $C \subset \mathbb{C}^\infty$. Tomemos otra circunferencia C' que no pase por ∞ y sean g y h homografías tales que $g[C'] = C$ y $h[f[C]] = C'$. Así, $f' = gfh$ es una transformación de Möbius que cumple $f'[C'] = C'$. Si probamos que $f'|_{C'}$ es una homografía, entonces $f|_C$ también lo será. Equivalentemente, no perdemos generalidad si suponemos que $f[C] = C$ y que $\infty \notin C$.

Por el teorema anterior f se corresponde con una homografía $\bar{f} : S \rightarrow S$, de modo que $f = \pi^{-1}\bar{f}\pi$, donde π es la proyección estereográfica.

Sea E el espacio tridimensional euclídeo que contiene a la esfera S , que podemos considerar contenido en un espacio proyectivo $P(E)$. Entonces \mathbb{C} se identifica con la parte finita de un plano afín $\Pi \subset P(E)$ que pasa por el centro de S . En particular, $C \subset \Pi$.

Sea $C' = p^{-1}[C]$, que es una circunferencia en S (que no pasa por su polo Norte N), contenida en un plano Π' de $P(E)$. Ahora observamos que la proyección perspectiva $p : \Pi \rightarrow \Pi'$ de centro N se restringe a $\pi^{-1}|_C : C \rightarrow C'$, luego $p \circ \bar{f}|_{\Pi'} \circ p^{-1} : \Pi \rightarrow \Pi$ es una homografía que extiende a $f|_C$, luego $f|_C$ es una homografía.

Dada una homografía entre dos circunferencias, tomamos tres puntos de la primera y consideramos la transformación circular directa que coincide sobre ellos con la homografía dada. Obviamente ambas han de coincidir sobre toda la circunferencia y la unicidad es clara. ■

El hecho de que podamos ver a los puntos de \mathbb{C}^∞ como coordenadas de los puntos de una esfera S (respecto de un sistema de referencia ortonormal en E

respecto de un producto escalar respecto al que S tenga radio 1 y respecto de una elección arbitraria de un “polo norte” en S) aporta una información esencial, y es que en la geometría circular de \mathbb{C}^∞ el punto ∞ es un punto “destacado”, mientras que en la geometría circular de S no hay puntos destacados, sino que hemos de destacar un punto arbitrariamente a la hora de asignar coordenadas en \mathbb{C}^∞ a sus puntos.

Capítulo XI

La geometría hiperbólica

La geometría hiperbólica es la geometría no euclídea más próxima a la geometría euclídea, en cuanto que satisface todos los axiomas de Hilbert salvo el axioma de las paralelas. Aunque podríamos definir espacios hiperbólicos de cualquier dimensión, nos limitaremos a estudiar el caso bidimensional por simplicidad. Un plano hiperbólico puede obtenerse a partir de un plano proyectivo por un proceso formalmente similar a como se obtiene un plano euclídeo. Para obtener un plano euclídeo primero seleccionamos una recta, lo que nos permite distinguir entre puntos finitos e infinitos y definir la noción de paralelismo, con lo que tenemos un espacio afín. Después fijamos una involución elíptica en la recta infinita, con lo que podemos definir la perpendicularidad, las semejanzas y las isometrías. Para obtener un plano hiperbólico basta seleccionar una cónica real en lugar de una recta. Esto nos divide los puntos del plano proyectivo en tres clases: puntos finitos (los interiores a la cónica), puntos infinitos (los de la cónica) y puntos ultrainfinitos (los exteriores a la cónica). El plano hiperbólico propiamente dicho está formado por los puntos finitos, de modo que dos rectas que se corten en un punto infinito o ultrainfinito son disjuntas desde el punto de vista hiperbólico. Es fácil ver entonces que por un punto exterior a una recta pasan infinitas paralelas.

Una diferencia importante respecto a la geometría afín es que ahora no necesitamos introducir ninguna estructura adicional para definir la perpendicularidad. El papel de la involución ortogonal lo desempeña aquí la polaridad de la cónica, y el hecho de que esté determinada por la propia cónica termina traduciéndose en que los conceptos análogos a los de afinidades, semejanzas e isometrías se confunden en uno solo en el caso hiperbólico. En particular no hay más semejanzas que las isometrías. Pasemos a detallar todas estas ideas.

11.1 El plano hiperbólico

Definición 11.1 Un *plano hiperbólico* es un plano proyectivo sobre un cuerpo ordenado euclídeo en el que se ha seleccionado una cónica real, a la que llamaremos *cónica infinita*. A los puntos interiores a la cónica infinita los llamaremos

puntos finitos, a los de la cónica los llamaremos *infinitos* y a los exteriores *ultra-infinitos*. Llamaremos *grupo hiperbólico* al grupo de las homografías que fijan a la cónica. Las *rectas finitas* serán las que contengan puntos finitos.

Salvo que indiquemos lo contrario, cuando hablemos de puntos y rectas en un plano hiperbólico sobrentenderemos que se trata de puntos y rectas finitas. Es claro que toda recta finita contiene exactamente dos puntos infinitos, los cuales determinan dos segmentos, uno formado por los puntos finitos y otro por los ultrainfinitos. Observemos que las aplicaciones del grupo hiperbólico hacen corresponder puntos finitos con puntos finitos, pues éstos se caracterizan por que por ellos no pasan tangentes a la cónica, y esta propiedad se conserva por homografías.

En lugar de definir las rectas paralelas como las que no tienen puntos (finitos) en común, conviene distinguir dos clases de paralelismo:

Diremos que dos rectas son *paralelas* o *ultraparalelas* según si se cortan en un punto infinito o ultrainfinito, respectivamente. De este modo, dos rectas sin puntos (finitos) comunes son paralelas o ultraparalelas. Convendremos en que toda recta es paralela a sí misma. En estos términos, por un punto exterior a una recta pasan exactamente dos rectas paralelas, e infinitas ultraparalelas. Por un punto de una recta pasa una única paralela (ella misma) y ninguna ultraparalela.

Diremos que dos rectas son *perpendiculares* si son conjugadas respecto a la cónica infinita.

Teorema 11.2 *Por cada punto pasa una única perpendicular a una recta dada. Dos rectas son ultraparalelas si y sólo si tienen una perpendicular común, y en tal caso es única.*

DEMOSTRACIÓN: La perpendicular a una recta r por un punto p es la recta que une P con el polo de r . Si dos rectas son ultraparalelas, se cortan en un punto ultrainfinito P , luego la polar de P es una perpendicular común. Recíprocamente, si r es una perpendicular común a dos rectas, éstas se cortan en el polo de r , que es ultrainfinito. ■

Diremos que las rectas perpendiculares a una dada forman un *haz* de rectas ultraparalelas, y es claro que tales haces están en correspondencia con los puntos ultrainfinitos.

Tenemos, pues, que un plano hiperbólico es un subconjunto de un plano proyectivo. Si dotamos a éste de estructura euclídea tendremos una representación del plano hiperbólico como subconjunto de un plano euclídeo, con lo que muchos de los teoremas euclídeos se traducirán a teoremas hiperbólicos. Veamos que esto se puede hacer de modo que la relación entre ambos planos sea especialmente simple:

Sea O cualquier punto del plano hiperbólico, sea r su polar, que será una recta ultrainfinita, consideremos el espacio euclídeo que resulta de tomar a r

como recta infinita con la involución (elíptica) inducida por la cónica infinita. Entonces dicha cónica se convierte en una circunferencia C de centro O , las rectas hiperbólicas coinciden con los segmentos euclídeos de extremos en C y las relaciones de orden hiperbólicas coinciden con las euclídeas.

Definición 11.3 Se llama *modelo de Klein* del plano hiperbólico al espacio formado por los puntos interiores de una circunferencia euclídea C y tomando como rectas las cuerdas de C . (Aunque en teoría es irrelevante, usualmente consideraremos, por simplicidad, la circunferencia de centro $(0,0)$ y radio 1.)

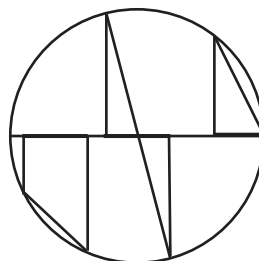
Todos los conceptos referentes al plano hiperbólico se definen en términos de la estructura proyectiva del plano que lo contiene, pero es útil relacionarlos con la geometría euclídea a través del plano de Klein. Por ejemplo, observamos que las perpendiculares hiperbólicas a una recta que pasa por el centro O del plano de Klein coinciden con las euclídeas.

Ejercicio: Describir en términos geométricos sencillos la construcción de la perpendicular a una recta por un punto dado en el plano de Klein.

Ejemplo: Como muestra de la utilidad del plano de Klein probaremos el hecho siguiente:

La proyección ortogonal de una recta en otra que no sea perpendicular a ella tiene por imagen una semirrecta si ambas rectas son paralelas y un segmento en caso contrario.

Podemos tomar uno de los puntos de la recta imagen como centro del plano de Klein, con lo que ésta es un diámetro del mismo. La proyección ortogonal desde la otra recta coincide entonces con la euclídea, pero dicha recta es un segmento, luego su imagen es otro segmento salvo en el caso de que ambas tengan un extremo común, es decir, salvo que se trate de rectas paralelas. ■



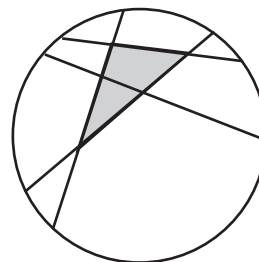
Observemos que todas las homografías del grupo hiperbólico conservan la perpendicularidad, luego todas ellas son el equivalente a las semejanzas de la geometría parabólica. Si queremos definir las isometrías como en el caso parabólico deberemos empezar por definir la reflexión respecto a una recta. Es claro que sólo hay una definición razonable: la reflexión respecto de una recta r ha de tener orden 2, luego ha de ser inducida por una involución en la cónica infinita, y además ha de fijar los puntos de r , luego dicha involución ha de ser la involución que fija a los dos puntos de intersección de r con \mathcal{C} . En definitiva, hemos de definir una *reflexión* como una involución hiperbólica de la cónica infinita. Ahora bien, los teoremas 8.63 4) y 8.74 implican que toda homografía del grupo hiperbólico es producto de reflexiones, luego hemos de considerarlas a todas como isometrías.

Definición 11.4 A las homografías del grupo hiperbólico las llamaremos también *isometrías*. Diremos que dos figuras son *congruentes* si existe una isometría que transforma una en otra.

Notemos que las isometrías que dejan fijo al centro O del plano de Klein fijan también a su recta polar, que es la recta infinita, luego son biyecciones afines. Puesto que fijan a la circunferencia, de hecho son isometrías euclídeas. El recíproco es obvio, luego tenemos que las isometrías hiperbólicas del plano de Klein que fijan a su centro son exactamente las euclídeas, es decir, los giros de centro O y las reflexiones respecto de rectas que pasan por O . En particular la reflexión respecto de una recta r que pasa por O fija a todos los puntos de r y, como O puede ser cualquier punto, esto vale para todas las reflexiones hiperbólicas.

Con esto tenemos definidos todos los conceptos que aparecen en los axiomas de la geometría absoluta excepto las relaciones de orden en una recta, pero éstas son fáciles de definir: si a una recta que quitamos un punto infinito o ultrainfinito obtenemos una recta afín con dos ordenaciones naturales, las restricciones de estas ordenaciones al segmento finito no dependen del punto que hemos quitado, por lo que tenemos dos ordenaciones naturales en porción finita de la recta. En el plano de Klein las relaciones de orden hiperbólicas coinciden con las euclídeas.

Ahora es fácil comprobar que el plano hiperbólico cumple los axiomas de la geometría absoluta. Los grupos A y B se comprueban sin dificultad alguna en el plano de Klein. Por ejemplo, un triángulo en el plano de Klein es un triángulo euclídeo contenido en la circunferencia infinita, si una recta hiperbólica corta a uno de sus lados y no pasa por ninguno de los vértices, entonces ha de cortar a otro de los lados, porque esta propiedad se cumple en el plano euclídeo. Esto demuestra el axioma B5 para el plano hiperbólico. Igualmente se prueban los demás axiomas. El axioma de Arquímedes equivale a que el cuerpo de escalares sea arquimediano.



Los axiomas de congruencia requieren mayor atención porque la congruencia hiperbólica no coincide con la euclídea, como tampoco lo hacen la perpendicularidad ni mucho menos el paralelismo. Pensemos, por ejemplo, que si el plano de Klein satisface realmente los axiomas del grupo C entonces una recta debe contener infinitos segmentos disjuntos de la misma longitud, pero dicha recta es un segmento euclídeo, y no contiene infinitos segmentos euclídeos de la misma longitud. Lo que sucede es que los segmentos hiperbólicos son más cortos desde el punto de vista euclídeo cuanto más cerca están del borde del círculo finito.

Pese a ello, vamos a reducir también al caso euclídeo la comprobación de los axiomas del grupo C . Sólo necesitamos este hecho adicional: dados dos puntos P y Q existe una isometría hiperbólica que transforma uno en otro. De hecho vamos a ver que existe una que los intercambia. Sea r la recta que los une. Sean A y B sus puntos infinitos. El teorema 8.78 nos da que existe un único par M, N

de puntos de r harmónicamente conjugados respecto de P y Q y respecto de A y B . Uno de los dos, digamos M será finito y el otro ultraintinito. Entonces la polar de N es una recta finita s que pasa por M . La reflexión respecto de s fija a M e intercambia a P y Q . Notemos que M es el punto medio hiperbólico del segmento \overline{PQ} .

Ahora vamos a probar los axiomas de congruencia. El axioma C1 es obvio. Para C2 partimos de tres puntos A, B, A' y una semirrecta s de origen A' , y hemos de probar que existe un único punto B' en s tal que $\overline{AB} \equiv \overline{A'B'}$. Aplicando una isometría podemos suponer que A' es el centro del plano de Klein, con lo que la semirrecta es un radio del círculo. Mediante otra isometría transformamos el segmento \overline{AB} en un segmento $\overline{A'B''}$. Ahora podemos aplicar un giro de centro A' , que es una isometría hiperbólica, para transformar $\overline{A'B''}$ en un segmento $\overline{A'B'}$ con B' en s .

Para probar la unicidad hemos de ver que si $\overline{A'B'} \equiv \overline{A'B''}$ con B'' en s entonces $B' = B''$. Ahora bien, una isometría entre los dos segmentos hace corresponder sus extremos. Componiéndola si es preciso con una isometría que intercambie A' y B'' podemos suponer que fija a A' . Pero si una isometría fija a A' y transforma B' en B'' (ambos en la misma semirrecta) necesariamente $B' = B''$, pues la isometría es también una isometría euclídea.

La prueba de C3 es simple: dados A, B, C colineales y A', B', C' colineales también tales que $\overline{AB} \equiv \overline{A'B'}$ y $\overline{BC} \equiv \overline{B'C'}$, entonces existe una isometría que transforma A en A' y B en B' . La imagen de C ha de ser un punto C'' en la misma semirrecta que C' con origen B' y de modo que $\overline{B'C''} \equiv \overline{BC} \equiv \overline{B'C'}$, luego por C2 ha de ser $C'' = C'$.

El axioma C4 se comprueba exactamente igual que C2: dado un ángulo L , una semirrecta s y un semiplano π cuya frontera sea la prolongación de s , suponemos que el origen de s es el centro del círculo, transportamos el ángulo para que su vértice este en dicho centro y aplicamos un giro para obtener un ángulo de lado s y, si es necesario, una simetría para que quede contenido en π . La unicidad se prueba de forma similar.

Antes de probar C5 hemos de probar que la definición de congruencia de triángulos que dimos en el capítulo I coincide con la que aquí hemos dado. Una implicación es evidente. Supongamos que $\widehat{ABC} \equiv \widehat{A'B'C'}$ en el sentido de que sus lados y ángulos son congruentes. Entonces existe una isometría que transforma A y B en A' y B' respectivamente. Componiéndola con una reflexión respecto a AB podemos suponer que ésta transforma C en un punto C'' en el mismo semiplano que C' respecto de $A'B'$. Por C4 se ha de cumplir $\widehat{B'A'C''} = \widehat{B'A'C'}$ y $\widehat{A'B'C''} = \widehat{A'B'C'}$. Por consiguiente $A'C' = A'C''$ y $B'C' = B'C''$, luego $C' = C''$ es el punto de corte de ambas rectas. Así pues, $\widehat{A'B'C'}$ es la imagen de \widehat{ABC} , luego son congruentes en el sentido general.

Ahora la prueba de C5 es completamente análoga a la de C2 y C4 (pero no hace falta probar la unicidad). Así pues:

Teorema 11.5 *El plano hiperbólico cumple los axiomas de la geometría absoluta (incluyendo el axioma de Arquímedes si el cuerpo de escalares es arquimediano), pero no cumple el axioma de las paralelas.*

Las diferencias entre la geometría euclídea y la hiperbólica se hacen notables en cuanto hablamos de perpendicularidad y paralelismo. Ya hemos visto que dos rectas paralelas no tienen perpendiculares comunes. Otro hecho insólito es que dos pares cualesquiera de rectas paralelas son congruentes (pues cada uno está determinado por tres puntos infinitos, y existe una isometría que transporta tres puntos infinitos cualesquiera en otros tres).

Ejercicio: Probar que dos pares de rectas ultraparalelas correspondientes a un mismo haz son congruentes.

11.2 Medida de segmentos y ángulos

Sabemos que el plano hiperbólico (sobre un cuerpo arquimediano) cumple todos los resultados que probamos en el capítulo II. En particular podemos asociar a cada segmento una longitud real, definida salvo un factor constante, de modo que dos segmentos son congruentes si y sólo si tienen la misma longitud. Como ya hemos comentado, aunque representemos el plano hiperbólico como un círculo euclídeo de centro O , la longitud hiperbólica no puede coincidir con la euclídea ni siquiera en el caso de segmentos con un extremo en O .

Vamos a buscar una expresión para la longitud hiperbólica de un segmento. Por el teorema 2.22, cualquier invariante que cumpla $m(u+v) = m(u) + m(v)$ para todo par de segmentos u y v nos da la longitud de los segmentos respecto a cierta unidad de medida. A la hora de buscar un invariante es lógico pensar en la razón doble.

Dados dos puntos P y Q , llamaremos P_∞ y Q_∞ a los puntos infinitos de la recta PQ de modo que el orden de los cuatro sea P_∞, P, Q, Q_∞ . La cantidad $\mathcal{R}(P, Q, Q_\infty, P_\infty)$ es ciertamente invariante por isometrías. Además, si un punto R está entre P y Q es inmediato comprobar la relación

$$\mathcal{R}(P, Q, Q_\infty, P_\infty) = \mathcal{R}(P, R; Q_\infty, P_\infty)\mathcal{R}(R, Q, P_\infty, Q_\infty).$$

Nos interesa una relación de este tipo, pero aditiva, no multiplicativa. Para ello basta tomar logaritmos. Notemos que como $PQ \parallel P_\infty Q_\infty$ la razón doble es positiva. De hecho es fácil ver que es mayor que 1, pues

$$|\mathcal{R}(P, Q, Q_\infty, P_\infty)| = \frac{\overline{PQ_\infty}}{\overline{PP_\infty}} : \frac{\overline{QQ_\infty}}{\overline{QP_\infty}}$$

y ciertamente $\overline{PQ_\infty} \overline{QP_\infty} > \overline{QQ_\infty} \overline{PP_\infty}$. Éste es el motivo por el que hemos considerado esta razón doble y no $\mathcal{R}(P, Q, P_\infty, Q_\infty)$, que es la inversa de la anterior, y por lo tanto menor que 1. Según esto podemos definir

$$d(P, Q) = \frac{1}{2} \log \mathcal{R}(P, Q, Q_\infty, P_\infty) \quad (11.1)$$

y tenemos que $d(P, Q)$ es un número real positivo invariante por isometrías, y cuando P, Q, R son tres puntos colineales, con R entre P y Q , se cumple $d(P, Q) = d(P, R) + d(R, Q)$. Convenimos además que $d(P, P) = 0$.

Así, $d(P, Q)$ es la *distancia hiperbólica* entre los puntos P y Q . En lugar de la constante $1/2$ podríamos haber elegido cualquier otro número real positivo. Modificar esta constante equivale a cambiar de unidad de longitud. Más adelante justificaremos esta elección.

Es clara la analogía entre la fórmula que hemos encontrado para la medida de segmentos hiperbólicos y la fórmula de Cayley que dimos en el capítulo anterior para la medida de ángulos euclídeos. En el caso de los ángulos la proximidad es más que formal, pues podemos reducir la medida de ángulos hiperbólicos a la de ángulos euclídeos. En efecto, sabemos que las isometrías que fijan al centro O del plano de Klein son las mismas para la geometría euclídea y para la hiperbólica, de donde se sigue que dos ángulos de vértice O son congruentes para una si y sólo si lo son para la otra. A su vez de aquí se sigue que la suma de ángulos de vértice O es la misma en ambos casos y la medida de ángulos, tal y como la construimos en el capítulo II no depende de nada más, luego la medida euclídea coincide con la hiperbólica en el caso de ángulos de vértice O . Según la fórmula de Cayley, si r_1 y r_2 son dos rectas que pasan por O el ángulo que forman mide

$$\widehat{r_1 r_2} = \frac{1}{2} \arg \mathcal{R}(r_1, r_2, I_1, I_2),$$

donde I_1 y I_2 son las rectas que pasan por O y por los puntos circulares.

Si obtenemos una expresión invariante por isometrías hiperbólicas y que coincida con ésta sobre las rectas que se cortan en O , tendremos una medida de ángulos hiperbólicos. Ahora bien, esto es fácil de conseguir:

Teorema 11.6 *El ángulo que forman dos rectas hiperbólicas secantes r_1 y r_2 viene dado por la expresión¹*

$$\widehat{r_1 r_2} = \frac{1}{2} \arg \mathcal{R}(r_1, r_2, I_1, I_2),$$

donde I_1 e I_2 son las tangentes a la cónica infinita por el punto $r_1 \cap r_2$.

Efectivamente, esta expresión es invariante por isometrías y si el punto de corte es el centro del plano de Klein entonces las tangentes pasan por los puntos circulares, luego el ángulo coincide con el euclídeo.

La fórmula anterior nos da el (menor) ángulo orientado determinado por dos rectas secantes, donde 'orientado' quiere decir que cambia de signo si cambiamos el orden de las rectas. Para eliminar esta dependencia del orden podemos tomar el argumento en $[-\pi, \pi]$ y tomar el valor absoluto. Esto nos da la medida (en el sentido del capítulo II) del ángulo agudo (o recto) que determinan las dos rectas. Para calcular el ángulo entre dos semirectas aplicamos la fórmula anterior y calculamos el complementario (restando el resultado de π) si el ángulo que queremos medir es obtuso.

¹Como en el caso de la fórmula de Cayley, debemos señalar que la expresión habitual de esta fórmula es $\widehat{r_1 r_2} = -\frac{i}{2} \log \mathcal{R}(r_1, r_2, I_1, I_2)$, donde \log representa al logaritmo complejo.

Vamos a deducir unas fórmulas muy útiles para calcular longitudes y ángulos hiperbólicos a partir de las coordenadas homogéneas de los puntos y rectas. Sea A la matriz de la cónica infinita en un sistema de referencia dado. Sea $f(X, Y) = XAY^t$, de modo que la ecuación de la cónica es $f(X, X) = 0$. Sean X e Y las coordenadas homogéneas de dos puntos interiores a la cónica. La recta que los une está formada por los puntos de coordenadas $X + \lambda Y$, donde $\lambda \in \mathbb{R} \cup \{\infty\}$. Esta recta corta a la cónica en los puntos correspondientes a los valores de λ que satisfacen la ecuación

$$0 = f(X + \lambda Y, X + \lambda Y) = \lambda^2 f(Y, Y) + 2f(X, Y)\lambda + f(X, X).$$

Concretamente:

$$\lambda = \frac{-f(X, Y) \pm \sqrt{f(X, Y)^2 - f(X, X)f(Y, Y)}}{f(Y, Y)}.$$

Podemos considerar a λ como la coordenada cartesiana de un punto de la recta XY respecto a un sistema de referencia de origen X y en el que Y es el punto infinito. Si llamamos λ_1 y λ_2 a los dos valores dados por la fórmula anterior y P_1, P_2 a los puntos correspondientes, entonces

$$\mathcal{R}(X, Y, P_1, P_2) = \mathcal{R}(0, \infty, \lambda_1, \lambda_2) = \frac{\lambda_1}{\lambda_2}.$$

Para calcular $d(X, Y)$ hemos de ordenar P_1 y P_2 de modo que la razón doble sea mayor que 1. Si los ponemos en orden equivocado lo único que sucede es que el logaritmo cambia de signo, por lo que en lugar de preocuparnos por determinar el orden correcto añadimos un valor absoluto a la fórmula. Así, la distancia hiperbólica entre los puntos de coordenadas X e Y es

$$\begin{aligned} d &= \frac{1}{2} \left| \log \frac{f(X, Y) + \sqrt{f(X, Y)^2 - f(X, X)f(Y, Y)}}{f(X, Y) - \sqrt{f(X, Y)^2 - f(X, X)f(Y, Y)}} \right| \\ &= \frac{1}{2} \left| \log \frac{\left(f(X, Y) + \sqrt{f(X, Y)^2 - f(X, X)f(Y, Y)} \right)^2}{f(X, X)f(Y, Y)} \right|. \end{aligned}$$

El coseno hiperbólico biyecta los números positivos con los números mayores que 1, luego es equivalente conocer d que $\cosh d$. Sin embargo vamos a ver que la expresión de $\cosh d$ es algo más simple. Calculamos

$$e^{\pm d} = \left| \frac{f(X, Y) + \sqrt{f(X, Y)^2 - f(X, X)f(Y, Y)}}{\sqrt{f(X, X)f(Y, Y)}} \right|,$$

de donde se obtiene fácilmente

$$\cosh d = \frac{e^d + e^{-d}}{2} = \frac{|f(X, Y)|}{\sqrt{f(X, X)f(Y, Y)}}. \quad (11.2)$$

Ahora particularizamos al caso en que la cónica es la circunferencia de radio 1 y los puntos son $X = (x, y, 1)$, $Y = (x', y', 1)$. Entonces $f(X, Y) = xx' + yy' - 1$, luego en total obtenemos:

Teorema 11.7 *La distancia hiperbólica d en el plano de Klein entre dos puntos cuyas coordenadas cartesianas son (x, y) y (x', y') viene dada por*

$$\cosh d = \frac{1 - xx' - yy'}{\sqrt{(1 - x^2 - y^2)(1 - x'^2 - y'^2)}}.$$

Los mismos cálculos nos dan una expresión similar para el ángulo entre dos rectas. Hemos de partir de la ecuación $F(X, Y) = XA^{-1}Y$ y las coordenadas homogéneas X e Y de dos rectas finitas que se corten en un punto P . Entonces $X + \lambda Y$ recorre todas las rectas que pasan por P y los valores de λ que obtenemos corresponden a las dos tangentes imaginarias por P a la cónica infinita. Si llamamos θ al ángulo entre las dos rectas, llegamos a que

$$\theta = \pm \frac{1}{2} \arg \frac{\left(F(X, Y) + \sqrt{F(X, Y)^2 - F(X, X)F(Y, Y)}\right)^2}{F(X, X)F(Y, Y)}.$$

Más aún, si recordamos la deducción de la fórmula de Cayley veremos que en realidad la razón doble que hemos calculado tiene módulo 1 y así

$$\cos 2\theta \pm i \sin 2\theta = \frac{\left(F(X, Y) + \sqrt{F(X, Y)^2 - F(X, X)F(Y, Y)}\right)^2}{F(X, X)F(Y, Y)}.$$

Ahora usamos que $\cos 2\theta + i \sin 2\theta = (\cos \theta + i \sin \theta)^2$, con lo que simplificamos el cuadrado y tomamos la parte real. El resultado es

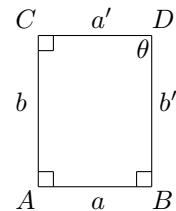
$$\cos \theta = \frac{F(X, Y)}{\sqrt{F(X, X)F(Y, Y)}}.$$

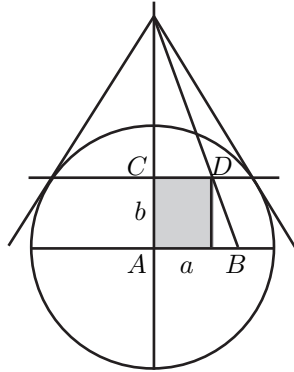
Si lo aplicamos al caso en que la cónica infinita es la circunferencia unidad, la ecuación tangencial es $F(X, Y) = xx' + yy' - zz'$, con lo que llegamos al teorema siguiente:

Teorema 11.8 *El ángulo hiperbólico θ en el plano de Klein entre dos rectas cuyas coordenadas homogéneas son (x, y, z) y (x', y', z') viene dado por*

$$\cos \theta = \frac{xx' + yy' - zz'}{\sqrt{(x^2 + y^2 - z^2)(x'^2 + y'^2 - z'^2)}}.$$

Ejemplo: Consideremos un cuadrilátero hiperbólico con tres ángulos rectos del que conocemos las longitudes a y b de dos de sus lados. Vamos a usar las fórmulas que hemos obtenido para calcular el cuarto ángulo. No perdemos generalidad si suponemos que sus vértices tienen coordenadas $A(0, 0)$, $B(x, 0)$, $C(0, y)$, $D(x, y)$ en el plano de Klein.





Lo primero que conviene observar es que el ángulo θ es agudo. Para ello no hay más que construir la perpendicular a CD por D y observar que forma con CD un ángulo mayor que θ . Por lo tanto, si calculamos el ángulo entre CD y BD estaremos calculando θ y no su complementario. Incidentalmente hemos probado que no existen rectángulos hiperbólicos.

El teorema 11.7 nos da

$$\cosh a = \frac{1}{\sqrt{1-x^2}}, \quad \cosh b = \frac{1}{\sqrt{1-y^2}},$$

de donde podemos despejar $x = \tanh a$, $y = \tanh b$.

Por otra parte el teorema 11.8 aplicado a las rectas $CD = (0, 1, -y)$ y $BD = (1, 0, -x)$ nos da

$$\cos \theta = \frac{xy}{\sqrt{(1-x^2)(1-y^2)}},$$

y al sustituir los valores de x e y en función de a y b llegamos a

$$\cos \theta = \sinh a \sinh b.$$

■

Ejercicio: ¿Qué sucede si $\sinh a \sinh b \geq 1$?

Ejercicio: Probar que la hipotenusa de un triángulo rectángulo es mayor que cualquiera de los catetos. Deducir que la distancia más corta entre un punto y una recta se alcanza en el pie de la perpendicular a la recta por el punto.

