TEMA 3. CONCEPTOS BÁSICOS DE ÁLGEBRA LINEAL

28 de octubre de 2002

CAPÍTULO 3

CONCEPTOS BÁSICOS DE ÁLGEBRA LINEAL

En este tema recordaremos los conceptos básicos de álgebra lineal que han sido expuestos al alumno en cursos anteriores e introduciremos una notación uniforme para los mismos [4, 5, 6]. Además, observaremos como muchos de dichos conceptos son también aplicables a espacios de funciones, que serán usados en este curso en los temas de aproximación de funciones.

Tras una aplicación simple del producto de matrices en ingeniería, para el modelado de las pérdidas (disipación) en líneas de transmisión, pasaremos a repasar el concepto de espacio vectorial, dependencia lineal y bases. Presentaremos, además de los espacios vectoriales $\mathbb{R}^n$ y $\mathbb{C}^n$, espacios de funciones como $\mathbb{P}^n(a,b)$, los polinomios de grado a lo sumo n definidos en el intervalo (a,b), y $\mathbf{C}^0[a,b]$, las funciones reales de variable real continuas en $[a,b]$. Estos espacios vectoriales tienen dimensión infinita. Es importante notar que todo espacio vectorial tiene base, tanto si es de dimensión finita como infinita, al menos si aceptamos el axioma de elección [1], y así haremos en este curso. Hay matemáticos, los constructivistas, que no aceptan dicho axioma, en cuyo caso sólo utilizan los espacios vectoriales para los que se puede construir explícitamente una base [2].

Desde el punto de vista de los métodos numéricos es muy importante medir el error de un algoritmo incluso cuando la respuesta es un vector. Para medir el tamaño de un vector se suele introducir una norma en un espacio vectorial, y se habla de espacios normados, destacando entre ellos los espacios de Banach. Todas las normas definidas en espacios normados de dimensión finita son equivalentes entre sí, no así en dimensión infinita.

Para medir tamaños, también se puede introducir un producto interior, que conduce automáticamente a una norma asociada, y que además nos permite calcular ángulos entre vectores. En espacios con producto interior podemos definir un concepto de ortogonalidad, y construir bases ortogonales. Los coeficientes de un vector respecto a una base ortogonal son sus coeficientes de Fourier, que utilizaremos mucho en teoría de la aproximación de funciones.

Introduciremos también las matrices, como representaciones de aplicaciones lineales. Estudiaremos sus tipos más importantes, cómo se opera con ellas, y sus propiedades. Estudiaremos la resolución de sistemas lineales, el determinante, la inversa de una matriz, su traza, sus autovalores y autovectores. Una matriz se puede escribir en diferentes formas canónicas, de entre las que destaca la descomposición de Schur, la forma de Jordan y la descomposición en valores singulares.

Finalmente introduciremos el concepto de normas matriciales, y presentaremos las normas más utilizadas. Muchos de los conceptos y teoremas del álgebra lineal que presentaremos en este tema se pueden encontrar en prácticamente todos los libros de análisis numérico, como [3], así como en la mayoría de los libros de álgebra y geometría lineal [4, 5, 6].

3.1 Aplicaciones en ingeniería

Hay muchos problemas físicos y aplicaciones en ingeniería que se modelan mediante problemas lineales y cuyo estudio requiere la solución de un sistema de ecuaciones lineales. Con objeto de concretar presentaremos un ejemplo de un circuito eléctrico pasivo, una línea de transmisión.

Para el modelado de la transmisión de electricidad en un cable se utilizan líneas de transmisión lineales tanto distribuidas (modelo continuo) como de parámetros concentrados (modelo discreto). En la figura 3.1 aparece una fotografía de un cable coaxial (línea de transmisión), una sección del cual se puede modelar mediante el circuito de parámetros concentrados que aparece en la figura 3.2, donde se ha considerado sólo la propagación de señales de baja frecuencia, por lo que se ha despreciado la capacitancia y la inductancia por unidad de longitud, y sólo se considera la resistencia y la conductancia por unidad de longitud. Aplicando las leyes de Kirchoff y la ley de Ohm ($V = I\,R$) se obtienen fácilmente las ecuaciones de esta línea con múltiples etapas.

Sin embargo, es usual modelar este problema mediante la técnica de la matriz de impedancias [Z], método por el cuál se consideran una serie de etapas elementales, denominadas multipuertos, cajas de dos entradas y dos salidas, que se concatenan para formar la línea completa. Para esta línea tenemos tres cajas como las mostradas en la figura 3.3, que son de dos tipos, multipuertos con una resistencia y con una conductancia.

La figura 3.4 muestra un multipuerto con una conductancia. De la ley de Ohm, sabemos que $V_1 = I\,R = V_2$, y de las leyes de Kirchoff de los nudos que

$$I_1 = I_2 + I, \qquad I_2 = I_1 - I = I_1 - V_1\,G,$$

Figura 3.1. Cable coaxial (línea de transmisión) modelo FLC78–50J de la compañía Harris Corporation. © Harris Corporation.

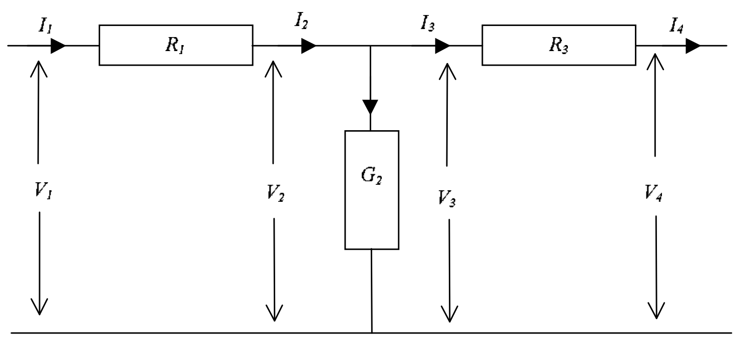

Figura 3.2. Modelo pasivo de parámetros concentrados de una línea de transmisión lineal que modela el cableado de una casa.

