
Una demostración geométrica del teorema de Perron-Frobenius

escrito por

GEMA MARÍA BAENA MARZO

Tutor: Roberto Canogar



Facultad de Ciencias
UNIVERSIDAD DE EDUCACIÓN A DISTANCIA

Trabajo presentado para la obtención del título de
Master Universitario en Matemáticas Avanzadas de la UNED.
Especialidad Geometría y Topología

SEPTIEMBRE 2015

ABSTRACT

Abstract en español:

En este documento se pretende desarrollar con el máximo detalle y el máximo rigor matemático la demostración geométrica del Teorema de Perron-Frobenius publicada por los matemáticos D.Alberto Borobia y D. Ujué R. Trías, en el año 1992, en la Revista Matemática de Universidad Complutense de Madrid (Vol.5, nº 1).

Abstract in English:

In this document we try to explain in detail and mathematical accuracy the geometric proof of Perron-Frobenius theorem published by D.Alberto Borobia and D. Ujué R. Trías, in 1992, in the Mathematic Magazine of the Complutense University of Madrid (Vol.5, nº 1).

DEDICATORIA Y AGRADECIMIENTOS

A Jose Carlos Cobo.

Teniendo en cuenta que muchas de las miles de horas que se necesitan se las he tenido que robar a él para que se quede con nuestro pequeño, considero que es un gran merecedor de esta dedicación.

A gradecimientos a Roberto Canogar, tutor y director este trabajo, por la paciencia y dedicación que ha tenido conmigo.

Es digna de admirar la gran capacidad de comunicación que tiene, la elegancia con la que ha resuelto todas mis dudas. Cuando yo daba una demostración por imposible, con sólo tres frases suyas quedaba resuelta.

Muchas gracias y enhorabuena por su trabajo.

TABLA DE CONTENIDOS

	Página
1 Introducción	1
1.1. Aplicaciones del Teorema de Perron-Frobenius	1
1.2. Breve biografía de Perron y de Frobenius	3
2 Teorema de Perron-Frobenius	5
2.1. Demostración del Teorema de Perron-Frobenius	6
3 Forma de Frobenius para matrices imprimitivas	17
3.1. Demostración de la existencia de la Forma de Frobenius	18
3.2. Ejemplo de cálculo de la Forma de Frobenius	31
A Apéndice	35
Conclusiones	39
Bibliografía	41

INTRODUCCIÓN

El teorema de Perron-Frobenius garantiza la existencia de un autovalor positivo, que acota en módulo el resto de autovalores de la matriz, y tiene asociado un autovector positivo. Este hecho juega un papel importante como instrumento para conseguir condiciones de existencia, unicidad, positividad y estabilidad de las soluciones en modelos lineales multisectoriales tanto estáticos como dinámicos.

En la literatura aparecen diversas pruebas alternativas a las iniciales de Perron (1907) y Frobenius (1908, 1909, 1912); entre ellas, la más conocida entre los economistas es probablemente la de Wielandt (1950) que utiliza el teorema del punto fijo de Brouwer, y que se popularizó por el trabajo de Debreu y Herstein (1953). Las pruebas de Karlin (1959) y Nikaido (1969) son elementales sin utilizar teoremas de punto fijo; Murata (1972) ofrece una prueba parecida a la inicial de Frobenius, mientras que la que aparece en el libro de Arrow y Hahn es muy similar a la de Karlin. También hay demostraciones construidas a partir de la Teoría de Juegos.

El objetivo de este trabajo es explicar con todo detalle la demostración geométrica que fue publicada, en 1992, por Alberto Borobia y Ujué R. Trías en la revista *Matemática* de la Universidad Complutense de Madrid, en el volumen 5, número 1.

En el capítulo 3 se trabaja con los conceptos de matrices primitivas e imprimitivas, se expone que las matrices primitivas recogen mucha más información que las matrices imprimitivas, y se demuestra que toda matriz imprimitiva se puede transformar en otra formada por bloques primitivos. En un último subapartado, se da un ejemplo de esta transformación.

1.1. Aplicaciones del Teorema de Perron-Frobenius

El teorema de Perron-Frobenius constituye uno de los ejemplos más significativos de la forma en la que las matemáticas, desarrolladas por motivos puramente teóricos (es decir, por

una motivación ajena a las aplicaciones), proporcionan una sólida fundamentación teórica a aplicaciones que no pertenecen al reino de la matemática pura.

Algunos ejemplos de sus múltiples aplicaciones es: el análisis de las características demográficas de una población, la distribución del poder en una red social, el desarrollo de motores de búsqueda como Google, el crecimiento de poblaciones de animales, la longitud de las colas en un supermercado, el análisis de la evolución de los precios de los bienes de consumo.

Una de las aplicaciones más celebradas del teorema de Perron-Frobenius está relacionada con la obtención de diferentes rankings, ordenaciones por importancia o clasificaciones, ya sean estas deportivas (qué jugador o equipo es el mejor), de páginas web o, incluso, de artículos, autores o revistas científicas (un autor es bueno si su trabajo es referenciado y reconocido por otros autores buenos, y lo mismo cabe decir de un artículo o, análogamente, de una revista científica).

Ejemplo: ¿Cómo determinar cuál es el mejor equipo de un campeonato en el que no se producen todos los emparejamientos posibles, es decir, en el que cada equipo no juega con todos los demás? Admitiendo, incluso, que haya equipos que se enfrenten entre sí varias veces a lo largo de un campeonato (como sucede, por ejemplo, en las ligas de baloncesto o béisbol americanas), determinar qué equipo es el mejor es una tarea complicada, pues lo «bueno» o «malo» que es un equipo no solo debe depender de los partidos que haya ganado, sino también de lo «bueno» o «malo» que sea el rival, lo que a su vez debe depender de lo «buenos» o «malos» que sean los equipos a los que ha ganado previamente.

Para ello, supondremos que la importancia a repartir entre los diferentes elementos participantes es igual a 1, y que dicha importancia se reparte entre los diferentes participantes asignándoles a cada uno un valor entre 0 y 1, de manera que el coeficiente que le corresponde a cada uno indica lo importante (o bueno) que es en relación con el resto de participantes.

De este modo, si tenemos que distribuir la importancia entre n participantes, nuestro objetivo es encontrar un vector $c = (c_1, \dots, c_n)$ tal que c_i represente la importancia del participante i , y, como consecuencia, que se satisfagan los dos requisitos siguientes:

$$0 \leq c_i \leq 1 \quad \forall i = 0, \dots, n \quad \text{y} \quad \sum_{i=1}^n c_i = 1$$

Para plantear estos problemas los vamos a representar mediante un grafo, cada elemento a ordenar en nuestro ranking va a ser un nodo, y las relaciones existentes entre los nodos las representamos por aristas dirigidas. Para nosotros va a ser fundamental la matriz de adyacencia asociada al grafo, donde elemento (i,j) es el número de aristas del nodo i al nodo j .

La importancia de cada nodo va a ser proporcional a los nodos con los que está relacionado. Sea λ la constante de proporcionalidad asociada al problema concreto, tenemos que: $c = (1/\lambda) A^t c$ lo que quiere decir que c es un autovector de A^t (traspuesta de la matriz de adyacencia de A) asociado al autovalor $1/\lambda$.

Perron, en 1907, dio la solución cuando la matriz A es positiva, pero en general estas matrices son no negativas (es decir, suelen tener ceros). Para ser positivas necesitaríamos que todos los

nodos estén comunicados entre sí y en ambos sentidos, además necesitaríamos que hubiese lazos del nodo en sí mismo. Frobenius, en 1912, extendió el resultado a matrices no-negativas siempre que satisfagan unas ciertas hipótesis adicionales, este nuevo resultado se conoce como el teorema de Perron-Frobenius.

En este trabajos analizamos detenidamente una de las muchas demostraciones que se han dado del Teorema de Perron Frobenius.

1.2. Breve biografía de Perron y de Frobenius

Perron, gran aficionado al alpinismo, fue un hombre de una vitalidad extraordinaria que continuó con su labor docente en la Universidad de Múnich hasta los ochenta años (aunque su jubilación formal tuvo lugar a los setenta y uno), siendo destacable que dieciocho de sus doscientos dieciocho artículos los publicó después de cumplir ochenta y cuatro años. Por otra parte, Ferdinand Georg Frobenius (1849-1917) a quien se recuerda por ser, junto con Cayley y Sylvester, uno de los fundadores del núcleo principal de la teoría de matrices, fue uno de los primeros que estudiaron esta teoría como una disciplina en sí misma. Al parecer, fue al leer en un volumen de 1907 de la revista *Mathematische Annalen* el artículo de Perron «Hacia la teoría de matrices» cuando Frobenius se interesó por el trabajo de este último. No es sorprendente que Frobenius, un experto en la teoría de matrices, respondiese al reto planteado por Perron de encontrar una demostración de su teorema que evitase utilizar su lema del límite. Frobenius encontró una demostración del teorema de Perron que evitaba el uso del lema del límite mencionado, demostración que publicó en 1908. Tras publicar una continuación de este trabajo sobre matrices positivas en 1909, se propuso resolver el siguiente problema: «dada una matriz no negativa A , determinar los autovalores de A para los que existen autovectores no negativos». Fue este problema el que le llevó a obtener su célebre teorema sobre matrices no negativas publicado en 1912

TEOREMA DE PERRON-FROBENIUS

Antes de enunciar el Teorema de Perron-Frobenius exponemos el tipo de matrices con las que trabajamos, y algunos conceptos previos.

Trabajamos con matrices cuadradas, endomorfismos de R^n , sin entradas negativas y que sean irreducibles.

Una matriz \mathbf{A} es **irreducible** si **no** existe ninguna matriz permutación \mathbf{P} tal que \mathbf{PAP}^t sea de la forma

$$\begin{pmatrix} M & 0 \\ P & N \end{pmatrix}$$

donde M y N son matrices cuadradas.

Existe una relación biunívoca entre las matrices $n \times n$ y los grafos dirigidos con n nudos, siendo el elemento a_{ij} la cantidad de aristas dirigidas que tienen el origen en el nodo i -ésimo y el extremo en el nodo j -ésimo. La matriz asociada a cada grafo se llama matriz de adyacencia. Además, si M es la matriz de adyacencia un grafo, M^k indica cuantos caminos de longitud k hay entre los nodos, su elemento de la posición (i,j) nos indica la cantidad de caminos de longitud k que van del nodo i -ésimo al nodo j -ésimo.

Con esta nomenclatura, una matriz es irreducible, si y sólo si, el grafo que representa es fuertemente conexo (es decir, si para todo par de vértices v_i, v_j hay algún camino de v_i a v_j y otro de v_j a v_i).