Ejercicio: Con la notación del ejemplo anterior, probar que $\sinh a' > \sinh a$ y por lo tanto $a' > a$. Deducir que la distancia más corta entre dos puntos de dos rectas ultraparalelas es la que media entre los pies de su única perpendicular común. Probar también que si una recta tiene dos puntos a la misma distancia de otra recta, entonces no pueden ser paralelas.

La distancia entre dos rectas paralelas Veamos otra aplicación del teorema 11.7. Consideremos dos rectas paralelas. No perdemos generalidad si suponemos que son las rectas $y = 0$ e $y = 1 - x$ del plano de Klein. La distancia a la primera del punto (x, y) de la segunda viene dada por

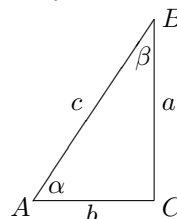
$$\cosh d = \frac{\sqrt{1-x^2}}{\sqrt{1-x^2-(1-x)^2}} = \sqrt{\frac{1+x}{2x}} = \sqrt{\frac{1}{2} \left(1 + \frac{1}{x}\right)}.$$

A la vista de esta fórmula es claro que si r y s son dos rectas paralelas no hay dos puntos en r a la misma distancia de s , así como que la recta r puede orientarse de modo que la distancia a s de un punto P de r se hace arbitrariamente pequeña si P se toma suficientemente a la derecha y arbitrariamente grande si P se toma suficientemente a la izquierda. En particular no hay una mínima distancia entre las dos rectas. ■

11.3 Trigonometría hiperbólica

Los triángulos hiperbólicos verifican relaciones trigonométricas algo más complejas que las de los triángulos euclídeos, pero igualmente manejables.

Consideremos un triángulo rectángulo hiperbólico. Digamos que sus vértices son A, B, C , sus lados a, b, c y sus ángulos α, β, γ , donde $\gamma = \pi/2$. No perdemos generalidad si suponemos que las coordenadas cartesianas de los vértices son $A(0, 0), B(x, y), C(x, 0)$. Como A es el centro del plano de Klein el ángulo hiperbólico α coincide con el ángulo euclídeo correspondiente. Llamemos $z = \sqrt{x^2 + y^2}$, es decir, z es la longitud euclídea de la hipotenusa del triángulo. El teorema 11.7 nos da las relaciones



$$\cosh a = \frac{\sqrt{1-x^2}}{\sqrt{1-z^2}}, \quad \cosh b = \frac{1}{\sqrt{1-x^2}}, \quad \cosh c = \frac{1}{\sqrt{1-z^2}}.$$

Es fácil despejar de las dos últimas

$$x = \tanh b, \quad z = \tanh c.$$

De la primera obtenemos

$$(1-x^2) \cosh^2 a - y^2 \cosh^2 a = 1-x^2 \quad \Rightarrow \quad y^2 \cosh^2 a = (1-\tanh^2 b) \sinh^2 a,$$

con lo que

$$y = \frac{\tanh a}{\cosh b}.$$

De aquí se obtienen todas las fórmulas trigonométricas hiperbólicas sobre triángulos rectángulos. Con ellas podemos calcular los cinco elementos del triángulo a partir de dos cualesquiera de ellos. En primer lugar vemos cómo calcular los ángulos si conocemos dos de los lados.

Conocidos un cateto y la hipotenusa Si conocemos b y c podemos aprovechar que el ángulo hiperbólico α coincide con el euclídeo, con lo que

$$\cos \alpha = \frac{x}{z} = \frac{\tanh b}{\tanh c}.$$

Por simetría se ha de cumplir la fórmula análoga con el otro cateto, luego

$$\tanh b = \cos \alpha \tanh c, \quad (11.3)$$

$$\tanh a = \cos \beta \tanh c. \quad (11.4)$$

Conocidos los catetos Si conocemos a y b usamos la tangente de α :

$$\tan \alpha = \frac{y}{x} = \frac{\tanh a}{\sinh b}.$$

Por simetría la misma fórmula ha de valer para el cálculo de β , con lo que

$$\tanh a = \tan \alpha \sinh b, \quad (11.5)$$

$$\tanh b = \tan \beta \sinh a. \quad (11.6)$$

El teorema de Pitágoras hiperbólico Veamos ahora cómo calcular un lado si conocemos los otros dos. Con esto podremos resolver cualquier triángulo rectángulo conocidos dos lados. Sustituimos los valores de x, y, z en la relación pitagórica euclídea $z^2 = x^2 + y^2$, con lo que obtenemos

$$\tanh^2 c = \tanh^2 b + \frac{\tanh^2 a}{\cosh^2 b}.$$

Usando la relación obvia entre cosenos y tangentes hiperbólicas llegamos a

$$1 - \frac{1}{\cosh^2 c} = 1 - \frac{1}{\cosh^2 b} + \frac{1}{\cosh^2 b} \left(1 - \frac{1}{\cosh^2 a}\right),$$

que se reduce a

$$\cosh c = \cosh a \cosh b. \quad (11.7)$$

Conocidos un lado y un ángulo Las fórmulas que hemos dado para los cosenos y las tangentes nos permiten ya resolver este caso. No obstante la fórmula anterior nos da también expresiones para los senos:

$$\operatorname{sen} \alpha = \frac{y}{z} = \frac{\tanh a}{\cosh b \tanh c} = \frac{\operatorname{senh} a \cosh c}{\cosh a \cosh b \operatorname{senh} c} = \frac{\operatorname{senh} a}{\operatorname{senh} c}.$$

Así pues:

$$\operatorname{senh} a = \operatorname{sen} \alpha \operatorname{senh} c, \quad (11.8)$$

$$\operatorname{senh} b = \operatorname{sen} \beta \operatorname{senh} c. \quad (11.9)$$

Conocidos dos ángulos Éste es el punto más peculiar de la trigonometría hiperbólica, y es que nos permite calcular los lados de un triángulo conocidos sus ángulos. De momento lo probamos sólo para triángulos rectángulos, pero luego veremos que vale en general. Dividimos (11.3) entre (11.9) y aplicamos (11.7):

$$\frac{\cos \alpha}{\operatorname{sen} \beta} = \frac{\tanh b \operatorname{senh} c}{\tanh c \operatorname{senh} b} = \frac{\cosh c}{\cosh b} = \cosh a.$$

Similarmente se prueba la fórmula análoga. Multiplicándolas y aplicando (11.7) obtenemos una tercera que nos permite calcular la hipotenusa:

$$\cos \alpha = \operatorname{sen} \beta \cosh a, \quad (11.10)$$

$$\cos \beta = \operatorname{sen} \alpha \cosh b, \quad (11.11)$$

$$1 = \tan \alpha \tan \beta \cosh c. \quad (11.12)$$

La suma de los ángulos Hay una fórmula de la trigonometría euclídea que ya no es válida en el caso hiperbólico, y es que la suma de los ángulos de un triángulo no es igual a π . De hecho siempre es menor que π . De momento lo probamos

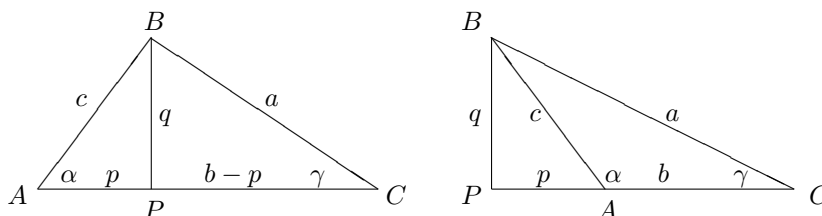
para triángulos rectángulos. Llamemos $\bar{\beta}$ al ángulo euclídeo correspondiente a β , de modo que

$$\cos \bar{\beta} = \frac{y}{z} = \frac{\tanh a}{\cosh b \tanh c} = \frac{\cos \beta}{\cosh b} < \cos \beta.$$

Puesto que los ángulos son agudos, se cumple $\beta < \bar{\beta}$, luego

$$\alpha + \beta + \frac{\pi}{2} < \alpha + \bar{\beta} + \frac{\pi}{2} = \pi.$$

Triángulos arbitrarios La resolución de triángulos arbitrarios se consigue descomponiéndolos en dos triángulos rectángulos. Para ello trazamos una de las alturas, y hay que distinguir dos casos según si el pie P queda entre los dos vértices opuestos o no.



De todos modos hay ocasiones en las que las pruebas se pueden reducir al primer caso (figura de la izquierda), pues por el teorema 2.28 sabemos que los ángulos de un triángulo no suman más de π , luego al menos dos de ellos son agudos. Si suponemos que A y C lo son, entonces el pie P de la altura por C está necesariamente entre A y C . Por ejemplo, teniendo en cuenta que los ángulos no rectos de los triángulos rectángulos suman menos de $\pi/2$, basta considerar este primer caso para concluir que los ángulos de cualquier triángulo \widehat{ABC} suman menos de π .

Teorema 11.9 *La suma de los ángulos de un triángulo hiperbólico es siempre menor que π .*

Vamos a ver que tres de los seis elementos de un triángulo determinan los otros tres. Para ello probaremos teoremas análogos hiperbólicos al teorema de los senos y el teorema del coseno.

Teorema 11.10 (Teorema de los senos) *Todo triángulos hiperbólico satisface la relación*

$$\frac{\text{sen } \alpha}{\text{senh } a} = \frac{\text{sen } \beta}{\text{senh } b} = \frac{\text{sen } \gamma}{\text{senh } c}.$$

DEMOSTRACIÓN: Supongamos la situación de la figura de la izquierda. Entonces

$$\text{senh } q = \text{sen } \gamma \text{ senh } a, \quad \text{senh } q = \text{sen } \alpha \text{ senh } c,$$

y al igualar obtenemos la relación buscada. Si la situación es la de la figura de la derecha la prueba es similar. ■

En geometría hiperbólica contamos con dos teoremas del coseno, uno para lados y otro para ángulos:

Teorema 11.11 (Teorema del coseno) *Todo triángulo hiperbólico satisface las relaciones*

$$\begin{aligned}\cosh a &= \cosh b \cosh c - \sinh b \sinh c \cos \alpha, \\ \cos \alpha &= -\cos \beta \cos \gamma + \sin \beta \sin \gamma \cosh a.\end{aligned}$$

DEMOSTRACIÓN: Como en el teorema anterior, suponemos que el pie de la altura por B queda entre A y C . El caso contrario es análogo.

$$\begin{aligned}\cosh a &= \cosh q \cosh(b-p) = \cosh q (\cosh b \cosh p - \sinh b \sinh p) \\ &= \cosh b \cosh c - \sinh b \cosh c \tanh p \\ &= \cosh b \cosh c - \sinh b \cosh c \cos \alpha \tanh c \\ &= \cosh b \cosh c - \sinh b \sinh c \cos \alpha.\end{aligned}$$

Sea $\delta = \widehat{PBC}$.

$$\begin{aligned}\cos \alpha &= \cosh q \sin(\beta - \delta) = \cosh q \sin \beta \cos \delta - \cosh q \cos \beta \sin \delta \\ &= \cosh q \cosh(v-p) \sin \gamma \sin \beta - \cos \beta \cos \gamma \\ &= \cosh a \sin \gamma \sin \beta - \cos \beta \cos \gamma.\end{aligned}$$

■

Es fácil ver que con estas fórmulas podemos resolver cualquier triángulo.² En particular, si conocemos los tres ángulos el teorema del coseno para ángulos nos permite calcular los tres lados. Por consiguiente:

Teorema 11.12 *Dos triángulos hiperbólicos son congruentes si y sólo si tienen iguales sus ángulos respectivos.*

Unidades de longitud absolutas En la geometría euclídea existe una asimetría entre la medida de longitudes y la medida de ángulos. En efecto, no es posible definir geoméricamente una unidad de longitud. El metro se define actualmente en términos del segundo y de la velocidad de la luz, y a su vez el segundo se define en términos de una propiedad física del átomo de kriptón; antiguamente el metro se definía como la longitud del patrón de platino que se hallaba en París. En cualquier caso, si alguien quiere construir un metro con precisión se ve obligado a observar átomos de kriptón, a viajar a París o algo similar. En cambio, no es necesario viajar a París para construir un ángulo recto, o un radián, o un grado sexagesimal con precisión. Podría haber un ángulo patrón en París, pero no es necesario porque existen unidades naturales de ángulo, en el sentido de que pueden definirse por medios puramente geométricos. Nosotros hemos definido el radián y no hemos tenido que aludir a ningún átomo.

Ahora podemos observar que en la geometría hiperbólica, no sólo podemos definir geoméricamente una unidad de ángulo (por ejemplo, el radián puede

²Como en el caso euclídeo, si conocemos dos lados y un ángulo distinto del que forman puede haber dos soluciones.

definirse como $2/\pi$ ángulos rectos), sino que también podemos definir una unidad de longitud. En efecto, un triángulo equilátero tiene sus tres ángulos iguales (por el teorema de los senos). Si el lado mide l , el teorema del coseno nos da que sus ángulos cumplen

$$\cos \alpha = \frac{\cosh l}{1 + \cosh l}.$$

Por lo tanto, podemos definir la unidad de longitud de un plano hiperbólico como la longitud del lado de un triángulo equilátero cuyos ángulos miden

$$\alpha = \arccos \frac{\cosh 1}{1 + \cosh 1} \approx 0.92 \text{ rad} \approx 52^\circ 38' 35.665''.$$

Esta definición determina de forma absoluta una longitud en el plano. Igual que podemos definir distintas unidades de ángulo (el ángulo recto, el radian, el grado sexagesimal, etc.) también podemos definir otras unidades de longitud cambiando el valor anterior de α por otro número real $< \pi/3$. Esto equivale a modificar la constante $1/2$ de la definición 11.1. Si en lugar de $1/2$ hubiéramos puesto otro valor $k/2$, en la fórmula del teorema 11.7 aparecería $\cosh kd$ y, en general, todas las fórmulas trigonométricas que hemos probado incluirían el factor k multiplicando a los argumentos hiperbólicos. Esto hace que $1/2$ sea la elección que da lugar a fórmulas “más simples”.

La geometría hiperbólica es localmente euclídea Las fórmulas trigonométricas que hemos obtenido para la geometría hiperbólica parecen muy diferentes de sus análogas euclídeas. Sin embargo, sucede que a pequeña escala la geometría hiperbólica se parece a la euclídea y, en cierto sentido, tiende a ella. Para precisar esta idea observamos que al truncar las series de Taylor correspondientes, obtenemos las aproximaciones:

$$\sinh x \approx x, \quad \cosh x \approx 1 + \frac{x^2}{2}, \quad \tanh x \approx x.$$

Así, por ejemplo, si en el teorema de Pitágoras hiperbólico (11.7) aproximamos los cosenos hiperbólicos, obtenemos

$$1 + \frac{c^2}{2} \approx \left(1 + \frac{a^2}{2}\right) \left(1 + \frac{b^2}{2}\right),$$

(donde a y b son los catetos de un triángulo rectángulo y c es su hipotenusa) luego $c^2 \approx a^2 + b^2 + 2a^2b^2$. Si suprimimos también el último factor³ (de orden 4) queda $c^2 \approx a^2 + b^2$, que es el teorema de Pitágoras euclídeo.

³Esta supresión de términos no es arbitraria, aunque no hemos definido el concepto de polinomio de Taylor de una función de dos variables, lo que hacemos es tomar el polinomio de Taylor de grado dos de ambos miembros, con lo que obtenemos

$$1 + \frac{c^2}{2} \approx 1 + \frac{a^2}{2} + \frac{b^2}{2}.$$

Por ejemplo si viviéramos en un espacio hiperbólico en el que la unidad natural de longitud equivaliera a 1 km, entonces la hipotenusa de un triángulo rectángulo cuyos catetos midieran 3 y 4 metros respectivamente mediría

$$1000 \operatorname{arg} \cosh(\cosh(3/1000) \cosh(4/1000)) \approx 5.0000048,$$

es decir, 4.8 micras más de lo que exige el teorema de Pitágoras euclídeo.

Similarmente, en el teorema de los senos podemos aproximar los senos hiperbólicos por sus argumentos y entonces tenemos el teorema de los senos euclídeo, la primera de las fórmulas del teorema del coseno se convierte en

$$1 + \frac{a^2}{2} \approx \left(1 + \frac{b^2}{2}\right) \left(1 + \frac{c^2}{2}\right) - bc \cos \alpha,$$

y eliminando el término de orden 4 queda $a^2 \approx b^2 + c^2 - 2bc \cos \alpha$, que es el teorema del coseno euclídeo. La segunda fórmula se convierte en

$$\cos \alpha \approx -\cos \beta \cos \gamma + \sin \beta \sin \gamma + \frac{a^2}{2} \sin \beta \sin \gamma,$$

que equivale a

$$\cos \alpha \approx -\cos(\beta + \gamma) = \cos(\pi - \beta - \gamma),$$

lo que a su vez equivale a que $\alpha + \beta + \gamma \approx \pi$. Así pues, cuanto menores son los lados de un triángulo, más se parece a π la suma de sus ángulos.

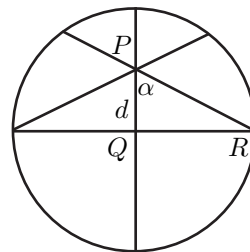
El lector puede comprobar que las fórmulas que hemos obtenido para resolver triángulos rectángulos también tienden a sus análogas euclídeas para longitudes pequeñas.

En resumen, mientras la geometría euclídea es invariante a escala (las figuras tienen las mismas propiedades independientemente de su tamaño, y no puede decirse que una figura sea “grande” o “pequeña” en ningún sentido absoluto), en la geometría hiperbólica sí que podemos distinguir entre figuras “pequeñas” (en relación con la unidad absoluta), cuyas propiedades son indistinguibles de las de las figuras euclídeas, y figuras “grandes”, en las que las diferencias se vuelven patentes.

Ejemplo: el ángulo de paralelismo Consideremos una recta cualquiera r y un punto exterior P . Sea s la perpendicular a r por P y sea $Q = r \cap s$. Podemos suponer que Q es el centro del plano de Klein y que r es el eje x . Sea t una de las paralelas a r por P . Sea $R = r \cap t$.

Entonces \widehat{PQR} es lo que se llama un triángulo rectángulo asintótico, con dos lados infinitos y un ángulo nulo. Las mismas técnicas que hemos empleado con los triángulos rectángulos usuales nos dan la relación entre sus dos únicos elementos finitos, el lado $d = \overline{PQ}$ y el ángulo $\alpha = \hat{P}$, a saber:

$$\cos \alpha = \tanh d.$$



También es fácil ver que α es el mismo para las dos paralelas por P y recibe el nombre de *ángulo de paralelismo* a r por P . Su interpretación es sencilla: las rectas que forman con s un ángulo inferior a α cortan a r , las dos que forman ángulo α son paralelas y las que forman ángulo superior a α son ultraparalelas. Hemos probado que el ángulo de paralelismo a una recta por un punto es agudo y depende únicamente de la distancia d del punto a la recta.

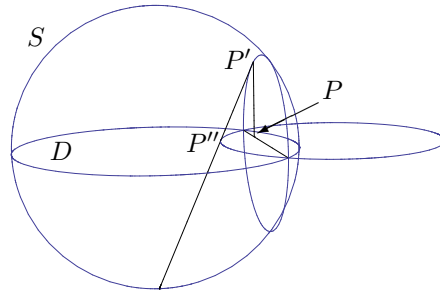
Ejercicio: Probar la relación de Lobachevski para el ángulo de paralelismo:

$$\tan \frac{\alpha}{2} = e^{-d}.$$

11.4 El modelo de Poincaré

Del mismo modo que el plano de Klein relaciona la geometría hiperbólica con la euclídea, hay otro modelo de plano hiperbólico, debido a Poincaré, que relaciona la geometría hiperbólica con la geometría circular.

Consideremos una esfera S y sea D el círculo delimitado por su ecuador C . Sea O su centro. Podemos considerar D como plano de Klein, en cuyo caso lo llamaremos D_K . Sea D_P el mismo círculo D , pero visto ahora como subconjunto de la recta proyectiva compleja. Sea $G_{KP} : D_K \rightarrow D_P$ la biyección definida como sigue: A cada punto P de D_K le asignamos el punto P' donde la perpendicular a D por P corta a una de las semiesferas de S . A continuación proyectamos P' mediante la proyección estereográfica desde el polo de la semiesfera opuesta, con lo que obtenemos un punto $P'' = G_{KP}(P)$.



Si consideramos la esfera de centro $(0, 0, 0)$ y radio 1, las expresiones explícitas para G_{KP} y para su inversa G_{PK} son

$$G_{KP}(x, y) = \left(\frac{x}{1 + \sqrt{1 - x^2 - y^2}}, \frac{y}{1 + \sqrt{1 - x^2 - y^2}} \right),$$

$$G_{PK}(x, y) = \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1} \right).$$

Consideramos a D_P como plano hiperbólico tomando como rectas las imágenes por G_{KP} de las rectas de D_K , entendiendo que dos rectas $G_{KP}[r]$ y $G_{KP}[s]$ son perpendiculares si y sólo si lo son r y s , y que las isometrías de D_P son las aplicaciones de la forma $G_{PK}fG_{KP}$, donde f es una isometría en D_K .

Obviamente, con esto hemos transportado a través de G_{KP} toda la geometría de D_K , y por consiguiente D_P , con sus rectas, cumple exactamente las mismas propiedades geométricas que D_K con las suyas. El interés de esta construcción radica en que la geometría de D_P puede describirse de forma natural en términos de la geometría circular de la recta compleja, sin ninguna referencia a D_K .

En primer lugar notamos que al levantar las rectas de D_K sobre la semiesfera éstas se transforman en los cortes de S con los planos verticales, que son las circunferencias perpendiculares a C . Más exactamente, las rectas hiperbólicas se transforman en las porciones de estas circunferencias contenidas en la semiesfera elegida. Al aplicar la proyección estereográfica obtenemos que las rectas de D_P son simplemente las intersecciones con D_P de las circunferencias de la recta compleja perpendiculares a C . Entre ellas están las rectas que pasan por O .

Según el teorema 10.33, a cada isometría f de D_K le podemos asignar la única transformación circular directa f' que coincide con f en C . Ésta puede fijar a D_P o bien intercambiarlo con el exterior de C . En el segundo caso llamamos f' a la composición de f' con la inversión respecto de C y así f' es la única transformación circular que fija a D_P y que coincide con f en C . Es claro que la correspondencia $f \mapsto f'$ biyecta las isometrías con las transformaciones circulares que fijan a D_P . Vamos a probar que $f' = G_{PK}fG_{KP}$. Si A y B son dos puntos de C , entonces f transforma la recta AB de D_K en $f(A)f(B)$. Por otra parte $G_{KP}[AB]$ es el arco de la circunferencia ortogonal a C que pasa por A y por B , luego $f'[G_{KP}[AB]]$ es el arco de la circunferencia ortogonal a C que pasa por $f(A)$ y $f(B)$, luego $G_{PK}[f'[G_{KP}[AB]]] = f(A)f(B)$. Así pues, f y $G_{KP}f'G_{PK}$ coinciden sobre rectas. Expresando cada punto de D_K como intersección de dos rectas vemos que también coinciden sobre puntos.

Por consiguiente, las isometrías de D_P son las transformaciones circulares que fijan a D .

Para acabar de caracterizar la geometría de D_P probaremos que los ángulos entre dos rectas (circunferencias) de D_P son los mismos que en la geometría circular. Esto se debe a que G_{KP} deja invariantes a las rectas de D_K que pasan por O , luego el ángulo entre dos rectas de D_P que pasan por O coincide con el ángulo entre estas mismas rectas en D_K , que coincide con el euclídeo, que a su vez coincide con el ángulo entre ellas en la geometría circular, luego efectivamente el ángulo entre dos rectas que pasan por O para la geometría de D_P y para la geometría circular es el mismo. Como las isometrías de D_P son transformaciones circulares, conservan ángulos para las dos geometrías, luego la igualdad de ángulos es válida para rectas que se corten en un punto arbitrario.

Por último notamos que, dado que la geometría de D_P está descrita completamente en términos de la geometría circular, no necesitamos suponer que D_P es un círculo euclídeo, sino que todo lo dicho vale en general tomando como D_K uno de los lados (círculo o semiplano) que determina una circunferencia en la recta proyectiva compleja. En resumen:

Definición 11.13 Llamaremos *modelo de Poincaré* del plano hiperbólico al conjunto de los puntos situados a un mismo lado D (círculo o semiplano) de

una circunferencia C en la recta proyectiva compleja, tomando como rectas las intersecciones con D de las circunferencias perpendiculares a C . Las isometrías hiperbólicas son entonces las transformaciones circulares que fijan a D . Los ángulos hiperbólicos coinciden con los de la geometría circular.

Hemos probado que la geometría del plano de Poincaré es exactamente la misma que la del plano de Klein. Aunque con las nociones de recta, isometría y perpendicular tenemos totalmente caracterizada la geometría hiperbólica, vamos ver que también hay una expresión sencilla para la longitud hiperbólica en el plano de Poincaré en términos de la geometría circular.

Veamos primero un caso particular. Tomemos como C el círculo de centro O y radio 1 en el plano $z = 0$ y consideremos el punto $(x, 0, 0)$ del plano de Klein. Su imagen en el plano de Poincaré se obtiene elevándolo, por ejemplo, a la semiesfera superior, lo que nos da el punto $(x, 0, \sqrt{1-x^2})$, y después calculando la intersección con $z = 0$ de la recta que une este punto con el polo sur $(0, 0, -1)$. Un cálculo sencillo nos da que dicha imagen es el punto $(x', 0)$, donde

$$x' = \frac{x}{1 + \sqrt{1-x^2}}.$$

En términos de números complejos dicha imagen es simplemente el número $x' + 0i = x'$.

Vamos a calcular la distancia hiperbólica en el plano de Poincaré entre los puntos $P(0, 0)$ y $Q(x, 0)$. Los puntos infinitos involucrados son $P_\infty(-1, 0)$ y $Q_\infty(1, 0)$. Entonces

$$\mathcal{R}(P, Q, Q_\infty, P_\infty) = \frac{1+x}{1-x}, \quad \text{luego} \quad d_K(P, Q) = \frac{1}{2} \log \frac{1+x}{1-x}.$$

Ésta ha de ser también la distancia hiperbólica entre $P' = 0$ y $Q' = x'$ en el plano de Poincaré. Vamos a expresarla en términos de 0 , x' y los puntos infinitos de la recta que determinan, que son $P'_\infty = -1$ y $Q'_\infty = 1$. Para ello observamos que

$$\begin{aligned} \mathcal{R}(P', Q', Q'_\infty, P'_\infty)^2 &= \left(\frac{1+x'}{1-x'} \right)^2 = \left(\frac{1+x+\sqrt{1-x^2}}{1-x+\sqrt{1-x^2}} \right)^2 \\ &= \frac{1+x+(1+x)\sqrt{1-x^2}}{1-x+(1-x)\sqrt{1-x^2}} = \frac{(1+x)(1+\sqrt{1-x^2})}{(1-x)(1+\sqrt{1-x^2})} \\ &= \frac{1+x}{1-x} = \mathcal{R}(P, Q, Q_\infty, P_\infty). \end{aligned}$$

Así pues, tomando logaritmos queda que la distancia hiperbólica en el plano de Poincaré entre P' y Q' viene dada por

$$d_P(P', Q') = \log \mathcal{R}(P', Q', Q'_\infty, P'_\infty).$$

En principio tenemos probada esta fórmula cuando $P' = 0$ y Q' está sobre el eje real, pero ambos miembros de la igualdad son invariantes por las isometrías

del plano de Poincaré (las transformaciones circulares conservan la razón doble), y todo par de puntos puede transformarse mediante una isometría en un par en las condiciones anteriores. Esto prueba el teorema siguiente:

Teorema 11.14 *La distancia hiperbólica entre dos puntos P y Q en el plano de Poincaré viene dada por*

$$d_P(P, Q) = \log \mathcal{R}(P, Q, Q_\infty, P_\infty),$$

donde P_∞ y Q_∞ son los puntos infinitos de la recta (circular) PQ , de modo que el orden entre los cuatro puntos sea P_∞, P, Q, Q_∞ .

Ejercicio: Transportar la estructura hiperbólica del plano de Poincaré a una semi-esfera mediante la proyección estereográfica. Describir las rectas, los ángulos y las isometrías.

11.5 Las isometrías hiperbólicas

Vamos a describir con más detalle los distintos tipos de isometrías hiperbólicas:

Reflexiones Recordemos que hemos definido la reflexión respecto a una recta r como la isometría que induce en la cónica infinita la involución que fija a los puntos infinitos de r . Al representarla en el plano de Klein con centro en un punto de r vemos que la reflexión ha de ser una isometría euclídea, por lo que ha de ser la simetría respecto de r y por consiguiente fija a todos los puntos de r . En el modelo de Poincaré las reflexiones coinciden con las de la geometría circular. Desde un punto de vista puramente hiperbólico, la reflexión respecto a la recta r es la isometría que a cada punto P lo lleva al punto situado sobre la perpendicular a r por P y a la misma distancia de ésta que P .

Como en el caso euclídeo, las isometrías son productos de reflexiones:

Teorema 11.15 *Toda isometría es producto de a lo sumo tres reflexiones.*

DEMOSTRACIÓN: Por el teorema 8.63 4) toda isometría es producto de dos involuciones, una de las cuales es hiperbólica, según se ve en la prueba. Si la otra es elíptica, la descomponemos en producto de dos involuciones hiperbólicas por 8.74. ■

Giros Un *giro* de centro O es la composición de dos reflexiones respecto de dos rectas que se cortan en O . Tomando a O como centro del plano de Klein vemos que los giros así definidos coinciden con los euclídeos, luego los giros hiperbólicos con un mismo centro O forman un grupo isomorfo a \mathbb{R}/\mathbb{Z} . Las imágenes de un punto P por todos los giros de centro O forman una circunferencia de centro O . Existe cierta tradición de llamar *ciclos* a las circunferencias hiperbólicas. En el plano de Klein de centro O las circunferencias hiperbólicas de centro O coinciden con las euclídeas.

Giros infinitos Definimos un *giro* alrededor de un punto infinito O como la composición de dos reflexiones cuyos ejes pasen por O , es decir, los giros infinitos son la composición de dos reflexiones respecto a dos rectas paralelas. La mejor representación de un giro infinito nos la da el semiplano de Poincaré. No perdemos generalidad si suponemos que el centro del giro O es el punto infinito en el sentido de la geometría circular. Entonces las rectas que pasan por O son las rectas verticales (suponiendo que representamos la frontera del semiplano como una recta horizontal), y la composición de dos reflexiones respecto a dos de estas rectas es simplemente una traslación (euclídea) en dirección horizontal. Vemos, pues, que los giros hiperbólicos alrededor de un punto infinito O forman un grupo isomorfo a \mathbb{R} .