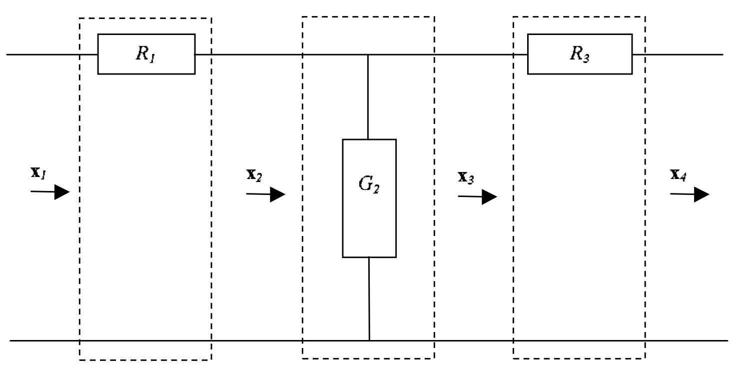

Figura 3.3. Modelo de una línea de transmisión lineal dividido en tres etapas elementales.

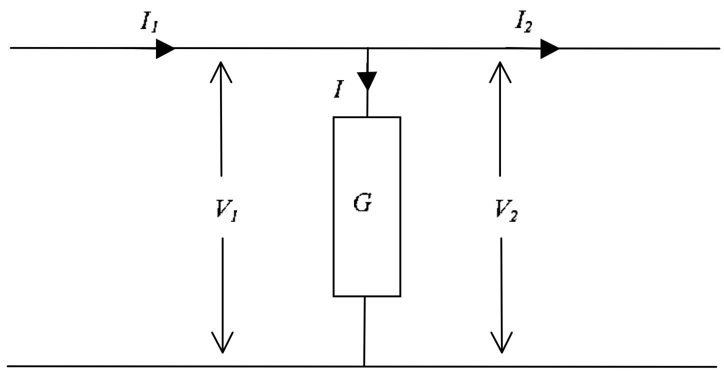

Figura 3.4. Etapa de una línea de transmisión con una conductancia.

que se puede escribir de forma matricial como

$$\begin{pmatrix} V_2 \\ I_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -G & 1 \end{pmatrix} \begin{pmatrix} V_1 \\ I_1 \end{pmatrix},$$

y vectorialmente como $\mathbf{x}_2 = P(G)\,\mathbf{x}_1$.

Por otro lado, la figura 3.5 muestra un multipuerto con una resistencia. Observamos que $I_1 = I_2$, $V_1 - V_2 = R\,I$, y $I_2 = V_1 - R\,I_1$, con lo que obtenemos en forma matricial

$$\begin{pmatrix} V_2 \\ I_2 \end{pmatrix} = \begin{pmatrix} 1 & -R \\ 0 & 1 \end{pmatrix} \begin{pmatrix} V_1 \\ I_1 \end{pmatrix},$$

y vectorialmente como $\mathbf{x}_2 = Q(R)\,\mathbf{x}_1$.

La red de la figura 3.3 se puede escribir fácilmente utilizando las expresiones de cada una de las etapas como

$$\mathbf{x}_2 = Q(R_1)\,\mathbf{x}_1, \qquad \mathbf{x}_3 = P(G_2)\,\mathbf{x}_2, \qquad \mathbf{x}_4 = Q(R_3)\,\mathbf{x}_3,$$

y finalmente, $\mathbf{x}_4 = Q(R_3)\,P(G_2)\,Q(R_1)\,\mathbf{x}_1$, que conduce a

$$\mathbf{x}_4 \equiv \begin{pmatrix} V_4 \\ I_4 \end{pmatrix} = \begin{pmatrix} 1 & -R_3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -G_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -R_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} V_1 \\ I_1 \end{pmatrix} \equiv \mathbf{x}_1$$

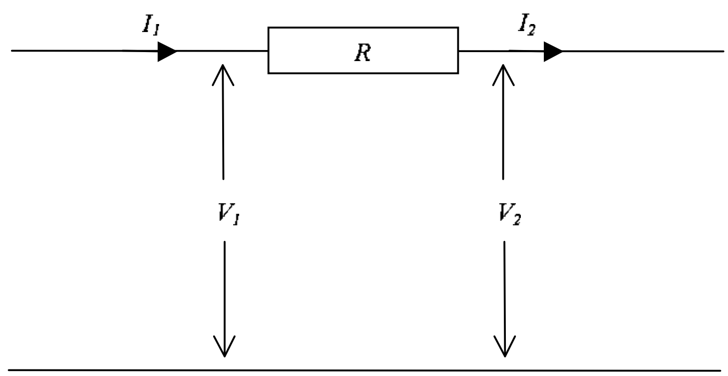

Figura 3.5. Etapa de una línea de transmisión con sólo una resistencia en serie.

que multiplicando las matrices nos da finalmente

$$\mathbf{x}_4 = \begin{pmatrix} 1 + G_2\, R_3 & -R_1 - R_3 - R_1\, G_2\, R_3 \\ -G_2 & 1 + R_1\, G_2 \end{pmatrix} \mathbf{x}_1.$$

3.2 Espacios vectoriales, métricos, normados y con producto interior

3.2.1 Espacios vectoriales

Un espacio vectorial V sobre un cuerpo de escalares $\mathbb{K}$, está formado por un conjunto de vectores, y dos operaciones, la suma de vectores, que dota a V de la estructura de grupo conmutativo, y la multiplicación de vectores por escalares, que es distributiva respecto de la suma[1].

Matemáticamente, V es un espacio vectorial sobre un cuerpo $\mathbb{K}$ (normalmente usaremos $\mathbb{R}$ o $\mathbb{C}$), dotado de dos operaciones binarias, la suma de vectores y el producto por escalar, y se

[1]Los espacios $\mathbb{R}^2$ y $\mathbb{R}^3$ fueron introducidos por los franceses Pierre de Fermat (1601–1665) y René Descartes (1596–1650) alrededor de 1636 en el marco de la geometría. El concepto de vector fue introducido por el polaco Bernard Placidus Johann Nepomuk Bolzano (1781–1848) en 1804, para formalizar la geometría. La definición axiomática de espacio vectorial, y con ella de álgebra lineal, prácticamente en su forma actual, es debida al italiano Giuseppe Peano (1858–1932) en 1888, quien indicó que basó sus ideas en los trabajos de los sajones (ahora serían alemanes) Gottfried Wilhelm von Leibniz (1646–1716) y August Ferdinand Möbius (1790–1868), del prusiano (ahora sería polaco) Hermann Günter Grassmann (1809–1877), y del irlandés Sir William Rowan Hamilton (1805–1865).