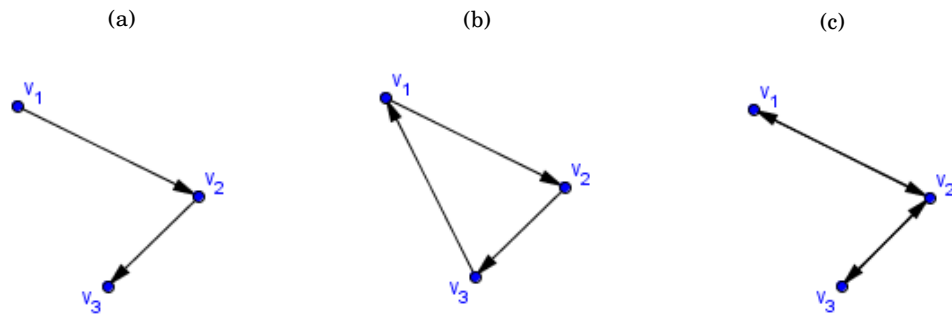


FIGURA 2.1. Irreducibilidad a través de grafos

El grafo a) no es fuertemente conexo porque por ejemplo no hay ningún camino para ir del vértice v_3 al vértice v_1 , entonces su matriz adjunta $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ es reducible. Sin embargo, tanto los grafos (b) como el (c) son fuertemente conexo, por ello sus matrices adjuntas que son respectivamente, $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ y $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ son irreducibles (Observad que para que sean reducibles los bloques de la diagonal tienen que ser cuadradas, el bloque nulo superior derecho debería ser 2×1 o 1×2 , pero no 1×1 ; no hay ninguna permutación que PAP^t lo cumpla).

Un **rayo** en R^n (en la dirección de $v \in R^n \setminus 0$) es el conjunto $r[v] = \{ \mu v \mid \mu > 0 \}$. Nosotros identificamos la colección de rayos en R^n con S^{n-1} .

Teorema I (Perron-Frobenius)

Sea A una matriz $n \times n$, no negativa e irreducible, entonces existe un autovalor positivo y simple λ de A que tiene asociado un autovector positivo (es decir, con todas sus coordenadas positivas), además el valor absoluto de este autovalor es el mayor entre los valores absolutos de todos los autovalores de A .

2.1. Demostración del Teorema de Perron-Frobenius

En este capítulo vamos a demostrar el teorema I de capítulo anterior.

Consideramos R^+ como el conjunto de rayos no-negativos, es decir, las semirectas con origen en el origen de coordenadas que están contenidas en el cuadrante C^+ , definido como:

$$C^+ = \{ (x_1, x_2, \dots, x_n) \in R^n : x_i \geq 0, i = 1, \dots, n \}$$

(i) A es una aplicación cerrada en el conjunto R^+ de rayos no-negativos, $A(R^+) \subseteq R^+$. Es decir, si A actúa sobre un rayo no negativo devuelve otro rayo no-negativo. Esto ocurre porque tanto los escalares de la matriz como las coordenadas del rayo son elementos no-negativos,

entonces, cada coordenada del rayo se puede multiplicar por un escalar o sumarle un múltiplo de otra coordenada pero nunca se va a volver negativa.

Es más, ningún rayo en R^+ se transforma en el elemento nulo.

Lo probamos por reducción al absurdo.

Supongamos que la imagen de $v=(0, \dots, x_i, \dots, 0)$ mediante \mathbf{A} es nula, $v'=\mathbf{A} \cdot v^t=(a_{1i}x_i, a_{2i}x_i, \dots, a_{ni}x_i) = (0, \dots, 0)^t$ obliga a que la columna i -ésima de \mathbf{A} sea nula, entonces \mathbf{A} es reducible (ya que esto implica que el grafo asociado a la matriz \mathbf{A} no tiene ninguna arista llegando al nodo i -ésimo, entonces desde ningún nodo se puede llegar al nodo i -ésimo, por tanto el grafo asociado no es fuertemente conexo, o lo que es lo mismo, la matriz sería reducible).

Si en vez de tener sólo una coordenada no-nula tuviese m , como tanto los elementos del vector como los de la matriz son no-negativos, esto conllevaría que cada coordenada del vector imagen es el resultado de m sumandos no negativos, obliga a que los m sumandos sean nulos en cada coordenada, es decir, obliga a que la matriz \mathbf{A} tenga m columnas de ceros, (el grafo correspondiente tendría m nodos a los que no se llega desde ninguno de los otros nodos, el grafo no es fuertemente conexo), y por tanto la matriz \mathbf{A} es reducible.

Hay una contradicción, ya que sabemos que \mathbf{A} es irreducible, y por tanto la hipótesis es falsa. Es decir, ningún rayo de R^+ se transforma mediante \mathbf{A} en el elemento nulo.

- (ii) No hay ningún rayo en ∂R^+ que quede invariante por \mathbf{A} . Lo razonamos por reducción al absurdo. Supongamos que existe un rayo del borde del cuadrante, $r= r[v]$, tal que $\mathbf{A}(r)=r$; entonces, existe k , $0 < k < n$, tal que las primeras k coordenadas de v son cero (después de re-ordenarlas si fuese necesario); La condición $\mathbf{A}(r)=r$ implica que \mathbf{A} es de la forma

$$\begin{pmatrix} M & 0 \\ P & N \end{pmatrix}$$

donde M es $k \times k$, y N es de orden $(n-k) \times (n-k)$. Contradicción con que \mathbf{A} es irreducible.

Si por ejemplo, trabajamos R^5 y sea $v=(0,0,0,a,b)$, si $\mathbf{A} \cdot v^t = \lambda v^t$, esto obliga a que \mathbf{A} sea de la forma:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \lambda a \\ \lambda b \end{pmatrix}$$

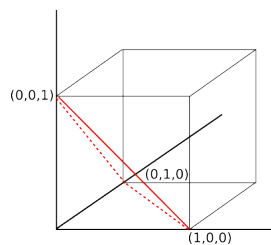
- (iii) En este apartado vamos a utilizar el Teorema del punto fijo de Brouwer, lo recordamos:

Teorema del punto fijo de Brouwer: En un espacio euclídeo, toda aplicación continua de un conjunto convexo, compacto y no vacío, en sí mismo, admite un punto fijo.

Para poder aplicar el Teorema de punto fijo de Brouwer a R^+ , necesitamos comprender que este espacio se puede ver como un conjunto convexo, compacto y no vacío. Para ello, basta identificar cada rayo con su punto perteneciente al n-símplice

$$\{(x_1, x_2, \dots, x_n) : x_1 + x_2 + \dots + x_n = 1, x_i \geq 0 \ \forall i = 0, \dots, n\}.$$

En R^3 lo podemos visualizar. Cada rayo de R^+ se identifica mediante el punto de intersección del propio rayo con el triángulo T, siendo T la intersección del plano $x+y+z=1$ con el octante positivo (ver figura 1, está señalado en rojo).



Hemos identificado R^+ con un n-símplice, por lo que cumple ser convexo, compacto y no vacío. Por tanto, como $\mathbf{A}(R^+) \subseteq R^+$ por (i), el Teorema del punto fijo de Brouwer asegura que hay un punto fijo de dicho triángulo, o equivalentemente, un rayo r invariante en R^+ . Por (ii) sabemos que ningún rayo del borde del cuadrante es fijo, lo cuál implica que r es un rayo positivo. Es decir, si tomamos un vector v en la dirección del rayo r (v es un autovector), v no tiene ninguna de sus coordenadas nula. Además, el autovalor asociado a v , λ , es positivo puesto que tanto v como su imagen pertenecen al cuadrante positivo, C^+ .

Para acabar la demostración vamos a representar cada rayo por su punto que dista una unidad del origen de coordenadas. De este modo, hay una correspondencia biunívoca entre los rayos de R^+ y la esfera S^{n-1} . Y vamos a estudiarlo particularizando a R^3 , es análogo en R^n .

En R^3 , para saber como actúa \mathbf{A} sobre un plano, p , basta con saber como actúa sobre la circunferencia S^1_p que obtenemos de la intersección de S^2 con el plano p .

Supongamos que π es un plano invariante que contiene a r , sea L el arco $R^+ \cap S^1_\pi$, por pertenecer L al plano invariante y por (i), se cumple que $\mathbf{A}(L) \subseteq L$:

La aplicación \mathbf{A} restringida a π , es una aplicación lineal y podemos encontrar su forma de Jordan. Sabemos que uno de sus autovalores es λ , por tanto hay tres posibilidades:

(a) $mult_{alg}(\lambda) = mult_{geom}(\lambda) = 2$, en cuyo caso la forma de Jordan sería:

$$A|\pi = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

(b) $mult_{alg}(\lambda) = 2$ y $mult_{geom}(\lambda) = 1$, en cuyo caso la forma de Jordan sería:

$$A|\pi = \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$$

(c) $mult_{alg}(\lambda) = mult_{geom}(\lambda) = 1$, en cuyo caso la forma de Jordan sería:

$$A|\pi = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$$

Analizamos cada uno de los tres casos y vamos a llegar a la conclusión de que sólo es posible el caso c).

(1) Vemos por reducción al absurdo que el caso (a) no es posible.

Supongamos que existen dos autovectores independientes asociados a λ . Tomando estos dos vectores como base del plano observamos que todos los vectores del plano son autovectores asociados al autovalor λ . Explico el por qué: sean v_1 y v_2 los autovectores asociados a λ , cualquier vector w del plano se puede expresar como $w = a \cdot v_1 + b \cdot v_2$, y por ser \mathbf{A} la matriz asociada a una aplicación lineal, tenemos que la imagen de w sería

$$\mathbf{A} \cdot w = a \cdot \mathbf{A}v_1 + b \cdot \mathbf{A}v_2 = a \cdot \lambda \cdot v_1 + b \cdot \lambda \cdot v_2 = \lambda \cdot (a \cdot v_1 + b \cdot v_2) = \lambda \cdot w.$$

Esto implica que \mathbf{A} deja invariantes todos los rayos del plano, en particular, deja invariantes los rayos del borde de R^+ , lo cuál contradice al apartado (ii).

Por tanto la hipótesis es incorrecta, no pueden existir dos autovectores independientes asociados a λ .

(2) Vemos ahora que el caso (b) no es posible.

Demostración. Reducción al absurdo:

Suponemos que la matriz $A|\pi$ es semejante a la matriz del caso b).

Sabemos que dos matrices semejantes pueden pensarse como dos descripciones de una misma transformación lineal, pero con respecto a bases distintas. Vamos a analizar la transformación que conlleva la matriz de Jordan del caso b) y, por tanto, como estamos suponiendo que la matriz de $A|\pi$ es semejante, va a realizar la misma transformación lineal pero respecto de otra base, es decir, dejando fijos otros vectores.

Observamos que:

$$\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a\lambda \\ a + b\lambda \end{pmatrix}$$

La aplicación lineal dada por $\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$

deja fijos los rayos que pasan por $(0,1)$ y $(0,-1)$, y todos los demás rayos giran en sentido anti-horario. Pero no todos los vectores sufren un giro de la misma amplitud, cuanto más cerca está el vector de uno de los vectores fijos menor es el giro, pues la imagen no puede sobrepasar al vector fijo.

Esto se comprueba fácilmente comparando el ángulo que forma el vector (a,b) con el semieje-positivo OX, con el ángulo que forma su imagen $(a\lambda, a + b\lambda)$ con ese mismo semieje.

Si la matriz de $A|\pi$ es semejante a esta matriz, va a realizar la misma transformación. Es decir, deja dos rayos fijos y, en este caso, todos los demás rayos giran en sentido anti-horario sufriendo un giro de menor amplitud cuánto más cerca esté de los rayos fijos.

Vemos la contradicción:

- Los rayos fijos, a través de $A|\pi$, no pueden ser del borde de R^+ (porque contradiría a (ii)).
- Los rayos fijos, a través de $A|\pi$, no pueden ser distintos del eje vertical (porque si no el vector que pasa por $(0,1)$ sale de R^+ , y esto contradice al apartado (i)).