Las imágenes de un punto por todos los giros de centro un punto infinito O constituyen una figura llamada *horociclo* (gr. = círculo en el borde) de centro O . En el caso concreto del semiplano de Poincaré, los horociclos de centro ∞ son las rectas horizontales. De aquí deducimos varios hechos que son invariantes por isometrías, y por lo tanto válidos para horociclos con centro arbitrario:

- *Por cada punto finito pasa un único horociclo de centro un punto infinito. Dos horociclos con el mismo centro son disjuntos.*
- *Si P es un punto de un horociclo H de centro O , las únicas rectas cuyo único punto en común con H es P son OP y su perpendicular por P .*
- *Dados dos puntos de un horociclo de centro O , existe un único giro de centro O que transforma uno en el otro.*

Consideremos una transformación circular que convierta el semiplano de Poincaré en un círculo D con circunferencia C de centro P . Sea O la imagen de ∞ . Entonces el diámetro OP es la imagen de una recta vertical, luego los horociclos de centro ∞ se transforman en las circunferencias contenidas en D que pasan por O y perpendiculares a OP , luego todas ellas tienen en común con C la recta tangente por O . En general, dos cónicas se dicen *osculantes* (lat. ‘que se besan’) si comparten un punto y su tangente por el mismo. Puesto que sólo hay una circunferencia en D osculante a C por P que pase por un punto dado de OP , concluimos que todas ellas son horociclos. Así pues:

Los horociclos en el plano de Poincaré circular son las circunferencias osculantes a la circunferencia infinita.

Los horociclos del plano de Klein serán las imágenes de estas circunferencias por la correspondencia entre ambos planos. Ésta se calcula proyectando en primer lugar la circunferencia sobre la esfera desde su polo Sur, lo que nos da una circunferencia en la esfera (pues la proyección estereográfica transforma circunferencias en circunferencias). Más precisamente, si el horociclo de partida tiene centro O , es una circunferencia simétrica respecto de OP , luego su imagen en la esfera es simétrica respecto del plano OPN (donde N es el polo Norte de la esfera). A continuación proyectamos la circunferencia de nuevo sobre el disco

unitario, pero esta vez ortogonalmente. Puesto que la proyección ortogonal es una proyección perspectiva, dicha imagen será una elipse con un punto en la circunferencia infinita y simétrica respecto de OP . Esto significa que OP es uno de sus ejes, y si la distancia entre los vértices correspondientes es d , es fácil ver que la circunferencia en la esfera tiene que tener diámetro $\sqrt{2d}$, así como que ésta es también la distancia entre los vértices del eje menor de la proyección. Por lo tanto:

Los horociclos del plano de Klein son las elipses osculantes a la circunferencia infinita cuyos ejes miden d y $\sqrt{2d}$, con $0 < d < 2$.

Aquí hemos usado también que por cada punto entre O y P pasa una única elipse en estas condiciones, luego todas ellas tienen que ser horociclos.

Ejercicio: Probar que los horociclos de centro O quedan fijos por las reflexiones respecto de rectas que pasan por O .

Ejercicio: Probar que si dos puntos P y Q están en un mismo horociclo de centro O , entonces la distancia de P a OQ es igual a la distancia de Q a OP . Deducir de aquí una caracterización de los horociclos en términos de la distancia hiperbólica.

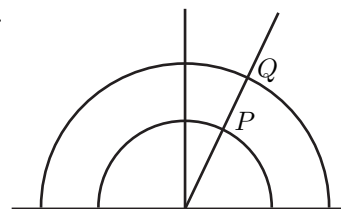
Traslaciones Definimos una *traslación* hiperbólica como la composición de dos reflexiones respecto a rectas hiperparalelas. El modelo que muestra más claramente las traslaciones es el semiplano de Poincaré. Si lo identificamos concretamente con el semiplano $y > 0$ en \mathbb{R}^2 , no perdemos generalidad si suponemos que la perpendicular común a los ejes de las dos reflexiones es el semieje $x = 0$. Entonces estos ejes son dos semicircunferencias de centro $(0, 0)$ y es fácil ver que la composición de las reflexiones respecto a ambas es una homotecia $(x, y) \mapsto (\alpha x, \alpha y)$ con $\alpha > 0$. Más aún, toda homotecia de centro $(0, 0)$ y razón positiva es una traslación hiperbólica.

También es claro que la única recta fijada es el semieje vertical. En general, cada traslación hiperbólica fija a una única recta, a la que llamaremos *eje* de la traslación, y que no es sino la perpendicular común a los ejes de las reflexiones que componen la traslación. Las traslaciones con un eje dado forman un grupo isomorfo a \mathbb{R}^* .

Las imágenes de un punto P a través de todas las traslaciones de eje r forman una figura llamada *curva equidistante* a r por P . En el semiplano de Poincaré, las equidistantes al semieje $x = 0$ son las semirrectas que pasan por el origen.

Su nombre se debe a que si P no está en r , la equidistante a r por P está formada por todos los puntos que están en el mismo semiplano que P respecto a r y cuya distancia a r es la misma que la de P .

En efecto, si Q está en la equidistante a r por P , la traslación de eje r que envía P a Q es una isometría que fija a r , luego ambos puntos están a la misma distancia de r . Por otra parte, si r es la semirrecta $x = 0$ en el semiplano de Poincaré, en cada semicircunferencia de centro

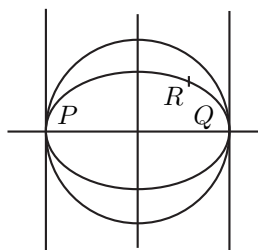


$(0,0)$ sólo puede haber un dos puntos a la misma distancia de r que P , y los que están al mismo lado de r que P forman precisamente la semirrecta que pasa por $(0,0)$ y P , que es la equidistante.

Dada una recta r y un número $d > 0$, existen dos equidistantes a una distancia d de r , una en cada uno de los semiplanos determinados por r . Si pasamos a un círculo el semiplano de Poincaré, las dos equidistantes a una recta se convierten en dos arcos de circunferencia que pasan por sus puntos infinitos. Al pasar al modelo proyectivo nos encontramos con que las dos equidistantes se convierten en una misma cónica.

En efecto: No perdemos generalidad si suponemos que r es un diámetro del plano de Klein. Por concretar supondremos que r es el eje $y = 0$. Sean P y Q los puntos infinitos de r . Las tangentes a C por P y Q son las rectas verticales $x = \pm 1$. Dado cualquier punto R , es claro que existe una única cónica \mathcal{C} osculante a C por P y Q y que pasa por R .

Es fácil obtenerla como imagen de C por una afinidad que fija a los ejes $x = 0$ e $y = 0$, por lo que \mathcal{C} es una elipse con estas rectas como ejes. Una reflexión respecto a una recta vertical intercambia los puntos P y Q así como las tangentes a C (luego también a \mathcal{C}) por estos puntos. Además fija a dos puntos de \mathcal{C} , luego de hecho fija a \mathcal{C} . Por consiguiente las traslaciones de eje r , que son composición de dos de estas reflexiones, también fijan a \mathcal{C} . Esto prueba que la equidistante a r por R está contenida en \mathcal{C} , más concretamente, en el arco de \mathcal{C} que está en el mismo semiplano que R respecto a r . Por otra parte, cada recta vertical ha de cortar a la equidistante a r por R y sólo corta a dicho arco de \mathcal{C} en un punto, luego tenemos la igualdad.



Falta probar que el arco opuesto de \mathcal{C} es la otra equidistante correspondiente a la misma distancia, pero esto es evidente, pues la reflexión respecto a r fija a \mathcal{C} (porque una elipse es simétrica respecto de sus ejes), luego cada punto del arco opuesto dista de r lo mismo que su simétrico. En resumen:

Las equidistantes a una recta r son las cónicas osculantes a la cónica infinita por los puntos infinitos de r . Más exactamente, son los dos arcos en que estos puntos dividen a la cónica.

Reflexión con traslación Una reflexión con traslación es simplemente la composición de una traslación con la reflexión respecto de su eje.

Teorema 11.16 *Toda isometría del plano proyectivo distinta de la identidad es de uno de los cinco tipos anteriores.*

DEMOSTRACIÓN: Recordemos que toda homografía en una cónica determina un centro y un eje (definición 9.35). De la propia definición se sigue que el eje pasa por los puntos fijos de la cónica si es que los hay. Por lo tanto, el eje es secante, tangente o exterior a la cónica según si la homografía es hiperbólica, parabólica o elíptica. En el último caso el eje corta a la cónica en dos puntos

imaginarios, fijados por la extensión de la homografía a la complexificación. Es claro entonces que la homografía ha de fijar al eje en cualquier caso, y por consiguiente también al centro, que es su polo. Recordemos también el teorema 9.38, según el cual el centro de la composición de dos involuciones es la intersección de los ejes de éstas.

Sea, pues, una isometría hiperbólica f de centro O y eje r . Si O es un punto finito entonces podemos tomarlo como centro del plano de Klein y concluir que f ha de ser un giro o una reflexión (porque al fijar a la circunferencia y a su centro, fija a la recta infinita, luego es una isometría euclídea). Como la restricción de f a la circunferencia infinita es elíptica (el eje es exterior) de hecho ha de ser un giro. Así pues, los giros son las isometrías hiperbólicas con centro finito. El centro del giro coincide con el centro en el sentido general.

Si O es infinito, podemos descomponer f en producto de dos involuciones cuyos ejes pasan por O y no pueden ser tangentes (porque no hay involuciones parabólicas) luego son rectas finitas y por consiguiente f es un giro infinito de centro O . Concluimos que los giros infinitos son las isometrías hiperbólicas de centro infinito. El centro del giro coincide con el centro en sentido general.

Si O es ultrainfinito entonces el eje r es una recta finita. Si f es una involución entonces es la reflexión respecto a r . En caso contrario expresamos f como composición de dos involuciones. La prueba del teorema 8.63 muestra que podemos exigir que una de ellas sea hiperbólica, es decir, una reflexión. Si ambas son reflexiones entonces f es una traslación, pues los ejes de las reflexiones se cortan en el punto ultrainfinito O . Notamos también que el eje de la traslación coincide con el eje en sentido general.

Si una de las involuciones es elíptica el determinante de f ha de ser negativo. Consideramos la reflexión g respecto a r . Entonces fg fija a los puntos infinitos de r , luego sigue teniendo eje r , pero tiene determinante positivo. Así pues, ha de descomponerse en producto de dos reflexiones y es, por consiguiente, una traslación. Por consiguiente, $f = (fg)g$ es una simetría con traslación. Es fácil ver que $fg = gf$. ■

Finalmente probamos que las isometrías pueden caracterizarse como en el caso euclídeo como las biyecciones que conservan las distancias.

Teorema 11.17 *Una biyección f del (conjunto de puntos finitos del) plano hiperbólico en sí mismo que cumpla $d(f(P), f(Q)) = d(P, Q)$, para todo par de puntos P y Q , es una isometría.*

DEMOSTRACIÓN: Dados dos puntos P y Q , existe una isometría que coincide con f sobre ambos, y por el mismo razonamiento anterior podemos suponer que f fija a ambos. Es fácil ver que f fija o intercambia los semiplanos respecto de PQ . Componiéndola si es preciso con la reflexión respecto de PQ , podemos suponer que los fija. Dado cualquier punto R exterior a PQ sea $R' = f(R)$. Los triángulos \widehat{PQR} y $\widehat{PQR'}$ son congruentes y están en el mismo semiplano respecto a PQ , luego han de ser iguales, luego $f(R) = R$, luego f es la identidad. ■

11.6 La geometría diferencial hiperbólica

En esta última sección vamos a estudiar el plano hiperbólico desde el punto de vista de la geometría diferencial. Para ello tomamos, por ejemplo,

$$H = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\},$$

y observamos que es una variedad diferenciable en el sentido de [An 7.1] tomando como carta de cualquier punto a la identidad $X_K : D_K \rightarrow H$, donde $D_K = H$, pero aquí lo llamamos D_K para enfatizar que la carta X_K identifica los puntos de H con los puntos del plano de Klein.

La aplicación $X_P = G_{PK} : D_P \rightarrow H$ construida en la sección 11.4 es también una carta para todos los puntos de H , a través de la cual dichos puntos se identifican con los del plano de Poincaré circular. Si componemos X_P con una homografía $S \rightarrow D_P$, donde $S = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$, obtenemos otra carta que nos permite identificar los puntos de H con los de S , considerado como semiplano de Poincaré.

De este modo, los distintos modelos del plano hiperbólico pueden identificarse con cartas de un mismo plano hiperbólico “abstracto”. El hecho de que desde un punto de vista conjuntista hayamos definido H como el círculo unitario en \mathbb{R}^2 es algo “anecdótico” que podemos “olvidar” a partir de este momento, lo que equivale a “olvidar” que X_K es la aplicación identidad y pasar a pensar que es una carta más, en pie de igualdad con X_P o cualquier otra carta de H .

En estos términos tenemos, por ejemplo, el concepto de “recta hiperbólica”, que técnicamente se corresponde con las rectas del modelo de Klein, pero que conviene pensar que es un concepto “abstracto” con traducciones distintas en cartas distintas. Así, las antiimágenes de las rectas hiperbólicas en el plano de Klein son los segmentos de recta (euclídea) que contiene, mientras que las antiimágenes de estas mismas rectas en el modelo circular de Poincaré son los arcos de circunferencia ortogonales a la circunferencia unidad, y sus antiimágenes en el semiplano de Poincaré S son las semicircunferencias con centro en la frontera del semiplano y las rectas verticales.

El tensor métrico Si insistiéramos en recordar que hemos definido H como un cierto abierto en \mathbb{R}^2 , entonces tendríamos que concluir, según [An 7.17] que su tensor métrico (respecto de la carta X_K) es $g = dx \otimes dx + dy \otimes dy$, pero usando este tensor obtendríamos la distancia euclídea entre dos puntos de H (visto como el disco unidad en \mathbb{R}^2), y no es esto lo que nos interesa. Vamos a ver que podemos definir un tensor métrico en H , distinto del inducido por su inclusión en \mathbb{R}^2 , que induce en H la distancia hiperbólica. Dicho tensor tendrá la forma

$$g = E dx \otimes dx + F(dx \otimes dy + dy \otimes dx) + G dy \otimes dy,$$

para ciertas funciones E, F, G que dependerán de la carta considerada. Equivalentemente, podemos determinarlas encontrando la expresión para el elemento de longitud hiperbólica:

$$ds^2 = E dx^2 + 2F dx dy + G dy^2.$$

Consideremos una parametrización $\alpha : [a, b] \rightarrow H$ de un segmento de recta hiperbólica, es decir, una biyección entre un intervalo en \mathbb{R} y un segmento hiperbólico de modo que si $t_1 < t_2 < t_3$ entonces $\alpha(t_1) - \alpha(t_2) - \alpha(t_3)$. Sea $(x(t), y(t))$ la expresión en coordenadas de α respecto del modelo de Klein. Todo segmento puede parametrizarse en estas condiciones. Por ejemplo, si $p, q \in D_K$ son las coordenadas de sus extremos, basta tomar $(x(t), y(t)) = (1-t)p + tq$, con $0 \leq t \leq 1$.

Llamemos $s : [a, b] \rightarrow \mathbb{R}$ a la longitud de arco, es decir, $s(t) = d(\alpha(a), \alpha(t))$, donde d representa a la distancia hiperbólica. Es claro entonces que

$$|s(t+h) - s(t)| = d(\alpha(t+h), \alpha(t)),$$

pues la distancia hiperbólica es aditiva sobre los segmentos de recta. Equivalentemente,

$$\frac{s(t+h) - s(t)}{h} = \frac{d(\alpha(t+h), \alpha(t))}{|h|},$$

luego

$$\frac{ds}{dt} = \lim_{h \rightarrow 0} \frac{d(\alpha(t+h), \alpha(t))}{|h|}. \quad (11.13)$$

Para calcular este límite usamos que, en general, en el modelo del plano hiperbólico determinado por la cónica de ecuación $f(X, X) = 0$, la distancia hiperbólica cumple (11.2), que equivale a

$$\sinh d(X, Y) = \sqrt{\frac{f^2(X, Y) - f(X, X)f(Y, Y)}{f(X, X)f(Y, Y)}}.$$

Preferimos la expresión en términos del seno hiperbólico, aunque es más complicada, porque nos permite eliminar dicho seno hiperbólico usando que

$$\lim_{x \rightarrow 0} \frac{\sinh x}{x} = 1.$$

En estos términos tenemos que considerar la expresión en coordenadas homogéneas $X(t) = (x(t), y(t), 1)$, y entonces

$$\begin{aligned} \frac{ds}{dt} &= \lim_{h \rightarrow 0} \frac{d(X(t+h), X(t))}{|h|} = \lim_{h \rightarrow 0} \frac{1}{\frac{\sinh d(X(t+h), X(t))}{d(X(t+h), X(t))}} \frac{\sinh d(X(t+h), X(t))}{|h|} \\ &= \lim_{h \rightarrow 0} \frac{1}{|h|} \sqrt{\frac{f^2(X(t), X(t+h)) - f(X(t), X(t))f(X(t+h), X(t+h))}{f(X(t), X(t))f(X(t+h), X(t+h))}}. \end{aligned}$$

Sea $X(t+h) = X(t) + \Delta X(t, h)$. Usando que f es bilineal la expresión se simplifica:

$$\frac{ds}{dt} = \lim_{h \rightarrow 0} \frac{1}{|h|} \sqrt{\frac{f^2(X, \Delta X) - f(X, X)f(\Delta X, \Delta X)}{f(X, X)(f(X, X) + 2f(X, \Delta X) + f(\Delta X, \Delta X))}}$$

$$\begin{aligned}
&= \lim_{h \rightarrow 0} \sqrt{\frac{f^2(X, \frac{\Delta X}{h}) - f(X, X)f(\frac{\Delta X}{h}, \frac{\Delta X}{h})}{f(X, X)(f(X, X) + 2f(X, \Delta X) + f(\Delta X, \Delta X))}} \\
&= \sqrt{\frac{f^2(X, X') - f(X, X)f(X', X')}{f^2(X, X)}}.
\end{aligned}$$

A partir de aquí particularizamos ya al caso de la forma bilineal que determina la circunferencia unitaria, es decir, $f(X_1, X_2) = z_1 z_2 - x_1 x_2 - y_1 y_2$. El resultado es:

$$\left(\frac{ds}{dt}\right)^2 = \frac{(xx' - yy')^2 + (1 - x^2 - y^2)(x'^2 + y'^2)}{(1 - x^2 - y^2)^2}.$$

Operando y multiplicando por dt^2 obtenemos una expresión para el elemento de longitud hiperbólica en el plano de Klein:

$$ds^2 = \frac{dx^2 + dy^2 - (x dy - y dx)^2}{(1 - x^2 - y^2)^2}.$$

Definición 11.18 Llamaremos tensor métrico hiperbólico al tensor en el plano hiperbólico H que en la carta correspondiente al plano de Klein es de la forma

$$g = E dx \otimes dx + F(dx \otimes dy + dy \otimes dx) + G dy \otimes dy,$$

donde

$$E = \frac{1 - y^2}{(1 - x^2 - y^2)^2}, \quad F = \frac{xy}{(1 - x^2 - y^2)^2}, \quad G = \frac{1 - x^2}{(1 - x^2 - y^2)^2}.$$

Hemos probado que si $\alpha : [a, b] \rightarrow H$ parametriza un segmento de recta, entonces la función $s(t) = d(\alpha(a), \alpha(t))$ verifica que $s(a) = 0$ y

$$\frac{ds}{dt} = \sqrt{g_{\alpha(t)}(\alpha'(t), \alpha'(t))} = \sqrt{Ex'^2 + 2Fx'y' + Gy'^2},$$

por lo que la longitud hiperbólica del segmento es

$$s(b) = \int_a^b \sqrt{Ex'^2 + 2Fx'y' + Gy'^2} dt = \int_a^b \sqrt{E dx^2 + 2F dx dy + G dy^2},$$

y que el tensor métrico que acabamos de definir es el único tensor en H para el que se cumple esto.

Puesto que la expresión del tensor métrico en el plano de Klein no es muy manejable, antes de analizarlo con más detalle conviene encontrar expresiones alternativas respecto de otras cartas. Consideremos, por ejemplo, las coordenadas polares. Si diferenciamos $(x, y) = (r \cos \theta, r \sin \theta)$ y sustituimos en la expresión de ds obtenemos

$$ds^2 = \frac{dr^2}{(1 - r^2)^2} + \frac{r^2}{1 - r^2} d\theta^2.$$

Si un punto se encuentra a una distancia hiperbólica ρ del centro del plano de Klein, la distancia euclídea es $r = \tanh \rho$. Al diferenciar esta relación y sustituir en la expresión anterior obtenemos el elemento de longitud hiperbólico en coordenadas polares hiperbólicas, que resulta ser

$$ds^2 = d\rho^2 + \sinh^2 \rho d\theta^2. \quad (11.14)$$

Esta fórmula es intrínseca, en el sentido de que las coordenadas (ρ, θ) de un punto P representan la distancia hiperbólica de P a un punto fijo O y el ángulo de la semirrecta \overrightarrow{OP} con una semirrecta fija de origen O , y nada de esto depende del plano de Klein. Por lo tanto la fórmula es válida también en el círculo de Poincaré. La relación entre la distancia euclídea r y la distancia hiperbólica ρ de un punto P al punto 0 en el círculo de Poincaré es

$$\rho = \log \frac{1+r}{1-r},$$

de donde

$$\sinh \rho = \frac{2r}{1-r^2}, \quad d\rho = \frac{2dr}{1-r^2}.$$

Por consiguiente, el elemento de longitud en coordenadas polares (euclídeas) en el círculo de Poincaré resulta ser

$$ds = 2 \frac{\sqrt{dr^2 + r^2 d\theta^2}}{1-r^2}.$$

Observemos que el numerador es el elemento de longitud euclídea en coordenadas polares. Si usamos la notación compleja para el arco $z(t) = x(t) + iy(t)$ y llamamos $|dz| = \sqrt{dx^2 + dy^2}$ al elemento de longitud euclídea, entonces tenemos

$$ds = \frac{2|dz|}{1-|z|^2}.$$

La transformación circular

$$z = \frac{iw + 1}{w + i}$$

convierte el círculo $|z| < 1$ en el semiplano $\text{Im } w > 0$. Se comprueba⁴ que

$$dz = \frac{-2dw}{(w+i)^2}, \quad |dz| = \frac{2|dw|}{|w+i|^2}, \quad 1-|z|^2 = \frac{4y}{|w+i|^2}.$$

De todo esto resulta que

$$ds = \frac{|dw|}{y}.$$

⁴La teoría de funciones de variable compleja justifica que es lícito calcular dz derivando la expresión anterior como si w fuera una variable real y tratando las constantes imaginarias como constantes reales. El lector puede, si lo desea, hacer los cálculos en términos de las dos variables reales x, y , pero está avisado de que llegará al mismo resultado.

En resumen, hemos encontrado las expresiones siguientes para el tensor métrico hiperbólico:

Carta	Coordenadas	E	F	G
Modelo de Klein	(x, y)	$\frac{1-y^2}{(1-x^2-y^2)^2}$	$\frac{xy}{(1-x^2-y^2)^2}$	$\frac{1-x^2}{(1-x^2-y^2)^2}$
en polares	(r, θ)	$\frac{1}{(1-r^2)^2}$	0	$\frac{r^2}{1-r^2}$
Coordenadas polares	(ρ, θ)	1	0	$\sinh^2 \rho$
Círculo de Poincaré	(x, y)	$\frac{4}{(1-x^2-y^2)^2}$	0	$\frac{4}{(1-x^2-y^2)^2}$
en polares	(r, θ)	$\frac{4}{(1-r^2)^2}$	0	$\frac{4r^2}{(1-r^2)^2}$
Semiplano de Poincaré	(x, y)	$1/y^2$	0	$1/y^2$

Ahora es claro (por ejemplo, observando la expresión para el semiplano de Poincaré) que, para cada $P \in H$, la forma bilineal $g_P : T_P H \times T_P H \rightarrow \mathbb{R}$ es un producto escalar, pues su matriz en la base de $T_P H$ determinada por la carta del semiplano de Poincaré es la matriz diagonal $[1/y^2, 1/y^2]$, que tiene estrictamente positivos los elementos de su diagonal, luego $g_P(\vec{v}, \vec{v}) > 0$ para todo vector $\vec{v} \in T_P H$ no nulo.

Esto hace que el elemento de longitud $ds(P) : T_P H \rightarrow \mathbb{R}$ dado por

$$ds = \sqrt{E dx^2 + 2F dx dy + G dy^2}$$

está bien definido para todo punto y todo vector y podemos definir la *longitud hiperbólica* de cualquier curva $\alpha : [a, b] \rightarrow H$ como

$$L(\alpha) = \int_a^b ds = \int_a^b \sqrt{E dx^2 + 2F dx dy + G dy^2},$$

donde $(x(t), y(t))$ es la expresión en coordenadas de α .

Hemos demostrado que si α parametriza un segmento hiperbólico, entonces la longitud así definida es la distancia hiperbólica entre sus extremos.

Más aún, podemos afirmar que, si $s(t) = L(\alpha|_{[a,t]})$, entonces $s(a) = 0$ y

$$\frac{ds}{dt} = \sqrt{E x'^2 + 2F x' y' + G y'^2} = \lim_{h \rightarrow 0} \frac{d(\alpha(t+h), \alpha(t))}{|h|}.$$

En efecto, la primera igualdad se cumple ahora por la definición de longitud y la segunda la hemos demostrado, pues en el argumento que hemos dado para parametrizaciones de segmentos sólo hemos usado realmente que α es una tal parametrización a la hora de justificar la igualdad (11.13), mientras que el resto del argumento (el que prueba la igualdad entre el límite y la expresión en términos del tensor métrico) es válido para curvas cualesquiera.

Esta expresión muestra que la longitud hiperbólica de una curva es invariante por isometrías hiperbólicas, pues si componemos α con una isometría el argumento del límite de la expresión anterior no se altera. A su vez, esto implica que las isometrías hiperbólicas son isometrías en el sentido de [An 7.18].

Posponemos hasta el final de la sección la discusión sobre por qué esta definición de longitud de una curva es razonable (para tratar a la vez la cuestión análoga respecto de otros conceptos que vamos a considerar).

Observemos que las expresiones diferenciales de la longitud hiperbólica muestran la naturaleza de la distancia hiperbólica mucho más claramente que la fórmula de la distancia entre dos puntos. Por ejemplo, en el caso del círculo de Poincaré:

$$ds = \frac{2\sqrt{dx^2 + dy^2}}{1 - x^2 - y^2}$$

vemos que, prescindiendo del 2 que aparece para normalizar la unidad de longitud, el elemento de longitud hiperbólica es el mismo que el euclídeo salvo un denominador que es el más simple posible que tiende a 0 cuando nos acercamos a la frontera del disco, de modo que la distancia hiperbólica entre dos puntos es más grande cuanto más cerca están de dicha frontera. Lo mismo se ve en la expresión para el semiplano de Poincaré, donde el elemento de longitud hiperbólica es el elemento de longitud euclídea dividido entre y , para que las longitudes se dilaten a medida que nos acercamos a la frontera del semiplano.

Ejemplo Usando la expresión en coordenadas polares es fácil calcular la longitud de una circunferencia de radio r :

$$\int_0^{2\pi} \sinh r \, d\theta = 2\pi \sinh r.$$

Como era de esperar, para valores pequeños de r el resultado se asemeja a la longitud euclídea $2\pi r$. Más en general, la expresión (11.14) para el elemento de longitud en coordenadas polares se aproxima al euclídeo $ds^2 = dr^2 + \rho^2 d\theta$ para valores pequeños de ρ , lo cual significa que la distancia hiperbólica se confunde con la euclídea en regiones suficientemente pequeñas del plano. ■

Observemos ahora que, según ya hemos señalado, el producto escalar asociado al tensor métrico en el semiplano de Poincaré es

$$g_P(\vec{v}, \vec{w}) = \frac{1}{y^2} \vec{v} \cdot \vec{w},$$

por lo que dos vectores son ortogonales si y sólo si su representación en la carta correspondiente al semiplano son ortogonales en el sentido euclídeo. Pero a su vez, la ortogonalidad para el producto euclídeo equivale a la ortogonalidad respecto de la geometría circular, que a su vez equivale a la ortogonalidad hiperbólica.

Por consiguiente, concluimos que dos vectores son ortogonales respecto a la geometría hiperbólica si y sólo si son ortogonales respecto del tensor métrico hiperbólico.

Hemos probado esto considerando la carta correspondiente al semiplano de Poincaré, pero tanto la ortogonalidad hiperbólica como la ortogonalidad respecto del tensor métrico son conceptos intrínsecos de la geometría hiperbólica,

independientes de la carta considerada, por lo que el resultado es cierto en general.

Ejemplo Veamos un sencillo ejemplo numérico para ilustrar la afirmación precedente. Consideremos el plano de Klein, tomemos un punto ultraintinito cualquiera, por ejemplo, $(1, 1)$. Calculamos su polar, que es $x + y = 1$. Tomemos un punto arbitrario en esta recta, por ejemplo, $P = (0.8, 0.2)$. Entonces, dos vectores perpendiculares en P son $(-1, 1)$ (el vector de la recta polar) y $(0.2, 0.8)$ (el vector que une P con el polo $(1, 1)$). Vemos que estos vectores no son ortogonales respecto del producto escalar euclídeo, pero sí que lo son respecto del producto escalar hiperbólico, que en dicho punto es el dado por la matriz

$$M_P = \begin{pmatrix} \frac{75}{8} & \frac{25}{16} \\ \frac{25}{16} & \frac{225}{64} \end{pmatrix}$$

(esto resulta de calcular E, F, G para $(x, y) = (0.8, 0.2)$ con las fórmulas de la primera fila de la tabla precedente). Ahora es fácil comprobar que

$$(-1, 1)M(0.8, 0.2) = 0.$$

Hemos probado que esto no es casual. ■

Geodésicas Hemos visto que el tensor métrico hiperbólico determina tanto la longitud hiperbólica como la perpendicularidad hiperbólica. Ahora vamos a demostrar que, de hecho, determina las propias rectas hiperbólicas, pues éstas resultan ser las geodésicas respecto del tensor métrico.