denota $(V,+,\cdot)$ donde

$$+ : V \times V \longrightarrow V, \qquad \cdot : \mathbb{K} \times V \longrightarrow V,$$

verificando las siguientes propiedades

1. $v_1 + v_2 = v_2 + v_1, \qquad \forall v_1, v_2 \in V,$
2. $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3), \qquad \forall v_1, v_2, v_3 \in V,$
3. $\exists\,!\,0 \in V, \qquad 0 + v = v + 0 = v, \qquad \forall v \in V,$
4. $\forall v \in V, \quad \exists\,! -v \in V, \quad v + (-v) = (-v) + v = 0,$
5. $1 \cdot v = v, \qquad 1 \in \mathbb{K}, \quad \forall v \in V,$
6. $\alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) \cdot v, \qquad \forall \alpha, \beta \in \mathbb{K}, \quad \forall v \in V,$
7. $\alpha \cdot (v_1 + v_2) = \alpha \cdot v_1 + \alpha \cdot v_2, \qquad \forall \alpha \in \mathbb{K}, \quad \forall v_1, v_2 \in V,$
8. $(\alpha + \beta) \cdot v) = \alpha \cdot v + \beta \cdot v, \qquad \forall \alpha, \beta \in \mathbb{K}, \quad \forall v \in V.$

Ejemplos típicos de espacios vectoriales son $\mathbb{R}^n$ sobre $\mathbb{R}$ ($\mathbb{C}^n$ sobre $\mathbb{C}$), es decir, los espacios de n-tuplas de números reales (complejos) con la adición componente a componente, y el producto por un escalar. También son espacios vectoriales sobre $\mathbb{R}$ (o sobre $\mathbb{C}$) el conjunto de los polinomios de grado menor o igual que n (sea $\mathbb{P}^n(a,b)$), el conjunto de las funciones continuas definidas en un intervalo $\mathbf{C}^0[a,b]$ y el conjunto de funciones de clase k en un intervalo, $\mathbf{C}^k(a,b)$, es decir, las funciones continuas con derivadas continuas hasta orden k, inclusive, en el intervalo abierto (a,b). Este resultado es fácil de demostrar. Más aún, el conjunto de funciones $L_p(a,b)$, definido como

$$L_p = \{f : \mathbb{R} \to \mathbb{R} : \int_a^b |f(x)|^p \, dx < \infty\}, \qquad p > 0,$$

también es un espacio vectorial. En este curso utilizaremos fundamentalmente L_2, el espacio de funciones de cuadrado integrable, L_1, el espacio de funciones de módulo integrable, y L_∞, que se puede demostrar que es el espacio de funciones acotadas (véanse los ejercicios resueltos).

Una combinación lineal del conjunto de n vectores v_1, v_2, $\ldots$, $v_n \in V$, denotado $\{v_i\} \subset V$, es una expresión de la forma

$$\sum_{i=1}^{n} \alpha_i \, v_i = \alpha_1 \, v_1 + \alpha_2 \, v_2 + \cdots + \alpha_n \, v_n,$$

donde $\{\alpha_i\} \in \mathbb{K}$.

Un conjunto de vectores $\{v_i\}$ se dice linealmente dependiente si existe alguna combinación lineal de los mismos igual al vector 0,

$$\alpha_1\, v_1 + \alpha_2\, v_2 + \cdots + \alpha_n\, v_n = 0,$$

con al menos un $\alpha_i \neq 0$. Sea $\alpha_i \neq 0$, entonces v_i se puede escribir como una combinación lineal de los demás v_j de la forma

$$v_i = -\frac{\alpha_1}{\alpha_i} v_1 - \cdots - \frac{\alpha_{i-1}}{\alpha_i} v_{i-1} - \frac{\alpha_{i+1}}{\alpha_i} v_{i+1} - \cdots - \frac{\alpha_n}{\alpha_i} v_n.$$

Un conjunto de vectores $\{v_i\}$ se dice linealmente independiente si

$$\sum_{i=1}^{n} \alpha_i\, v_i = 0, \quad \Rightarrow \quad \alpha_i = 0, \quad \forall i.$$

Un conjunto de vectores $\{e_i\} \subset V$, linealmente independiente, se dice que forma una base de V si todo vector $v \in V$ se puede escribir de forma única como

$$v = \sum \alpha_i\, e_i.$$

Se puede demostrar que todas las bases de un determinado espacio vectorial tienen el mismo número de vectores. El cardinal de cualquiera de estas bases se denomina dimensión del espacio vectorial. Existen espacios vectoriales de dimensión finita ($\mathbb{R}^n$ de dimensión n, o $\mathbb{P}^n(a,b)$ de dimensión $(n+1)$), y de dimensión infinita ($\mathbf{C}^0[a,b]$, $\mathbf{C}^k(a,b)$ o $L_p(a,b)$) [1].

Un concepto importante relacionado con espacios vectoriales es el concepto de dualidad. Se denomina espacio dual, V^*, del espacio vectorial V al espacio vectorial definido por todas las formas lineales en V, es decir,

$$V^* = \{f : V \to \mathbb{K} : f(\alpha_1\, v_1 + \alpha_2\, v_2) = f(\alpha_1\, v_1) + f(\alpha_2\, v_2), \forall v_1, v_2 \in V, \forall \alpha_1, \alpha_2 \in \mathbb{K}\}.$$

Para V de dimensión finita, se puede demostrar que $\dim(V) = \dim(V^*)$ y además, $V = (V^*)^*$. Más aún, ambos espacios son isomorfos [1][2]. Sin embargo, si V es de dimensión infinita, ninguna de las afirmaciones es cierta[3].

[2] Aunque el alumno debe conocer el concepto de isomorfismo, no entremos en detalles en este curso.

[3] Inciso técnico: Estos espacios se denominan espacios vectoriales topológicos. El ejemplo más conocido es el espacio vectorial topológico de las funciones, reales de variable real, continuas e infinitamente derivables de forma continua, con soporte compacto, llamado $\mathcal{S} = \mathbf{C}_c^\infty$; su dual $\mathcal{S}^*$ es el espacio de Schwartz de distribuciones, que incluye a funciones generalizadas tan extrañas como la función delta de Dirac y todas sus infinitas derivadas. A estas funciones, el francés Laurent Schwartz (1915-2002) las denominó distribuciones.