Por tanto, la matriz de $A|\pi$ no puede ser semejante a la del caso b). □

Nota: Si en vez de tomar como forman de Jordan la matriz $\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$ tomo la matriz $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, la transformación que conlleva es similar, la diferencia es que deja fijo el eje horizontal, y que los vectores giran en sentido horario. En $A|\pi$ se llegaría a contradicción viendo que el semieje-positivo OX, si es fijo contradice a (ii); y si no es fijo, hay rayos de R^+ que salen de r^+ .

Por (1) deducimos que la multiplicidad geométrica de λ es 1, y por (2) deducimos que la multiplicidad algebraica de λ es también 1. En conclusión, sólo se puede dar el caso

$$(c) A|\pi = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$$

Resumiendo, en este apartado hemos visto que el autovalor λ , correspondiente al rayo fijo, es simple (tiene multiplicidad algebraica 1), es positivo y tiene asociado un autovector estrictamente positivo.

(iv) Solo nos queda mostrar que $|\lambda| \geq |\mu|$ para el otro autovalor μ .

Razonamos por reducción al absurdo. Supongamos que el autovalor μ cumple que $|\lambda| < |\mu|$, distinguimos dos casos:

Caso1.- μ es real.

Sea v_μ el autovector asociado al autovalor μ tal que su rayo $r_\mu \notin R^+$ (porque si $r_\mu \in R^+$ ya estaría demostrado el teorema). Entonces \mathbf{A}^n actúa en la circunferencia S^1 que representa todos los rayos del plano generado por v_λ y v_μ , y fija al conjunto de rayos $\{\pm r_\lambda, \pm r_\mu\}$. Como $\lambda < |\mu|$, el movimiento que realiza \mathbf{A}^n en S^1 tiene dos puntos de atracción, $\{\pm r_\mu\}$, y dos puntos de repulsión, $\{\pm r_\lambda\}$. Entonces r_μ atrae a uno de los dos puntos de $\partial R^+ \cap S^1$ fuera de R^+ , lo cual no es posible.

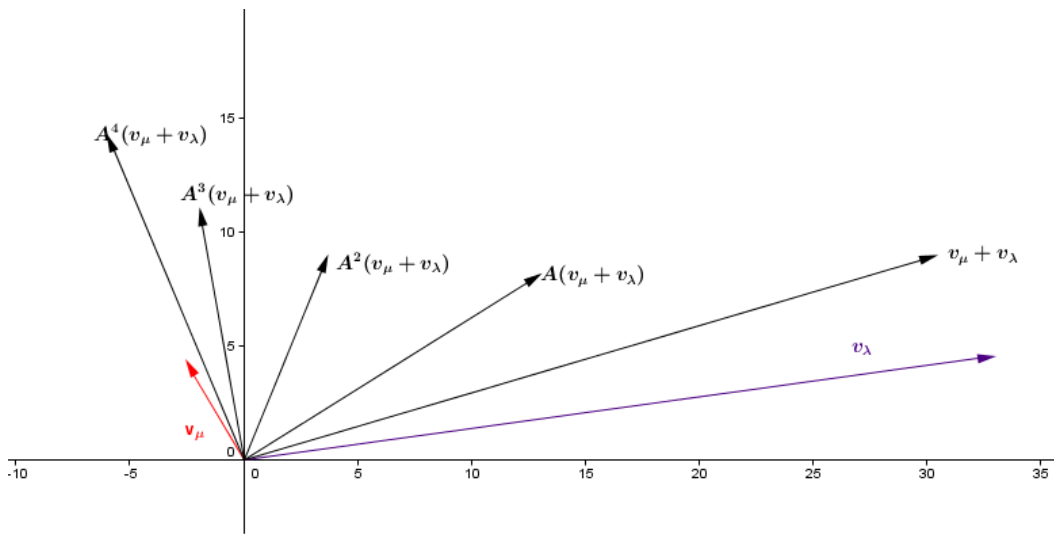


Figura 3.1: El dibujo muestra cómo tras sucesivas aplicaciones de \mathbf{A} , los rayos se aproximan al rayo asociado al autovalor de mayor valor absoluto (en este ejemplo hemos tomado $\lambda=1/2$ y $\mu=4/3$).

Caso2.- μ es complejo.

Comenzamos recordando que si μ es autovalor de \mathbf{A} , entonces $\bar{\mu}$ también es autovalor de \mathbf{A} . Además si v es un autovector asociado a μ , se cumple que \bar{v} es autovector asociado a $\bar{\mu}$.

Esto se puede explicar por dos vías, una sería utilizando que como la matriz es real, el polinomio característico tiene coeficientes reales. Nosotros vamos a razonarlo desde el punto de vista matricial, que nos va a dar a la vez información sobre los autovectores asociados a estos autovalores. Sabemos que la matriz conjugada $\bar{\mathbf{A}}$ consiste en escribir los conjugados de cada uno de los elementos de \mathbf{A} , y que se conservan las propiedades de los conjugados, por ejemplo que $\overline{\mathbf{A}\mathbf{B}} = \bar{\mathbf{A}}\bar{\mathbf{B}}$; además el conjugado de una matriz real es ella misma. Entonces, si μ es un autovalor complejo de una matriz real \mathbf{A} y v es su autovector asociado, se verifica:

$$(\mathbf{A} - \mu\mathbf{I})v = 0, \text{ tomando conjugados}$$

$$\overline{(\mathbf{A} - \mu\mathbf{I})v} = (\bar{\mathbf{A}} - \bar{\mu}\bar{\mathbf{I}})\bar{v} = (\mathbf{A} - \bar{\mu}\mathbf{I})\bar{v} = 0$$

con lo que no sólo tenemos que $\bar{\mu}$ es también autovalor, si no también que el conjugado del vector es autovector asociado a $\bar{\mu}$.

Vamos a reflexionar sobre que significa que un vector sea autovector. Esto significa que el transformado de ese vector es el resultado de multiplicar el vector por el autovalor. Cuando el autovalor es real, la imagen del autovector es otro vector que mantiene la dirección, si el autovalor es positivo mantiene el sentido y si el autovalor es negativo tiene sentido contrario, y que su módulo es el producto del módulo del vector inicial por el valor absoluto del autovalor. Pero cuando trabajamos con autovalores complejos esto cambia. Si multiplicamos dos números complejos el resultado es otro número complejo que tiene como módulo el producto de los módulos y como argumento la suma de los argumentos. Es decir, si tomo la aplicación $f(z)=wz$, siendo $z \in \mathbb{C}$ y w un complejo fijo, esta aplicación multiplica todos los puntos del plano complejo por w , es decir, cada punto lo manda a un nuevo punto tal que:

$$\arg(f(z)) = \arg(z) + \arg(w)$$

$$|f(z)| = |z| |w|$$

Observamos que multiplicar por un complejo geoméricamente equivale a realizar una rotación y una dilatación.

La multiplicación por un número complejo, se puede expresar de forma matricial como sigue:

sea $z = x + iy$

$w = a + bi$

Teniendo en cuenta que: $f(x + iy) = (a + bi)(x + iy) = (ax - by) + i(bx + ay)$, tenemos que su expresión matricial es:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Al utilizar la forma matricial sólo se trabaja con números reales.

Multiplicar un vector complejo, v , por un número complejo $\mu = a + bi$, se puede descomponer en multiplicar tanto la parte real del vector (el vector u_1 formada por la parte real de cada coordenada de v) como su parte imaginaria (el vector u_2 formada por la parte imaginaria de cada coordenada de v) por la matriz 2×2 que corresponde al número complejo. Ver como se transforma su parte real y su parte imaginaria equivale a ver como se transforma el vector complejo, y de esta forma trabajamos con números reales.

Si el autovalor $\mu = a + bi$, a y b números reales, y $v = u_1 + iu_2$ el autovector asociado, u_1 y u_2 son vectores, siendo u_1 la parte real de v , y u_2 la parte imaginaria de v , entonces $\bar{v} = u_1 - iu_2$ es autovector de A asociado a $\bar{\mu} = a - bi$ y, por tanto, tenemos las igualdades

$$Av = \mu v = (a + bi)(u_1 + iu_2) \longrightarrow Au_1 + iAu_2 = (au_1 - bu_2) + i(bu_1 + au_2)$$

$$A\bar{v} = \bar{\mu}\bar{v} = (a - bi)(u_1 - iu_2) \longrightarrow Au_1 - iAu_2 = (au_1 - bu_2) - i(bu_1 + au_2)$$

Por tanto, identificando las partes real e imaginaria en cualquiera de las dos igualdades anteriores tenemos,

$$\begin{cases} Au_1 = au_1 - bu_2 \\ Au_2 = bu_1 + au_2 \end{cases}$$

Tenemos

$$A(\dots v | \bar{v} | \dots) = (\dots v | \bar{v} | \dots) \begin{pmatrix} \ddots & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 \\ 0 & 0 & \bar{\mu} & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix}$$

$$A(\dots u_1 | u_2 | \dots) = (\dots u_1 | u_2 | \dots) \begin{pmatrix} \ddots & 0 & 0 & 0 \\ 0 & a & b & 0 \\ 0 & -b & a & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix}$$

Cuando la matriz es diagonalizable (cuando coincide la multiplicidad algebraica y geométrica de cada autovalor) existe una matriz de paso P tal que $P^{-1}AP = D$, siendo D una matriz diagonal formada por los autovalores (cada autovalor aparece tantas veces como indique su multiplicidad algebraica), y P es la matriz que tiene por columna i -ésima el autovector asociado al autovalor que ocupa la posición d_{ii} en D .

Pero si la matriz tiene autovalores complejos, tanto D como P tiene elementos complejos. Existe una descomposición alternativa $Q^{-1}AQ = B$, donde Q y B son matrices reales. B no es una matriz diagonal, es una matriz diagonal por bloque, cada bloque será 1×1 si se corresponde con un autovalor real o 2×2 si se corresponde con un autovalor complejo (cambiando el autovalor $a+bi$, por la matriz $2 \times 2 \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$). Las columnas de Q son autovectores reales o vectores obtenidos a partir de la parte real y de la parte imaginaria (por parejas) de un autovector complejo. El orden para escribir las columnas de Q , es el mismo que sigan los autovalores en B .

En el trabajo de menciona un plano invariante asociado a un autovalor complejo, vamos a explicar a que plano nos referimos.

La matriz A actúa de C^n en C^n . El espacio propio de un autovalor complejo (conjunto de autovectores asociados a él) está contenido en C^n , pero a nosotros nos interesa su corte con R^n . Si intersecamos el espacio generado por un autovector complejo, $v = u_1 + iu_2$, con R^n nos da un plano de R^n que contiene a los vectores u_1 y u_2 . Llamamos P a este plano, $P = \langle \text{re}(v), \text{im}(v) \rangle$

Sabemos que el resultado de multiplicar un espacio complejo por un número complejo fijo es una rotación y una dilatación. Es decir, la transformación que sufre el espacio propio de un

autovalor complejo al transformarlo por A es una rotación y una dilatación. En particular, como el espacio propio es invariante, él y su imagen por A tienen la misma parte real, que es el plano P , por lo que podemos afirmar que este plano es invariante por A .