La definición [An 7.29] de derivada covariante no tiene sentido en nuestro contexto, pues tiene en cuenta la forma en que la variedad está sumergida en \mathbb{R}^m , mientras que nosotros estamos prescindiendo de dicha inmersión, pues nos llevaría a la métrica euclídea. En su lugar, vamos a tomar como definición de los símbolos de Christoffel a las ecuaciones

$$\begin{aligned} \Gamma_{11}^1 &= \frac{E_x}{2E}, & \Gamma_{11}^2 &= -\frac{E_y}{2G}, & \Gamma_{12}^1 &= \frac{E_y}{2E}, \\ \Gamma_{12}^2 &= \frac{G_x}{2G}, & \Gamma_{22}^1 &= -\frac{G_x}{2E}, & \Gamma_{22}^2 &= \frac{G_y}{2G}. \end{aligned}$$

Al final de esta sección discutiremos por qué esta definición es razonable. Admitiendo estas igualdades por definición, recordamos que la condición general (en términos de coordenadas) para que una curva $\alpha(t) = (x_1(t), \dots, x_n(t))$ sea una geodésica es que su aceleración geodésica

$$(x_k'' + \sum_{i,j=1}^n \Gamma_{ij}^k x_i' x_j')$$

sea en cada punto proporcional a $\alpha' = (x_1', \dots, x_n')$.

Si consideramos el semiplano de Poincaré, donde el tensor métrico tiene la expresión más simple, obtenemos que todos los símbolos de Christoffel son nulos excepto

$$\Gamma_{11}^2 = \frac{1}{y}, \quad \Gamma_{12}^1 = \Gamma_{21}^1 = -\frac{1}{y}, \quad \Gamma_{22}^2 = -\frac{1}{y},$$

y la aceleración geodésica de una curva $(x(t), y(t))$ es

$$\left(x'' - 2\frac{x'y'}{y}, y'' + \frac{x'^2 - y'^2}{y} \right).$$

Con esto podemos probar:

Teorema 11.19 *Las geodésicas del plano hiperbólico son las rectas.*

DEMOSTRACIÓN: Continuamos trabajando en el semiplano de Poincaré, donde tenemos dos tipos de rectas: las rectas verticales y las semicircunferencias con centro en el eje real. Las rectas verticales se parametrizan como $(x, y) = (x_0, t)$, luego $(x', y') = (0, 1)$ y $(x'', y'') = (0, 0)$. La aceleración geodésica es entonces $(0, -1/y)$, que ciertamente es proporcional a $(0, 1)$, luego las rectas verticales son geodésicas.

Las semicircunferencias se parametrizan como $(x, y) = (x_0 + r \cos t, r \sin t)$, de modo que $(x', y') = (-r \sin t, r \cos t)$ y un simple cálculo muestra que la aceleración geodésica es

$$\left(r \cos t, \frac{-r \cos^2 t}{\sin t} \right) = -\frac{\cos t}{\sin t} (-r \sin t, r \cos t),$$

que ciertamente es múltiplo de (x', y') . Con esto hemos probado que todas las rectas son geodésicas. Teniendo en cuenta que por cada punto pasa una única geodésica en cada dirección (véase el ejemplo tras el teorema [An 7.10]) y que por cada punto pasa una recta en cada dirección, concluimos que las rectas son las únicas geodésicas del plano hiperbólico. ■

El área hiperbólica En la sección [An 7.4] se muestra que el tensor métrico de una superficie diferenciable determina también su elemento de área, que en términos de las coordenadas de una carta es

$$d\sigma = \sqrt{EG - F^2} dx dy.$$

Una vez más, provisionalmente vamos a aceptar esto como definición del área hiperbólica por analogía, y dejamos para más adelante la justificación de tal definición.

Las expresiones más simples para el elemento de área las encontramos en el plano de Klein en coordenadas polares, en el que

$$d\sigma = \sinh \rho d\rho d\theta$$

y en el semiplano de Poincaré, en el que $d\sigma = (1/y^2) dx dy$.

Ejemplo Vamos a usar la primera para calcular el área de un círculo hiperbólico de radio r :

$$A = \int_0^{2\pi} \int_0^r \sinh \rho \, d\rho \, d\theta = 2\pi \int_0^r \sinh \rho \, d\rho = 2\pi(\cosh r - 1) = 4\pi \sinh^2 \frac{r}{2}.$$

Como era de esperar, para valores pequeños de r tenemos que $A \approx \pi r^2$. ■

Consideremos ahora un triángulo rectángulo como indica la figura. Tomemos como origen de las coordenadas polares el vértice A . Para hallar su área hemos de integrar $d\sigma$ variando θ entre 0 y α y, para un θ dado, en virtud de la relación trigonométrica

$$\tanh b = \cos \theta \tanh \rho.$$

vemos que ρ ha de llegar hasta

$$\operatorname{argtanh} \frac{\tanh b}{\cos \theta}.$$

Por lo tanto el área es

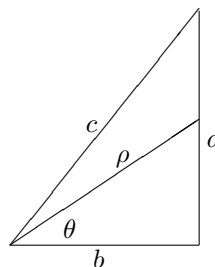
$$\begin{aligned} \int_0^\alpha \int_0^{\operatorname{argtanh}(\tanh b / \cos \theta)} \sinh \rho \, d\rho \, d\theta &= \int_0^\alpha \left(\frac{1}{\sqrt{1 - \frac{\tanh^2 b}{\cos^2 \theta}}} - 1 \right) d\theta \\ &= \int_0^\alpha \left(\frac{\cos \theta}{\sqrt{(\cos^2 \theta - 1) + (1 - \tanh^2 b)}} - 1 \right) d\theta \\ \int_0^\alpha \left(\frac{\cosh b \cos \theta}{\sqrt{1 - (\cosh b \sin \theta)^2}} - 1 \right) d\theta &= \operatorname{arcsen}(\cosh b \sin \alpha) - \alpha \\ &= \operatorname{arcsen} \cos \beta - \alpha = \frac{\pi}{2} - \beta - \alpha = \pi - \alpha - \beta - \gamma. \end{aligned}$$

Todo triángulo hiperbólico T se puede expresar como la unión o la diferencia de dos triángulos rectángulos, de donde es fácil concluir que en general que el área de un triángulo de ángulos α, β, γ es

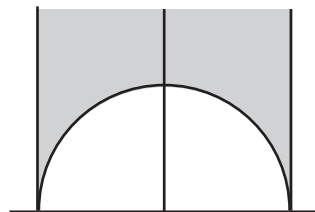
$$\pi - \alpha - \beta - \gamma.$$

Esto nos da otra prueba de que la suma de los ángulos de un triángulo hiperbólico es siempre menor que π , y se aproxima más a π cuanto menor es el triángulo. ■

Ahora vamos a usar el semiplano de Poincaré para calcular el área de un triángulo con sus tres vértices infinitos. En principio, nada garantiza que dicha área tenga que ser finita, pero vamos a ver que es así.



No perdemos generalidad si suponemos que los vértices son $(-1, 0)$, $(1, 0)$ y el punto infinito. La recta que une los dos puntos finitos es la dada por $y = \sqrt{1-x^2}$, con lo que tenemos que integrar $d\sigma = (1/y^2) dx dy$ en el recinto sombreado en la figura.



$$\int_{-1}^1 \int_{\sqrt{1-x^2}}^{+\infty} \frac{1}{y^2} dy dx.$$

Para calcular la integral interior usamos el teorema de la convergencia monótona:

$$\int_{\sqrt{1-x^2}}^{+\infty} \frac{1}{y^2} dy = \lim_{t \rightarrow +\infty} \int_{\sqrt{1-x^2}}^t \frac{1}{y^2} dy = \lim_{t \rightarrow +\infty} \left[-\frac{1}{y} \right]_{\sqrt{1-x^2}}^t = \frac{1}{\sqrt{1-x^2}}.$$

Por lo tanto:

$$\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx = [\arcsen x]_{-1}^1 = \pi.$$

Concluimos que todos los triángulos con vértices infinitos tienen área π (lo que viene a decir que sus ángulos son nulos).

Ejercicio: Probar que el área de un triángulo con vértices finitos o infinitos es igual a π menos sus ángulos finitos.

La curvatura de Gauss La curvatura de Gauss mide la forma en que una superficie “se curva” en el espacio que la contiene. Esto no tiene sentido en nuestro contexto, puesto que estamos considerando una métrica no relacionada con el espacio \mathbb{R}^2 en el que tenemos sumergido el plano hiperbólico. No obstante, en la sección [An 7.6] vimos que la curvatura de Gauss de una superficie puede medirse exclusivamente a partir del tensor métrico, sin hacer referencia a nada externo a la superficie, luego tiene sentido plantearse qué obtendremos si empleamos dicho procedimiento de medida en nuestro contexto.

Concretamente, tras [An 7.42] vimos que en una carta en la que $F = 0$ y $E = G = A$, la curvatura de Gauss puede calcularse mediante la fórmula

$$K = -\frac{1}{2A} \left(\frac{\partial^2 \log A}{\partial x^2} + \frac{\partial^2 \log A}{\partial y^2} \right).$$

Si aplicamos esto a la carta correspondiente al semiplano de Poincaré, en la cual $A = 1/y^2$, obtenemos $K = -1$. Así pues, el plano hiperbólico tiene curvatura -1 en todos sus puntos. Aquí influye la elección de la unidad de longitud. Con otra elección, la curvatura resultaría igualmente constante y negativa, pero su valor sería otro.

Esto significa que si existiera una superficie en \mathbb{R}^3 isométrica al plano hiperbólico, su curvatura tendría que ser constante igual a -1 . Puede probarse que no existe tal superficie, pero sí que conocemos superficies con curvatura constante igual a -1 . Una de ellas es la pseudoesfera y ahora vamos a analizar su relación con el plano hiperbólico.

La pseudoesfera En la sección [An 7.6] hemos visto que la pseudoesfera (con $l = 1$) admite cartas

$$X :]x_0 - \pi, x_0 + \pi[\times]1, +\infty[\longrightarrow \mathbb{R}^3$$

(para cualquier elección de x_0) respecto de las cuales el tensor métrico tiene la forma

$$ds^2 = \frac{1}{y^2}(dx^2 + dy^2),$$

que es exactamente el tensor métrico del semiplano de Poincaré. Por lo tanto, la carta X es una isometría entre un trozo del semiplano de Poincaré y media pseudoesfera menos un meridiano.

Esto se traduce en que si interpretamos el dominio de X como un subconjunto del semiplano de Poincaré o, mejor aún, si interpretamos a X como una biyección $f : U \longrightarrow V$ entre un abierto en el plano hiperbólico y un abierto en la pseudoesfera, la longitud hiperbólica de cada curva contenida en U coincide con la longitud (euclídea) de su imagen en V , el área hiperbólica de cualquier región contenida en U coincide con el área (euclídea) de su imagen en V , una curva contenida en U es una geodésica (respecto a la definición *ad hoc* que hemos dado para el plano hiperbólico) si y sólo si su imagen es una geodésica euclídea, etc.

Para entender la relación que este hecho establece entre el plano hiperbólico y la pseudoesfera, conviene pensar en la relación que hay entre el plano euclídeo usual y el cilindro, pues es casi la misma. Una carta para un cilindro es

$$X(x, y) = (\cos x, \operatorname{sen} x, y),$$

y el elemento de longitud correspondiente es $ds^2 = dx^2 + dy^2$, el mismo que el euclídeo, lo cual se traduce en que X es una isometría entre un trozo de plano y el cilindro menos un meridiano. Así, el cilindro y el plano son localmente isométricos, pero no son isométricos, en el sentido de que no es posible establecer una isometría entre todo el plano y todo el cilindro. La isometría local significa que alguien que “viva en un cilindro” no puede darse cuenta de ello estudiando la geometría de su entorno, pues ésta es la misma que la del plano. Lo único que puede hacer para detectar que no “vive en un plano” es constatar que si camina en línea recta (= geodésica) en la dirección adecuada, al cabo de un tiempo vuelve al punto de partida, cosa que nunca podría ocurrir en un plano.

Lo mismo sucede con la pseudoesfera: alguien que “viva en ella” se encuentra con que la geometría de su entorno es exactamente la geometría hiperbólica, pero puede descubrir que no vive en un plano hiperbólico constatando que si camina por una geodésica adecuada regresa al punto de partida.⁵

⁵En realidad, en este caso hay una segunda diferencia, y es que la pseudoesfera se interrumpe bruscamente cuando llegamos a $y = 0$, lo cual no es el equivalente a llegar a $y = 0$ en el semiplano de Poincaré, sino a llegar en éste a $y = 1$. La diferencia es sustancial: si caminamos verticalmente hacia $y = 0$ por el semiplano de Poincaré con pasos de igual longitud (hiperbólica), nunca llegaremos a $y = 0$ (porque la longitud euclídea de los pasos es cada vez menor), pero sí que rebasaremos $y = 1$. En la pseudoesfera sucede lo segundo: si caminamos hacia abajo, en un número finito de pasos llegamos a su borde, cosa que no ocurriría en un plano hiperbólico.

Discusión de las definiciones de longitud y área hiperbólicas La isometría local entre el plano hiperbólico, con los conceptos geométricos que hemos definido en él “arbitrariamente”, y una superficie en \mathbb{R}^3 , con los conceptos geométricos “naturales”, justifica que las definiciones que hemos dado son de hecho “las correctas”. Esto ha de entenderse como sigue: aunque técnicamente hay que definir los conceptos de longitud de un segmento o área de una región, lo cierto es que tales definiciones no pueden (o no deben) ser arbitrarias, sino que el hecho de que la longitud de los segmentos hiperbólicos esté determinada hace que cada curva tenga una longitud “de verdad”, la que podríamos calcular aproximando la curva por una sucesión de pequeños segmentos de recta que nos den idea de cuántos pasos tendría que dar para recorrerla un “viajante” tan pequeño que sus pasos rectos fueran indistinguibles de arcos de la curva, e igualmente, cada región del plano tiene un área “de verdad”, la que podríamos calcular cubriéndola con muchos pequeños fragmentos indistinguibles de cuadrados euclídeos, que nos informaran de “cuánta tela” hace falta trocear en cuadraditos para cubrir con ellos la región.

Consideremos de nuevo la isometría $f : U \rightarrow V$ entre un abierto del plano hiperbólico y un abierto de la pseudoesfera. Las longitudes de los segmentos de recta son los mismos en U y en V , pero en V las longitudes “de verdad” de las curvas, las coherentes con la longitud hiperbólica de los arcos de geodésica (que en V es también la longitud “de verdad” de dichos arcos), es la dada por el tensor métrico de V , que es el mismo que el de U , por lo que la longitud que corresponde “de verdad” a las curvas de U es la que hemos definido como tal, la determinada por el tensor métrico hiperbólico, y lo mismo se aplica al área.

Concluimos así que las definiciones que hemos dado de longitud y de área hiperbólica son “las correctas” al menos en U . Pero ambas son invariantes por isometrías, y existe una isometría hiperbólica que transforma cualquier punto en cualquier otro, luego podemos afirmar, más aún, que todo punto de H tiene un entorno en el que nuestras definiciones de longitud y de área son “las correctas”. Por último, el hecho de que las definiciones sean “localmente correctas” implica que lo son globalmente. Por ejemplo, podemos dividir cualquier curva que una dos puntos de H en un número finito de arcos lo suficientemente pequeños como para que podamos asegurar que su longitud, tal y como la hemos definido es “la correcta”, y entonces la longitud total, que es la suma de dichas longitudes, es también “la que debe ser”. Lo mismo vale para el área.

En cuanto al concepto de geodésica hiperbólica, ya hemos visto que se corresponde con el concepto de “recta hiperbólica”, pero esto a su vez tiene una interpretación. Puede probarse que en cualquier variedad diferencial las geodésicas son las curvas que unen dos puntos por el camino más corto.⁶ Por lo tanto, el hecho de que las rectas hiperbólicas sean geodésicas se interpreta como que las rectas minimizan las distancias entre los puntos que unen.

⁶En realidad esto sólo es cierto localmente, pues, por ejemplo, podemos unir dos puntos cercanos de un cilindro mediante una geodésica que dé la vuelta al cilindro, y su longitud será mayor que la de otras curvas que unan los dos puntos sin ser geodésicas, pero sin rodear el cilindro. No obstante, en el caso del plano hiperbólico —pero no en el de la pseudoesfera— es cierto también globalmente.

Capítulo XII

La geometría elíptica

La geometría elíptica es la que resulta de dotar de una métrica al plano proyectivo completo, considerando a todos los puntos finitos. Así como la geometría hiperbólica incumple el axioma de las paralelas porque por un punto exterior a una recta pasa más de una paralela, la geometría elíptica lo incumple en sentido contrario: como bien sabemos, en un plano proyectivo no existen paralelas.

Formalmente, la geometría elíptica se puede desarrollar de forma completamente análoga a la hiperbólica. El único cambio es que vamos a tomar como cónica infinita una cónica imaginaria, de modo que no dé lugar a ninguna división de los puntos del plano. Al igual que en los otros casos, podríamos definir espacios elípticos de cualquier dimensión, pero nosotros nos limitaremos al caso plano porque es completamente representativo y más simple en la práctica.

12.1 El plano elíptico

Definición 12.1 Un *plano elíptico* es un plano proyectivo real en el que hemos seleccionado una cónica imaginaria (cónica infinita). Diremos que dos rectas son *perpendiculares* si son conjugadas respecto a la cónica infinita. Llamaremos *grupo elíptico* o *grupo de isometrías* al grupo de las homografías que fijan a la cónica infinita. Dos figuras son *congruentes* si existe una isometría que transforma una en otra.

Vamos a introducir un modelo que nos permita relacionar la geometría elíptica con la euclídea. El plano proyectivo es de la forma $P(V)$, donde V es un espacio vectorial tridimensional real. Existe un sistema de referencia proyectivo respecto al cual la cónica infinita tiene ecuación $x^2 + y^2 + z^2 = 0$. Este sistema de referencia se corresponde con una base de V . Dotamos a V del producto escalar que en dicha base tiene matriz identidad, es decir, el producto respecto al cual dicha base es ortonormal. Más en general, podemos considerar a V como un espacio afín euclídeo en el que hemos fijado un sistema de referencia ortonormal de modo que los puntos del plano proyectivo son las rectas que pasan por el origen O y las coordenadas homogéneas de un punto son las coordenadas en

el sistema de referencia fijado de uno de los puntos (que no sea O) de la recta correspondiente.

Dos puntos de coordenadas X e Y son conjugados respecto de la cónica infinita si y sólo si $XY^t = 0$, es decir, si y sólo si vistos como rectas de V son ortogonales.

Las rectas proyectivas están formadas por los puntos que cumplen una ecuación $ax + by + cz = 0$, es decir, son los planos de V que pasan por O . También es claro que un punto y una recta son conjugados si y sólo si, vistos como recta y plano, uno es perpendicular al otro.

Todas estas relaciones se pueden simplificar si escogemos representantes concretos en V para los puntos proyectivos. Sea S la esfera en V de centro O y radio 1. Cada recta que pasa por O corta a S en dos puntos antípodos. En términos proyectivos, cada punto tiene dos ternas de coordenadas homogéneas sujetas a la restricción $x^2 + y^2 + z^2 = 1$. Podemos, pues, identificar el plano proyectivo con el conjunto de los pares de puntos antípodos de S . De ahora en adelante y salvo que indiquemos lo contrario, cuando hablemos de un punto de S nos referiremos a un par de puntos antípodos considerados como un solo objeto. Las rectas proyectivas se corresponden con las intersecciones con S de los planos que pasan por su centro, es decir, son los llamados *círculos máximos* de la esfera, las circunferencias cuyo radio coincide con el de la esfera. Dos circunferencias son ortogonales (en el sentido de la geometría elíptica) si la perpendicular por el origen del plano de una pasa por la otra. En otros términos, tomando a una como ecuador de la esfera, sus perpendiculares son las circunferencias que pasan por los polos.

Consideremos ahora una isometría elíptica f . Podemos verla también como un automorfismo de V (visto como espacio vectorial). Sean X e Y las coordenadas de dos puntos y sean X' e Y' las coordenadas de sus imágenes. Puesto que f ha de conservar la polaridad de la cónica infinita, los puntos serán conjugados si y sólo si lo son sus imágenes, es decir, $XY^t = 0$ es equivalente a $X'Y'^t = 0$. Esto significa que f conserva la ortogonalidad en V , luego es una semejanza, luego $f = kg$, donde k es un número real no nulo y g es una isometría lineal de V . Ahora bien, es claro que f y g inducen la misma homografía en el plano proyectivo, luego podemos concluir que las isometrías elípticas están inducidas por las isometrías lineales de V . El recíproco es cierto: si f es una isometría lineal y A es su matriz en el sistema de referencia que estamos considerando, un punto está en la cónica infinita si sus coordenadas (imaginarias) cumplen $XX^t = 0$. Su imagen por f tiene coordenadas XA y, como A es una matriz ortogonal, $XAA^tX^t = XX^t = 0$, luego la imagen sigue en la cónica. En resumen: las isometrías elípticas son exactamente las homografías inducidas por las isometrías lineales de V o, equivalentemente, por las isometrías de V (visto como espacio afín) que fijan a O .

Definición 12.2 Llamaremos *modelo esférico* de la geometría elíptica al conjunto S formado por los pares de puntos antípodos de una esfera euclídea. Las rectas elípticas son los círculos máximos. Dos rectas son perpendiculares si lo

son sus tangentes en los puntos de corte. Una isometría elíptica es la restricción a S de una isometría del espacio euclídeo que fija al centro de S .

Además hemos visto que el polo de una recta en el modelo esférico es el par de polos (en el sentido geográfico) cuando consideramos a la recta como ecuador.

Las isometrías del modelo esférico son en particular homografías, luego conservan ángulos en el sentido de la sección A.3. Podemos, pues, definir el ángulo entre dos rectas elípticas en la esfera como el ángulo euclídeo entre sus tangentes por los puntos de corte, y es claro entonces que dos pares de rectas son congruentes si y sólo si forman el mismo ángulo. Veamos que el ángulo entre dos rectas puede definirse sin hacer referencia al modelo esférico exactamente como en la geometría hiperbólica, es decir,

$$\widehat{r_1 r_2} = \frac{1}{2} |\arg \mathcal{R}(r_1, r_2, I_1, I_2)|,$$

donde I_1 e I_2 son las tangentes a la cónica infinita por el punto $r_1 \cap r_2$ y el argumento se toma en $] -\pi, \pi[$.

En efecto, el ángulo definido así es sin duda invariante por isometrías. Basta probar que ambas definiciones coinciden sobre rectas que pasan por el polo norte $(0, 0, 1)$. Sus coordenadas homogéneas serán de la forma $(a, b, 0)$ y $(a', b', 0)$. Sus tangentes por el polo norte serán las intersecciones de los planos $ax + by = 0$ y $a'x + b'y = 0$ con el plano $z = 1$. Ahora bien, podemos identificar el plano $z = 1$ con el plano afín que resulta de eliminar la recta $z = 0$ en el plano proyectivo de partida. Entonces las tangentes cuyo ángulo hemos de medir son las propias rectas iniciales (sin su punto infinito). Con más detalle: en este plano afín tenemos ahora una estructura euclídea, de modo que las coordenadas cartesianas (x, y) correspondientes a las coordenadas homogéneas $(x, y, 1)$ están asociadas a un sistema de referencia ortonormal. Hasta aquí tenemos que el ángulo entre las rectas definido en el modelo esférico corresponde con el ángulo entre ellas mismas respecto a esta estructura euclídea. Este ángulo puede calcularse por la fórmula de Cayley, que es formalmente idéntica a la que queremos probar. Sólo hemos de asegurarnos de que las tangentes a la cónica infinita por el punto de corte $(0, 0, 1)$ coinciden con las rectas isotropas por dicho punto, pero esto es obvio, pues ambas son las rectas $x \pm iy = 0$. En efecto, ambas pasan por los puntos circulares $(1, \pm i, 0)$ y sus coordenadas homogéneas cumplen $x^2 + y^2 + z^2 = 0$.

Podemos plantearnos la fórmula análoga a la distancia hiperbólica entre dos puntos P y Q en términos de la razón doble $\mathcal{R}(P, Q, P_\infty, Q_\infty)$, donde P_∞ y Q_∞ son los puntos de corte entre la cónica infinita y la recta PQ . Ahora bien, si r_1 y r_2 son las polares de P y Q y se cortan en un punto R , la recta polar de R es PQ , luego las tangentes I_1 e I_2 a la cónica infinita que pasan por R cortan a PQ en P_∞ y Q_∞ , luego $\mathcal{R}(P, Q, P_\infty, Q_\infty) = \mathcal{R}(r_1, r_2, I_1, I_2)$. Por lo tanto, esta razón doble es un número complejo de módulo 1 cuyo argumento (dividido entre 2) es el ángulo entre dichas polares. Esto nos lleva a la definición siguiente, donde recogemos también la definición de ángulo que hemos discutido anteriormente.

Definición 12.3 Llamaremos *distancia elíptica* entre dos puntos P y Q como

$$d(P, Q) = \frac{1}{2} |\arg \mathcal{R}(P, Q, P_\infty, Q_\infty)|,$$

donde P_∞ y Q_∞ son los puntos en que la recta PQ corta a la cónica infinita y el argumento se toma en $] -\pi, \pi[$.

El *ángulo elíptico* entre dos rectas r_1 y r_2 es

$$\widehat{r_1 r_2} = \frac{1}{2} |\arg \mathcal{R}(r_1, r_2, I_1, I_2)|,$$

donde I_1 e I_2 son las tangentes a la cónica infinita por el punto $r_1 \cap r_2$ y el argumento se toma en $] -\pi, \pi[$.

Ambas magnitudes pueden variar en $[0, \pi/2[$, con el convenio de que son nulas si los puntos o las rectas son iguales. Ambas se conservan por isometrías. Hemos probado que la distancia entre dos puntos es el ángulo entre sus polares. En el modelo esférico la distancia entre dos puntos P y Q coincide con el ángulo entre las rectas OP y OQ , donde O es el centro de la esfera. El ángulo entre dos rectas es el definido en la sección A.3.

También es claro que dos pares de puntos/rectas son congruentes si y sólo si su distancia/ángulo es el mismo.

Exactamente el mismo argumento que en el caso de la geometría hiperbólica nos lleva a las fórmulas:

$$\cos d = \frac{xx' + yy' + zz'}{\sqrt{x^2 + y^2 + z^2} \sqrt{x'^2 + y'^2 + z'^2}},$$

$$\cos \theta = \frac{xx' + yy' + zz'}{\sqrt{x^2 + y^2 + z^2} \sqrt{x'^2 + y'^2 + z'^2}},$$

donde (x, y, z) y (x', y', z') son las coordenadas homogéneas de dos puntos/rectas. No obstante, estas fórmulas se siguen inmediatamente del modelo esférico usando las propiedades del producto escalar en V .

12.2 Biláteros y triángulos

La ordenación circular de la geometría elíptica confiere al plano propiedades muy distintas de las que se siguen de la ordenación lineal de las geometrías euclídea e hiperbólica. Para empezar hemos de tener presente que dos puntos no determinan un segmento, sino dos segmentos complementarios. De este modo, cuando hablemos de un segmento \overline{AB} habremos de entender que se trata de uno de los dos segmentos de extremos A y B . Recordemos que habíamos adoptado el convenio de llamar \overline{ACB} al segmento de extremos A y B que contiene a C .

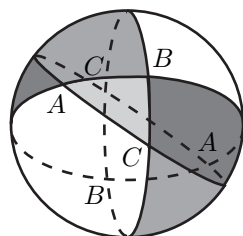
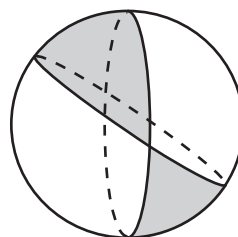
Es fácil ver que un segmento y su complementario no son congruentes salvo que los extremos disten $\pi/2$. Por ello no podemos definir la longitud de un segmento como la distancia entre sus extremos, ya que en tal caso estaríamos asignando la misma longitud a segmentos complementarios no congruentes. Para

definir adecuadamente la longitud de un segmento nos basaremos en el hecho siguiente, que se comprueba sin dificultad en el modelo esférico:

Cada segmento elíptico \overline{AB} contiene un único punto C con la propiedad de que $d(A, C) = d(C, B)$. Lo llamaremos punto medio del segmento.

Ahora definimos la *longitud* de un segmento como el doble de la distancia de sus extremos a su punto medio. La idea es que en el modelo esférico un segmento es un arco de circunferencia de amplitud menor que π (en realidad es un par de arcos antípodos de amplitud menor que π), por lo que la mitad de un segmento es un arco de amplitud menor que $\pi/2$ y así la amplitud de este arco sí coincide con la distancia entre sus extremos, luego la longitud del segmento completo así definida coincide con su amplitud como arco. Ahora sí podemos afirmar que dos segmentos son congruentes si y sólo si tienen la misma longitud. Es fácil probar que la longitud de un segmento y la de su complementario suman π . Podemos expresar esto diciendo que las rectas tienen longitud π .

Una recta no divide al plano en dos partes. Dos rectas no lo dividen en cuatro ángulos, sino en dos regiones que podemos llamar *biláteros*. Un bilátero tiene un vértice y dos lados. Dos puntos están en el mismo bilátero si existe un segmento que los une sin cortar a los lados. Una forma fácil de probar estos hechos es observar que los biláteros son simplemente los semiplanos que determina una recta en el espacio afín que resulta de tomar a la otra como recta infinita. Podemos asignar una medida al ángulo de cada bilátero del mismo modo que hemos hecho con los segmentos. La suma de los ángulos de dos biláteros complementarios es π .



Consideremos tres puntos no colineales A, B, C . Tomemos AB como recta infinita. Entonces los biláteros de vértice A y lados AB y AC se convierten en los dos semiplanos determinados por AC . La recta BC es secante a AC , luego divide al plano en cuatro ángulos. Dos puntos están en el mismo ángulo si y sólo si el segmento (afín) que los une no corta a BC ni a AC , o equivalentemente, si y sólo si existe un segmento que los une y no corta a ninguna de las rectas AB, AC, BC .

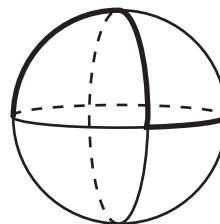
Esta caracterización es simétrica y no depende de la recta que hemos tomado como infinita ni del orden de los puntos, luego podemos afirmar que tres puntos no colineales A, B, C dividen al plano en cuatro partes que llamaremos *triángulos* de vértices A, B, C . Es claro por las definiciones que dos puntos están en el mismo triángulo de vértices A, B, C si y sólo si están en el mismo bilátero de vértice A y lados AB, AC , en el mismo bilátero de vértice B y lados BA, BC y en el mismo bilátero de vértice C y lados CA y CB . Llamaremos \hat{A}, \hat{B} y \hat{C} para un triángulo dado T a los biláteros de vértices A, B, C que contienen a T , y los llamaremos *ángulos* de T .