3.2.2 Distancias y espacios métricos

Sea X un conjunto no vacío a cuyos elementos llamaremos puntos. Una distancia en X es una aplicación $d : X \times X \longrightarrow \mathbb{R}$ que verifica

1. $d(x, y) = 0 \Leftrightarrow x = y$, $\forall x, y \in X$,
2. $d(x, y) + d(x, z) \geq d(y, z)$, $\forall x, y, z \in X$,

donde a 2 se le llama desigualdad triangular. Al par $M = (X, d)$ se le llama espacio métrico[4].

Consecuencia de la definición son estas otras propiedades (que el lector puede verificar fácilmente)

3. $d(x, y) \geq 0$, $\forall x, y \in X$,
4. $d(x, y) = d(y, x)$, $\forall x, y \in X$.

Para los espacios métricos que son espacios vectoriales, cuyos puntos son vectores, se suele utilizar el término espacios vectoriales topológicos. En este curso todos los espacios métricos que estudiaremos serán espacios vectoriales.

Como primer ejemplo de espacio métrico tomemos $\mathbb{R}^n$ con la distancia p discreta[5]

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}.$$

Demostremos que es realmente una distancia. Claramente cumple con la primera propiedad

$$d(x, y) = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n} |x_i - y_i|^p \quad \Leftrightarrow \quad |x_i - y_i| = 0 \quad \Leftrightarrow \quad x_i = y_i, \quad \forall i.$$

La segunda requiere el uso de varios lemas previos, la desigualdad de Young[6], que nos permite demostrar la desigualdad de Hölder, con la que finalmente demostraremos la desigualdad de Minkowski[7]. Ésta última nos da directamente el resultado a probar.

[4]El concepto de espacio métrico se desarrolló entre 1900 y 1910, cristalizando ideas que ya habían surgido en el siglo XIX. Sus hitos más importantes son la teoría de ecuaciones integrales del sueco Erik Ivar Fredholm (1866–1927) en 1900, la tesis sobre integración del francés Henri Léon Lebesgue (1875–1941) en 1902, la teoría espectral del prusiano (ahora sería ruso) David Hilbert (1862–1943) en 1906, y la tesis sobre espacios métricos del francés Maurice René Fréchet (1878–1973) en 1906.

[5]También se denomina distancia de Hölder, en honor al alemán Otto Ludwig Hölder (1859–1937). Para $p = 2$ coincide con la distancia euclídea, en honor a Euclides de Alejandría (≈325 AC–≈265 AC). Sorprendentemente, de Euclides se sabe muy poco: trabajó en Alejandría (Egipto), fue pupilo de Platón y escribió su famoso libro, en trece volúmenes, "Los Elementos", en el que axiomatiza la geometría (euclídea).

[6]Por el inglés William Henry Young (1863–1942).

[7]En honor al ruso (ahora sería lituano) Hermann Minkowski (1864–1909).

Lema 3.1 *Desigualdad de Young. Sean $\alpha, \beta \geq 0$, $p > 1$, y q conjugado a (o dual de) p, que significa que $1/p + 1/q = 1$, es decir, $q = p/(p-1)$. Entonces se verifica*

$$\alpha\,\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}.$$

Demostración. Utilizando cálculo elemental, fijando β, el máximo de la función

$$f(\alpha) = \alpha\,\beta - \frac{\alpha^p}{p},$$

se obtiene haciendo

$$f'(\alpha) = 0, \qquad \alpha = \beta^{1/(p-1)} = \beta^{q/p},$$

con lo que obtenemos

$$\alpha\,\beta - \frac{\alpha^p}{p} \leq \beta^{(p+q)/p} - \frac{\beta^q}{p} = \beta^q - \frac{\beta^q}{p} = \frac{\beta^q}{q},$$

con lo que queda demostrado.

Lema 3.2 *Desigualdad de Hölder en $\mathbb{R}^n$. Si $a, b \in \mathbb{R}^n$, entonces*

$$\sum_{i=1}^n |a_i\,b_i| \leq \left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \left(\sum_{i=1}^n |b_i|^q\right)^{1/q},$$

con $p > 1$ y $1/p + 1/q = 1$.

Demostración. Basta aplicar la desigualdad de Young a unos α y β convenientemente elegidos. Sean

$$\alpha = \frac{|a_i|}{\left(\sum_{i=1}^n |a_i|^p\right)^{1/p}}, \qquad \beta = \frac{|b_i|}{\left(\sum_{i=1}^n |b_i|^q\right)^{1/q}},$$

con lo que la desigualdad de Young conduce a

$$\frac{|a_i|}{\left(\sum_{i=1}^n |a_i|^p\right)^{1/p}} \frac{|b_i|}{\left(\sum_{i=1}^n |b_i|^q\right)^{1/q}} \leq \frac{1}{p} \frac{|a_i|^p}{\sum_{i=1}^n |a_i|^p} + \frac{1}{q} \frac{|b_i|^q}{\sum_{i=1}^n |b_i|^q},$$

que sumando para todos los índices

$$\frac{\sum_{i=1}^n |a_i\,b_i|}{\left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \left(\sum_{i=1}^n |b_i|^q\right)^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

y el resultado queda demostrado.

Lema 3.3 *Desigualdad de Minkowski en $\mathbb{R}^n$. Para $\mathbb{R} \ni a_i, b_i > 0$ y $p > 1$, entonces*

$$\left(\sum_{i=1}^{n}(a_i + b_i)^p\right)^{1/p} \leq \left(\sum_{i=1}^{n}(a_i)^p\right)^{1/p} + \left(\sum_{i=1}^{n}(b_i)^p\right)^{1/p}$$