Vamos a demostrar que si P es el plano invariante correspondiente al autovalor μ , P no contiene ningún rayo de R^+ :

En un principio tengo tres posibilidades para $P \cap R^+$:

- a) Que se corten en un único rayo.
- b) Que parte de P esté contenido en R^+ .
- c) Que P no corte a R^+ .

Vemos que a) es imposible:

Para que $P \cap R^+$ sea un rayo es necesario que dicho rayo pertenezca al borde de R^+ . Entonces la imagen de ese rayo sería el mismo, ya que $A(R^+) \subseteq R^+$, el plano es invariante y su único elemento perteneciente a R^+ es este rayo. Esto es imposible porque contradice al apartado (ii) que dice que ningún rayo del borde de R^+ es invariante. En conclusión, $P \cap R^+$ no es un rayo.

Vemos que b) es imposible por reducción al absurdo:

Teniendo en cuenta que la circunferencia S^1 representa a todos los rayos de P , cada punto de la circunferencia representa al rayo que pasa por el origen y por ese punto, la colección de rayos de R^+ que está contenida en P se corresponde un arco L de dicha circunferencia. Como A es una rotación en S^1 (porque es como actúa A en todo el espacio propio de μ), $A(L)$ no estaría contenida en L , lo cuál contradice que $A(R^+) \subseteq R^+$. Por tanto, la opción b) también es imposible.

Hemos llegado a la conclusión de que sólo se puede dar la opción c), es decir, P no corta a R^+ .

Demostramos que $|\lambda| \geq |\mu|$ por reducción al absurdo:

Supongamos que $|\lambda| < |\mu|$. Llamamos E^3 al espacio generado por P y por v_λ .

Por ser $|\lambda| < |\mu|$ la acción de A en E^3 muestra repulsión a la línea generada por v_λ , y una atracción al plano P (similar a lo que ocurriría en la figura 3.1). Esto implica que los rayos de $R^+ \setminus r_\lambda$ en E^3 se aproximan a P tanto como queramos si aplicamos repetidamente la aplicación de A , y por tanto nosotros concluimos que todos los rayos se salen de R^+ después de aplicar un número de veces, lo cuál no es posible porque sabemos que $A(R^+) \subseteq R^+$. (una demostración formal de esto es la siguiente: sea $r[x]$ un rayo de E^3 en $R^+ \setminus r_\lambda$ con $x = v_\lambda + w$, siendo $w \in P$ y $w \neq 0$. Entonces

$$A^n\left(\frac{x}{|\mu|^n}\right) = \frac{A^n(v\lambda) + A^n(w)}{|\mu|^n} = \frac{\lambda^n}{|\mu|^n}v\lambda + \frac{\mu^n}{|\mu|^n}w.$$

Cuando $n \rightarrow \infty$ entonces $\frac{\lambda^n}{|\mu|^n} \rightarrow 0$, y por tanto el rayo $A^n(r[x])$ se acerca a P tanto como queramos, ya que el módulo de $\frac{\mu^n}{|\mu|^n}w \in P$ es constante).

Por tanto la hipótesis es falsa. Podemos afirmar que $\lambda \geq |\mu|$.

FORMA DE FROBENIUS PARA MATRICES IMPRIMITIVAS

Como hemos expuesto en la introducción el Teorema de Perron-Frobenius es una generalización a matrices no negativas, del Teorema de Perron que era sólo para matrices positivas. Pero hay una propiedad que se cumple para matrices positivas, pero no siempre se cumple para matrices no negativas, y es que para matrices positivas existe un único autovalor cuyo valor absoluto alcanza el máximo entre los valores absolutos de todos los autovalores de A .

Cuando trabajamos con matrices no-negativas, decimos que la matriz es **primitiva** si tiene dicha propiedad (si existe un único autovalor cuyo valor absoluto alcanza el valor máximo entre los valores absolutos de todos los autovalores de la matriz). Decimos que la matriz es **imprimitiva**, con índice de imprimicidad h ($h > 1$), si existen h autovalores tales que su valor absoluto es el máximo de los valores absolutos de todos los autovalores.

Otra caracterización de las matrices primitivas es que existe un número natural $j \geq 1$ tal que B^j es positiva (es decir, todas sus entradas son positivas).

En realidad una matriz primitiva nos da mucha más información que una matriz imprimitiva, por ello va a ser de vital importancia poder expresar las matrices imprimitivas en función de algunas matrices primitivas, y esto es lo que nos permite nuestro segundo teorema. La forma de Frobenius para matrices imprimitivas con índice de imprimicidad h , nos da una nueva matriz por bloques tal que su h -ésima potencia es suma directa de h matrices primitivas.

Los bloques de la nueva matriz se llaman bloques primitivos. Por ejemplo, si estamos trabajando en un proceso de Markov, conocer estos bloques nos da una visión general del proceso mucho más completa.

Teorema II: Forma de Frobenius para matrices imprimitivas

Sea \mathbf{A} una matriz $n \times n$, no negativa e irreducible, entonces re-ordenando los miembros de la base canónica de R^n , \mathbf{A} puede expresarse de la forma

$$\hat{A} = \begin{pmatrix} 0 & 0 & \dots & 0 & A_d \\ A_1 & 0 & \dots & 0 & 0 \\ 0 & A_2 & \dots & 0 & 0 \\ \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & \dots & A_{d-1} & 0 \end{pmatrix}$$

Donde los bloques de la diagonal son matrices cuadradas de ceros posiblemente de distinto tamaño y \hat{A}^d es suma directa de d matrices primitivas.

NOTA: “ d ” es índice de imprimicidad, es decir, el número de autovalores que hay sobre el círculo (en realidad sobre la circunferencia) espectral.

3.1. Demostración de la existencia de la Forma de Frobenius

Esta demostración vamos a suponer que la matriz \mathbf{A} tiene rango máximo (en otro caso los argumentos son ligeramente más elaborados).

Que el rango de \mathbf{A} sea máximo supone, desde el punto de vista de geometría clásica, que la aplicación lineal que define \mathbf{A} es inyectiva y sobreyectiva, no se pierde ni se aumenta dimensión en sentido clásico (dimensión 1, rectas; dimensión 2, planos; dimensión 3, volúmenes). Es decir, la imagen de una recta es otra recta, la imagen de un plano es otro plano,...

Sabemos que $A(R^+) \subseteq R^+$, esta demostración va a consistir en analizar cómo varía la posición de los semiejes coordenados R^n que delimitan a R^+ al hacer su imagen a través de la función lineal representada por la matriz \mathbf{A} .

Vamos a definir un nuevo concepto que dado un rayo de R^+ nos diga que posición ocupa, es decir, si coincide con un eje coordenado, si está apoyado en algún plano coordenado,...

Este nuevo concepto lo llamamos dimensión de un rayo.

Definimos la **dimensión de un rayo como el número de coordenadas no nulas que tiene su vector director**.

Por ejemplo, sea r un rayo perteneciente a R^3 :

$\dim(r)=1$, implica que el rayo está sobre un eje coordenado.

$\dim(r)=2$, implica que el rayo está sobre uno de los planos coordenados (me refiero a los planos XY , XZ , YZ) pero sin coincidir con ninguno de los ejes coordenados.

$\dim(r)=3$, implica que el rayo pertenece al interior de $(R^3)^+$.

Nota: Como ahora la palabra dimensión tiene dos significados distintos, lo que vamos a hacer es que si sólo decimos dimensión nos referimos a este nuevo concepto, y si queremos referirnos a dimensión con su significado clásico lo denotamos con un asterisco (*dimensión** 1, rectas; *dimensión** 2, planos; *dimensión** 3, volúmenes).

Como es más cómodo trabajar con puntos que con rayos, vamos a representar cada rayo de $(R^n)^+$ por su punto de intersección con el n-símplice que se forma al cortar $(R^n)^+$ con $x_1 + x_2 + \dots + x_n = 1$.

Observaciones:

1. En R^3 el 3-símplice es el triángulo de vértices $(1,0,0), (0,1,0)$ y $(0,0,1)$.
En R^4 el 4-símplice es el tetraedro de vértices $(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$.
En general, un n-símplice tiene n vértices $\{e_1, e_2, \dots, e_n\}$ donde e_i es el punto de R^n que tiene todas las coordenadas nulas excepto la de la posición i-ésima que tiene valor 1.
2. Los puntos de un n-símplice son las distintas combinaciones baricéntricas de sus vértices, es decir, combinaciones lineales $\lambda_1 e_1 + \dots + \lambda_n e_n$, tales que $0 \leq \lambda_i \leq 1$ y $\sum_{i=1}^n \lambda_i = 1$.
Para identificar cada punto del n-símplice, basta con conocer sus coeficientes $\lambda_1, \dots, \lambda_n$, al conjunto de estos coeficientes $(\lambda_1, \dots, \lambda_n)$ se conoce como coordenadas baricéntricas del punto.
3. Cada rayo del R^+ viene representado por un único punto del n-símplice. (Es decir, la intersección de un rayo y el n-símplice siempre es un único punto).
4. Cada punto del n-símplice representa a un único rayo.
Cada punto P del n-símplice representa al rayo que pasa por el origen de coordenadas y por dicho punto, es decir, al rayo que tiene como vector director el vector \overrightarrow{OP} , siendo O el origen de coordenadas.
Por tanto, la dimensión de un rayo coincide con el número de coordenadas no nulas del punto que lo representa en el n-símplice.
5. En todo n-símplice las caras 1-dimensionales* son segmentos (combinación baricéntrica de dos de sus vértices), las caras 2-dimensionales* son triángulos (combinación baricéntrica de tres de sus vértices), las caras 3-dimensionales* son tetraedros (combinación baricéntrica de cuatro de los vértices) y en dimensiones* mayores ya no se pueden visualizar pero una cara (t-1)-dimensional* es la combinación baricéntrica de t de sus vértices.
6. Los puntos interiores de la cara (t-1)-dimensional* (combinación baricéntrica de t de sus vértices) tienen t coordenadas distintas de cero, es decir, se corresponden con rayos de dimensión t.

7. Por abuso del lenguaje, decimos que un rayo está contenido en una cara $(t-1)$ -dimensional* si el punto del n -símplice que lo representa está en esa cara.

Llamamos **rayos de la base**, y los denotamos por $\{r_1, \dots, r_n\}$, a los rayos que se corresponden con los vértices del n -símplice. Es decir, los rayos que tienen por vector director un vector de la base canónica de R^n .