Sea T un triángulo de vértices A, B, C . Podemos tomar una recta que pase por A pero no esté contenida en \hat{A} . Si la convertimos en recta infinita las rectas AB y AC se vuelven paralelas y T se convierte en el conjunto de los puntos comprendidos entre ambas y que están en uno de los semiplanos determinados por BC . Vemos entonces que de los dos segmentos que B y C determinan en BC , uno está contenido en el ángulo \hat{A} y el otro es exterior a él.

En general, llamaremos *lado opuesto* al vértice A al segmento de BC contenido en \hat{A} . Lo mismo vale para los otros vértices. Dado un triángulo T , cada uno de los otros tres triángulos con sus mismos vértices comparte con T un ángulo y su lado opuesto. Los otros dos lados y ángulos son los complementarios de los de T .

Si, al igual que antes, representamos a un triángulo T como el conjunto de puntos afines comprendidos entre dos rectas paralelas y situados en un mismo semiplano respecto a una tercera recta r , veremos claramente que existe una recta que no corta a ninguno de sus lados (ni pasa por sus vértices). Basta tomar una recta paralela a r en el semiplano opuesto a T . Tenemos, pues, que siempre existe una recta que no corta a ninguno de los lados de un triángulo. Si tomamos esta recta como infinita T se convierte en un triángulo afín ordinario. De aquí podemos traducir al plano elíptico algunos teoremas sobre la ordenación lineal. Por ejemplo, una recta no puede cortar a los tres lados de un triángulo. Sin embargo, el hecho de que exista una recta que no corta a los vértices y lados de un triángulo implica que corta a los complementarios de los lados, luego estos segmentos no son los lados de ningún triángulo.

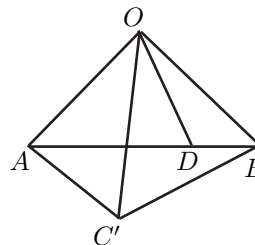
Esto muestra que tres segmentos pueden conectar tres puntos y pese a ello no ser los lados de ningún triángulo. Con más detalle: digamos que un triángulo tiene lados a, b, c y llamemos $-a, -b, -c$ a los complementarios. Entonces los otros triángulos tienen lados $(a, -b, -c)$, $(-a, b, -c)$ y $(-a, -b, c)$. Las otras cuatro combinaciones son imposibles. Vemos entonces que si en una combinación posible de segmentos cambiamos uno por su complementario obtenemos una combinación imposible y viceversa.



Teorema 12.4 *Un lado de un triángulo es menor que la suma de los otros dos.*

DEMOSTRACIÓN: Consideremos un triángulo elíptico \widehat{ABC} . Supongamos que \widehat{AB} es el lado mayor. Si el triángulo tiene dos lados iguales mayores o iguales que el tercero la desigualdad es obvia. Bastará probar que $\widehat{AB} < \widehat{AC} + \widehat{CB}$. Trabajamos en el modelo esférico. Sea O el centro de la esfera. En cuanto sigue \widehat{AB} representará el segmento euclídeo de extremos A y B y en lugar del segmento elíptico correspondiente.

Sea D el punto sobre \widehat{AB} que hace $\widehat{AOD} = \widehat{AOC}$. Sea C' el punto sobre la semirrecta \overrightarrow{OC} que cumple $\overline{OC'} = \overline{OD}$. Entonces $\widehat{AOC'} = \widehat{AOD}$, pues tienen dos lados iguales y el ángulo entre ellos, luego $\widehat{AC'} = \widehat{AD}$.



Ahora, $\overline{BC'} > \overline{AB} - \overline{AC'} = \overline{AB} - \overline{AD} = \overline{DB}$. Los triángulos $\widehat{OBC'}$ y \widehat{OBD} tienen dos lados iguales y un lado desigual $\overline{BC'} > \overline{DB}$, luego $\widehat{BOC'} > \widehat{BOD} = \widehat{AOB} - \widehat{AOD}$, es decir, $\widehat{BOC} > \widehat{AOB} - \widehat{AOC}$, que en términos del triángulo elíptico original es $\overline{AB} < \overline{AC} + \overline{CB}$, como había que probar. ■

Ejercicio: Probar que un lado de un triángulo elíptico es siempre mayor que la diferencia de los otros dos.

Teorema 12.5 *La suma de los lados de un triángulo es menor que 2π .*

DEMOSTRACIÓN: Si el triángulo tiene lados a, b, c , consideramos uno de los triángulos adyacentes, digamos el de lados $\pi - a, \pi - b, c$. Por el teorema anterior $c < \pi - a + \pi - b$, luego $a + b + c < 2\pi$. ■

Al aplicar la polaridad de la cónica infinita a las rectas que pasan por el vértice de un bilátero y están contenidas en él obtenemos un segmento. Es fácil ver que si el ángulo del bilátero mide a lo sumo $\pi/2$ lo mismo le ocurre al segmento, y en tal caso el ángulo del bilátero coincide con el que forman sus lados y la longitud del segmento coincide con la distancia entre sus extremos, y además sabemos que ambas coinciden. Es decir, la polaridad transforma biláteros rectos y agudos en segmentos de la misma longitud. Pasando por los complementarios vemos que también es cierto para biláteros obtusos.

En particular la polaridad transforma los tres ángulos de un triángulo T en tres segmentos que conectan los polos de sus lados. La longitud de cada segmento es igual a la amplitud del ángulo del que proviene. Sin embargo estos segmentos no son los lados de ningún triángulo. Para probarlo observamos que un punto P del triángulo está contenido en una recta contenida en cada ángulo, luego su recta polar pasa por un punto de cada segmento. Por consiguiente, los complementarios de estos segmentos sí son los lados de un triángulo (al tomar complementarios tres veces pasamos de una combinación imposible a otra posible).

Definición 12.6 Llamaremos *triángulo polar* de un triángulo elíptico T al triángulo que tiene por vértices los polos de las prolongaciones de sus lados y por lados los complementarios de los segmentos formados por los polos de las rectas contenidas en sus ángulos.

Hemos probado que si T' es el triángulo polar de T y los ángulos de T miden α, β y γ , entonces los lados de T' miden $\pi - \alpha, \pi - \beta, \pi - \gamma$.

Veamos que los ángulos de T' son los complementarios de los biláteros formados por las polares de los puntos de los lados de T . Digamos que los vértices de T son A, B, C y sus lados opuestos son a, b, c . Tomemos un punto Q de la recta BC . Entonces Q estará en a si y sólo si la recta AQ está en \hat{A} , si y sólo si $p(AQ)$ no está en el lado de T' de extremos $p(AB), p(AC)$, si y sólo si $p(Q)$ no está en el bilátero de vértice $p(BC)$, si y sólo si $p(Q)$ está en el complementario de dicho bilátero.

Con esto es inmediato el teorema siguiente:

Teorema 12.7 *Si T es un triángulo elíptico y T' es su triángulo polar, entonces T es el triángulo polar de T' . Si los ángulos de T miden α, β, γ y sus lados miden a, b, c , entonces los lados de T' miden $\pi - \alpha, \pi - \beta, \pi - \gamma$ y sus ángulos miden $\pi - a, \pi - b, \pi - c$.*

Este teorema nos permite extender a la geometría elíptica el principio de dualidad de la geometría proyectiva. Veamos una primera aplicación:

Teorema 12.8 *La suma de los ángulos de un triángulo es mayor que π y menor que 3π .*

DEMOSTRACIÓN: Si los ángulos del triángulo son α, β, γ , entonces el teorema anterior aplicado al triángulo polar nos da $0 < \pi - \alpha + \pi - \beta + \pi - \gamma < 2\pi$, luego $\pi < \alpha + \beta + \gamma < 3\pi$. ■

Ejercicio: Probar que un triángulo coincide con su polar si y sólo si todos sus lados y ángulos miden $\pi/2$.

12.3 Isometrías elípticas

Las isometrías elípticas son mucho más simples que las euclídeas o las hiperbólicas. Consideremos el modelo esférico. Una isometría elíptica está inducida por una isometría euclídea que fija a la esfera, y sabemos que en un sistema de referencia adecuado su matriz ha de ser de la forma

$$\begin{pmatrix} \cos \theta & \operatorname{sen} \theta & 0 \\ -\operatorname{sen} \theta & \cos \theta & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}.$$

Si multiplicamos todos los coeficientes por -1 obtenemos otra isometría euclídea que induce la misma isometría elíptica (pues estamos componiéndola con la isometría que intercambia los puntos antípodas). Por consiguiente podemos suponer que la última entrada es un $+1$. En definitiva, cada isometría elíptica está inducida por un único giro en \mathbb{R}^3 .

El eje del giro corta al plano elíptico en un único punto O . Es claro que O es un punto fijo, al igual que lo es su recta polar r . Si el ángulo θ es distinto de π entonces O es el único punto fijo y r la única recta fija. Si $\theta = \pi$ entonces la isometría fija también a todos los puntos de r y a todas las rectas que pasan por O . En definitiva:

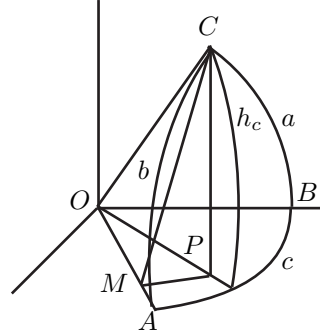
Teorema 12.9 *Si f es una isometría del plano elíptico distinta de la identidad, existen un punto O y una recta r unívocamente determinados, a los que llamaremos centro y eje de f de modo que los únicos puntos fijados por f son O y quizá los puntos de r , y las únicas rectas fijadas por f son r y quizá las rectas que pasan por O . Además r es la recta polar de O .*

Por razones obvias las isometrías elípticas se llaman también *giros*.

12.4 Trigonometría elíptica

La trigonometría elíptica es completamente análoga a la hiperbólica. Consideremos un triángulo de vértices A, B, C , lados a, b, c y ángulos α, β, γ , relacionados como es habitual. Comenzaremos demostrando el teorema del coseno.

Supongamos primeramente que α y b son agudos. Sea P la proyección ortogonal de C sobre el plano AOB y sea M su proyección ortogonal sobre OA . El plano CPM es perpendicular a OA , pues contiene las rectas CP y CM , ambas perpendiculares a OA . Por consiguiente, las rectas MP y MC son paralelas a las tangentes a la esfera por M contenidas en los planos OAB y OAC respectivamente, es decir, $\widehat{CMP} = \alpha$. Claramente se tiene además que $\overline{OM} = \cos b$, $\overline{MC} = \sin b$, $\overline{MP} = \sin b \cos \alpha$, $\overline{PC} = \sin b \sin \alpha$.



Consideremos la suma de vectores $\overrightarrow{OM} + \overrightarrow{MP} + \overrightarrow{PC} = \overrightarrow{OC}$ y proyectémosla ortogonalmente sobre la recta OB . Si llamamos p a la proyección es fácil comprobar que

$$\begin{aligned} p(\overrightarrow{OM}) &= \cos b \cos c, \\ p(\overrightarrow{MP}) &= \sin b \cos \alpha \cos(\pi/2 - c) = \sin b \sin c \cos \alpha, \\ p(\overrightarrow{PC}) &= 0, \\ p(\overrightarrow{OC}) &= \cos a. \end{aligned}$$

Por lo tanto

$$\cos a = \cos b \cos c + \sin b \sin c \cos \alpha. \tag{12.1}$$

Si $\alpha > \pi/2$ entonces cambia el sentido de \overrightarrow{MP} , pero también cambia el signo de $\cos \alpha$, luego la fórmula sigue siendo válida. Si $b > \pi/2$ entonces aplicamos la fórmula al triángulo adyacente de lados $a, \pi - b, \pi - c$ y el mismo ángulo α y obtenemos la misma fórmula para el triángulo dado. Si $\alpha = \pi/2$ entonces $M = P$ y si $b = \pi/2$ entonces $O = M$. En ambos casos la fórmula se deduce sin dificultad.

Volviendo al caso en que α y b son agudos, consideremos ahora la proyección sobre el eje perpendicular al plano AOB . Entonces

$$\begin{aligned} p(\overrightarrow{OM}) &= 0, \\ p(\overrightarrow{MP}) &= 0, \\ p(\overrightarrow{PC}) &= \sin b \sin \alpha, \\ p(\overrightarrow{OC}) &= \sin h_c, \end{aligned}$$

donde h_c es la altura del triángulo que une C con c . Tenemos, pues,

$$\operatorname{sen} h_c = \operatorname{sen} b \operatorname{sen} \alpha,$$

y como antes se comprueba que esta relación es válida en el caso general. Aplicándola a β obtenemos $\operatorname{sen} h_c = \operatorname{sen} a \operatorname{sen} \beta$. Despejando $\operatorname{sen} h_c$ e igualando obtenemos el teorema de los senos. Así pues,

$$\frac{\operatorname{sen} a}{\operatorname{sen} \alpha} = \frac{\operatorname{sen} b}{\operatorname{sen} \beta} = \frac{\operatorname{sen} c}{\operatorname{sen} \gamma}. \quad (12.2)$$

Las fórmulas (12.1) y (12.2) se conocen como *fórmulas de Bessel*.

La fórmula (12.1) muestra que si dos triángulos tienen los lados iguales entonces también tienen los ángulos iguales, y de aquí se sigue fácilmente que son congruentes. Por dualidad, si dos triángulos tienen los ángulos iguales también son congruentes. De hecho, aplicando la fórmula (12.1) al triángulo polar obtenemos el teorema del coseno para un ángulo:

$$\cos \alpha = -\cos \beta \cos \gamma + \operatorname{sen} \beta \operatorname{sen} \gamma \cos a. \quad (12.3)$$

Dejamos a cargo del lector la comprobación de que todas las fórmulas de la trigonometría hiperbólica son válidas para triángulos elípticos cuando las razones hiperbólicas se sustituyen por las correspondientes circulares. Todas las comprobaciones son fáciles a partir de las fórmulas que ya tenemos. Por ejemplo, el teorema del coseno cuando $\alpha = \pi/2$ se convierte en el teorema de Pitágoras elíptico:

$$\cos a = \cos b \cos c.$$

Cada relación tiene su dual, resultante de aplicarla al triángulo polar. Por ejemplo, todo triángulo rectilátero (con un lado recto, digamos a) cumple

$$\cos \alpha = -\cos \beta \cos \gamma.$$

Como en el caso de la geometría hiperbólica, todas las fórmulas de la trigonometría elíptica se convierten en las equivalentes euclídeas si sustituimos los senos y cosenos de las longitudes de segmentos por sus aproximaciones de orden 2:

$$\operatorname{sen} x \approx x, \quad \cos x \approx 1 - \frac{x^2}{2}.$$

Así pues, la geometría elíptica es también localmente euclídea.

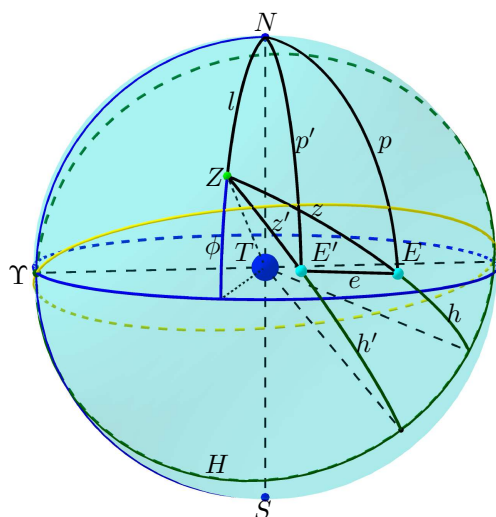
También es claro que es posible definir geoméricamente una unidad de longitud, sin recurrir a ninguna característica física. Por ejemplo, con la distancia que hemos definido es fácil ver que los lados de un triángulo con tres ángulos rectos miden $\pi/2$, por lo que podemos definir la unidad “natural” de longitud elíptica como $2/\pi$ veces la longitud del lado de un triángulo con tres ángulos rectos.

El problema de las dos altitudes de Gauss Como ejemplo de aplicación de la trigonometría elíptica vamos a resolver el problema de las dos altitudes, que hemos mencionado en la introducción:

Determinar la posición y la hora a partir de las altitudes de dos estrellas cuya posición es conocida.

Vamos a explicar el enunciado y el método de resolución considerando un ejemplo concreto:

El 15 de enero de 2021, en un punto del océano Atlántico, a la 1:15 hora de Greenwich, las altitudes de las estrellas Procyon y Biegelgeuse en la esfera celeste resultaron ser de 40° y 60° , respectivamente. Determinar las coordenadas y la hora local del punto de observación.



La figura representa la esfera celeste con la Tierra en su centro. El eje de rotación de la Tierra corta a la esfera en los polos Norte y Sur celestes. El plano perpendicular a la esfera la corta en el ecuador celeste. El punto Z es el zénit del punto de observación, es decir, el punto de la esfera celeste que está sobre su vertical. El plano perpendicular al eje TZ corta a la esfera celeste en el horizonte astronómico H y el meridiano celeste que pasa por el zénit es la línea horaria del punto de observación, porque el mediodía solar en un punto es el momento en el que el Sol pasa por su línea horaria.

A lo largo del año, el Sol recorre una circunferencia sobre la esfera celeste, que se conoce como la eclíptica, que no coincide con el ecuador celeste, sino que forma con él un ángulo de 23.5° . Las constelaciones situadas sobre la eclíptica son las doce constelaciones del zodiaco. La eclíptica corta al ecuador celeste en dos puntos diametralmente opuestos, conocidos como el punto de Aries y el punto de Libra, porque cuando Hiparco los definió estaban en dichas constelaciones, pero, en virtud de la precesión de los equinoccios, estos puntos

avanzan aproximadamente un grado cada 72 años, y actualmente el punto de Aries (que en la figura está marcado con el signo Υ de Aries) se encuentra en la constelación de Piscis, cerca del borde con Acuario.

Los puntos E y E' representan las posiciones de las estrellas Procyon y Betelgeuse, respectivamente. (Son la octava y la novena estrellas más brillantes del cielo.) El problema supone que un observador ha medido sus *altitudes*, es decir, los ángulos que forman con el horizonte astronómico, y han resultado ser $h = 40^\circ$ y $h' = 60^\circ$.

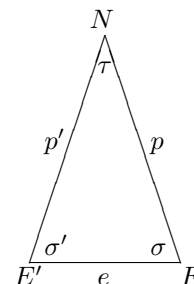
Hay que suponer también que conocemos las posiciones de ambas estrellas en la esfera celeste. La forma usual en que las expresan los astrónomos es ésta:

	Ascensión recta	Declinación
Procyon	7h 39m 12.8s	$+5^\circ 13' 30.0''$
Begelgeuse	5h 55m 10.307s	$+7^\circ 24' 25.35''$

La *declinación* δ es simplemente la latitud en la esfera celeste, mientras que la *ascensión recta* α es el nombre con el que se conoce tradicionalmente a la longitud medida desde el punto de Aries, y habitualmente se mide en horas, minutos y segundos, de modo que $0 \leq \alpha < 24\text{h}$, por lo que 24h equivalen a 360° o, lo que es lo mismo, $1\text{h} = 15^\circ$. Si traducimos las posiciones de las estrellas a grados sexagesimales son, para Procyon $(\alpha, \delta) = (114.826, 5.225)$ y para Betelgeuse $(\alpha', \delta') = (88.7929, 7.40704)$.

Consideremos ahora el triángulo esférico de vértices N , E , E' . Conocemos dos de sus lados, $p = 90^\circ - \delta$, $p' = 90^\circ - \delta'$, y también el ángulo $\tau = \alpha' - \alpha = 26.0326$, como es fácil deducir teniendo en cuenta que el plano tangente a la esfera celeste por N es paralelo al ecuador celeste. Aplicamos el teorema del coseno:

$$\cos e = \cos p \cos p' + \sin p \sin p' \cos \tau,$$



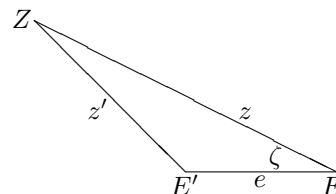
del que deducimos que $e = 25.962^\circ$. A su vez, el teorema de los senos:

$$\frac{\sin p'}{\sin \sigma} = \frac{\sin e}{\sin \tau} = \frac{\sin p}{\sin \sigma'}$$

nos permite calcular $\sigma = 83.8056^\circ$, $\sigma' = 86.7197^\circ$.

Ahora consideramos el triángulo esférico de vértices Z , E , E' , del que conocemos los lados $z = 90^\circ - h = 50^\circ$ y $z' = 90^\circ - h' = 30^\circ$, además del lado e que hemos calculado antes. Aplicamos el teorema del coseno:

$$\cos z' = \cos z \cos e + \sin z \sin e \cos \zeta$$



del que deducimos que $\zeta = 30.7839^\circ$.

Por último, consideramos el triángulo de vértices N , Z , E , del cual conocemos los lados p y z , así como el ángulo $\eta = \sigma - \zeta = 53.0217^\circ$. El teorema del coseno

$$\cos l = \cos p \cos z + \sin p \sin z \cos \eta$$

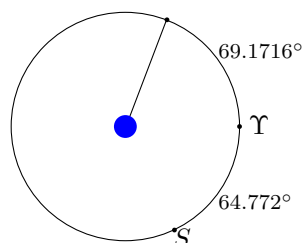
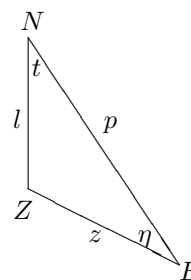
nos da que $l = 58.8415^\circ$, de donde concluimos que la latitud del punto de observación es $\phi = 90^\circ - l = 31.1585^\circ$ (ϕ es el ángulo que forma el zénit con el ecuador celeste, que es el mismo que forma el punto de observación con el ecuador terrestre, es decir, su latitud). Por otra parte,

$$\cos z = \cos l \cos p + \sin l \sin p \cos t,$$

de donde obtenemos $t = 45.6539^\circ$. Ahora bien, usando de nuevo que el plano tangente por N es paralelo al ecuador celeste, vemos que $t = \alpha - \alpha_0$, donde α_0 es la ascensión recta del punto de observación, que resulta ser $\alpha_0 = 69.1716^\circ$. Esto es la longitud del punto de observación medida respecto del punto de Aries. Para determinar la longitud geográfica necesitamos conocer la ascensión recta del Sol. Esto es más delicado, pues, mientras las estrellas se pueden considerar fijas en la esfera celeste, con lo que tienen coordenadas constantes, el Sol se mueve sobre ella a lo largo del año, así que su ascensión recta depende del tiempo.

En realidad, lo que queremos calcular es la hora solar local, pues si sabemos la hora, comparando con la hora en Greenwich (que nos la da cualquier reloj) podemos calcular la longitud. Entender plenamente la situación requiere una digresión, de modo que para, no interrumpir la resolución del problema, vamos a aceptar de momento que, en la fecha dada en el enunciado, la ascensión recta del Sol era $\alpha_S = 295.228^\circ$

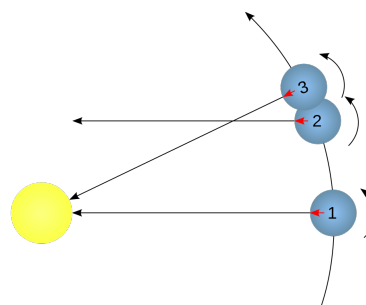
La figura muestra la esfera celeste vista desde arriba (desde el norte). La línea horaria del observador forma un ángulo de 69.1716° con el punto de Aries, y el meridiano en el que se encuentra el Sol forma un ángulo de $360^\circ - \alpha_S = 64.772^\circ$. La Tierra gira en sentido horario (aunque en este contexto es mejor considerar que la esfera celeste gira en sentido antihorario) y da una vuelta completa cada día. En el punto de observación será mediodía cuando el Sol esté en su línea horaria. Para que esto suceda, la esfera tiene que girar $69.1716 + 64.772 = 133.9436^\circ$ (donde suponemos que Υ y S están fijos, pues se mueven muy lentamente sobre la esfera en comparación con su movimiento de rotación). Así pues, para que sea mediodía en el punto de observación, la esfera celeste tiene que girar $133.9436^\circ = 8\text{h } 55\text{m } 46\text{s}$, luego éste es el tiempo que falta para que sea mediodía. Esto significa que en el momento de la observación eran las 3:04, hora local. Puesto que la hora de Greenwich era 1:15, la diferencia de 109 minutos se traduce (a razón de 15° por hora) en que la latitud del punto de observación es de -27.25° , con lo que las coordenadas del punto son $31^\circ 9' N$ $27^\circ 15' O$.





El día solar medio Aunque no aporta nada al uso de la trigonometría esférica que hemos hecho en el problema anterior, por completitud vamos a aclarar algunas sutilezas sobre la hora solar que hemos calculado. El tiempo que tarda la Tierra en dar una vuelta completa alrededor de sí misma es el *día sidéreo*, que se llama así porque coincide con el tiempo que tarda cualquier estrella en dar una vuelta completa alrededor de los polos celestes. Su duración es de 23h 56m 3.998s. En cambio, el *día solar* es el tiempo comprendido entre dos tránsitos del Sol por un mismo meridiano horario, y su duración es de alrededor de 24h, pero con varios aumentos y disminuciones a lo largo del año, de modo que las diferencias de duración pueden llegar a ser casi de un minuto.

La diferencia del día solar con respecto al día sidéreo se debe a que en el tiempo en que la Tierra da una vuelta completa alrededor de sí misma, avanza un poco en su órbita alrededor del Sol (de la posición 1 a la posición 2), por lo que los puntos del meridiano que veían en 1 al Sol en su meridiano horario de la esfera celeste, un día sidéreo después lo ven en 2 un poco hacia el este, de modo que la Tierra tiene que girar un poco más hasta (hasta 3) para que el Sol vuelva a pasar por el mismo meridiano.



Ese “un poco más” depende de lo que la Tierra haya avanzado en el día sidéreo, y sucede que ese avance es muy irregular, en parte porque la velocidad de la Tierra alrededor del Sol es mayor cuanto más cerca se encuentra del perihelio, y en parte debido a la inclinación de su eje de rotación respecto del plano de la eclíptica. El día de 24 horas exactas que marcan los relojes es lo que se llama *día solar medio*, y no depende de la posición real del Sol en el cielo, sino en la posición de un Sol ideal llamado Sol medio.

Para definir el Sol medio definimos primero el *Sol dinámico*, que es un punto ideal que se mueve sobre la eclíptica dando una vuelta completa en el mismo tiempo que el Sol real, pero a velocidad angular constante, es decir, aproximadamente a $365.25 / (360 \cdot 24 \cdot 60)$ grados por minuto, y coincide con el Sol real en dos ocasiones, cuando pasa por el perihelio y por el afelio. Así, en 2020, el Sol real (y el dinámico) pasaron por el perihelio el 5 de enero a las 7:48 (hora de Greenwich). El Sol real siguió a su ritmo por la eclíptica y llegó al punto

de Aries el 20 de marzo a las 3:50, lo que se corresponde con el equinoccio de primavera en el hemisferio norte (o de otoño en el hemisferio sur), pero el Sol dinámico llegó al punto de Aries el 22 de marzo a las 12:27 (el punto de Aries está a 77.07° del perihelio). Aunque la velocidad angular del Sol dinámico sea constante, la de su ascensión recta no lo es, debido a la inclinación de la eclíptica.

Ejercicio: Calcular la relación entre el ángulo que forma un punto de la eclíptica respecto del punto de Aries y su ascensión recta.

Por ello, es necesario introducir otro punto ideal, el *Sol medio*, que se mueve por la eclíptica de modo que su ascensión recta da una vuelta completa alrededor del ecuador celeste en el mismo tiempo que el sol Dinámico (o que el Sol real), pero a velocidad constante, de modo que el Sol medio y el dinámico coinciden en el punto de Aries. Así, teniendo en cuenta que el Sol medio tenía ascensión recta nula el 22 de marzo a las 12:27, podemos calcular que su ascensión recta en la fecha del enunciado era la que hemos usado en la resolución del problema, $\alpha_S = 295.228^\circ$. El *día solar medio* es el tiempo que transcurre entre dos tránsitos sucesivos del Sol medio por el mismo meridiano horario, cosa que, según el día, sucederá antes o después del tránsito del Sol real.

El hecho de que la ascensión recta del Sol medio se mueva a velocidad angular constante sobre el ecuador celeste se traduce en que el día solar medio mantiene la misma duración a lo largo del año, y la unidad de tiempo, el “segundo”, está definido de modo que el día solar medio sea exactamente de 24 horas (pero no se define en términos del día solar medio, sino en términos de la física atómica, de modo que tiene sentido medir variaciones del día solar medio a lo largo del tiempo).

Más aún, la *hora solar media* de un lugar se define como 12 horas más el tiempo (positivo o negativo) que ha pasado desde el tránsito del Sol medio por el meridiano local. De este modo, en un lugar son las 0:00 si faltan exactamente 12 horas para el mediodía solar medio, y son las 12:00 en el momento del mediodía solar medio, cuando el Sol medio pasa por el meridiano horario. Si medimos la longitud en horas, entonces la diferencia de longitud entre dos puntos de la Tierra coincide con la diferencia de la hora solar media entre ambos puntos, cosa que hemos usado en el cálculo de la longitud en el problema de las dos altitudes. En la práctica, la hora solar media no es la hora que marcan los relojes debido a que la hora se mantiene constante en cada uno de los 24 husos horarios en que se divide la esfera terrestre (y a causa de los horarios estacionales que adoptan algunos países). Así por ejemplo, la hora en la España peninsular es una o dos horas más (según la estación) de la hora solar media del meridiano de Greenwich. ■

12.5 La geometría diferencial elíptica

Veamos ahora que podemos considerar al plano elíptico como una variedad diferenciable. Intuitivamente, la situación es mucho más simple que la análoga correspondiente al plano hiperbólico, pero formalmente nos encontramos con

un problema técnico no trivial. En efecto, desde un punto de vista intuitivo, el modelo esférico nos muestra que la geometría del plano elíptico E menos una recta r es exactamente la de una semiesfera de radio 1, pues la recta r se corresponde con un círculo máximo de la esfera, y cada punto de $E \setminus r$ se corresponde con un único punto de una de las semiesferas determinadas por r , de modo que la longitud de un segmento de recta es su longitud euclídea como arco en la esfera (es decir, el ángulo que forman los radios de la esfera correspondientes a los extremos del segmento), por lo que “lo razonable” es considerar que la longitud de una curva arbitraria y el área de una región arbitraria en $E \setminus r$ son las calculadas a partir de la geometría diferencial de la esfera.

En otros términos, el plano elíptico es localmente isométrico a una esfera igual que el plano hiperbólico es localmente isométrico a una pseudoesfera. En particular, podemos afirmar que la curvatura del plano elíptico es 1 en todos los puntos (donde el valor unitario depende de la elección canónica que hemos hecho de la unidad de longitud), y que las geodésicas del plano elíptico son las rectas elípticas, pues unas y otras coinciden con los círculos máximos de la esfera.