Demostración. Con objeto de poder aplicar la desigualdad de Hölder, factorizaremos de la forma

$$\sum_{i=1}^{n}(a_i + b_i)^p = \sum_{i=1}^{n} a_i\,(a_i + b_i)^{p-1} + \sum_{i=1}^{n} b_i\,(a_i + b_i)^{p-1},$$

donde, para $1/p + 1/q = 1$, obtendremos

$$\sum_{i=1}^{n} a_i\,(a_i + b_i)^{p-1} \leq \left(\sum_{i=1}^{n}(a_i)^p\right)^{1/p} \left(\sum_{i=1}^{n}(a_i + b_i)^{(p-1)\,q}\right)^{1/q},$$

$$\sum_{i=1}^{n} b_i\,(a_i + b_i)^{p-1} \leq \left(\sum_{i=1}^{n}(b_i)^p\right)^{1/p} \left(\sum_{i=1}^{n}(a_i + b_i)^{(p-1)\,q}\right)^{1/q},$$

por lo tanto

$$\sum_{i=1}^{n}(a_i + b_i)^p \leq \left(\left(\sum_{i=1}^{n}(a_i)^p\right)^{1/p} + \left(\sum_{i=1}^{n}(b_i)^p\right)^{1/p}\right) \left(\sum_{i=1}^{n}(a_i + b_i)^{(p-1)\,q}\right)^{1/q},$$

que podemos simplificar notando que

$$\frac{1}{p} + \frac{1}{q} \quad \Rightarrow \quad p + q = p\,q \quad \Rightarrow \quad p = pq - q = (p-1)\,q,$$

y dividiendo por

$$\left(\sum_{i=1}^{n}(a_i + b_i)^p\right)^{1/q}$$

para obtener

$$\left(\sum_{i=1}^{n}(a_i + b_i)^p\right)^{1-1/q} \leq \left(\sum_{i=1}^{n}(a_i)^p\right)^{1/p} + \left(\sum_{i=1}^{n}(b_i)^p\right)^{1/p},$$

que es la desigualdad de Minkowski que queríamos demostrar.

Otro ejemplo de espacio métrico es el espacio de funciones continuas definidas en el intervalo $[a, b]$ valuadas en los reales, $\mathbf{C}([a, b], \mathbb{R})$, con la distancia p continua

$$d(f, g) = \left(\int_a^b |f(x) - g(x)|^p\, dx\right)^{1/p}.$$

El lector notará que esta definición es correcta ya que el teorema de Riemann[8] garantiza que las funciones continuas son integrables. En general podemos definir el espacio $L_p([a, b], \mathbb{R}) \equiv$

[8] En honor a Georg Friedrich Bernhard Riemann (1826–1866) que nació en Breselenz, Hanover (ahora sería alemán).

$L_p([a,b])$ como el espacio de funciones donde esta distancia está bien definida, tiene sentido (las funciones son integrables, incluso si no son continuas) y es finita para cualquier par de funciones de ese espacio[9]. Se demuestra que d es una distancia utilizando la desigualdad de Hölder para integrales

$$\int_a^b |f(x)\cdot g(x)|\,dx \le \left(\int_a^b |f(x)|^p\,dx\right)^{1/p} \cdot \left(\int_a^b |g(x)|^q\,dx\right)^{1/q}.$$

3.2.3 Sucesiones de Cauchy y espacios métricos completos

Se llama sucesión en un espacio métrico M a una aplicación $f : \mathbb{N} \longrightarrow M$. Normalmente se denota el elemento $f(n)$ de la forma x_n y a la sucesión como $\{x_n\}$.

El conjunto de las sucesiones de números reales $x \equiv \{x_i\}$ tales que

$$\sum_{i=1}^{\infty} |x_i|^p < \infty,$$

es un espacio métrico con la distancia

$$d(x,y) = \left(\sum_{i=1}^{\infty} |x_i - y_i|^p\right)^{1/p}.$$

Se denomina $l_p(\mathbb{R}) \equiv l_p$. Para demostrar que d es una distancia se utiliza la desigualdad de Hölder para sucesiones

$$\sum_{i=1}^{\infty} |x_i \cdot y_i| \le \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p} \cdot \left(\sum_{i=1}^{\infty} |y_i|^q\right)^{1/q}.$$

Se dice que la sucesión $\{x_n\} \subset M$ es convergente y tiene por límite a $x_0 \in M$, si para todo $\varepsilon > 0$, existe un $n_0 \in \mathbb{N}$, tal que

$$\forall n \ge n_0, \qquad d(x_n, x_0) < \varepsilon.$$

Se denota como $x_0 = \lim_{n\to\infty} x_n$ o como $x_n \to x_0$.

Una sucesión $\{x_n\} \subset M$ es de Cauchy[10] si dado $\epsilon > 0$, existe un $n_0 \in \mathbb{N}$ tal que

$$\forall n, m \ge n_0, \qquad d(x_n, x_m) < \varepsilon.$$

[9] Inciso técnico: aquí se debe utilizar la integración en el sentido de Lebesgue, quien generalizó la teoría de la integración de Riemann.

[10] En honor al francés Augustin-Louis Cauchy (1789–1857).

Claramente, toda sucesión convergente es de Cauchy, no así al contrario, ya que el "límite" podría estar fuera de M.

Un espacio métrico M es completo si toda sucesión de Cauchy es convergente, es decir, tiene un límite en M. La gran importancia de los espacios completos en análisis numérico es que nos permiten determinar si un método numérico iterativo converge comparando solamente sus iterados, sin necesidad de conocer el límite, la solución exacta que estamos calculando.

3.2.4 Normas de vectores y espacios normados

Para medir la magnitud (tamaño) de un escalar (número) en un cuerpo $\mathbb{K}$ se utiliza una valoración, que es una aplicación $|\cdot| : \mathbb{K} \longrightarrow \mathbb{R}$ verificando

1. $|\alpha| \geq 0, \qquad (|\alpha| = 0 \quad \Leftrightarrow \quad \alpha = 0), \qquad \forall \alpha \in \mathbb{K},$
2. $|\alpha + \beta| \leq |\alpha| + |\beta|, \qquad \forall \alpha, \beta \in \mathbb{K},$
3. $|\alpha \cdot \beta| \leq |\alpha| \cdot |\beta|, \qquad \forall \alpha, \beta \in \mathbb{K}.$

Un cuerpo $\mathbb{K}$ con una valoración se denomina valorado. Se denomina valor absoluto a una valoración tal que $|\alpha \cdot \beta| = |\alpha| \cdot |\beta|$.

Para medir la longitud (tamaño) de un vector en un espacio vectorial se suele utilizar una norma. Una función $\|\cdot\| : V \longrightarrow \mathbb{R}$ en un espacio vectorial V sobre un cuerpo valorado $\mathbb{K}$ es una norma si cumple[11]

1. $\|x\| \geq 0, \qquad (\|x\| = 0 \quad \Leftrightarrow \quad x = 0), \qquad \forall x \in V,$
2. $\|x + y\| \leq \|x\| + \|y\|, \qquad \forall x, y \in V,$ (llamada desigualdad triangular),
3. $\|\alpha\, x\| = |\alpha|\, \|x\|, \qquad \forall \alpha \in \mathbb{K}, \quad \forall x \in V.$

Un espacio vectorial dotado con una norma se denomina espacio normado.