Los lemas siguientes dan la demostración:

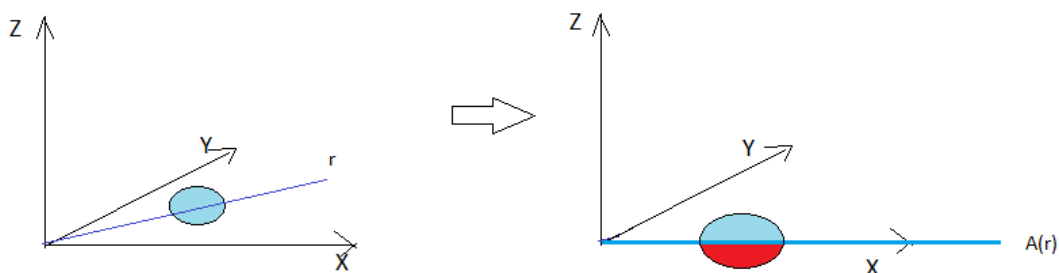
Lema 3.1 (Lema 1 de la desmotración original). *Sea r un rayo en R^+ , entonces $\dim[A(r)] \geq \dim[r]$. Por tanto existe un entero $s \geq 1$ tal que $\forall s' \geq s$ y $\forall r \in R^+$ $\dim[A^{s'}(r)] = \dim[A^s(r)]$.*

Demostración. La desglosamos en tres partes:

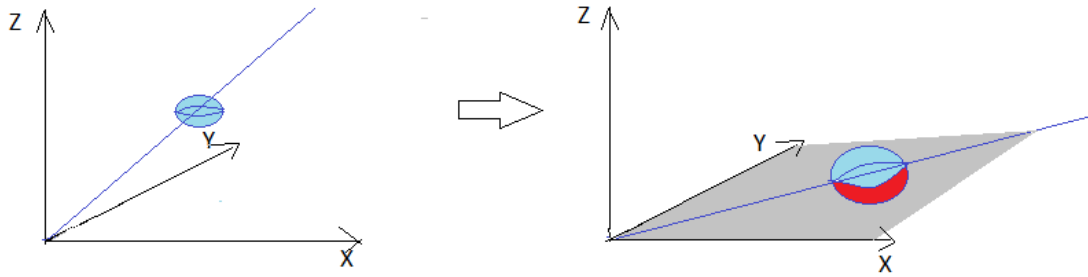
- (i) Vemos que $\dim[A(r)] \geq \dim[r]$ por reducción al absurdo:

Supongamos que $\dim[A(r)] < \dim[r]$ (recordamos que dimensión es número de coordenadas no nulas) entonces un entorno de r en R^+ se sale de R^+ bajo la acción de A , por ser $\text{Rg}(A) = n$, lo cuál no es posible por el apartado (i) del punto 2.1 de este documento.

La idea intuitiva de lo que dice el párrafo anterior es la siguiente: Si por ejemplo $\dim(r) = 2$ y $\dim[A(r)] = 1$, esto significa que r está en uno de los planos coordenados y su imagen es uno de los ejes coordenados. Si tomamos en el plano coordenado de r un entorno de un punto del rayo, de forma que el entorno esté contenido en R^+ , cuando haga la imagen de ese entorno como el punto se ha trasladado al borde de R^+ parte de la imagen del entorno sale de R^+ . (En el dibujo señalo en rojo la parte que sale de R^+).



Otro ejemplo sería, si $\dim(r) = 3$ y $\dim[A(r)] = 2$, esto significa que r pertenece al interior de R^+ y su imagen está en uno de los planos coordenados. Tomamos un entorno en R^+ (contenido en R^+) de un punto del rayo, la imagen de ese entorno tiene una parte fuera de R^+ . Lo vemos gráficamente. (En el dibujo señalo en rojo la parte que sale de R^+).



(ii) Demostramos que para cualquier entero $j \geq 1$ la función asociada al entero $\dim[A^j(r)]$ para los rayos $r \in R^+$ es constante en el interior de cada cara de R^+ .

Tenemos que:

$$A \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} r_1 a_{11} + r_2 a_{12} + \dots + r_n a_{1n} \\ r_1 a_{21} + r_2 a_{22} + \dots + r_n a_{2n} \\ \vdots \\ r_1 a_{n1} + r_2 a_{n2} + \dots + r_n a_{nn} \end{pmatrix}$$

Si dos rayos, r y t , pertenecen a la misma cara de R^+ sus vectores directores tienen las mismas coordenadas nulas (es decir, $r_i = 0$ si y sólo si $t_i = 0$ para todo $i=1, \dots, n$, siendo (r_1, \dots, r_n) el vector director de r y (t_1, \dots, t_n) el vector director del rayo t). Por tanto su imágenes también tienen las mismas coordenadas nulas, ya que la coordenada i -ésima de $A(r)$ viene dada por $r_1 a_{i1} + r_2 a_{i2} + \dots + r_n a_{in}$, si se anula es porque todos los elementos a_{ij} son nulos si $r_i \neq 0$, y en consecuencia la coordenada i -ésima de $A(t)$ que es $t_1 a_{i1} + t_2 a_{i2} + \dots + t_n a_{in}$ también sería nula. Todo esto es cierto porque tanto los elementos de la matriz como las coordenadas de los vectores son elementos no negativos.

(iii) Dado un rayo r cualquiera, probamos que existe un $s \geq 1$ tal que $\forall s' \geq s$ entonces $\dim[A^{s'}(r)] = \dim[A^s(r)]$.

Si tomamos la sucesión

$$\dim[A(r)], \dim[A^2(r)], \dim[A^3(r)], \dots, \dim[A^k(r)], \dots,$$

tenemos una sucesión monótona creciente (puesto que hemos visto que $\dim[A(r)] \geq \dim[r]$) y acotada superiormente por n (ya que $\dim[\text{rayo}] = "n^\circ \text{ de coordenadas no nulas}" \leq n$), por tanto, la sucesión es convergente. Como los términos de esta sucesión son números naturales, a partir de un cierto término la sucesión se hace constante, es decir, existe un s tal que $\forall s' \geq s$ entonces $\dim[A^{s'}(r)] = \dim[A^s(r)]$.

Es cierto para todos los rayos pero el valor de s y el valor de $\dim[A^s(r)]$ dependerá de cada rayo. A continuación veremos que existe un valor de s que sirve para todos los rayos de R^+ .

(iv) Existe un s que sirve para todos los rayos de R^+ .

Tenemos que:

(1) Por (iii) tenemos que para cada rayo existe una potencia s de A para la que la dimensión de las imágenes del rayo se estabiliza. Además por (ii) tenemos que la dimensión que alcanzan los rayos de una misma cara es la misma. Entonces, dado un rayo el valor de s para el que se estabiliza la dimensión de las imágenes de r es el mismo para todos los rayos del interior de esa cara. Es decir, hemos llegado a que existe un valor de s para cada cara del n -símplice.

(2) Un n -símplice tiene $2^n - 1$ cara. Lo vemos:

Coger una cara de *dimensión** $i-1$ consiste en elegir i vértices de los n posibles. Eso se puede hacer de $\binom{n}{i}$ formas. El total de cara de todas las dimensiones posibles será

$$\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n - 1$$

De (1) y (2) tenemos que existe un número finito de caras, y cada cara tiene asociado un s . Tomando el mayor de estos " s ", tenemos un valor de s tal que $\forall s' \geq s, \dim[A^{s'}(r)] = \dim[A^s(r)] \forall r \in R^+$.

□

Nota: En el siguiente lema veremos que los rayos que coinciden con los ejes coordenados alcanzan todos la misma dimensión.

Lema 3.2 (Lema 2 de la desmotración original). .

Existe un entero $t \geq 1$ tal que $\dim[A^s(r_i)] = t$ para todo $i=1, \dots, n$.

Demostración. La clave de esta demostración está en ver que si los rayos r_i alcanzan dimensiones distintas entonces la matriz A es reducible (lo demostramos a continuación).

Como nosotros trabajamos con matrices irreducibles, no podemos llegar a que A es reducible, y por tanto, todos los rayos r_i tienen que alcanzar la misma dimensión.

Demostramos que si los rayos r_i alcanzan dimensiones distintas entonces A es reducible.

Suponemos que los rayos r_i alcanzan dimensiones distintas. Llamamos " t " a la dimensión que alcanza el rayo que alcanza la menor dimensión y " s " a la potencia de A en la se estabiliza la dimensión (sabemos que se estabiliza porque el $\text{rg}(A) = n$).

Sin pérdida de generalidad podemos suponer que los rayos que alcanzan dimensión mayor que t son los p primeros (si no fuese así basta con realizar una permutación de filas y columnas, de forma que la nueva matriz PAP^t si lo cumpliría).

Hasta ahora tenemos que:

$$\dim(A(r_i)) > t \text{ para } \{r_1, r_2, \dots, r_p\}$$

$$\dim(A(r_i)) = t \text{ para } \{r_{p+1}, \dots, r_n\}$$

Vamos a ver que esto obliga a que la matriz A (o la matriz PAP^t , si he tenido que reorganizar la base) tenga la siguiente estructura:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} & 0 & \dots & 0 \\ a_{(p+1)1} & & \dots & & a_{(p+1)(p+1)} & \dots & a_{(p+1)n} \\ \vdots & & & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & & a_{n(p+1)} & \dots & a_{nn} \end{pmatrix}$$

Lo que equivale a decir que A es reducible.

Lo razonamos por reducción al absurdo.

Supongamos que un elemento de la matriz de ceros que queda en la esquina superior derecha no fuese nulo, es decir, $a_{ij} > 0$ con $1 \leq i \leq p$ y $(p+1) \leq j \leq n$. Observamos que, la imagen del r_j (es decir la imagen de su vector director, que es el vector que tiene nulas todas las coordenadas menos la j-ésima que es un 1) es la columna j-ésima de la matriz:

$$A(r_j) = (0, \dots, a_{ij}, 0, \dots, 0, a_{(p+1)j}, \dots, a_{nj})^T = a_{ij}r_i + a_{(p+1)j}r_{p+1} + \dots + a_{nj}r_n$$

Como la matriz A se corresponde con una aplicación lineal, respeta la suma de vectores y la multiplicación por escalares, podemos afirmar que:

$$A^2(r_j) = A(A(r_j)) = A(a_{ij}r_i + a_{(p+1)j}r_{p+1} + \dots + a_{nj}r_n) = a_{ij}A(r_i) + a_{(p+1)j}A(r_{p+1}) + \dots + a_{nj}A(r_n)$$

y así sucesivamente hasta llegar a A^{s+1} , en cuyo caso tendríamos:

$$A^{s+1}(r_j) = A^s(A(r_j)) = A^s(a_{ij}r_i + a_{(p+1)j}r_{p+1} + \dots + a_{nj}r_n) = a_{ij}A^s(r_i) + a_{(p+1)j}A^s(r_{p+1}) + \dots + a_{nj}A^s(r_n)$$

Vamos a analizar la dimensión de $A^{s+1}(r_j)$, y vamos a ver que llegamos a una contradicción. Por un lado, sabemos que la dimensión de $A^{s+1}(r_j)$ es exactamente t, porque hemos tomado j mayor que p (por hipótesis estos rayos alcanzaban dimensión t en la potencia s-ésima), y por ser $\text{rg}(A) = n$ a partir de la potencia s-ésima la dimensión es constante.

Pero por otro lado

$$\dim A^{s+1}(r_j) = \dim [a_{ij}A^s(r_i) + a_{(p+1)j}A^s(r_{p+1}) + \dots + a_{nj}A^s(r_n)] \geq \dim [a_{ij}A^s(r_i)] > t$$

Porque recordamos que $a_{ij} > 0$ y $\dim(A^s(r_i)) > t$ (porque $i \leq p$), y sabemos que $\dim(A^s(r_i)) > t$ significa que tiene más de t coordenadas no nulas. Como los coeficientes a_{kl} son todos no-negativos y la imagen de un rayo no-negativos también es otro rayo no-negativo, el vector suma también va a tener al menos esas coordenadas no nulas, lo que equivale a decir,

que va a tener al menos la dimensión de $A^s(r_i)$.