Ahora bien, existe un problema técnico que nos impide considerar como rigurosas las afirmaciones precedentes, y es que no estamos en condiciones de afirmar que el plano elíptico sea una variedad diferenciable en el sentido que conocemos, es decir, en el de la definición [An 7.1]. En efecto, dicha definición requeriría considerar a E como un subconjunto de \mathbb{R}^m , para cierto m y lo máximo que tenemos es que los puntos de E pueden identificarse con pares de puntos antípodos en una esfera de radio 1, pero esto no nos permite, con rigor, considerar que $E \subset \mathbb{R}^3$.

La topología proyectiva Antes de considerar el problema más a fondo vamos a precisar lo que acabamos de decir. De momento podemos razonar en dimensión arbitraria sin ningún esfuerzo adicional. Llamamos

$$S^n = \{x \in \mathbb{R}^{n+1} \mid x_1^2 + \dots + x_n^2 = 1\}$$

y consideramos la aplicación $\pi : S^n \rightarrow \mathbb{P}^n(\mathbb{R})$ dada por $\pi(x) = \langle x \rangle$, es decir, la aplicación que a cada $x \in S^n$ le asigna el punto de $\mathbb{P}^n(\mathbb{R})$ que tiene a x por vector de coordenadas homogéneas. Claramente π es suprayectiva y cada punto de $\mathbb{P}^n(\mathbb{R})$ tiene dos antiimágenes de la forma $\pm x$. Conviene considerar también el homeomorfismo $a : S^n \rightarrow S^n$ dado por $a(x) = -x$, que a cada punto le asigna su antípoda. Obviamente $a \circ \pi = \pi$.

Podemos usar π para definir una topología en $\mathbb{P}^n(\mathbb{R})$: diremos que un conjunto $V \subset \mathbb{P}^n(\mathbb{R})$ es abierto si $\pi^{-1}[V]$ es abierto en S^n . Es inmediato comprobar que los abiertos así definidos forman realmente una topología en el espacio proyectivo. En lo sucesivo consideraremos siempre como espacios topológicos a los espacios proyectivos reales con esta topología.¹

¹Más en general, todo esto vale igualmente para espacios proyectivos complejos sin más que cambiar \mathbb{R} por \mathbb{C} en todo momento.

Observemos que los espacios proyectivos son espacios de Hausdorff. En efecto, dados dos puntos distintos $P, Q \in \mathbb{P}^n(\mathbb{R})$, cada uno de ellos tendrá dos antiimágenes distintas $\pm x, \pm y$, respectivamente. Claramente, existe un $\epsilon > 0$ tal que las bolas $B_\epsilon(\pm x)$ y $B_\epsilon(\pm y)$ son disjuntas dos a dos, y entonces $U = \pi[B_\epsilon(x)]$ y $V = \pi[B_\epsilon(y)]$ son dos entornos disjuntos de P y Q en $\mathbb{P}^n(\mathbb{R})$. En efecto, son abiertos porque $\pi^{-1}[U] = B_\epsilon(x) \cup B_\epsilon(-x)$ es abierto en S^n , e igualmente se razona con V . También es fácil ver que son disjuntos.

Una consecuencia trivial de la definición de la topología proyectiva es que la aplicación $\pi : S^n \rightarrow \mathbb{P}^n(\mathbb{R})$ es continua, y como la esfera S^n es compacta, concluimos que los espacios proyectivos son compactos.

Más aún, conviene observar que π es abierta, pues si U es cualquier abierto en S^n , entonces $\pi[U]$ es abierto en $\mathbb{P}^n(\mathbb{R})$ porque $\pi^{-1}[\pi[U]] = U \cup a[U]$ es abierto en S^n .

Si P es cualquier punto de $\mathbb{P}^n(\mathbb{R})$ y $x \in S^n$ es una de sus antiimágenes por π , podemos considerar

$$U_x = \{y \in S^n \mid x \cdot y > 0\},$$

que no es sino la semiesfera abierta que tiene a x “en la cima”. Si llamamos $H = \pi[\{y \in S^n \mid x \cdot y = 0\}]$ al hiperplano polar de x , es inmediato que π se restringe a un homeomorfismo $\pi|_{U_x} : U_x \rightarrow \mathbb{P}^n(\mathbb{R}) \setminus H$. Así pues:

El abierto que resulta que eliminar un hiperplano del espacio proyectivo de dimensión n es homeomorfo a una semiesfera de dimensión n .

A partir de aquí nos centramos en el caso bidimensional.

Inmersiones de $\mathbb{P}^2(\mathbb{R})$ Acabamos de probar que, cuando al plano elíptico $\mathbb{P}^2(\mathbb{R})$ le quitamos una recta proyectiva, el abierto resultante es homeomorfo a una semiesfera abierta de radio 1. Observemos que el radio es irrelevante, porque todas las semiesferas son homeomorfas entre sí, e igualmente podríamos haber dicho un disco abierto en lugar de una semiesfera. No obstante, es fácil probar que el homeomorfismo que hemos encontrado es una isometría de espacios métricos (y ahora sí que es relevante que estemos considerando una esfera de radio 1). Ésta es una de las propiedades que señalábamos al principio de esta sección: que la distancia entre dos puntos del plano elíptico coincide con su distancia euclídea cuando los identificamos con puntos de una semiesfera.

El problema que tenemos planteado es que querríamos probar que es una isometría de variedades diferenciales, pues eso nos permitiría hablar de la longitud elíptica de una curva arbitraria, o del área elíptica de una región arbitraria (razonable). Y aquí retomamos el problema de que no podemos aplicar la definición [An 7.1] a $\mathbb{P}^2(\mathbb{R})$ porque no es un subconjunto de \mathbb{R}^m , para ningún m .

De hecho, se puede demostrar que $\mathbb{P}^2(\mathbb{R})$ no es homeomorfo a ningún subespacio de \mathbb{R}^3 (ni, por consiguiente, de \mathbb{R}^2).

En principio, podemos decir que $P^2(\mathbb{R})$ es una *variedad topológica* de dimensión 2 (o una *superficie topológica*), lo cual significa que es un espacio topológico tal que cada uno de sus puntos tiene un entorno homeomorfo a un abierto en \mathbb{R}^2 (porque todo punto tiene un entorno homeomorfo a una semiesfera abierta, la cual es a su vez homeomorfa a un disco abierto).

La solución “moderna” a este problema consiste en introducir un concepto de variedad diferenciable más abstracto que el presentado en [An 7.1], de modo que cualquier variedad topológica X puede dotarse de estructura de variedad diferenciable (sin necesidad de ser un subespacio de ningún \mathbb{R}^m) tomando como cartas ciertos homeomorfismos entre abiertos de X y abiertos de \mathbb{R}^n que cumplan cierta condición de diferenciabilidad.

Así, el lector familiarizado con la geometría diferencial abstracta no tendrá dificultades en comprobar que las restricciones de π a distintas semiesferas definen un sistema de cartas que dotan a $P^2(\mathbb{R})$ de estructura de superficie diferenciable. Sin embargo, para mantenernos en el contexto del capítulo VII de [An] vamos a probar que $P^2(\mathbb{R})$ es homeomorfo a un subespacio de \mathbb{R}^4 .

Para ello consideramos la aplicación $F : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ dada por

$$F(x, y, z) = (xy, xz, yz, x^2 - y^2).$$

Claramente es continua y cumple que $F(p) = F(-p)$, para todo $p \in \mathbb{R}^3$. Llamamos $E = F[S^2]$ y $f : S^2 \rightarrow E$ a la restricción de F , que es una aplicación continua y suprayectiva. Vamos a probar que $f(p) = f(q)$ si y sólo si $q = \pm p$.

En efecto, una implicación es obvia, y si $p = (x_1, y_1, z_1)$ y $q = (x_2, y_2, z_2)$ cumplen $f(p) = f(q)$, tenemos que

$$(x_1 + y_1i)^2 = x_1^2 - y_1^2 + 2x_1y_1i = x_2^2 - y_2^2 + 2x_2y_2i = (x_2 + y_2i)^2,$$

luego $x_1 + y_1i = \pm(x_2 + y_2i)$, es decir, $(x_1, y_1) = \pm(x_2, y_2)$. Si $(x_1, y_1) = (0, 0)$, entonces $p = \pm(0, 0, 1)$ y $q = \pm(0, 0, 1)$, luego $p = \pm q$. En caso contrario, las relaciones $x_1z_1 = x_2z_2$, $y_1z_1 = y_2z_2$ nos permiten concluir igualmente que $p = \pm q$.

Esto implica que f induce una biyección $\bar{f} : P^2(\mathbb{R}) \rightarrow E$, caracterizada por la conmutatividad del diagrama

$$\begin{array}{ccc} P^2(\mathbb{R}) & \xrightarrow{\bar{f}} & E \\ \uparrow \pi & \nearrow f & \\ S^2 & & \end{array}$$

Observemos que \bar{f} es trivialmente continua, pues si U es abierto en E , entonces $\bar{f}^{-1}[U]$ es abierto en $P^2(\mathbb{R})$ porque $\pi^{-1}[\bar{f}^{-1}[U]] = f^{-1}[U]$ es abierto en la esfera. Como $P^2(\mathbb{R})$ es compacto, tenemos que \bar{f} es un homeomorfismo. Identificando los puntos de $P^2(\mathbb{R})$ con sus imágenes por \bar{f} podemos considerar que $P^2(\mathbb{R}) \subset \mathbb{R}^4$.

Ahora podríamos demostrar que E cumple la definición [An 7.1] de variedad diferenciable, pero sucede que con la métrica que hereda de \mathbb{R}^4 no es localmente isométrica a S^2 . Esto nos obligaría a hacer lo mismo que hemos hecho con el plano hiperbólico: prescindir del tensor métrico definido a partir del de \mathbb{R}^4 y definir otro alternativo a partir de la distancia hiperbólica (en nuestro caso de la distancia elíptica) para que la longitud de los segmentos de recta elíptica sea la que requiere la geometría elíptica. No obstante, hay una solución mejor, y es que, aunque hemos presentado la inmersión \tilde{f} por ser la más simple posible que nos permite identificar el plano proyectivo con un subespacio de un \mathbb{R}^n , retocándola ligeramente podemos conseguir una inmersión que sea localmente isométrica a la esfera con el tensor métrico natural, sin necesidad de cambiarlo por otro. El precio que tenemos que pagar, aparte de introducir algún coeficiente en la definición de F , es considerar \mathbb{R}^5 en lugar de \mathbb{R}^4 . Concretamente, ahora consideramos la función $F : \mathbb{R}^3 \rightarrow \mathbb{R}^5$ dada por

$$F(x, y, z) = (xy, xz, yz, \frac{1}{2}(x^2 - y^2), \frac{1}{2\sqrt{3}}(x^2 + y^2 - 2z^2)).$$

Claramente $F(p) = F(-p)$, por lo que la misma comprobación precedente (sin más que no tener en cuenta para nada la quinta función coordenada de F) muestra que la restricción $f = F|_{S^2}$ cumple $f(p) = f(q)$ si y sólo si $p = \pm q$.

Ahora llamamos $E = F[S^2]$, de modo que $f : S^2 \rightarrow E$ es igualmente continua y suprayectiva e induce una biyección $\tilde{f} : \mathbb{P}^2(\mathbb{R}) \rightarrow E$ que trivialmente es un homeomorfismo.

Así estamos en una situación análoga a la precedente, que nos permite considerar $\mathbb{P}^2(\mathbb{R}) \subset \mathbb{R}^5$ identificando sus puntos con los de E a través de f . Sin embargo, vamos a ver que con esta identificación $\mathbb{P}^2(\mathbb{R})$ resulta ser una superficie diferenciable localmente isométrica a la esfera, tal y como queríamos.

Para ello observamos en primer lugar que la matriz jacobiana de F es

$$\begin{pmatrix} y & z & 0 & x & x/\sqrt{3} \\ x & 0 & z & -y & y/\sqrt{3} \\ 0 & x & y & 0 & -2z/\sqrt{3} \end{pmatrix}.$$

Es fácil comprobar que su rango es 3 en todos los puntos $(x, y, z) \neq (0, 0, 0)$. Por ejemplo, el determinante de las tres primeras columnas es $-2xyz$, luego para que el rango fuera < 3 tendría que ser 0 una de las tres coordenadas, y una comprobación rutinaria en los tres casos muestra que en todos ellos el rango es 3. Por ejemplo, en el caso $z = 0$ la matriz se reduce a

$$\begin{pmatrix} y & 0 & 0 & x & x/\sqrt{3} \\ x & 0 & 0 & -y & y/\sqrt{3} \\ 0 & x & y & 0 & 0 \end{pmatrix}.$$

El determinante de las columnas 1, 2, 4 sólo puede ser 0 si $x = 0$, y el de las columnas 1, 3, 4 sólo puede ser 0 si $y = 0$, con lo que llegamos a que el punto tiene que ser el $(0, 0, 0)$, que habíamos descartado.

Ahora tomamos un punto $P \in E$ arbitrario y sea $x \in S^2$ tal que $F(x) = P$. Sea $X : U \rightarrow S^2$ una carta de S^2 alrededor de x , lo cual significa que $X[U]$ es abierto en S^2 y que JX tiene rango 2 en todo punto de U . Entonces, la composición $\bar{X} = X \circ F : U \rightarrow E$ cumple que $\bar{X}[U] = \bar{f}[X[U]]$ es un abierto en E y $J\bar{X} = JX \cdot JF$ tiene rango 2 en cada punto p , porque es la matriz de la composición de dos monomorfismos $\mathbb{R}^2 \rightarrow \mathbb{R}^3 \rightarrow \mathbb{R}^5$ (concretamente, es la matriz de $dX(p) \circ dF(X(p))$). Esto prueba que $X \circ F$ es una carta de E alrededor de P , luego E es una variedad diferenciable,² de acuerdo con la definición [An 7.1]. Además, la aplicación $f : S^2 \rightarrow E$ es obviamente diferenciable, según la definición [An 7.11].

Ahora vamos a comparar los tensores métricos de S^2 y de E . Por simplicidad estudiaremos los elementos de longitud correspondientes, pues el tensor métrico y el elemento de longitud se determinan mutuamente. Para ello consideramos las restricciones $x, y, z : S^2 \rightarrow \mathbb{R}$ de las funciones coordenadas de \mathbb{R}^3 , de modo que el elemento de longitud de S^2 es $ds^2 = dx^2 + dy^2 + dz^2$.

Fijada una semiesfera abierta U en S^2 y su imagen $U' = f[U]$, tenemos que $f|_U : U \rightarrow U'$ es biyectiva y diferenciable, por lo que podemos considerar las aplicaciones $\bar{x}, \bar{y}, \bar{z} : U' \rightarrow \mathbb{R}$ que resultan de componer $(f|_U)^{-1}$ con x, y, z . Obviamente cumplen la relación

$$\bar{x}^2 + \bar{y}^2 + \bar{z}^2 = 1. \quad (12.4)$$

A su vez, calculando la diferencial de ambos miembros resulta que

$$\bar{x} d\bar{x} + \bar{y} d\bar{y} + \bar{z} d\bar{z} = 0. \quad (12.5)$$

Por otra parte, podemos considerar las restricciones $x_1, \dots, x_5 : E \rightarrow \mathbb{R}$ de las funciones coordenadas de \mathbb{R}^5 , de modo que el elemento de longitud de E viene dado por

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 + dx_5^2.$$

La definición de f nos da las relaciones siguientes:

$$x_1 = \bar{x}\bar{y}, \quad x_2 = \bar{x}\bar{z}, \quad x_3 = \bar{y}\bar{z}, \quad x_4 = \frac{1}{2}(\bar{x}^2 - \bar{y}^2), \quad x_5 = \frac{1}{2\sqrt{3}}(\bar{x}^2 + \bar{y}^2 - 2\bar{z}^2).$$

De ellas deducimos las relaciones entre las diferenciales:

$$\begin{aligned} dx_1 &= \bar{y} d\bar{x} + \bar{x} d\bar{y}, & dx_2 &= \bar{z} d\bar{x} + \bar{x} d\bar{z}, & dx_3 &= \bar{z} d\bar{y} + \bar{y} d\bar{z}, \\ dx_4 &= \bar{x} d\bar{x} - \bar{y} d\bar{y}, & dx_5 &= \frac{1}{\sqrt{3}}(\bar{x} d\bar{x} + \bar{y} d\bar{y} - 2\bar{z} d\bar{z}) = -\sqrt{3} \bar{z} d\bar{z}, \end{aligned}$$

donde hemos simplificado dx_5 usando (12.5).

²Si el lector trata de probar que la inmersión en \mathbb{R}^4 que hemos definido anteriormente también es una variedad diferenciable se encontrará con que la matriz jacobiana de F no tiene rango 3 en todos los puntos, lo que obliga a tomar cartas de S^2 y comprobar directamente que las composiciones $\mathbb{R}^2 \rightarrow \mathbb{R}^3 \rightarrow \mathbb{R}^3$ de las diferenciales tienen rango 2 en todos los puntos.

Esto nos permite expresar el elemento de longitud de E en términos de $\bar{x}, \bar{y}, \bar{z}$. Como la expresión es larga operamos por separado cada cuadrado:

$$\begin{aligned} dx_1^2 &= \bar{y}^2 d\bar{x}^2 + \bar{x}^2 d\bar{y}^2 + 2\bar{x}\bar{y} d\bar{x} d\bar{y}, \\ dx_2^2 &= \bar{z}^2 d\bar{x}^2 + \bar{x}^2 d\bar{z}^2 + 2\bar{x}\bar{z} d\bar{x} d\bar{z}, \\ dx_3^2 &= \bar{z}^2 d\bar{y}^2 + \bar{y}^2 d\bar{z}^2 + 2\bar{y}\bar{z} d\bar{y} d\bar{z}, \\ dx_4^2 &= \bar{x}^2 d\bar{x}^2 + \bar{y}^2 d\bar{y}^2 - 2\bar{x}\bar{y} d\bar{x} d\bar{y}, \\ dx_5^2 &= 3\bar{z}^2 d\bar{z}^2. \end{aligned}$$

Al sumar todos los términos, usando (12.4), obtenemos

$$ds^2 = d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2 + 2(\bar{x} d\bar{x} + \bar{y} d\bar{y} + \bar{z} d\bar{z})d\bar{z} = d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2.$$

Equivalentemente, el tensor métrico de E en U' es

$$g_E = d\bar{x} \otimes d\bar{x} + d\bar{y} \otimes d\bar{y} + d\bar{z} \otimes d\bar{z},$$

mientras que el de S^2 en U es

$$g_{S^2} = dx \otimes dx + dy \otimes dy + dz \otimes dz.$$

Esto implica que $f|_U : U \rightarrow U'$ es una isometría, pues, para todo punto $p \in U$ y todos los vectores $v, w \in T_p S$, se cumple que

$$\begin{aligned} g_{f(p)}(df(p)(v), df(p)(w)) &= d\bar{x}(f(p))(df(p)(v)) d\bar{x}(f(p))(df(p)(w)) \\ &+ d\bar{y}(f(p))(df(p)(v)) d\bar{y}(f(p))(df(p)(w)) + d\bar{z}(f(p))(df(p)(v)) d\bar{z}(f(p))(df(p)(w)) \\ &= d(f \circ \bar{x})(p)(v) d(f \circ \bar{x})(p)(w) + d(f \circ \bar{y})(p)(v) d(f \circ \bar{y})(p)(w) \\ &\quad + d(f \circ \bar{z})(p)(v) d(f \circ \bar{z})(p)(w) = \\ &dx(p)(v) dx(p)(w) + dy(p)(v) dy(p)(w) + dz(p)(v) dz(p)(w) = g_p(v, w). \end{aligned}$$

A partir de aquí consideraremos el plano elíptico $P^2(\mathbb{R})$ con la estructura de variedad diferencial que resulta de identificar sus puntos con los de E a través del homeomorfismo \bar{f} . El teorema siguiente resume lo que hemos obtenido:

Teorema 12.10 *El plano elíptico $E = P^2(\mathbb{R})$ es una superficie diferenciable tal que existe una aplicación diferenciable $f : S^2 \rightarrow E$ suprayectiva con la propiedad de que $f(p) = f(q)$ si y sólo si $p = \pm q$ y de modo que $r \subset E$ es una recta elíptica si y sólo si $f^{-1}[r]$ es un círculo máximo en S^2 , y si U es una de las semiesferas en que F divide la esfera, entonces $U' = f[U] = E \setminus r$ y $f|_U : U \rightarrow U'$ es una isometría.*

En efecto, observemos que una recta de $r' \subset P^2(\mathbb{R})$ se identifica (por definición) con $r = \bar{f}[r']$ como subconjunto de E , de modo que $f^{-1}[r] = \pi^{-1}[r']$ es un círculo máximo en S^2 . Es claro entonces que $\pi|_U : U \rightarrow P^2(\mathbb{R}) \setminus r'$ es un homeomorfismo, que compuesto con \bar{f} nos da el homeomorfismo $f|_U : U \rightarrow E \setminus r$, y acabamos de probar que es una isometría.

A partir de este punto ya podemos “olvidarnos” del modo concreto en que hemos sumergido $P^2(\mathbb{R})$ en \mathbb{R}^5 . Nos basta con saber que $P^2(\mathbb{R})$ es localmente isométrico a S^2 en las condiciones del teorema anterior. Si hubiéramos contado con el concepto abstracto de variedad diferenciable podríamos haber definido directamente una estructura de variedad en $P^2(\mathbb{R})$ (junto con un tensor métrico adecuado) de modo que tendríamos la situación del teorema sin necesidad de haber buscado una forma de sumergir el plano elíptico en ningún espacio \mathbb{R}^m .

Conviene hacer una precisión adicional, y es que el teorema anterior no nos obliga a fijar un sistema de referencia en \mathbb{R}^3 , sino que, tomando un sistema de referencia ortonormal arbitrario con centro en el centro de la esfera S^2 , podemos considerar coordenadas x, y, z respecto a las cuales cualquier punto prefijado tenga cualquier terna de coordenadas prefijada. En particular, dado un punto $P \in E$, siempre podemos tomar un sistema de referencia en \mathbb{R}^3 de modo que las antiimágenes de P tengan coordenadas $(0, 0, \pm 1)$, con lo que tomando como U la semiesfera $z > 0$, una carta alrededor de P está formada por la carta $(x, y) \mapsto (x, y, \sqrt{1 - z^2})$ de S^2 compuesta con $f|_U$. En particular, es una carta que asigna al punto prefijado las coordenadas $(0, 0)$.

El elemento de longitud El modelo esférico ya muestra que el elemento de longitud en $P^2(\mathbb{R})$ respecto al cual la longitud de un segmento de recta es su longitud elíptica es precisamente el de la estructura de superficie diferenciable que hemos considerado, pues la longitud de un segmento de recta cuando se identifica con un arco de círculo máximo en la esfera es precisamente (de acuerdo con el modelo esférico) su longitud euclídea, que es la que obtenemos con la estructura diferencial que hemos dado a $P^2(\mathbb{R})$, en virtud de la isometría local con S^2 . Ahora bien, es interesante observar que podemos llegar a ello por un camino totalmente análogo al que empleamos en el capítulo anterior para determinar el elemento de longitud hiperbólica.

Partimos de una cónica imaginaria en $P^2(R)$ de ecuación $f(X, X) = 0$ en coordenadas homogéneas y consideramos una curva $X(t) = (x(t), y(t), z(t))$. La expresión que hemos obtenido para la distancia elíptica equivale a

$$\operatorname{sen} d(X, Y) = \sqrt{\frac{-f^2(X, Y) + f(X, X)f(Y, Y)}{f(X, X)f(Y, Y)}},$$

donde hemos cambiado el coseno por un seno para usar que $\frac{\operatorname{sen} x}{x} \rightarrow 1$. Exactamente el mismo razonamiento que en el caso hiperbólico nos lleva a que

$$\lim_{h \rightarrow 0} \frac{d(X(t+h), X(t))}{|h|} = \sqrt{\frac{-f^2(X, X') + f(X, X)f(X', X')}{f^2(X, X)}}.$$

Si suponemos, más concretamente, que $f(X_1, X_2) = x_1x_2 + y_1y_2 + z_1z_2$ y que $f(X(t), X(t)) = 1$, es decir, que elegimos las coordenadas homogéneas de cada punto de la curva en la esfera unitaria, entonces al derivar queda $f(X, X') = 0$, luego

$$\lim_{h \rightarrow 0} \frac{d(X(t+h), X(t))}{|h|} = \sqrt{x'^2 + y'^2 + z'^2}.$$

En el caso en que $X : [a, b] \rightarrow \mathbb{P}^2(\mathbb{R})$ parametriza un segmento de recta, como en el caso hiperbólico se razona que el límite es la derivada de la longitud elíptica $s(t) = L(X|_{[0,t]})$, por lo que la longitud del segmento es

$$L = \int_a^b \sqrt{x'^2 + y'^2 + z'^2} dt = \int_a^b \sqrt{dx^2 + dy^2 + dz^2},$$

es decir, la longitud de la curva determinada por el elemento de longitud

$$s^2 = dx^2 + dy^2 + dz^2.$$

Esto justifica que, una vez tenemos una estructura de superficie diferenciable en $\mathbb{P}^2(\mathbb{R})$, para que su métrica sea la elíptica es necesario que el tensor métrico sea el anterior, es decir, el mismo que el de la esfera, y por ello la estructura diferencial que hemos definido sobre el plano proyectivo (eligiendo adecuadamente su inmersión en \mathbb{R}^5) es la que proporciona ciertamente el tensor métrico correspondiente a la geometría elíptica.

Por ello, las longitudes de curvas arbitrarias y las superficies de regiones arbitrarias (razonables) calculadas con la estructura diferencial de $\mathbb{P}^2(\mathbb{R})$ son las “correctas” para la geometría elíptica, en el sentido de que son las únicas compatibles con la longitud elíptica definida al principio de este capítulo (sin consideraciones de geometría diferencial). Más detalladamente, la situación es la misma que teníamos con el plano hiperbólico y la pseudoesfera: alguien que “viva” en un plano elíptico y que sepa calcular longitudes de segmentos (y las use para aproximar longitudes de curvas y áreas de regiones) no puede distinguir la geometría de su entorno de la geometría de una semiesfera, por lo que las longitudes y las áreas “correctas” correspondientes con su forma de medir segmentos de recta son las que sabemos que son correctas sobre la esfera.

Es interesante expresar el elemento de longitud elíptica en términos de coordenadas polares. Dado un punto arbitrario O del plano elíptico, podemos identificarlo con el polo Norte $(0, 0, 1)$ de la esfera, y es claro que los puntos que se encuentran a una distancia ρ de dicho polo Norte son los de la forma

$$(x, y, z) = (\text{sen } \rho \cos \theta, \text{sen } \rho \text{sen } \theta, \cos \rho).$$

Cada punto de $\mathbb{P}^2(\mathbb{R})$ distinto de O y que no esté sobre la recta polar de O está determinado por una única coordenada $\rho > 0$ y por un ángulo θ unívocamente determinado salvo múltiplos de 2π . Éstas son las coordenadas polares elípticas. Diferenciando obtenemos que

$$dx = \cos \rho \cos \theta d\rho - \text{sen } \rho \text{sen } \theta d\theta, \quad dy = \cos \rho \text{sen } \theta d\rho + \text{sen } \rho \cos \theta d\theta,$$

$$dz = -\text{sen } \rho d\rho,$$

lo cual nos da a su vez que el elemento de longitud es

$$ds^2 = d\rho^2 + \text{sen}^2 \rho d\theta^2.$$

Así es fácil ver que la longitud de una circunferencia de radio elíptico r es

$$L = 2\pi \text{sen } r.$$

Curvatura y geodésicas Ahora ya podemos afirmar con rigor que la curvatura del plano elíptico es constante igual a 1 y que sus geodésicas son las rectas, tal y como afirmábamos “intuitivamente” al principio de la sección.

El área elíptica El elemento de área elíptico es el mismo que el de la esfera. Si lo expresamos en coordenadas polares, del elemento de longitud obtenemos que

$$E = 1, \quad F = 0, \quad G = \operatorname{sen}^2 \rho,$$

luego $d\sigma = \sqrt{EG - F^2} d\rho d\theta = \operatorname{sen} \rho d\rho d\theta$. De aquí se sigue, por ejemplo, que el área de un círculo elíptico de radio r es

$$A = \int_0^{2\pi} \int_0^r \operatorname{sen} \rho d\rho d\theta = 2\pi(1 - \cos r) = 4\pi \operatorname{sen}^2 \frac{r}{2}.$$

En particular, al igual que las rectas elípticas tienen longitud finita π , el área del plano elíptico es finita y vale

$$A = 4\pi \operatorname{sen}^2 \frac{\pi}{4} = 2\pi$$

(la mitad del área de la esfera, como era de esperar). Es fácil ver que el área de un bilátero de ángulo α es 2α . Consideremos ahora un triángulo elíptico T de lados a, b, c y ángulos α, β, γ . El plano elíptico es la unión de T y sus cuatro triángulos adyacentes. Si llamamos T_a al triángulo adyacente que comparte con T el lado a , tenemos que $T \cup T_a$ es un bilátero de ángulo α , luego

$$m(T) + m(T_a) = 2\alpha,$$

(donde m representa el área elíptica). Sumando las ecuaciones análogas para los otros triángulos adyacentes obtenemos que

$$3m(T) + m(T_a) + m(T_b) + m(T_c) = 2\alpha + 2\beta + 2\gamma,$$

pero $m(T) + m(T_a) + m(T_b) + m(T_c) = 2\pi$, porque los cuatro triángulos cubren el plano, luego

$$m(T) = \alpha + \beta + \gamma - \pi.$$

Esto nos da una prueba analítica de que los ángulos de un triángulo elíptico suman siempre más de π .

Apéndice A

La geometría inversiva

En [IGE 5.12] definimos el concepto de inversión respecto de una circunferencia y comprobamos su utilidad a la hora de demostrar numerosos resultados geométricos. En [IGE 6.6] definimos las transformaciones circulares, que incluyen a las inversiones como caso particular y que se estudian con más naturalidad si añadimos al plano un punto infinito. Aquí vamos a generalizar estos conceptos y a relacionarlos con la geometría afín y proyectiva. Como los resultados que vamos a exponer están relacionados con diferentes capítulos de este libro, la exponemos en un apéndice para reunir así resultados que, de otro modo, tendrían que quedar dispersos en el libro.