En base a los axiomas de una norma se puede demostrar la siguiente desigualdad triangular inversa válida en todo espacio normado,

$$|\, \|x\| - \|y\| \,| \leq \|x - y\|, \qquad \forall x, y \in V.$$

[11] Inciso técnico: Cuando no se cumple la propiedad $\|x\| = 0 \Leftrightarrow x = 0$, decimos que $\|\cdot\|$ es una seminorma y hablamos de espacios semi-normados. Muchos espacios vectoriales topológicos de interés no son normados, pero sí semi-normados.

Demostración: Hay que probar que

$$-\|x-y\| \leq \|x\| - \|y\| \leq \|x-y\|.$$

La segunda desigualdad se deduce fácilmente

$$\|x\| = \|x-y+y\| \leq \|x-y\| + \|y\|.$$

La primera desigualdad se cambia de signo

$$\|x-y\| \geq \|y\| - \|x\|,$$

y también se demuestra fácilmente,

$$\|y\| = \|y-x+x\| \leq \|y-x\| + \|x\| = \|x-y\| + \|x\|.$$

En un espacio normado $(V, \|\cdot\|)$, la aplicación $d : V \times V \longrightarrow \mathbb{R}$, $d(u,v) = \|u-v\|$ es una distancia, que se denomina distancia asociada y que verifica $\forall u, v, w \in V$ y $\forall \alpha \in \mathbb{K}$,

1. $d(x,y) = d(x-z, y-z)$,
2. $d(\alpha\, u, \alpha\, v) = |\alpha|\, d(u,v)$.

Por ello, todo espacio vectorial normado es un espacio métrico, aunque no es cierto lo contrario, de hecho, muchos espacio métricos ni siquiera son espacios vectoriales.

Algunos ejemplos de normas en $\mathbb{R}^n$ son la norma euclídea o norma dos que se define como

$$\|x\|_2 = \sqrt{\sum_i |x_i|^2},$$

la norma uno

$$\|x\|_1 = \sum_i |x_i|,$$

la norma infinito o del máximo

$$\|x\|_\infty = \max_i |x_i|,$$

y, en general, la norma p ($p > 0$)

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}.$$

Entre los ejercicios resueltos se encuentra la demostración de que la norma $\|x\|_p$, cumple los axiomas de norma y además que

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty.$$

Se llama sucesión de vectores en un espacio vectorial V a una aplicación $f : \mathbb{N} \longrightarrow V$. Al elemento $f(n)$ se le suele denotar como x_n y a la sucesión como $\{x_n\}$. Normalmente se considera que los números naturales son 1, 2, ..., es decir, no se incluye el cero (0). El espacio de las sucesiones reales (o complejas) es un espacio vectorial. Cuando se le dota de una norma p, sea $\|\cdot\|_p$, se le denomina espacio l_p; cuando queremos destacar el cuerpo, se usa la notación $l_p(\mathbb{R})$ o $l_p(\mathbb{C})$.

El espacio de las funciones continuas $\mathbf{C}^0[a,b]$ (y por ende $\mathbb{P}^n(a,b)$ y $\mathbf{C}^k(a,b)$) es un espacio normado con la norma

$$\|f\|_\infty = \|f\|_{\max} = \max_{[a,b]} |f(x)|,$$

que está bien definida dado que el teorema de Weierstrass que garantiza que toda función continua tiene máximo y mínimo en un intervalo compacto, y además, cumple los axiomas de norma, como el lector puede comprobar fácilmente (basta utilizar las propiedades del valor absoluto).

Los espacios $L_p(a,b)$, son también espacios normados con la norma

$$\|f\|_p = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}, \qquad f \in L_p(a,b).$$

Una reescritura adecuada de la demostración de que las normas $\|x\|_p$ están bien definidas en $\mathbb{R}^n$, con ligeros cambios, se puede utilizar para demostrar que los espacios $L_p(a,b)$ son espacios normados, y que se cumple que

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty = \max_{[a,b]} |f(x)|, \qquad \forall f \in L_\infty(a,b).$$

El lector notará que $f \in L_\infty(a,b)$ implica que $f \in L_p(a,b)$.

Se define la bola unidad en norma-p, sea $B_{1,p}$, como el conjunto de vectores

$$B_{1,p} = x \in V \;:\; \|x\|_p \leq 1,$$

y esfera unidad en norma-p, sea $S_{1,p}$, como

$$S_{1,p} = \delta B_{1,p} = x \in V \;:\; \|x\|_p = 1.$$

El lector puede dibujar fácilmente la bola unidad en $\mathbb{R}^2$ para las normas 1, 2, ∞, 1/2 y 4.

Se dice que dos normas $\|\cdot\|_p$ y $\|\cdot\|_q$ son equivalentes si existen dos constantes positivas m y M tales que

$$m\,\|x\|_p \leq \|x\|_q \leq M\,\|x\|_p, \qquad \forall x \in V.$$

Se puede demostrar que en un espacio vectorial de dimensión finita todas las normas son equivalentes entre sí. Entre los ejercicios resueltos se encuentra la demostración de la equivalencia entre las normas dos, uno e infinito. La equivalencia entre normas garantiza que un vector pequeño en una norma lo es en cualquier otra, y lo mismo para uno grande. Sin embargo, es muy importante que el lector note que la equivalencia entre normas no es cierta en un espacio normado de dimensión infinita, una norma puede indicar que una función es pequeña, y otra que es grande. Esto será tenido en cuenta cuando estudiemos la teoría de aproximación de funciones.

Se dice que la sucesión $\{x_n\} \subset V$, donde V es un espacio normado con norma $\|\cdot\|$, tiene por límite (o converge) a $x \in V$ si

$$\forall \epsilon > 0, \quad \exists n_0 \in \mathbb{N}, \quad \text{tal que} \quad \forall n \geq n_0, \quad \|x - x_n\| < \epsilon.$$

Normalmente se escribe

$$x = \lim_{n\to\infty} x_n,$$

o a veces $x_n \to x$, y se dice que la sucesión converge a x.