Hemos llegado a $\dim(A^s(r_j)) = t$ y $\dim(A^s(r_j)) > t$, lo cuál es imposible. Es decir, la hipótesis es falsa. No puede ser $a_{ij} > 0$ con $1 \leq i \leq p$ y $(p+1) \leq j \leq n$.

Por tanto, si los rayos r_i alcanzan dimensiones distintas A tiene la estructura expuesta, A es reducible.

□

Lema 3.3 (Lema 3 de la demostración original). *Las caras $(t-1)$ -dimensionales* del n -símplice que representa los rayos de $(R^n)^+$ o contienen en su interior los puntos que corresponden con exactamente t rayos del conjunto $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$, o no contienen a ninguno. Las caras que contienen t elementos son disjuntas.*

Demostración. Primero vamos a ver que los vectores de la base se agrupan de t en t , según las posiciones no nulas de su imagen s -ésima; en segundo lugar, vamos a ver que en estos grupos no puede haber más de t rayos; en un tercer apartado demostramos que todos los rayos de la base están en alguno de estos grupos; y finalizamos viendo como a partir de aquí se deduce la demostración de este lema.

Advertencias:

.) Si decimos que un rayo está “contenido” en una cara $(t-1)$ -dimensional*, en realidad significa que el punto del n -símplice que lo representa está en esa cara. (Ponemos la palabra “contenido” entre comillas para saber que se trata de abuso de lenguaje)

.) También por abuso del lenguaje, si decimos que un rayo tiene t coordenadas no nulas, lo que queremos decir es que su vector director tiene t coordenadas no nulas, o equivalentemente, que el punto del n -símplice que lo representa tiene t coordenadas no nulas.

Comenzamos la demostración:

(i) Vamos a ver que los vectores se agrupan de t en t .

Sabemos que la imagen s -ésima de los rayos de la base alcanzan dimensión t , tiene t coordenadas no nulas. Por tanto, su vector director se puede expresar como combinación lineal de t vectores de la base canónica, por abuso del lenguaje decimos que:

$$A^s(r_i) = \lambda_{i_1} r_{i_1} + \lambda_{i_2} r_{i_2} + \dots + \lambda_{i_t} r_{i_t} \quad (1)$$

El subíndice de los sumando indica las posiciones donde estaban los elementos no nulos, y el λ_{i_k} es el valor de la coordenada k -ésima de las coordenadas baricéntricas de $A^s(r_i)$.

Aplicamos A^s a ambos lados de la igualdad (1).

Teniendo en cuenta que la matriz A se corresponde con una aplicación lineal (respeta la suma de vectores y el producto por escalar) llegamos a:

$$A^{2s}(r_i) = A^s(A^s(r_i)) = \lambda_{i_1} A^s(r_{i_1}) + \lambda_{i_2} A^s(r_{i_2}) + \dots + \lambda_{i_t} A^s(r_{i_t}) \quad (2)$$

Sabemos que $\dim(A^{2s}(r_i))=t$ y $\dim(A^s(r_k))=t \forall k=1, \dots, n$ (por el lema 3.2) Es decir, en la expresión (2) todos los vectores de la derecha tienen exactamente t coordenadas no nulas (además sabemos que son elementos positivos), si el vector suma $(A^{2s}(r_i))$ sólo tiene t coordenadas no nulas es porque en todos los vectores de la derecha coinciden las t posiciones de los elementos no nulos. (Si no coincidiesen estas posiciones no nulas, al tener todos coordenadas positivas y no poderse contrarrestar, el vector suma tendría más de t posiciones no nulas).

Hemos llegado a que los vectores $\{A^s(r_{i_1}), A^s(r_{i_2}), \dots, A^s(r_{i_t})\}$, tienen las mismas t coordenadas no nulas, es decir, están en la misma cara $(t-1)$ -dimensional*.

Este razonamiento es cierto para la imagen s -ésima de todos los rayos de la base.

En resumen, como la imagen s -ésima de todos los rayos de la base tienen dimensión t , cada una de estas imágenes se expresa como combinación lineal de exactamente t rayos de la base, y utilizando el razonamiento de este apartado, las imágenes de estos t rayos de la base a través de A^s tienen las mismas t posiciones no nulas.

Observación 1: En un principio tendríamos n grupos (uno para cada vector $A^s(r_i)$ con $i=1, \dots, n$) de t rayos de la base, y en cada grupo las imágenes s -ésimas de esos rayos tienen las mismas posiciones no nulas. Es decir, esos t rayos están en una misma cara $(t-1)$ -dimensional* .

(ii) Demostramos que no puede haber más de t elementos del conjunto $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$ con las mismas t -posiciones no nulas.

Lo demostramos por reducción al absurdo:

Primero observamos que A^s es la matriz que tiene por columnas a los vectores $A^s(r_i)$ con $i=1, \dots, n$, $A^s = (A^s(r_1) \quad \dots \quad A^s(r_n))$, en particular, sus columnas tienen exactamente " t " elementos no nulos.

Ahora suponemos que tenemos $t+1$ vectores de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$ en los que coinciden sus t posiciones no nulas, si formamos una matriz en la que sus columnas sean estos $t+1$ vectores, tengo una matriz de orden $n \times (t+1)$ en la que sólo hay t filas distintas de cero, es decir, esta matriz como máximo tiene rango t , lo que equivale a decir que esos $(t+1)$ vectores no son linealmente independientes (sólo hay t vectores independientes). Entonces los vectores $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$ no son linealmente independientes, lo cual implica que A^s no tiene todas sus columnas independientes, es decir, A^s no tiene rango máximo.

Pero esto es imposible, puesto que A tiene rango máximo y todas sus potencias tienen rango máximo, (ya que si $|A| \neq 0$, entonces $|A^s| = |A|^s \neq 0$, sabemos que determinante no nulo, equivale, a rango máximo).

Hemos llegado a una contradicción, por tanto, la hipótesis es falsa. Concluimos que no puede haber más de t vectores de la base en los que coincidan las posiciones de las t coordenadas no nulas de su imagen a través de A^s .

Observación 2: Esto quiere decir que en la cara $(t-1)$ -dimensional* en la que estaban $\{A^s(r_{i_1}), A^s(r_{i_2}), \dots, A^s(r_{i_t})\}$, no puede estar la imagen s -ésima de ningún rayo más de la base. Es decir, las **caras $(t-1)$ -dimensionales* no pueden tener más de t rayos de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$.**

Observación 3: De aquí deducimos que los n grupos que hemos formado en el apartado (i) son coincidentes o disjuntos. Porque si hubiese dos grupos del apartado (i) que comparten elementos pero no son el mismo grupo, entonces, los rayos de ambos grupos tendrían las mismas t posiciones no nulas, lo cual implica que habría más de t vectores tales sus imágenes a través de A^s tendrían las mismas t posiciones no nulas, y sabemos que esto es imposible por (ii).

(iii) Vamos a ver que todos los rayos de la base pertenecen a algún grupo de los formados en el apartado (i).

Lo probamos por reducción al absurdo:

Supongamos que el rayo k -ésimo de la base no pertenece a ningún grupo de los que se formaron en el apartado (i). Esto implica que todos los rayos de $\{A^s(r_{i_1}), A^s(r_{i_2}), \dots, A^s(r_{i_n})\}$ tienen nula la coordenada k -ésima, entonces, en la matriz $A^s = (A^s(r_1) \vdots \dots \vdots A^s(r_n))$ todos los elementos de la fila k -ésima son nulos. Al tener una fila de ceros, el rango de A^s no es máximo. Pero esto contradice que A sea de rango máximo (como vimos en (ii) si A tiene rango máximo, todas sus potencias también).

Por lo tanto, la hipótesis es falsa. Concluimos que todos los rayos de la base pertenecen a algunos de los grupos formados en (i).

Concluimos la demostración:

1. Vamos a ver que sobre las caras $(t-1)$ -dimensionales* del n -símplice que representa los rayos de $(R^n)^+$ o contienen en su interior los puntos que corresponden con exactamente t rayos del conjunto $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$, o no contienen a ninguno.

Si la cara $(t-1)$ -dimensional* que tiene por vértices $\{e_{j_1}, e_{j_2}, \dots, e_{j_t}\}$ contiene el punto que representa a $A^s(r_j)$, entonces, contiene los puntos que representan la imágenes s -ésima de todos los rayos del grupo de r_j (de los grupos formados en (i)), porque todos estos rayos tienen las mismas posiciones no nulas (es decir, son combinaciones baricéntricas de los mismos vértices). Por tanto, contiene al menos la imagen s -ésima de t rayos de la base. Y por la observación 2, sabemos que ninguna cara $(t-1)$ -dimensional* puede contener la imagen s -ésima de más de t rayos de la

base. Por ello, podemos concluir que si sobre una cara $(t-1)$ -dimensional* está el punto correspondiente a un elemento de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$, entonces, esta cara contiene los puntos que corresponden con exactamente t rayos de dicho conjunto.

Por ello, dada una cara $(t-1)$ -dimensional* o no “contiene” ningún elemento de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$ o “contiene” exactamente t.

2. Vamos a ver que las caras que “contienen” t elementos de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$ son disjuntas.

Existe una correspondencia biunívoca entre los grupos de rayos (formados en (i)) y las caras $(t-1)$ -dimensional* que “contienen” t elementos de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$.

Lo vemos:

Si $A^s(r_i) = \lambda_{i_1} r_{i_1} + \lambda_{i_2} r_{i_2} + \dots + \lambda_{i_t} r_{i_t}$, esto quiere decir que las posiciones no nulas de $A^s(r_i)$ son i_1, i_2, \dots, i_t . Es decir, los subíndices de los rayos que forman cada grupo se corresponde con las posiciones no nulas de los rayos de otro grupo.

El punto que representa a $A^s(r_i)$ está en la cara $(t-1)$ -dimensional* dada por los vértices $\{e_{i_1}, e_{i_2}, \dots, e_{i_t}\}$

Agrupandos los n vértices del n-símplice de t en t, de manera que los subíndices de los t vértices que forman cada grupo sean los mismo t subíndices de los rayos que formaban cada grupo, tenemos que cada grupo de t vértices me define una cara $(t-1)$ -dimensional* que “contiene” elementos de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$ (de hecho por el punto 1 “contiene” justo t de esos elementos).

Por la forma en la que hemos agrupado los vértices, y por lo que hemos estudiado para los rayos, sabemos que todos los vértices están en alguno de los grupos, que todos los grupos tienen exactamente t vértices, y que los grupos son disjuntos dos a dos.

Que los grupos sean disjuntos dos a dos, implican que no hay dos caras (de las que “contienen” t elementos) compartan ningún vértice (o lo que es lo mismo, que no coinciden ninguna de las posiciones no nulas de dos rayos de grupos distintos). Por tanto, las caras $(t-1)$ -dimensional* que “contienen” imágenes s-ésimas de los rayos de la base son disjuntas dos a dos.

□

Una vez demostrados estos lemas estamos en condiciones de ver que la matriz A se puede expresar como dice el teorema, vamos a demostrar que existe una permutación P tal que

$$\hat{A} = P^t A P = \begin{pmatrix} 0 & 0 & \dots & 0 & A_d \\ A_1 & 0 & \dots & 0 & 0 \\ 0 & A_2 & \dots & 0 & 0 \\ \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & \dots & A_{d-1} & 0 \end{pmatrix}$$

Donde los bloques de la diagonal son matrices cuadradas de ceros.

Vamos a demostrarlo en **dos pasos**.