Al final de la sección [An 2.5] hemos visto que si

$$S = \{x \in \mathbb{R}^{n+1} \mid \|x\|^2 = 1\}$$

es la esfera unitaria en \mathbb{R}^{n+1} y $E^\infty = \mathbb{R}^n \cup \{\infty\}$ es la compactificación de Alexandroff de \mathbb{R}^n (definida en la sección [An 3.2]), la proyección estereográfica induce un homeomorfismo $\pi : S^n \rightarrow E^\infty$ dado por

$$\pi(x) = \left(\frac{x_1}{1 - x_{n+1}}, \dots, \frac{x_n}{1 - x_{n+1}} \right),$$

entendiendo que el polo norte $N = (0, \dots, 0, 1)$ tiene imagen $\pi(N) = \infty$. Su inversa es

$$\pi^{-1}(y) = \left(\frac{2y_1}{\|y\|^2 + 1}, \dots, \frac{2y_n}{\|y\|^2 + 1}, \frac{\|y\|^2 - 1}{\|y\|^2 + 1} \right).$$

Como todas las consideraciones que vamos a hacer van a ser de carácter geométrico y no topológico, técnicamente podemos trabajar en $E = R^n$, donde R es cualquier cuerpo ordenado euclídeo (y, correspondientemente, con $S \subset R^{n+1}$).

La idea fundamental de la geometría inversiva consiste en pensar que, los puntos de E son “en realidad” coordenadas de puntos de S , pero que “hay” un punto más, el punto ∞ , que no tiene coordenadas, pero que es “tan real” como los demás puntos que consideramos.

A.1 La proyección estereográfica

En esta primera sección mostraremos que la proyección estereográfica conserva ciertas propiedades geométricas, que son precisamente las propiedades significativas en la geometría inversiva.

Consideremos en primer lugar un hiperplano Π en R^{n+1} , que estará formado por los puntos $x \in R^{n+1}$ que satisfacen una ecuación de la forma $ax = \alpha$, para cierto $a \in R^{n+1}$ no nulo. Multiplicando la ecuación por una constante podemos suponer que $\|a\| = 1$ y $\alpha \geq 0$, con lo que $\alpha a \in \Pi$ y

$$\Pi = \{x \in R^{n+1} \mid a(x - \alpha a) = 0\}.$$

Por consiguiente,

$$S \cap \Pi = \{x \in \Pi \mid \|x - \alpha a\|^2 = 1 - \alpha^2\}.$$

En efecto, si $x \in S \cap \Pi$, entonces

$$\|x - \alpha a\|^2 = (x - \alpha a)(x - \alpha a) = \|x\|^2 + \alpha^2 - 2\alpha ax = 1 + \alpha^2 - 2\alpha^2 = 1 - \alpha^2$$

y, recíprocamente, si $x \in \Pi$ cumple $\|x - \alpha a\|^2 = 1 - \alpha^2$, entonces

$$\|x\|^2 = (x - \alpha a + \alpha a)(x - \alpha a + \alpha a) = \|x - \alpha a\|^2 + \alpha^2 = 1,$$

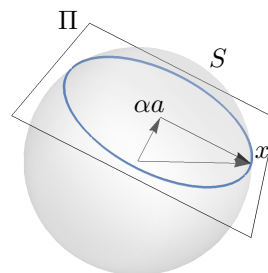
luego $x \in S \cap \Pi$.

Concluimos así que $S \cap \Pi = \emptyset$ si $\alpha > 1$, $S \cap \Pi = \{a\}$ si $\alpha = 1$, mientras que, para $0 \leq \alpha < 1$, tenemos que $S \cap \Pi$ es una esfera de radio $\sqrt{1 - \alpha^2}$ en Π .

Definición A.1 Llamaremos *hiperesferas* en S a las intersecciones no triviales de S con hiperplanos, es decir, las intersecciones que consten de más de un punto.

Hemos probado que las hiperesferas son, concretamente las intersecciones de S con hiperplanos de ecuación $ax = \alpha$, con $\|a\| = 1$ y $0 \leq \alpha < 1$. Notemos que si $n = 2$ las hiperesferas son simplemente las circunferencias contenidas en S .

Veamos ahora cómo se comporta la proyección estereográfica sobre las hiperesferas:



Teorema A.2 La proyección estereográfica transforma las hiperesferas de S que no pasan por el polo norte N en las esferas de E , y las hiperesferas de S que pasan por el polo norte en los hiperplanos de E^∞ (definidos como los hiperplanos de E a los que añadimos el punto ∞).

DEMOSTRACIÓN: Sea Π un hiperplano que defina una hiperesfera en S , con ecuación $ax = \alpha$. La hiperesfera pasa por N si y sólo si $a_{n+1} = \alpha$. Conviene

representar $a = (\tilde{a}, a_{n+1})$. La imagen de la hiperesfera está formada por los puntos $y \in E^\infty$ que cumplen $a\pi^{-1}(y) = \alpha$. Si y es finito esto, equivale a

$$\frac{2a_1y_1}{\|y\|^2 + 1} + \dots + \frac{2a_ny_n}{\|y\|^2 + 1} + a_{n+1} \frac{\|y\|^2 - 1}{\|y\|^2 + 1} = \alpha,$$

o también a $(\alpha - a_{n+1})\|y\|^2 - 2\hat{a}y + a_{n+1} + \alpha = 0$.

Si la hiperesfera pasa por N , los puntos finitos de su imagen cumplen la ecuación $2\hat{a}y = a_{n+1} + \alpha$, luego es un hiperplano en E . Recíprocamente, los puntos de un hiperplano en E cumplen una ecuación $ay = \alpha$, con a no nulo, y sus imágenes por π^{-1} son los puntos de $S \setminus \{N\}$ que cumplen la ecuación

$$\frac{a_1x_1}{1 - x_{n+1}} + \dots + \frac{a_nx_n}{1 - x_{n+1}} = \alpha,$$

que equivale a $a_1x_1 + \dots + a_nx_n + \alpha x_{n+1} = \alpha$, luego son los puntos de una hiperesfera en S que contiene además a N .

Por lo tanto, π pone en correspondencia biunívoca a las hiperesferas de S que pasan por N y los hiperplanos de E^∞ (considerando que contienen a ∞).

Si consideramos ahora una hiperesfera en S que no pase por N , los puntos de su imagen (necesariamente finitos) son los que cumplen

$$\|y\|^2 - \frac{2\hat{a}}{\alpha - a_{n+1}}y = -\frac{\alpha + a_{n+1}}{\alpha - a_{n+1}},$$

o también

$$\left\|y - \frac{\hat{a}}{\alpha - a_{n+1}}\right\|^2 = -\frac{\alpha + a_{n+1}}{\alpha - a_{n+1}} - \frac{\|\hat{a}\|^2}{(\alpha - a_{n+1})^2},$$

que es la ecuación de una esfera en E .

Recíprocamente, si $\|y - c\|^2 = r^2$ es la ecuación de una esfera en E , su imagen por π^{-1} está formada por los puntos $x \in S$ que cumplen

$$\sum_{i=1}^n \left(\frac{x_i}{1 - x_{n+1}} - c_i \right)^2 = r^2,$$

que equivale a

$$\sum_{i=1}^n (x_i - (1 - x_{n+1})c_i)^2 = r^2(1 - x_{n+1})^2,$$

$$\sum_{i=1}^n x_i^2 + (1 - x_{n+1})\|c\|^2 - 2(1 - x_{n+1}) \sum_{i=1}^n c_i x_i = r^2(1 - x_{n+1})^2.$$

El primer sumatorio es $1 - x_{n+1}^2 = (1 + x_{n+1})(1 - x_{n+1})$, luego la ecuación equivale a

$$1 + x_{n+1} + \|c\|^2 - 2 \sum_{i=1}^n c_i x_i = r^2(1 - x_{n+1}),$$

o también a

$$\sum_{i=1}^n 2c_i x_i - (r^2 + 1)x_{n+1} = \|c\|^2 + 1 - r^2,$$

que es la ecuación de una hiperesfera en S (que necesariamente no pasa por N , pues ∞ no está en la esfera de partida).

Concluimos que π hace corresponder las hiperesferas que no pasan por N con las esferas de E . ■

Así pues, si “queremos pensar” que E es una esfera menos un punto ∞ , “debemos pensar” que no hay ninguna diferencia esencial entre las esferas y los hiperplanos de E , pues “pensados como subconjuntos de “la esfera”, no son sino las hiperesferas (unas las que no pasan por N y otras las que pasan por N , pero no hay ninguna diferencia esencial entre las hiperesferas que pasan o no pasan por N , la diferencia la introduce la proyección estereográfica al enviar a N al infinito). Para concretar esta idea damos la definición siguiente:

Definición A.3 Llamaremos *hiperesferas* de E^∞ a los hiperplanos de E a los que añadimos el punto ∞ y a las esferas.

En estos términos, los hiperplanos son las hiperesferas que pasan por ∞ y las esferas las hiperesferas que no pasan por ∞ , y el teorema anterior afirma que la proyección estereográfica hace corresponder biunívocamente las hiperesferas de S con las hiperesferas de E^∞ .

Podemos considerar que las hiperesferas son los objetos geométricos básicos que estudia la geometría inversiva.

A.2 Transformaciones esféricas

En esta sección E representará siempre a un espacio afín euclídeo de dimensión $n \geq 2$ sobre un cuerpo euclídeo R . Fijando un sistema de referencia ortonormal en E tenemos un isomorfismo $E \cong R^n$ que a su vez nos permite considerar a la proyección estereográfica como aplicación $\pi : S \rightarrow E^\infty$ (donde S sigue siendo la esfera unitaria en R^{n+1}). Hay que tener presente que π depende de la elección del sistema de referencia en E , pero los conceptos que vamos a definir en esta sección no van a depender de dicha elección.

Si $f : E \rightarrow E$ es una biyección afín, la consideraremos extendida a E^∞ con el convenio de que $f(\infty) = \infty$.

Ejercicio: Probar que si $R = \mathbb{R}$, la extensión de una biyección afín es continua en ∞ , con lo que, de hecho, es un homeomorfismo de E^∞ en E^∞ .

El teorema siguiente muestra que las semejanzas “respetan” las hiperesferas:

Teorema A.4 *Las semejanzas en E transforman hiperesferas en hiperesferas.*

DEMOSTRACIÓN: Por el teorema 5.20, toda semejanza es composición de una isometría y una homotecia, luego basta probar el teorema para ambos tipos de biyecciones afines. Para el caso de las isometrías es inmediato, y a tu vez toda homotecia de centro p se expresa en coordenadas como $f(x) = p + k(x-p)$, luego es composición de dos traslaciones y una homotecia de la forma $f(x) = kx$, luego basta considerar una homotecia de este tipo. Ahora bien, f transforma el hiperplano de ecuación $ax = \alpha$ en el de ecuación $ax = k\alpha$, y la esfera de ecuación $\|x - c\|^2 = r^2$ en la de ecuación $\|x - kc\|^2 = (kr)^2$. ■

Veremos que las semejanzas son las únicas biyecciones de E que transforman hiperesferas en hiperesferas, pero no son las únicas biyecciones de E^∞ con esta propiedad, sino que, si permitimos que ∞ no quede fijo, aparece una nueva familia de aplicaciones que respetan las hiperesferas. Para encontrarlas fijamos un sistema de referencia ortonormal en E y consideramos la proyección estereográfica

$$\pi_+ : S \longrightarrow E^\infty.$$

Ahora la llamamos π_+ porque vamos a considerar también la proyección análoga $\pi_- : S \longrightarrow E^\infty$ definida desde el polo sur $-N = (0, \dots, 0, -1)$ en lugar de desde el polo norte N . Es fácil ver que su expresión en coordenadas es

$$\pi_-(x) = \frac{1}{1 + x_{n+1}}(x_1, \dots, x_n),$$

entendiendo que $\pi_-(-N) = \infty$.

Evidentemente, π_- también hace corresponder las hiperesferas de S con las de E^∞ . Esto nos lleva a considerar la biyección $J : E^\infty \longrightarrow E^\infty$ dada por $J = \pi_+^{-1} \circ \pi_-$, que transformará hiperesferas en hiperesferas, pero ahora $J(\infty) = \pi_+^{-1}(0) = 0$ y $J(0) = \pi_+^{-1}(\infty) = \infty$. Es inmediato comprobar que la expresión en coordenadas de J es

$$J(y) = \frac{y}{\|y\|^2}.$$

Obviamente esta expresión depende de la elección del sistema de referencia, pero podemos dar una definición intrínseca:

Definición A.5 Si E es un espacio afín euclídeo de dimensión $n \geq 2$, la *inversión* respecto de la esfera de centro p y radio r se define como la aplicación $J_{p,r} : E^\infty \longrightarrow E^\infty$ dada por

$$J_{p,r}(x) = p + \frac{r^2}{\|px\|^2} \overrightarrow{px},$$

con el convenio de que $J_{p,r}(p) = \infty$ y $J_{p,r}(\infty) = p$.

La interpretación geométrica¹ es relativamente simple: para $x \neq p$ finito, $J_{p,r}(x)$ es el punto q situado en la semirrecta de origen p determinada por x y que cumple $\|\overrightarrow{px}\| \|\overrightarrow{pq}\| = r^2$.

¹Compárese con la definición 10.17.

De aquí se sigue que $J_{p,r}$ es biyectiva y, más aún, que es su propia inversa, así como que fija a todos los puntos de la esfera de centro p y radio r , mientras que los puntos de la bola de centro p y radio r se transforman en los del exterior de la bola, y viceversa. Su expresión coordenada respecto de un sistema de referencia ortonormal es

$$J_{p,r}(y) = p + \frac{r^2(y-p)}{\|y-p\|^2},$$

con lo que la aplicación $J = J_{0,1}$ que hemos obtenido antes es la inversión respecto al origen de coordenadas de la esfera unitaria. Además, estas expresiones coordenadas muestran que

$$J_{p,r}(y) = p + rJ(r^{-1}(y-p)),$$

lo que en términos intrínsecos significa que toda inversión $J_{p,r}$ es la composición de la traslación que envía p al origen de coordenadas, con la homotecia de centro el origen y razón r^{-1} , con J , con la homotecia de razón r y con la traslación que envía el origen a p . Esto permite reducir muchos resultados sobre inversiones al caso particular de J .

Por ejemplo, si $R = \mathbb{R}$, el hecho de que $J = \pi^{-1}_+ \circ \pi_-$ implica que J es un homeomorfismo de E^∞ en sí mismo, pues tanto π_+ como π_- lo son. Las observaciones precedentes implican que lo mismo vale para toda inversión. Más en general, lo mismo vale para las transformaciones esféricas, que definimos a continuación:

Definición A.6 Si E es un espacio afín euclídeo de dimensión $n \geq 2$, definimos las *transformaciones esféricas* de E como los elementos del subgrupo $\mathcal{M}(E)$ de las biyecciones de E^∞ generado por las semejanzas y las inversiones.

Claramente $\mathcal{M}(E)$ está generado también por las isometrías, las homotecias y la inversión J (respecto de un sistema de referencia prefijado).

Teorema A.7 *Las transformaciones esféricas transforman las hiperesferas de E en hiperesferas de E .*

DEMOSTRACIÓN: El hecho de que (respecto de un sistema de referencia adecuado) se cumpla que $J = \pi_+^{-1} \circ \pi_-$ implica, como ya hemos observado, que J transforma hiperesferas en hiperesferas, y teniendo en cuenta el teorema A.4, lo mismo vale para toda transformación esférica. ■

Podríamos pensar que, del mismo modo en que "nos hemos encontrado" con las inversiones, tal vez podríamos encontrar otras biyecciones que transformen hiperesferas en hiperesferas, pero en realidad no es posible:

Teorema A.8 *Sea E un espacio afín euclídeo de dimensión $n \geq 2$ sobre un cuerpo euclídeo arquimediano y $f : E^\infty \rightarrow E^\infty$ una biyección que transforma hiperesferas en hiperesferas. Entonces f es una transformación esférica, y si además $f(\infty) = \infty$, se trata de una semejanza de E .*

DEMOSTRACIÓN: Supongamos en primer lugar que $f(\infty) = \infty$. Entonces f se restringe a una biyección $f : E \rightarrow E$ que transforma hiperplanos en hiperplanos, y como toda recta se puede expresar como intersección finita de hiperplanos, resulta que f transforma rectas en rectas. Por el teorema fundamental de la geometría afín concluimos que f es una biyección afín, que además transforma esferas en esferas. Componiéndola con una traslación y una homotecia, obtenemos otra biyección afín que deja invariante a la esfera unitaria S_1 . Si probamos que esta composición es una semejanza, también lo será f , luego no perdemos generalidad si suponemos que $f[S_1] = S_1$. Ahora bien, esto implica que f es, de hecho, una isometría, pues su aplicación lineal asociada cumple $\|f(\vec{v})\| = 1$ siempre que $\|\vec{v}\| = 1$, luego, aplicando esto a $\vec{v} = \vec{w}/\|\vec{w}\|$, resulta que $\|f(\vec{w})\| = \|\vec{w}\|$ para todo vector no nulo, y trivialmente para $\vec{0}$, lo que implica que f conserva las distancias entre los puntos, luego el teorema 3.27 nos da que es una isometría.

Si $f(\infty) = p \neq \infty$, entonces $g = f \circ J_{p,1}$ cumple que $g(\infty) = \infty$ y también es una biyección que transforma hiperesferas en hiperesferas, luego por la parte ya probada es una semejanza de E , luego f es una transformación esférica. ■

Observemos que nos encontramos con una “asimetría”, desde el momento en que tenemos definida la inversión respecto de una esfera de E , pero no respecto de un hiperplano de E . Cabría esperar que existieran también inversiones respecto de hiperplanos, para que tuviera sentido hablar de la inversión respecto de cualquier hiperesfera de E . Esto se arregla definiendo la inversión respecto de un hiperplano H como la simetría respecto de H , en el sentido usual. Estas simetrías son semejanzas, luego ya están incluidas en el grupo de transformaciones esféricas. Del mismo modo que las inversiones respecto a esferas intercambian el interior con el exterior de la esfera que fijan, las inversiones respecto de hiperplanos intercambian sus dos semiespacios.

En virtud del teorema 5.36, toda isometría de E es composición de reflexiones (simetrías respecto de hiperplanos), luego de inversiones respecto de hiperplanos. Por lo tanto, el grupo esférico está generado por las inversiones (respecto de hiperesferas) y las homotecias. Pero el teorema siguiente muestra que las homotecias son prescindibles:

Teorema A.9 *La composición de dos inversiones respecto de esferas con el mismo centro es una homotecia.*

DEMOSTRACIÓN: Tomamos un sistema de referencia en E que tenga origen en el centro común de las inversiones. Entonces las expresiones coordenadas son

$$y \mapsto \frac{r_i^2 y}{\|y\|^2},$$

para $i = 1, 2$, y al componerlas obtenemos

$$y \mapsto \frac{r_2^2}{r_1^2} y.$$

Más concretamente, vemos que toda homotecia de razón positiva se expresa como composición de dos inversiones. ■

Teniendo en cuenta que la homotecia de razón -1 es una isometría, tenemos probado el teorema siguiente:

Teorema A.10 *El grupo de las transformaciones esféricas está generado por las inversiones respecto de las hiperesferas.*

A.3 La geometría circular

Nos restringimos ahora a la geometría inversiva de dimensión 2, que se conoce también como *geometría circular*, porque ahora las esferas son circunferencias (y los hiperplanos son las rectas, pero nos referiremos conjuntamente a las hiperesferas como “circunferencias”). Por simplicidad trabajaremos sobre el cuerpo \mathbb{R} de los números reales. En este caso podemos valernos de dos “trucos” que resultan ser muy fructíferos:

El primero es que podemos identificar $E = \mathbb{R}^2$ con el cuerpo \mathbb{C} de los números complejos, con lo que \mathbb{C}^∞ es la esfera de Riemann que describimos al final de la sección 4.1.

El segundo consiste en identificar \mathbb{C}^∞ con la recta proyectiva compleja $P^1(\mathbb{C})$. Así, sobre \mathbb{C}^∞ tenemos definidas las homografías, que en coordenadas homogéneas tienen la forma

$$f(z, w) = (az + bw, cz + dw), \quad ad - bc \neq 0,$$

pero vamos a usar la forma afín:

$$f(z) = \frac{az + b}{cz + d},$$

entendiendo que $f(-d/c) = \infty$ y que $f(\infty) = a/c$ (hechos que se comprueban con la expresión de f en coordenadas homogéneas). Equivalentemente:

$$f(z) = \frac{a}{c} + \frac{b - \frac{ad}{c}}{cz + d},$$

lo que muestra que f se expresa como composición de cuatro homografías, de la forma $z \mapsto a + z$, $z \mapsto az$, $z \mapsto 1/z$.

La homografía $z \mapsto 1/z = \bar{z}/|z|^2$ puede expresarse como composición de la conjugación compleja $z \mapsto \bar{z}$ (que geoméricamente es la reflexión respecto del eje real) y la inversión $J(z) = z/|z|^2$.

El producto $z \mapsto az$ (con $a \neq 0$) puede descomponerse en $z \mapsto (a/|a|)z$ seguida de $z \mapsto |a|z$. La primera aplicación es un giro, que se descompone en dos simetrías respecto de rectas. La segunda es una homotecia, que puede expresarse como composición de dos inversiones de centro 0. Finalmente, la traslación $z \mapsto a + z$ puede expresarse como composición de dos reflexiones, la dada por $z \mapsto a/2 + z$ y $z \mapsto 3a/2 - z$.

Con esto hemos probado la mitad del teorema siguiente:

Teorema A.11 *Las homografías de \mathbb{C}^∞ son las aplicaciones $\mathbb{C}^\infty \rightarrow \mathbb{C}^\infty$ expresables como composición de un número par de inversiones.*

DEMOSTRACIÓN: Hay que probar que un producto de dos inversiones es una homografía. El producto de dos reflexiones (inversiones respecto de rectas) es un giro o una traslación. Las traslaciones son de la forma $z \mapsto z + a$, luego son homografías. Si f es un giro de centro O , entonces $z \mapsto f(z + O) - O$ es un giro de centro 0 , luego existe un $a \in \mathbb{C}$ de módulo 1 tal que $f(z + O) - O = az$, luego $f(z) = a(z - O) + O$, luego es una homografía.

Consideremos ahora una inversión f respecto de una recta y otra g respecto de una circunferencia. Supongamos primero que el eje de f pasa por el centro de la circunferencia fijada por g . Consideremos una traslación que lleve dicho centro al punto 0 , seguida de un giro de centro 0 que transforme el trasladado del eje de f en el eje real, seguido de una homotecia de centro 0 que transforme la traslación de la circunferencia de g en la circunferencia unidad. Si llamamos h a esta composición, es claro que h es una homografía. Es fácil ver que $h^{-1}fh$ es la reflexión respecto al eje real y que $h^{-1}gh$ es la inversión respecto de la circunferencia unidad. Por consiguiente $h^{-1}fh h^{-1}gh$ y $h^{-1}gh h^{-1}fh$ son ambos iguales a la homografía $z \mapsto 1/z$, luego fg y gf también son homografías.

Supongamos ahora que el eje de f no pasa por el centro de la circunferencia de g . Sea f' una reflexión respecto a una recta que pase por dicho centro. Entonces $fg = ff'f'g$ es el producto de dos homografías, luego es una homografía, e igualmente se prueba que gf lo es.

Si tenemos una composición de dos inversiones respecto de circunferencias f y g , tomamos como h la inversión respecto a la recta que pasa por los centros de sus circunferencias (o una recta cualquiera si ambos coinciden). Entonces $fg = fhhg$ es un producto de dos homografías, luego es una homografía, y lo mismo vale para gf . ■

En particular hemos probado que todas las homografías de $P^1(\mathbb{C})$ son transformaciones circulares (es decir, transformaciones esféricas del plano).

Definición A.12 Llamaremos $\mathcal{M}(\mathbb{C})$ al grupo de las transformaciones circulares de \mathbb{C}^∞ , es decir, el grupo generado por las inversiones, y llamaremos $\mathcal{M}_0(\mathbb{C})$ al subgrupo formado por las transformaciones circulares que pueden expresarse como producto de un número par de inversiones. A sus elementos los llamaremos *transformaciones circulares directas* o *transformaciones de Möbius*, mientras que las que se expresan como producto de un número impar de inversiones se llaman *transformaciones circulares inversas*.²

En estos términos hemos probado que las transformaciones de Möbius de \mathbb{C}^∞ son simplemente las homografías de $P^1(\mathbb{C})$.

Observemos que una transformación circular no puede ser a la vez directa e inversa, pues en tal caso podríamos expresar una inversión como producto de

²Las transformaciones de Möbius son las transformaciones circulares que conservan la orientación, en un sentido que no estamos en condiciones de definir aquí.

un número par de ellas, luego sería una homografía, pero esto es imposible, pues las inversiones fijan a infinitos puntos, mientras que una homografía distinta de la identidad no puede fijar a más de dos puntos.

Si f es cualquier transformación circular inversa, entonces su conjugada \bar{f} es una homografía, luego el grupo de las homografías es un subgrupo normal de índice 2 en el grupo de todas las transformaciones circulares (pues toda transformación circular es congruente módulo el subgrupo de las homografías con la identidad o con la conjugación compleja).

Si una transformación circular f es directa y deja fijos a tres puntos (finitos o no) entonces es la identidad, porque es una homografía de una recta proyectiva compleja. Si f es inversa esto ya no es cierto, pero sólo hay dos posibilidades. En efecto, sea g la inversión respecto de la recta que pasa por los tres puntos fijos si son colineales o respecto de la circunferencia que pasa por ellos si no lo son. Entonces fg es una homografía que fija a los tres puntos, luego es la identidad, luego $f = g$. Así hemos probado el teorema siguiente:

Teorema A.13 *Si una transformación circular de \mathbb{C}^∞ fija a tres puntos (finitos o no), entonces es la identidad o una inversión.*

Teorema A.14 *Dadas dos circunferencias en \mathbb{C}^∞ , existe una homografía (en particular, una transformación de Möbius) que transforma una en otra.*

DEMOSTRACIÓN: Dadas dos circunferencias C y C' , basta considerar una homografía f que transforme tres puntos de C en tres puntos de C' . Entonces $f[C]$ es una circunferencia con tres puntos en común con C' , luego $f[C] = C'$. ■

El teorema siguiente ilustra una técnica muy útil, consistente en convertir en ∞ un punto finito dado o convertir ∞ en un punto finito, de modo que determinadas circunferencias se conviertan en rectas o viceversa:

Teorema A.15 *Dos circunferencias distintas tienen cero, uno o dos puntos en común. Según el caso se llaman disjuntas, tangentes o secantes. Dado un punto A en una circunferencia C y un punto B que no esté en C , existe una única circunferencia C' que pasa por B y es tangente a C en A .*

DEMOSTRACIÓN: Aplicando una transformación circular si es preciso, podemos suponer que una de las circunferencias pasa por ∞ y la otra no, con lo que tenemos una recta y una circunferencia (en el sentido usual), luego efectivamente se cortan en a lo sumo dos puntos.

Respecto a la segunda parte, aplicando una transformación circular podemos suponer que A es el punto infinito, con lo que el resultado equivale a que por un punto exterior a una recta pasa una única paralela. ■

En lo sucesivo convendremos que la tangente a una circunferencia por dos de sus puntos es ella misma, y así, si P es un punto de una circunferencia C y Q es otro punto cualquiera (en C o fuera de C), existe una única circunferencia

tangente a C que pasa por P y Q . En estos términos la recta tangente a una circunferencia por uno de sus puntos es la circunferencia tangente por dicho punto y por ∞ .

Dos rectas secantes forman dos pares de ángulos opuestos por el vértice. Llamaremos *ángulo* entre ambas a la medida del menor de ellos. El *ángulo* entre una circunferencia y una recta secante es el ángulo entre la recta y la recta tangente por cualquiera de los puntos de corte. No importa la tangente que se escoja, pues la reflexión respecto a la perpendicular a la recta por el punto medio de los puntos de corte deja invariante a la recta y transforma una tangente en otra. El *ángulo* entre dos circunferencias (usuales) secantes es el ángulo entre sus tangentes por los puntos de corte. También es claro que no importa el punto de corte elegido.

Con esto tenemos definido el ángulo entre dos circunferencias secantes cualesquiera. La definición puede resumirse así: el ángulo entre dos circunferencias es el ángulo entre sus rectas tangentes por ∞ en uno de los puntos de corte.

Teorema A.16 *Las transformaciones circulares conservan los ángulos entre circunferencias.*

DEMOSTRACIÓN: Sean C_1 y C_2 dos circunferencias secantes y sea P un punto de corte (que podemos suponer finito). Sea t_i la tangente a C_i por P o bien $t_i = C_i$ si C_i es una recta. Entonces el ángulo entre C_1 y C_2 es por definición el ángulo entre t_1 y t_2 . Una inversión respecto de una circunferencia (usual) de centro P deja invariantes a t_1 y t_2 y transforma cada circunferencia C_i en una recta sin más punto en común con t_i que ∞ (o bien $t_i = C_i$). En cualquier caso transforma a C_i en una recta paralela a t_i , de modo que estas rectas forman entre sí el mismo ángulo que las circunferencias de partida.

Sean ahora C'_i las imágenes de cada C_i por una transformación circular arbitraria. Del mismo modo podemos transformarlas en dos rectas que forman el mismo ángulo que ellas.

El teorema se reduce, pues, a probar que si una transformación circular hace corresponder dos pares de rectas secantes, entonces el ángulo que forman es el mismo.

Sea f una transformación circular que haga corresponder dos rectas secantes r_1 y r_2 con dos rectas secantes r'_1 y r'_2 . Sea O el punto de corte de las primeras y O' el punto de corte de las segundas. Entonces o bien $f(O) = O'$ y $f(\infty) = \infty$ o bien $f(O) = \infty$ y $f(\infty) = O'$. En el segundo caso podemos componer f con una inversión respecto a una circunferencia de centro O' , lo cual deja inalteradas a r'_1 y r'_2 pero hace que ∞ quede fijo. Por lo tanto los dos pares de rectas se corresponden por una semejanza, y las semejanzas conservan los ángulos. ■

Ejercicio: Probar que si dos circunferencias son ortogonales, cada una de ellas queda fija por la inversión respecto de la otra.

Teorema A.17 *Sean C_1 y C_2 dos circunferencias secantes. Sea P un punto de corte y Q un punto cualquiera. Sea C'_1 la tangente a C_1 por P y Q . Entonces el ángulo entre C_1 y C_2 es el mismo que entre C'_1 y C'_2 .*

DEMOSTRACIÓN: Aplicamos una transformación circular f que convierta Q en ∞ . Basta probar que $f[C_1]$ y $f[C_2]$ forman el mismo ángulo que $f[C'_1]$ y $f[C'_2]$, pero esto es cierto por definición de ángulo entre circunferencias. ■

Ahora probamos una caracterización de la separación armónica 8.59:

Teorema A.18 Sean P, Q, R y S cuatro puntos colineales y sea C una circunferencia que tenga a PQ por diámetro. Entonces $\mathcal{H}(P, Q; R, S)$ si y sólo si toda circunferencia que pasa por R y S es ortogonal a C .