Comprobar la convergencia de una sucesión mediante la definición anterior es complicado ya que requiere conocer el límite. Es mejor comprobar si la sucesión es de Cauchy. Una sucesión $\{x_n\} \subset V$ es de Cauchy si

$$\forall \epsilon > 0, \quad \exists n_0 \in \mathbb{N}, \quad \text{tal que} \quad \forall n, m \geq n_0, \quad \|x_n - x_m\| < \epsilon.$$

Toda sucesión convergente es de Cauchy, sin embargo, no así al contrario. Un espacio normado se dice completo si toda sucesión de Cauchy en dicho espacio es convergente (tiene límite). Se denomina espacio de Banach[12] a un espacio normado completo. En un espacio de Banach es fácil verificar si una sucesión converge, basta estudiar la distancia entre sus elementos conforme el índice de éstos crece.

3.2.5 Productos internos y espacios con producto interno

Otra manera de medir la distancia entre dos vectores o la tamaño de un vector es mediante un producto interior, también llamado producto interno o escalar[13]. Un producto interno (o escalar) en el espacio vectorial V sobre los complejos es una función, sea $\langle\cdot,\cdot\rangle : V \times V \longrightarrow \mathbb{C}$,

[12] En honor al austro-húngaro (ahora sería polaco) Stefan Banach (1892–1945).

[13] Los espacios de Hilbert fueron introducidos como espacios de funciones en 1904. La versión más abstracta fue introducida en 1908 por el alemán (ahora sería estonio) Erhard Schmidt (1876–1959) alumno de Hilbert. La versión axiomática actual es de la tesis doctoral de Banach de 1920.

que cumple las siguientes propiedades

1. $\langle x, x\rangle \geq 0, \qquad (\langle x, x\rangle = 0 \quad \Leftrightarrow \quad x = 0), \quad \forall x \in V,$

2. $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle, \qquad \forall x, y, z \in V,$

3. $\langle \alpha\, x, y\rangle = \overline{\alpha}\,\langle x, y\rangle, \qquad \forall x, y \in V, \quad \forall \alpha \in \mathbb{C}$

4. $\langle x, y\rangle = \overline{\langle y, x\rangle}, \quad \forall x, y \in V,$

donde $\overline{\alpha}$ indica el número complejo conjugado de α. Un producto interno (o escalar) sobre los reales, se define igual pero sin los complejos conjugados.

Asociada a todo producto interior podemos definir una norma $\|x\|^2 = \langle x, x\rangle$ que se denomina asociada o subordinada a dicho producto. Es fácil comprobar que la función así definida cumple los axiomas que caracterizan a una norma, por lo que se puede afirmar que todo espacio vectorial con producto interior es un espacio normado. Sin embargo, el recíproco no es cierto.

Como ejemplo de espacio con producto interior tenemos $\mathbb{C}^n$ con el denominado producto escalar hermítico,

$$\langle x, y\rangle = \sum_{i=1}^{n} \overline{x_i}\, y_i.$$

Este producto interior cumple todas las propiedades anteriores, como se puede comprobar fácilmente. En $\mathbb{R}^n$ se puede definir un producto interior de manera similar no aplicando el operador conjugado complejo, es decir, utilizando el producto escalar euclídeo

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i\, y_i,$$

que tiene asociada la norma euclídea

$$\|x\|_2 = \sqrt{\sum |x_i|^2} = \sqrt{\langle x, x\rangle}.$$

Los axiomas de un producto interior garantizan la verificación de las siguientes propiedades (donde la norma es la asociada a dicho producto interior),

1. $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle.$

2. $\langle x, \alpha\, y\rangle = \alpha\,\langle x, y\rangle.$

3. Desigualdad de Cauchy-Schwarz[14]

$$|\langle x, y\rangle|^2 \leq \|x\|_2^2 \, \|y\|_2^2.$$

Demostración en $\mathbb{C}$: Para todo $\alpha, \beta \in \mathbb{C}$ se cumple que

$$\begin{aligned} 0 &\leq \langle \alpha\, x + \beta\, y, \alpha\, x + \beta\, y\rangle \\ &= \alpha\,\overline{\alpha}\,\langle x, x\rangle + \overline{\alpha}\,\beta\,\langle x, y\rangle + \alpha\,\overline{\beta}\,\overline{\langle x, y\rangle} + \beta\,\overline{\beta}\,\langle y, y\rangle, \end{aligned}$$

tomando dos valores particulares de α y β,

$$\alpha = \langle y, y\rangle, \qquad \beta = -\overline{\langle x, y\rangle},$$

tenemos

$$0 \leq \|y\|_2^4\,\|x\|_2^2 - 2\,\|y\|_2^2\,|\langle x, y\rangle|^2 + \|y\|_2^2\,|\langle x, y\rangle|^2,$$

$$\|y\|_2^2\,|\langle x, y\rangle|^2 \leq \|y\|_2^4\,\|x\|_2^2;$$

para $y = 0$ la desigualdad original se cumple trivialmente y para $y \neq 0$, la obtenemos dividiendo por $\|y\|_2^2$.

Demostración en $\mathbb{R}$: Para todo $\alpha \in \mathbb{R}$ se cumple que

$$0 \leq \langle \alpha\, x + y, \alpha\, x + y\rangle = \alpha^2\,\langle x, x\rangle + 2\,\alpha\,\langle x, y\rangle + \langle y, y\rangle,$$

parábola en α cuyo mínimo (vértice) se encuentra en

$$\alpha = -\frac{2\,\langle x, y\rangle}{2\,\langle x, x\rangle},$$

que sustituyendo en la expresión anterior nos da

$$0 \leq \frac{\langle x, y\rangle^2}{\langle x, x\rangle} - 2\frac{\langle x, y\rangle^2}{\langle x, x\rangle} + \langle y, y\rangle,$$

o lo que es lo mismo, el resultado a demostrar

$$\langle x, y\rangle^2 \leq \langle x, x\rangle^2\,\langle y, y\rangle.$$

[14]Hermann Amandus Schwarz (1843–1921) la publicó para integrales sobre superficies en un artículo en 1885; sin embargo, la desigualdad para integrales ya había sido publicada en 1821 por Cauchy, y en 1859 por Viktor Yakovlevich Bouniakowsky (1804–1899).