1. Vamos a ver que existe una matriz semejante a A , $P_1^t A P_1$, que es una matriz por “permutación por bloques”. Es decir, que $P_1^t A P_1$ es una matriz formada por filas y columnas de bloques, donde en cada fila y en cada columna hay un único bloque no-nulo.
2. Vamos a ver que los bloques del paso 1, se pueden reordenar de forma que lleguemos a la estructura de \hat{A} , es decir, que en las primeras $d-1$ columnas su bloque no-nulo sea el de debajo de la diagonal, y que en la columna d -ésima su bloque no-nulo sea el de la primera fila.

Antes de detallar estos dos pasos, vamos a recordar lo que tenemos hasta ahora y vamos a realizar una observación más.

Analizamos lo que tenemos hasta ahora:

- Tenemos grupos de t en t elementos que podemos interpretarlos como grupos de rayos (la similitud de los rayos de cada grupo es que sus imágenes s -ésimas tienen las mismas posiciones no nulas), o como grupo de vértices (donde cada grupo forma una cara $(t-1)$ -dimensional* que contiene t elementos de $\{A^s(r_1), A^s(r_2), \dots, A^s(r_n)\}$).
- Tenemos $n/t=d$ grupos diferentes (más adelante veremos que justo este número es el número de bloques no nulos tiene \hat{A}). Esto ocurre porque los grupos son disjuntos y cada uno de los n rayos (o de los n vértices) está incluido en alguno de los grupos.
- Hemos numerado estas caras de la 1 a la d y las denoto por C_1, \dots, C_d . (Los d grupos de rayos respetan la misma numeración, es decir, los grupos k -ésimos tanto de vértices como de rayos estén formados por los mismos subíndices).

Observación: Dado un grupo de rayos de los formados en el lema 3.3., las imágenes k -ésimas de estos rayos van todas a una misma cara de C_1, \dots, C_d para todo número natural k .

Demostramos esta observación por reducción al absurdo:

Demostración. Sean r_i y r_j rayos de un mismo grupo de modo que $A^k(r_i)$ y $A^k(r_j)$ van a caras distintas. Sin pérdida de generalidad podemos suponer que $A^k(r_i)$ va a C_1 y $A^k(r_j)$ va a C_2 .

Si r_i y r_j están en un mismo grupo es porque hay un tercer rayo r_m tal que $A^s(r_m)$ tiene no nulas las posiciones i -ésima y j -ésima, es decir, $A^s(r_m) = \lambda_i r_i + \lambda_j r_j + \dots$

Observamos que $A^{s+k}(r_i) = A^s(A^k(r_i))$ va ser combinación lineal de imágenes s -ésimas de los vértices C_1 , y $A^{s+k}(r_j) = A^s(A^k(r_j))$ va ser combinación lineal de imágenes s -ésimas

de los vértices C_2 . Tanto $A^{s+k}(r_i)$ como $A^{s+k}(r_j)$, tienen t coordenadas no nulas, pero en distintas posiciones.

Entonces si calculamos $A^{2s+k}(r_m) = A^{s+k}(A^s(r_m)) = A^{s+k}(\lambda_i r_i + \lambda_j r_j + \dots) = \lambda_i A^{s+k}(r_i) + \lambda_j A^{s+k}(r_j) + \dots$

Como tanto $A^{s+k}(r_i)$ como $A^{s+k}(r_j)$, tienen t coordenadas no nulas, pero en distintas posiciones, tendríamos que $\dim(A^{2s+k}(r_m)) > t$, lo cuál es imposible por el lema 2.

Por tanto la hipótesis es falsa. Podemos concluir que la imágenes de todos los rayos de un mismo grupo a través de cualquier potencia de A van todas a una misma cara de $\{C_1, \dots, C_d\}$.

□

Pasamos a demostrar el paso 1: Definimos una matriz permutación P_1 , de forma que reagrupe las columnas y filas de A (según se multiplica a derechas o a izquierdas, respectivamente). De modo, que la imagen de rayos de un mismo grupo queden como columnas consecutivas y, análogamente, que las filas correspondientes a las posiciones no nulas de cada cara $(t-1)$ -dimensional* queden consecutivas.

Definimos esta matriz permutación P_1 como sigue: Sus primeras t columnas sean las coordenadas traspuestas (es decir, puestas en vertical) de los vértices de la cara C_1 ; las columnas de la $t+1$ a las $2t$, sean las coordenadas traspuestas de los vértices de la cara C_2 ; y así sucesivamente, de modo que las últimas t columnas son las coordenadas traspuestas de los vértices de la cara C_d . (Llegaríamos a la misma matriz si hubiésemos razonado con los grupos de rayos).

Tenemos que:

- A es la matriz que tiene por columna a los rayos $(A(r_1))^t, \dots, (A(r_n))^t$.
- AP_1 es la matriz que tiene por columnas a los vectores $(A(r_1))^t, \dots, (A(r_n))^t$, pero reagrupados de forma que las imágenes de los t rayos que forman cada grupo están consecutivas.
- La matriz $P_1^t A$ reagrupa las filas de la matriz A , de forma que las t primeras filas de $P_1^t A$ son las t filas que se corresponden con las t posiciones no nulas de la cara C_1 ; las filas de la $t+1$ a las $2t$ de $P_1^t A$, son las t filas que corresponden a las t posiciones no nulas de la cara C_2 ; así sucesivamente, de modo que las últimas t filas de $P_1^t A$, son las t filas correspondientes a las posiciones no nulas de C_d . (Equivale a renombrar los vértices de modo que los que aparece en C_1 sean $\{e_1, \dots, e_t\}$, los que aparece en C_2 sean $\{e_{t+1}, \dots, e_{2t}\}$, y así sucesivamente).

Por tanto, la matriz $P_1^t AP_1$ es una matriz por bloques:

$$P_1^t A P_1 = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1d} \\ B_{21} & B_{22} & \dots & B_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ B_{d1} & B_{d2} & \dots & B_{dd} \end{pmatrix}$$

Vamos a ver que efectivamente hemos conseguido lo que buscábamos en este primer paso. Es decir:

- Cada bloque B_{ij} es de orden $t \times t$.
- Cada columna de bloques está formada por las t imágenes a través de A de los rayos de un grupo. Cada columna va a tener un único bloque no nulo, si su bloque no nulo es el k -ésimo significa que la imagen a través de A de los rayos de ese grupo están “contenidos” en la cara C_k .
- Cada fila de bloques se corresponde con una de las caras $\{C_1, \dots, C_d\}$. En cada fila hay un único bloque no nulo. Si en la fila k -ésima el único bloque no nulo es el j -ésimo significa que la imagen a través de A de los rayos del grupo j -ésimo están “contenidos” en la cara C_k .

En conclusión, en cada fila y en cada columna hay un único bloque no nulo. Si cambiamos los bloques nulos por ceros y los bloques no nulos por unos, tenemos una matriz permutación.

Pasamos a demostrar el paso 2: Probamos que la matriz “permutación por bloques” $P_1^t A P_1$ no tiene varios ciclos. Lo demostramos por reducción al absurdo:

Demostración. Supongamos que existe un ciclo de q caras, siendo q un número $1 \leq q < d$. (Es decir, que los elementos que están en esas q caras cuando le vuelvo a aplicar A cambian de cara pero a otra de ese grupo de q caras).

Sin pérdida de generalidad, podemos suponer que las caras que forman este ciclo son las q últimas. (En otro caso bastaría con permutar las filas y columnas de la matriz $P_1^t A P_1$, de forma que llegaríamos a otra matriz semejante que lo cumpliría).

Pero si esto es cierto obliga a que la matriz $P_1^t A P_1$, y por tanto la matriz A , sea **reducible**:

$$P_1^t A P_1 = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1d} \\ B_{21} & B_{22} & \dots & B_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ B_{d1} & B_{d2} & \dots & B_{dd} \end{pmatrix} = \begin{pmatrix} M_{(d-q) \times (d-q)} & \theta \\ M_{q \times (d-q)} & M_{q \times q} \end{pmatrix}$$

Es decir, todos los bloques B_{ij} con $1 \leq i \leq (q-1)$ y $(d-q) \leq j \leq d$, son nulos.

Hemos llegado a una contradicción porque sabemos que A **no es reducible**. Por tanto, la hipótesis es falsa. Si $q < d$, no puede existir un ciclo de q caras que se permuten entre ellas.

Por tanto, tenemos que las d caras se permutan entre ellas sin poderse descomponer en varios ciclos. Si las renombramos de modo que cada vez que aplicamos A , los elementos de la cara C_1 vayan a C_2 , los elementos de la cara C_2 vayan a C_3, \dots , los elementos de la cara C_{d-1} vayan a C_d , y los elementos de la cara C_d vayan a C_1 , llegamos a que los bloques no nulos de la matriz $P^t A P$ son los que están debajo de la diagonal principal y el de arriba a la derecha, que es justo lo que queríamos demostrar. \square

Observación: \hat{A}^d es una matriz diagonal por bloques.

Lo vemos:

Por la estructura de \hat{A} , como la primera imagen del los rayos del grupo i van a la cara C_{i+1} si $i < d$, y los rayos del grupo d van a la cara C_1 . Y como sabemos que al aplicar A de nuevo, los elementos de la cara C_1 vayan a C_2 , los elementos de la cara C_2 vayan a C_3, \dots , los elementos de la cara C_{d-1} vayan a C_d , y los elementos de la cara C_d vayan a C_1 . Y así sucesivamente, cuando hayamos aplicado A d veces, resulta que hemos llevado los rayos del grupo k a la cara C_k para todo $k = 1, \dots, d$. Por tanto, \hat{A}^d es una matriz diagonal por bloques.

Observación: Los bloques diagonales de \hat{A}^d son matrices primitivas.

Lo vemos:

Por ser una permutación cíclica de d grupos en d caras, podemos asegurar que siempre que se aplique un múltiplo d veces $\hat{A}^d, \hat{A}^{2d}, \dots, \hat{A}^{wd}$, se llega a una matriz diagonal por bloques. Como hacer la potencia de una matriz diagonal equivale a hacer la potencia de los elementos de su diagonal, tenemos que los bloques de la diagonal de \hat{A}^{wd} son los correspondientes bloques de \hat{A}^d elevados a w . Como va a llegar un momento que wd es mayor que s (recordamos que s es la menor potencia de A en la que los todos los rayos de la base han alcanzado dimensión máxima), entonces todos elementos de los bloques diagonales de \hat{A}^{wd} son positivos. Es decir, los bloques diagonales de \hat{A}^d elevados a w , dan una matriz positiva, esto quiere decir que los bloques diagonales de \hat{A}^d son matrices primitivas.