DEMOSTRACIÓN: Supongamos $\mathcal{H}(P, Q; R, S)$. Aplicando una inversión podemos suponer que P es el punto infinito. Entonces Q es el punto medio de R y S y la circunferencia C es la recta perpendicular a RS por Q . En estas condiciones es fácil ver que toda circunferencia que pasa por R y S es ortogonal a C .

Supongamos ahora que se cumple esta propiedad. Tomemos una circunferencia C' cualquiera que pase por R y S . Por hipótesis corta ortogonalmente a C en dos puntos A y B . Consideremos ahora la circunferencia C'' que pasa por A, R y el conjugado armónico de R respecto de P y Q . Por la parte anterior C' también es ortogonal a C , ahora bien, por dos puntos A y R pasa una única circunferencia ortogonal a C , luego $C' = C''$, de donde S ha de ser el conjugado armónico de R respecto de P y Q . ■

Ahora vamos a definir el ángulo entre dos circunferencias en una esfera y probaremos que la proyección estereográfica conserva los ángulos.

Definición A.19 Si dos circunferencias en una esfera S se cortan en un punto P , definimos el *ángulo* que forman como el ángulo entre sus rectas tangentes por P .

El teorema siguiente prueba en particular que el ángulo entre dos circunferencias secantes en una esfera no depende de cuál de los dos puntos de corte se toma para calcularlo.

Teorema A.20 La proyección estereográfica conserva ángulos.

DEMOSTRACIÓN: Observemos en primer lugar que si C es una circunferencia en una esfera S y C' es una circunferencia tangente a C por un punto P , entonces ambas tienen la misma recta tangente por P , concretamente, la intersección de los planos de C y C' . Por lo tanto, si tenemos dos circunferencias C_1 y C_2 y C'_1, C'_2 son sus tangentes por P y por otro punto Q , el ángulo entre C_1 y C_2 es el mismo que el ángulo entre C'_1 y C'_2 . Por el teorema A.17, lo mismo vale para las proyecciones, luego basta probar el teorema para circunferencias que se corten en un par de puntos antípodos P y Q .

Sean p_1 y p_2 las proyecciones estereográficas de una esfera en dos planos π_1 y π_2 . Sea $f: \pi_1 \rightarrow \pi_2$ una isometría entre ellos. Entonces $p_2^{-1}p_1f$ es una biyección de π_2 en sí mismo que transforma circunferencias en circunferencias, luego es una transformación circular, luego conserva ángulos, luego $p_2^{-1}p_1$ también conserva ángulos.

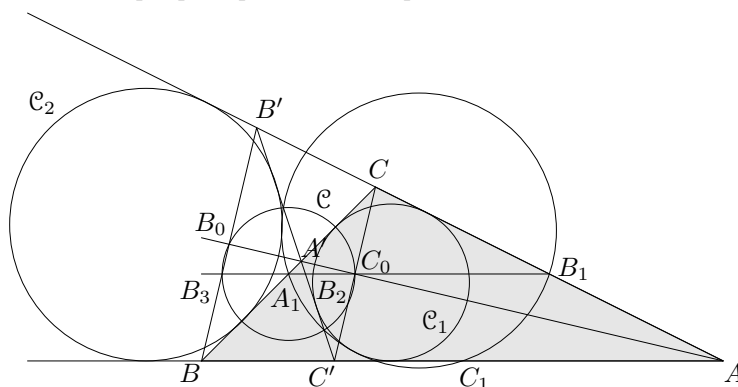
Esto implica que al proyectar dos circunferencias de una esfera en un plano obtenemos dos circunferencias que forman el mismo ángulo independientemente del plano sobre el que proyectemos. En particular, podemos utilizar la proyección respecto al plano perpendicular a la recta PQ .

En estas condiciones es fácil ver que las rectas tangentes a C_1 y C_2 tanto por P como por Q son paralelas a las proyecciones de C_1 y C_2 , luego los ángulos son los mismos. ■

Para terminar mostramos una aplicación no trivial de la geometría circular a la geometría euclídea [IGE 5.56]:

Teorema A.21 (Teorema de Feuerbach) *El círculo de los nueve puntos de un triángulo es tangente a sus cuatro circunferencias tritangentes.*

DEMOSTRACIÓN: Consideremos un triángulo \widehat{ABC} . Sean A_1 , B_1 y C_1 los puntos medios de sus lados, etiquetados como indica la figura. El círculo de los nueve puntos es el que pasa por estos tres puntos medios.



Consideremos la circunferencia inscrita C_1 y una de las circunferencias circunscritas C_2 , concretamente la que tiene su centro en la bisectriz de \hat{A} . Si el lado BC es perpendicular a esta bisectriz, entonces es fácil ver que el triángulo es isósceles, ambas circunferencias tocan a BC en su punto medio A_1 y el círculo de los nueve puntos tiene también a BC por tangente, con lo que efectivamente toca a C_1 y C_2 .

Supongamos que BC no es perpendicular a la bisectriz de \hat{A} . Entonces la reflexión respecto a ella fija a C_1 y C_2 , luego transforma a BC en otra recta tangente a ambas circunferencias. Digamos que ésta corta a los lados del triángulo en los puntos A' , B' y C' . El punto A' está sobre la bisectriz de \hat{A} .

Las rectas AB , AC , BC y $B'C'$ son los lados de un cuadrilátero completo dual circunscrito tanto a C_1 como a C_2 , luego su triángulo diagonal es autopolar para ambas (el polo de cada lado es la intersección de los otros dos). Estos lados son las rectas AA' , BB' , CC' . Ahora bien, como el lado AA' pasa por los centros de las circunferencias, su polo es infinito, luego las rectas BB' y CC' son paralelas y perpendiculares a AA' . Sean B_0 y C_0 los puntos de corte, es decir, los vértices finitos del triángulo. Notemos que el cuadrilátero completo de vértices B , B' , C , C' prueba que $\mathcal{H}(B_0, C_0; A, A')$.

Teniendo en cuenta que A_1 es el punto medio de \overline{BC} , es fácil ver que B_0 y C_0 están a la misma distancia de A_1 (si A'_1 es su imagen por la simetría respecto de BC , entonces $A_1A'_1$ es la mediatriz de B_0C_0), luego podemos trazar una circunferencia \mathcal{C} de centro A_1 que pase por B_0 y C_0 .

Los puntos B_0 y C_0 son conjugados respecto de las circunferencias \mathcal{C}_1 y \mathcal{C}_2 y son colineales con sus centros, luego están armónicamente separados por los dos pares de puntos donde éstas cortan a AA' . Podemos aplicar el teorema A.18 y concluir que \mathcal{C} es ortogonal tanto a \mathcal{C}_1 como a \mathcal{C}_2 . Por lo tanto la inversión respecto a \mathcal{C} deja fijas a cada una de ellas. Para probar el teorema bastará ver que esta inversión transforma el círculo de los nueve puntos en la recta $B'C'$, pues ésta es tangente a las circunferencias \mathcal{C}_i , luego el círculo de los nueve puntos también lo será.

La recta A_1B_1 es paralela a AB , luego no es paralela a $B'C'$. Sea B_2 el punto de corte. Notemos que como AA' es la bisectriz de \hat{A} , el punto C_0 es el punto medio del segmento $\overline{CC'}$. La recta A_1B_1 pasa por el punto medio de \overline{CB} y es paralela a BC' , luego contiene a C_0 . Por otra parte, del hecho de que BB' y CC' sean paralelas junto con que A_1 es el punto medio de \overline{BC} se sigue que A_1 es el punto medio del segmento que une C_0 con el punto B_3 en que A_1B_1 corta a BB' . Por consiguiente B_3 también está en \mathcal{C} . Ahora observamos que la proyección perspectiva desde B' hace corresponder $B_0 \mapsto B_3$, $C_0 \mapsto C_0$, $A \mapsto B_1$, $A' \mapsto B_2$, y como $\mathcal{H}(B_0, C_0; A, A')$, también $\mathcal{H}(B_3, C_0; B_1, B_2)$, luego la involución que \mathcal{C} induce en la recta A_1B_1 por inversión (que fija a C_0 y B_3) conjugua B_1 con B_2 , es decir, transforma B_1 en un punto de la recta $B'C'$. Simétricamente se prueba que el inverso de C_1 está también sobre $B'C'$ y ciertamente el inverso de A_1 es el punto infinito, luego el inverso del círculo de los nueve puntos es la recta $B'C'$. ■

Apéndice B

El código de Hamming

Vamos a ver aquí un ejemplo sencillo que ilustra el interés de las geometrías finitas. Se trata del código de corrección de errores de Hamming del que hemos hablado en la introducción. Aunque vamos a tratar con un único ejemplo concreto, merece la pena introducir algunos conceptos generales de la teoría de códigos.

En general, la teoría de códigos considera un *alfabeto*, que no es sino conjunto finito k , pero en la teoría algebraica de códigos se toma concretamente como k un cuerpo finito, y nosotros vamos a considerar únicamente el caso del cuerpo $k = \{0, 1\}$ de dos elementos. Así $V = k^n$ es un k -espacio vectorial, a cuyos elementos los llamaremos *palabras* de longitud n .

En el caso binario (en el que k es el cuerpo de 2 elementos) resulta útil tener en cuenta que las palabras pueden reinterpretarse como subconjuntos de un conjunto dado $X = \{P_1, \dots, P_n\}$ de n elementos. Concretamente, cada palabra x se corresponde con el conjunto A_x formado por los puntos P_i tales que $x(i) = 1$.

Así, la suma en el espacio $V = k^n$ se corresponde con la *diferencia simétrica* de conjuntos:

$$A \triangle B = (A \cup B) \setminus (A \cap B).$$

En efecto, basta tener en cuenta que $(x+y)_i = 1$ si y sólo si $x_i = 1$ o $y_i = 1$, pero no se dan los dos casos a la vez, luego $P_i \in A_{x+y}$ si y sólo si P_i pertenece a uno solo de los conjuntos A_x, A_y .

Esto se traduce en que la diferencia simétrica dota al conjunto $\mathcal{P}X$ de estructura de grupo abeliano.

Llamaremos $0, 1 \in V$ a las palabras constantes, de modo que

$$x + 0 = x, \quad x + x = 0$$

y $1+x$ es la palabra que resulta de intercambiar los ceros y los unos de x , a la que nos referiremos como *palabra complementaria* de x . Claramente $A_{1+x} = X \setminus A_x$.

En V podemos considerar la *distancia de Hamming*, que es la aplicación

$$d : V \times V \longrightarrow \mathbb{N}$$

para la cual $d(x, y)$ es el número de componentes en que difieren x, y .

Es fácil ver que se trata de una distancia en el sentido de [An 2.4]. La desigualdad triangular $d(x, z) \leq d(x, y) + d(y, z)$ se cumple porque si x, z difieren en la coordenada i -ésima, necesariamente lo hacen también x, y o bien y, z .

Notemos también que

$$d(x, y) = d(x + z, y + z),$$

y en particular se cumple que

$$d(x, y) = d(x + y, 0).$$

El *peso* de una palabra se define como $d(x, 0)$, es decir, como el número de unos que contiene, por lo que la relación precedente expresa que la distancia entre dos palabras es el peso de su suma.

Un *código lineal (binario) de corrección de errores* de tipo $[n, l, d]$ es un subespacio vectorial $C \leq k^n$ de dimensión l con la propiedad de que si $x, y \in C$ son palabras distintas, entonces $d(x, y) \geq d$.

El hecho de que C sea un subespacio vectorial se traduce en que la última condición equivale a que el peso de todas las palabras no nulas de C sea $\geq d$.

Un código de tipo $[n, l, d]$ consta de 2^l palabras, que son los mensajes posibles, de modo que si un mensaje $x \in C$ se corrompe con un máximo de $d - 1$ errores, el mensaje recibido x' no podrá ser un elemento de C , con lo que se detectará el error. Más aún, si el número de errores no es mayor que $e = E[(d - 1)/2]$, entonces el error no sólo podrá detectarse, sino también corregirse, pues el mensaje original x será la única palabra del código que distará a lo sumo e unidades del mensaje recibido x' , ya que si hubiera otra y , tendríamos que

$$d(x, y) \leq d(x, x') + d(x', y) \leq 2e \leq d - 1,$$

en contradicción con la definición de código.

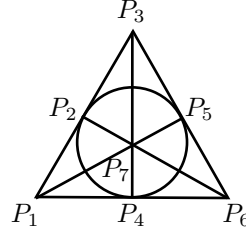
Un código es *perfecto* si toda palabra dista a lo sumo e unidades de una palabra del código, lo que significa que es imposible añadirle palabras al código sin perder su capacidad de corregir e errores.

Vamos a construir un código perfecto de tipo $[7, 4, 2]$, y para ello nos basamos en el plano de Fano F que hemos descrito en la sección 8.8.

Para ello identificamos las palabras de k^7 con los subconjuntos de F y definimos el *código de Hamming* como el subespacio vectorial \mathcal{C} generado por las siete rectas proyectivas. Vamos a probar que es, en efecto, un código binario de tipo $[7, 4, 2]$.

La figura muestra las rectas del plano de Fano y de ella podemos extraer sus palabras correspondientes, que son las filas de la matriz M indicada:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$



Estas filas son un sistema generador del código \mathcal{C} . Si consideramos las tres rectas que pasan por un mismo punto, por ejemplo, P_1 , tenemos que P_1 aparece tres veces, mientras que los demás puntos aparecen necesariamente una vez cada uno, por lo que la suma de las tres filas tiene que ser

$$1 = M_1 + M_4 + M_5 \in \mathcal{C}.$$

Esto implica que el complementario de una palabra de \mathcal{C} está en \mathcal{C} , y esto nos da otras 7 palabras (todas nuevas, pues los 7 generadores tienen 3 unos cada uno, y sus complementarios son palabras con 4 unos cada una), y si añadimos la palabra nula 0 tenemos un total de 16 palabras. Vamos a razonar que estas 16 palabras ya forman un subespacio vectorial de V , por lo que son el código \mathcal{C} completo.

En efecto, basta probar que al sumar dos cualesquiera de estas 16 palabras obtenemos nuevamente una de ellas. Esto es obvio si uno de los sumandos es 0 o 1, y los demás casos posibles son:

1. Que dos palabras correspondan a dos rectas L y L' de F , en cuyo caso $L + L' = 1 + L''$, donde L'' es la tercera recta que pasa por $L \cap L''$.
2. Que las palabras correspondan a una recta L y al complementario de otra $1 + L'$, en cuyo caso $L + 1 + L' = 1 + 1 + L'' = L''$, que es otra recta.
3. Que las palabras correspondan a dos complementarios de rectas, $1 + L$ y $1 + L'$, en cuyo caso $1 + L + 1 + L' = L + L'$, que es el complementario de una recta.

Así pues, el código de Hamming consta de 16 palabras, luego tiene dimensión 4. Ahora es inmediato que se trata de un código de tipo $[7, 4, 3]$, pues las palabras correspondientes a rectas tienen peso 3, los de los complementarios de rectas tienen peso 4 y la palabra 1 tiene peso 7.

Según hemos visto, esto implica que ninguna palabra de V dista a lo sumo $e = 1$ unidades de dos palabras de \mathcal{C} , luego cada palabra de \mathcal{C} tiene 8 palabras (distintas) de V a una distancia ≤ 1 (contándola a ella misma), luego la unión de todas las palabras que distan a lo sumo 1 unidad de una palabra de \mathcal{C} tiene $2^4 \cdot 8 = 2^7$ palabras, que son todas las palabras de V . Esto prueba que el código es perfecto.

Para operar con él consideramos la matriz generadora

$$G = \left(\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array} \right),$$

que está formada por las tres primeras filas de M y la complementaria de la cuarta. No es casual que G contenga la matriz identidad. Ello se debe a que hemos numerado los puntos del plano F de modo que P_1, P_2, P_3 forman una recta M_4 (la correspondiente a la cuarta fila de M). Esto hace que siempre podemos tomar una recta M_i que pase por P_i , para $i = 1, 2, 3$, que no sea M_4 y que no pase por P_4 (hay tres rectas que pasan por P_i , que son M_4 , la que pasa por P_4 y otra, que es la que tomamos como M_i). Así tenemos las cuatro primeras filas de la matriz M , con las tres primeras formando la matriz identidad y la cuarta es el complementario del vector necesario para completar la identidad. Por lo tanto, el código \mathcal{C} tenía que contener necesariamente cuatro palabras que formen una matriz G con la matriz identidad en cuatro de sus columnas (que hemos hecho que sean las primeras numerando adecuadamente los puntos y las rectas del plano). Claramente, las filas de G son linealmente independientes, luego son una base de \mathcal{C} .

En la práctica, esto se traduce en que cada mensaje formado por una sucesión de cuatro ceros y unos se prolonga hasta una única palabra del código de Hamming, pues basta sumar las filas de la matriz G correspondientes a las posiciones de los unos. Por ejemplo, la codificación del mensaje 0110 es

$$G_2 + G_3 = 0110|101.$$

Más prácticamente, la codificación del mensaje $m \in k^4$ es $x = mG$.

Para decodificar el mensaje basta eliminar los tres últimos dígitos. Hemos probado que cualquier error de 1 dígito puede ser corregido, pues sólo hay una palabra del código a distancia ≤ 1 de una palabra dada. Vamos a ver cómo puede detectarse un error en la práctica. Para ello conviene llamar

$$d_1 d_2 d_3 d_4 | p_1 p_2 p_3$$

a los dígitos de una palabra, de modo que los cuatro primeros son los datos y tres últimos son “comprobaciones de paridad”. En efecto, observando la matriz generadora G es fácil convencerse de que, dados unos datos $d_1 d_2 d_3 d_4$, el dígito p_1 se elige de modo que en las posiciones d_1, d_3, d_4, p_1 haya un número par de unos, y análogamente sucede con los otros dos, de acuerdo con la tabla siguiente:

	d_1	d_2	d_3	d_4
p_1	*		*	*
p_2		*	*	*
p_3	*	*		*

Por lo tanto, para comprobar si una palabra recibida $x \in V$ es correcta (supuesto que no haya sufrido más de un error) basta calcular Hx^t , donde H es la matriz de paridades:

$$H = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Si $x \in \mathcal{C}$ se cumplirá que $Hx^t = 0$. Más aún, si vemos esta igualdad como un sistema de 3 ecuaciones con 7 incógnitas, sus soluciones son un subespacio vectorial de dimensión 4 que contiene a \mathcal{C} , luego podemos asegurar que una palabra x está en \mathcal{C} si y sólo si $Hx^t = 0$.

Si en la transmisión se ha producido un único error, por ejemplo en la posición d_1 , esto alterará el producto Hx^t en las posiciones primera y tercera, luego obtendremos el vector $(1, 0, 1)$, que es la primera columna de H . En general, si el resultado de Hx^t no es el vector nulo, será una de las columnas de H , precisamente la del dígito erróneo.

Por ejemplo, si queremos transmitir el mensaje 0110, ya hemos visto que su codificación es 0110|101 y si recibimos este mensaje podemos comprobar que

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

En cambio, si el mensaje se corrompe y se recibe 0100|101, el resultado será:

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

que es la tercera columna de H , lo que indica que el tercer dígito del mensaje es incorrecto y debe cambiarse para subsanar el error.

En cambio, si el mensaje ha sufrido dos errores, la palabra recibida x no estará en \mathcal{C} (porque está a distancia 2 de una palabra de \mathcal{C}) y el cálculo Hx^t no dará 0, con lo que detectaremos el error, pero no podemos corregirlo. Por ejemplo, si nos llega el mensaje 0100|111, la comprobación nos da

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

que es la suma de las columnas tercera y sexta, correspondientes a los dos errores, y así sabemos que ha habido un error, pero si pensamos que el error está en el quinto dígito y rectificamos el mensaje a 0100|011, con lo que decodificamos 0100, no habremos acertado.

Notemos que no podemos saber si se han producido uno o dos errores, con lo que no podemos saber si nuestra corrección es correcta. Sin embargo, es posible extender el código con un dígito más tal que, en caso de que se hayan producido dos errores, nos avise de que no es posible corregir el mensaje. Para ello consideramos palabras de 8 dígitos y extendemos cada palabra de \mathcal{C} con el valor necesario para que tenga en total un número par de unos. Tenemos así el *código de Hamming extendido*, que claramente es de tipo $[9, 4, 4]$.

En la práctica, esto supone considerar las matrices

$$G = \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array} \right), \quad H = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Al calcular Hx^t , las tres primeras filas no tienen en cuenta el octavo dígito, por lo que darán el mismo resultado de antes tanto si éste se ha corrompido o no, es decir, darán 0 si y sólo si el mensaje sin el último dígito está en \mathcal{C} . Si el último dígito también es correcto, tendremos $Hx^t = 0$.

Si sólo ha habido un error en los 7 primeros dígitos, en la cuarta fila obtendremos un 1, luego seguiremos obteniendo la columna de H que indica el dígito incorrecto, y si el dígito corrupto es el octavo, igualmente obtendremos $(0, 0, 0, 1)$, que es la octava columna, luego en principio el código extendido funciona igual que el original.

Pero la ventaja es que si se producen dos errores, al multiplicar por H obtendremos un resultado de la forma $(x, y, z, 0)$, con (x, y, z) no nulo, y el resultado no será ninguna columna de H , lo que nos informa de que hay un error, pero que no es posible corregirlo (así, a diferencia de antes, ahora sabemos que estamos en ese caso y no daremos por buena una corrección falsa).

Si renunciamos a corregir los errores de un dígito, podemos detectar también errores de tres dígitos, pues dan lugar a vectores de comprobación Hx^t no nulos.

En efecto, si los tres dígitos erróneos no incluyen al último, el cuarto dígito de la comprobación será no nulo, mientras que si uno de los dígitos corruptos es el último, los tres primeros no serán todos nulos. Pero no podemos distinguir si una comprobación de la forma $(x, y, z, 0)$ corresponde a un error de un dígito (corregible) o a un error de tres dígitos (no corregible).

Índice de Materias

- adyacentes (ángulos), 22
- afínmente independientes, 96
- afín
 - aplicación, 101
 - espacio, 95
 - geometría, 243
- afinidad, 101
- agudo (ángulo), 34
- amplitud, 63, 130
 - de un arco, 66
- ángulo, 21
 - agudo, 34
 - de un triángulo, 24
 - entre circunferencias, 487, 488
 - entre dos vectores, 107
 - generalizado, 133
 - llano, 31
 - obtuso, 34
 - recto, 34
 - suplementario, 30
- ángulos
 - adyacentes, 22
 - alternos, 74
 - internos/externos, 74
 - opuestos por el vértice, 22
- antisimétrica (forma), 312
- aplicación afín, 101
- arco, 65
- arguesiano (plano), 262
- argumento, 122
- Arquímedes (propiedad de), 48
- asíntota, 362
- axioma
 - de las circunferencias, 41
 - de las paralelas, 3
- axiomas de tridimensionalidad, 20
- baricentro, 106
- Bessel (fórmulas), 462
- bilátero, 457
- bisectriz, 33
- cateto, 34
- Cauchy-Schwarz (desigualdad de), 112
- Cayley (fórmula), 398
- centro, 362
 - de perspectiva, 315
 - de una homografía en una cónica, 379
 - de una semejanza, 173
- circular (punto), 390
- círculo, 37
- circuncentro, 84
- circunferencia, 37, 389
 - circunscrita, 84
 - de los nueve puntos, 166
 - graduada, 122
- circunradio, 84
- colineación, 167, 298
- colineales (puntos), 2, 13
- compleción proyectiva, 278
- complemento ortogonal, 115
- concoide de Nicomedes, 222
- congruencia, 24, 25, 116, 453
 - de matrices, 179
- cónica, 343
 - no trivial, 354
 - real/imaginaria, 348
- conjugación
 - de cuaternios, 146
 - harmónica, 324
 - respecto a una involución, 324
 - respecto de una polaridad, 385
- constructible

- con regla, 209
 - a partir de una circunferencia, 214
 - con regla marcada, 220
 - con regla y compás, 193
 - con regla y transportador, 233
- convexo, 8
- coordenadas
 - baricéntricas, 107
 - cartesianas, 103
 - polares, 123, 134
- coplanares (puntos), 14
- correlación, 309
 - proyectiva, 309
 - simétrica, 312
- coseno, 82
- cruce (de rectas), 3
- cuádrica, 343
 - degenerada, 345
 - esférica, 387
 - reglada, 352
- cuadrado, 77
- cuadrangular (conjunto), 326
- cuadrilátero completo, 322
- cuaternios, 144
- cuerda, 37, 65

- diámetro, 37
 - de una cónica, 400
- dirección, 252
- distancia, 57, 109
 - elíptica, 456
 - hiperbólica, 423
- división de segmentos (propiedad de), 46
- dualidad, 313

- eje
 - de perspectiva, 315
 - de una cónica, 400
 - de una homografía en una cónica, 379
- elación, 319
- elipse, 360
- entre, 30
- equilátero (triángulo), 28

- equipolentes (vectores), 90
- escaleno, 28
- esférica (transformación), 482
- esfera, 389
 - de Riemann, 142
- espacio, 2
 - afín, 95
 - euclídeo, 109
 - orientado, 186
 - proyectivo, 303
 - director, 96
 - euclídeo, 109, 385
 - parabólico, 385
 - proyectivo, 280, 282, 287
- euclídeo
 - cuerpo, 119
 - espacio, 109, 385
 - espacio afín, 109
- excentricidad, 407
- exterior
 - a una cuádrica, 350
 - punto, 2

- Fano
 - axioma de, 323
 - plano de, 334
- foco, 400
- forma
 - bilineal
 - antisimétrica, 312
 - ortosimétrica, 312
 - simétrica, 312
 - canónica, 182
- frontera
 - de un ángulo, 21
 - de un círculo, 37
 - de un semiplano, 9
 - de un triángulo, 24

- geometría
 - afín, 243
 - de Hilbert, 2
 - euclídea, 73
 - plana, 13
 - tridimensional, 20
 - métrica, 24

- giro, 124, 152, 175, 436
 - infinito, 437
- grupo
 - afín, 157
 - de traslaciones, 158
 - elíptico, 453
 - hiperbólico, 418
 - lineal, 157
 - parabólico, 390
 - proyectivo, 289
- haz de rectas, 252
- hexágono, 262
- hipérbola, 360
- hiperesfera, 478
- hiperplano, 96
 - tangente, 350
- hipotenusa, 34
- homografía, 289
 - elíptica, parabólica, hiperbólica, 330
 - en una esfera, 413
 - entre haces de rectas, 368
 - entre recta y cónica, 376
 - entre recta y haz, 369
- homología, 319
- homotecia, 159, 250
 - lineal, 159, 254
- imaginaria (cónica), 348
- independencia afín, 14
- índice, 348
- ínfimo, 51
- inversión, 415, 481
- inversos (puntos), 397
- involución, 324
 - ortogonal, 385
- isósceles (triángulo), 28
- isótropa (recta), 385
- isometría, 113, 392, 420, 453
 - directa, inversa, 184
- Klein (plano de), 419
- Lagrange (fórmula de), 150
- llano (ángulo), 31
- longitud, 54
- mediatriz, 35
- Möbius (transformación de), 485
- movimiento, 184
- norma, 89, 109
- número real, 50
- obtuso (ángulo), 34
- ordenación débil, 268
- orientación, 87, 186
- origen (de una semirrecta), 8
- ortogonal
 - grupo, 171
 - matriz, 171
- ortogonalidad, 108, 109
- ortonormal, 109
- ortosimétrica (forma), 312
- osculante, 437
- parábola, 360
- parabólico (espacio), 385
- paralelismo, 20, 97, 418
- paralelogramo, 76
- perpendicular, 34, 36, 418, 453
- perspectividad, 319
- pie (de una perpendicular), 35, 37
- Pierpont (primo de), 226
- pitagórico (cuerpo), 82
- plano, 2, 96
 - elíptico, 453
 - hiperbólico, 417
 - proyectivo, 277
- Poincaré (plano de), 434
- polaridad, 313
 - de una cónica, 351
 - ortogonal, 385
- principio de dualidad, 310
- producto escalar, 108, 109
- proporción, 49
- proyección
 - paralela, 160, 267
 - perspectiva, 319, 371
- proyectivamente independiente, 290
- proyectivo (espacio), 287
- punto, 2, 95
 - finito, 278
 - infinito, 278, 418

- medio, 32, 176
 - ultrafinito, 418
- raíz de la unidad (primitiva), 138
- razón doble, 300, 311
 - entre rectas, 398
- real (cónica), 348
- rectángulo, 77
- recta, 2, 96
 - de Euler, 165
 - graduada, 47
 - orientada, 87
- reflexión, 176, 391, 419
- rombo, 77
- secante (a una cuádrica), 350
- secantes
 - planos, 20
 - rectas, 3
- segmento, 5, 327
- semejantes (triángulos), 77
- semejanza, 171, 390
 - lineal, 172
- semicircunferencia, 65
- semihaz de semirrectas, 30
- semihomografía, 298
- semilineal (aplicación), 170
- semiplano, 9
- semirrecta, 8
- seno, 82
- separación, 329
 - armónica, 323
- simétrica (forma), 312
- simetría, 176
 - axial, 124
 - puntual, 126
 - puntual,axial,especular, 176
- sistema de referencia, 102
 - canónico, 105
 - proyectivo, 291
- soporte
 - de un ángulo, 22
 - de un círculo, 37
- suma
 - de ángulos, 31
 - de segmentos, 25
- ortogonal, 115
 - suplementarias (variedades), 160
 - suplementarios (ángulos), 30
 - supremo, 51
- Tales (posición de), 77
- tangente
 - a una cuádrica, 350
 - recta, circunferencia, 38
 - trigonométrica, 83
- Teorema
 - de Brianchon, 374
 - de Desargues, 163, 259, 285
 - de Feuerbach, 489
 - de Gram-Schmidt, 110
 - de los senos, 84, 129
 - de Papos-Pascal, 262
 - de Pascal, 372
 - de Pasch, 4
 - de Pitágoras, 80
 - de Poncelet-Steiner, 216
 - de Steiner, 371
 - de Tales, 79, 101, 161
 - del coseno, 84, 430
 - fund. de la geom. afín, 170
 - fund. de la geom. proyectiva, 298
- traslación, 158, 250, 438
- traza, 252
- triángulo, 24, 284
 - elíptico, 457
 - equilátero, 28
 - escaleno, 28
 - isósceles, 28
 - medial, 164
 - polar, 459
- ultraparalelismo, 418
- variedad
 - afín, 244
 - lineal, 96
 - proyectiva, 287
- vector
 - de posición, 103
 - director, 91
 - fijo, 89
 - libre, 90

vectores ortogonales, 108, 109

vértice

de un ángulo, 21

de un triángulo, 24

de una cónica, 400