3.2. Ejemplo de cálculo de la Forma de Frobenius

Vamos a ver que la siguiente matriz es imprimitiva y vamos a calcular su Forma de Frobenius.

$$A = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 \end{pmatrix}$$

En primer lugar calculamos su polinomio característico:

$$P(\lambda) = |A - \lambda I| = \lambda^6 - 38\lambda^3 + 32 = \lambda^6 - 38\lambda^{6-3} + 32\lambda^{6-6}$$

El índice de imprimitividad viene dado por el máximo común divisor de los números señalados en negrita. Tenemos $h = \text{M.c.d}(6,3) = 3$

Por tanto, la estructura de la Forma de Frobenius que vamos a obtener es

$$F = \begin{pmatrix} 0 & 0 & A_3 \\ A_1 & 0 & 0 \\ 0 & A_2 & 0 \end{pmatrix} \text{ donde los ceros de la diagonal principal representan bloques de ceros.}$$

Calculamos las sucesivas imágenes de los vectores de la base canónica:

$$\begin{aligned} Ae_1^t &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} & A^2e_1^t &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 3 \end{pmatrix} & A^3e_1^t &= \begin{pmatrix} 11 \\ 0 \\ 5 \\ 0 \\ 0 \\ 0 \end{pmatrix} & A^4e_1^t &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 5 \\ 31 \\ 0 \end{pmatrix} \\ Ae_2^t &= \begin{pmatrix} 2 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix} & A^2e_2^t &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 10 \\ 0 \end{pmatrix} & A^3e_2^t &= \begin{pmatrix} 0 \\ 16 \\ 0 \\ 0 \\ 0 \\ 32 \end{pmatrix} & A^4e_2^t &= \begin{pmatrix} 128 \\ 0 \\ 64 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Observamos que la máxima dimensión que se alcanza es dos. Por tanto, $t=2$. Vamos a ver que todos los vectores de la base canónica han alcanzado dimensión 2 en su imagen a través de A^2 , lo que equivale a decir que se estabiliza para $s=2$.

$$\begin{aligned} A^2e_3^t &= \begin{pmatrix} 0 \\ 7 \\ 0 \\ 0 \\ 0 \\ 13 \end{pmatrix} & A^2e_4^t &= \begin{pmatrix} 9 \\ 0 \\ 7 \\ 0 \\ 0 \\ 0 \end{pmatrix} & A^2e_5^t &= \begin{pmatrix} 11 \\ 0 \\ 5 \\ 0 \\ 0 \\ 0 \end{pmatrix} & A^2e_6^t &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 7 \\ 0 \end{pmatrix} \end{aligned}$$

Busco una matriz permutación P de manera que los vectores $A^2e_i^t$ que tienen las mismas coordenadas no nulas sean consecutivos.

$$P = (e_1^t | e_3^t | e_2^t | e_6^t | e_4^t | e_5^t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

De este modo llegamos a que:

$$F = P^t A P = \begin{pmatrix} 0 & 0 & 2 & 3 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 0 \end{pmatrix}$$

La traspuesta de esta matriz es la Forma de Frobenius que buscamos:

$$F^t = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 & 0 & 0 \end{pmatrix}$$

Observación: Los vectores $A^2 e_i^t$ se agrupan de 2 en 2 según la posición de sus elementos no nulos y los bloques no nulos de F son de dimensión 2×2 . El hecho de que se agrupen de 2 en 2 no es casualidad, este “2” es debido a que la máxima dimensión que alcanzan los vectores de la base canónica es 2. Si por ejemplo, la máxima dimensión que alcanzara fuese 3 entonces se agruparían de 3 en 3, y los bloques no nulos de F serían de dimensión 3×3 .

Comprobamos que efectivamente F^3 es una matriz diagonal por bloques, y que sus bloques son matrices primitivas (por ser matrices positivas).

$$F^3 = \begin{pmatrix} 11 & 53 & 0 & 0 & 0 & 0 \\ 5 & 27 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 10 & 0 & 0 \\ 0 & 0 & 32 & 22 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 & 5 \\ 0 & 0 & 0 & 0 & 37 & 31 \end{pmatrix}$$



APÉNDICE

En toda la sección 3.1 hemos supuesto A tiene rango máximo ($\text{rg}(A)=n$ si A es de dimensión $n \times n$), e indicamos que en otro caso la demostración sería más elaborada.

Aunque no vamos a demostrar el por qué pero si vamos a exponer que ocurre cuando la dimensión no es máxima. En este caso la dimensión de los vectores de la base no se estabiliza, si no que entra en un ciclo, va bajando y subiendo, de hecho las dimensiones que va alcanzando varían justo en “ $n-\text{rg}(A)$ ” unidades.

Vemos dos ejemplos:

EJEMPLO 1: Dada la matriz

$$B = \begin{pmatrix} 0 & 0 & 0,7 & 0,3 & 0 & 0 \\ 0 & 0 & 0,6 & 0 & 0,4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0,8 & 0,2 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Por su Forma de de Frobenius sabemos que tiene índice de impricidad 3, aunque eso no afecta a lo que queremos estudiar.

Analizo sólo las imágenes del vector e_1 , pero lo mismo ocurre con todos.

$$\begin{array}{ccccc}
 B e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0,8 \end{pmatrix} & B^2 e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0 \end{pmatrix} & B^3 e_1^t = \begin{pmatrix} 0,8 \\ 0,8 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & B^4 e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0,8 \end{pmatrix} & B^5 e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0 \end{pmatrix} \\
 B^6 e_1^t = \begin{pmatrix} 0,8 \\ 0,8 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & B^7 e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0,8 \end{pmatrix} & B^8 e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0 \end{pmatrix} & B^9 e_1^t = \begin{pmatrix} 0,8 \\ 0,8 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & B^{10} e_1^t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0,8 \end{pmatrix}
 \end{array}$$

Observamos que:

$$\begin{aligned}
 \dim(B^2 e_1^t) &= \dim(B^5 e_1^t) = \dim(B^8 e_1^t) = 3 \\
 \dim(B^3 e_1^t) &= \dim(B^6 e_1^t) = \dim(B^9 e_1^t) = 2 \\
 \dim(B e_1^t) &= \dim(B^4 e_1^t) = \dim(B^7 e_1^t) = \dim(B^{10} e_1^t) = 1
 \end{aligned}$$

Las dimensiones de las imágenes va rotando 1,3,2,1,3,2,1,.. La diferencia entre la mayor y menor dimensión es justo $n - \text{rg}(A) = 6 - 4 = 2$

EJEMPLO 2: Dada la matriz

$$C = \begin{pmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 9 & 2 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Aunque el índice de imprimicidad no nos afecta a lo que estamos analizando en este apartado, lo calculamos:

$$P(\lambda) = |C - \lambda I| = \lambda^5 - 71\lambda^3 + 240\lambda = \lambda^5 - 71\lambda^{5-2} + 240\lambda^{5-4}$$

$$\text{Índice de imprimicidad} = \text{M.c.d}(2,4) = 2$$

Analizamos lo que nos interesa, que es ver que ocurre con la dimensión de las imágenes de los vectores de la base. Como $n=5$ y $\text{rg}(A)=4$, los vectores de la base no alcanzan una dimensión fija, si no que está constantemente oscilando, subiendo y bajando una unidad. Lo vemos:

$$\begin{array}{ccccc}
C e_1^t = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 0 \end{pmatrix} & C^2 e_1^t = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 4 \\ 0 \end{pmatrix} & C^3 e_1^t = \begin{pmatrix} 0 \\ 28 \\ 8 \\ 0 \\ 4 \end{pmatrix} & C^4 e_1^t = \begin{pmatrix} 44 \\ 0 \\ 0 \\ 284 \\ 0 \end{pmatrix} & C^5 e_1^t = \begin{pmatrix} 0 \\ 1988 \\ 88 \\ 0 \\ 234 \end{pmatrix} \\
C e_2^t = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 9 \\ 0 \end{pmatrix} & C^2 e_2^t = \begin{pmatrix} 0 \\ 63 \\ 2 \\ 0 \\ 9 \end{pmatrix} & C^3 e_2^t = \begin{pmatrix} 67 \\ 0 \\ 0 \\ 607 \\ 0 \end{pmatrix} & C^4 e_2^t = \begin{pmatrix} 0 \\ 4249 \\ 134 \\ 0 \\ 607 \end{pmatrix} & \\
C e_3^t = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \\ 0 \end{pmatrix} & C^2 e_3^t = \begin{pmatrix} 0 \\ 14 \\ 4 \\ 0 \\ 2 \end{pmatrix} & C^3 e_3^t = \begin{pmatrix} 22 \\ 0 \\ 0 \\ 142 \\ 0 \end{pmatrix} & C^4 e_3^t = \begin{pmatrix} 0 \\ 994 \\ 44 \\ 0 \\ 142 \end{pmatrix} & \\
C e_4^t = \begin{pmatrix} 0 \\ 7 \\ 0 \\ 0 \\ 1 \end{pmatrix} & C^2 e_4^t = \begin{pmatrix} 7 \\ 0 \\ 0 \\ 67 \\ 0 \end{pmatrix} & C^3 e_4^t = \begin{pmatrix} 0 \\ 469 \\ 14 \\ 0 \\ 67 \end{pmatrix} & C^4 e_4^t = \begin{pmatrix} 497 \\ 0 \\ 0 \\ 4517 \\ 0 \end{pmatrix} & \\
C e_5^t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4 \\ 0 \end{pmatrix} & C^2 e_5^t = \begin{pmatrix} 0 \\ 28 \\ 0 \\ 0 \\ 4 \end{pmatrix} & C^3 e_5^t = \begin{pmatrix} 28 \\ 0 \\ 0 \\ 268 \\ 0 \end{pmatrix} & C^4 e_5^t = \begin{pmatrix} 0 \\ 1876 \\ 56 \\ 0 \\ 268 \end{pmatrix} &
\end{array}$$

Efectivamente, como esperabamos la dimensión no se hace fija si no que oscila 2,3,2,3,2,3....

CONCLUSIONES

Hemos obtenido una demostración del teorema de Perron- Frobenius a partir de conocimientos elementales de geometría, usando el teorema del punto fijo de Brouwer y estudiando como actúa A sobre los subconjuntos de R^n . Posteriormente, hemos demostrado que toda matriz imprimitiva se puede expresar en la forma dada por Frobenius, para ello hemos definido un nuevo concepto de dimensión, y hemos visto como va variando la posición de los vectores de la base canónica de R^n al aplicarle A sucesivas veces.

Una de las continuaciones naturales de este trabajo sería extender a matrices de rango no máximo la demostración de cómo se llega a su Forma de Frobenius. Para ello, habría que analizar como actúa A sobre los vectores de la base canónica cuando A no es de rango máximo, lo cuál no se ha incluido porque supone trabajar con argumentos más elaborados. No obstante, si hemos incluido como apéndice de este trabajo una breve idea de lo que ocurre en estos casos.

Otra de las posibles ampliaciones del trabajo sería analizar la relación existente entre el número de bloques no nulos que tiene la Forma de Frobenius de una matriz y las características de la matriz. Llegaríamos a la conclusión de que este número de bloques coincide con el índice de imprimitividad de la matriz.

Para mi ha sido muy gratificante realizar este trabajo. Considero que en los últimos años me había alejado bastante del verdadero mundo de las matemáticas y me he disfrutado mucho con este reencuentro.

BIBLIOGRAFÍA

- [1] Alberto Borobia y Ujué R. Trías, *Revista Matemática de la Universidad Complutense de Madrid* vol.5 n°1
- [2] Carl D. Meyer, *Matrix Analysis and applied Linear Algebra*(ISBN-0: 89871-454-0)
- [3] Regino Criado, Miguel Romance y Luis E. Solá, *La Gaceta de la RSME*, Vol. 17 (2014), Núm. 3, Págs. 485 - 514
- [4] Pablo Fernández, *Bol. Soc. Esp. Mat. Apl.* n°30(2004), Págs.115-141.