4. Desigualdad triangular
$$\|x\|_2 + \|y\|_2 \geq \|x+y\|_2.$$

Demostración en $\mathbb{C}$:

$$\begin{aligned}\|x+y\|_2^2 &= \langle x+y, x+y\rangle = \langle x,x\rangle + \langle x,y\rangle + \overline{\langle x,y\rangle} + \langle y,y\rangle \\ &= \|x\|_2^2 + \|y\|_2^2 + 2\,\mathrm{Re}\,\{\langle x,y\rangle\} \leq \|x\|_2^2 + \|y\|_2^2 + 2\,|\langle x,y\rangle| \\ &\leq \|x\|_2^2 + \|y\|_2^2 + 2\,\|x\|_2\,\|y\|_2 = (\|x\|_2 + \|y\|_2)^2,\end{aligned}$$

donde se ha utilizado la desigualdad de Cauchy-Schwarz.

5. Ley del paralelogramo (válida para normas asociadas a productores interiores)
$$\|x+y\|_2^2 + \|x-y\|_2^2 = 2\,\|x\|_2^2 + 2\,\|y\|_2^2.$$

Demostración:

$$\begin{aligned}\|x+y\|_2^2 + \|x-y\|_2^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle + \langle x,x\rangle - \langle y,x\rangle - \langle x,y\rangle + \langle y,y\rangle \\ &= 2\,\langle x,x\rangle + 2\,\langle y,y\rangle = 2\,\|x\|_2^2 + 2\,\|y\|_2^2.\end{aligned}$$

Los espacios con producto interior que con su norma asociada son de Banach se denominan espacios de Hilbert. En estos espacios todas las sucesiones de Cauchy son convergentes y por ello son los espacios preferidos en análisis numérico (siempre y cuando puedan ser utilizados).

La existencia en $\mathbb{R}^n$ de un producto interior nos permite definir el ángulo entre dos vectores en dicho espacio, que se define como
$$\theta = \arccos \frac{\langle x,y\rangle}{\|x\|_2\,\|y\|_2}.$$
La desigualdad de Cauchy-Schwarz garantiza la corrección de esta definición
$$|\cos\theta| = \frac{|\langle x,y\rangle|}{\|x\|_2\,\|y\|_2} \leq 1.$$

Los productos interiores permiten introducir el concepto de ortogonalidad en espacios vectoriales. Dos vectores son ortogonales si $\langle x,y\rangle = 0$, y si además son unitarios $\|x\|_2 = \|y\|_2 = 1$ se dice que son ortonormales. Un sistema (conjunto) de vectores de V es ortogonal (ortonormal) si sus elementos son ortogonales (ortonormales) dos a dos.

Se puede demostrar que un conjunto ortogonal es un conjunto de vectores linealmente independientes. Además, en un espacio vectorial de dimensión finita n, un conjunto de n vectores linealmente independientes definen unívocamente una base de dicho espacio y, por tanto, un conjunto de n vectores ortogonales también.

Dada una base ortonormal $\{e_i\}$ de V, se puede escribir cualquier vector $x \in V$ en la forma

$$x = \sum_i \alpha_i \, e_i = \sum_i \langle e_i, x \rangle e_i,$$

ya que $\langle e_j, x \rangle = \alpha_j$ por ser $\langle e_i, e_j \rangle = \delta_{ij}$, la delta de Kronecker[15]. A este desarrollo se le denomina, en forma general, desarrollo de Fourier y a los coeficientes α_i, coeficientes de Fourier.

Dada una base ortonormal $\{e_i\}$ y un vector $v \in V$, sus coeficientes de Fourier verifican la desigualdad de Bessel

$$\sum_i \alpha_i^2 \le \|v\|^2,$$

y además si el espacio es de dimensión finita se verifica la identidad de Parseval[16]

$$\sum_i \alpha_i^2 = \|v\|^2.$$

En un espacio de dimensión infinita, un sistema ortonormal se dice completo o total si verifica la identidad de Parseval.

El ejemplo clásico de serie de Fourier es la basada en polinomios trigonométricos. Sea $f : (0, T) \to \mathbb{R}$ una función real periódica de periodo T, $f(x+T) = f(x)$, que además sea de cuadrado integrable $f \in L_2(0, T)$. Tomemos la siguiente base de $L_2(0, T)$

$$\{w_n\} \equiv \{e^{\mathrm{i}\, k_n\, x}\} \equiv \{1, \cos k_n x, \sin k_n x\},$$

entonces se puede escribir

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \, e^{\mathrm{i}\, k_n\, x} = a_0 + \sum_{n=1}^{\infty} (a_n \, \cos k_n + b_n \, \sin k_n), \qquad k_n = \frac{2\pi}{T}\, n,$$

donde los coeficientes de Fourier son, en forma compleja,

$$c_n = \frac{1}{T} \int_0^T f(x)\, e^{-\mathrm{i}\, k_n\, x}\, dx, \qquad -\infty < n < \infty,$$

[15]Del prusiano (actualmente sería polaco) Leopold Kronecker (1823–1891), quien dijo “Dios hizo los enteros, el resto lo hizo el hombre”.

[16]También llamado teorema de Parseval, que el francés Marc-Antoine Parseval des Chênes (1755–1836), presentó en su tesis en 1799, y publicó en 1801.

o en forma real,

$$a_n = \frac{1}{T}\int_0^T f(x)\,\cos k_n x\,dx, \qquad 0 \leq n < \infty,$$

$$b_n = \frac{1}{T}\int_0^T f(x)\,\sin k_n x\,dx, \qquad 0 < n < \infty.$$

Se puede demostrar que el sistema ortogonal de polinomios trigonométricos es completo, por lo que se verifica la desigualdad de Parseval

$$\|f\| = \frac{1}{T}\int_0^T f^2\,dx = \sum_{i=-\infty}^{\infty} c_n^2$$

Todo sistema de vectores linealmente independientes en un espacio dotado de producto interior, se puede ortogonalizar (ortonormalizar) por el procedimiento de ortogonalización de Gram-Schmidt[17]. Sea un sistema de vectores $\{x_n\} \subset V$ linealmente independientes entre sí, el sistema $\{y_n\}$ construido de la forma

$$y_1 = x_1, \qquad y_n = x_n - \sum_{i=1}^{n-1} \frac{\langle x_n, y_i\rangle}{\|y_i\|^2}\, y_i, \quad n > 1,$$

es ortogonal, por construcción, ya que si tenemos calculados los y_i, $i = 1, 2, \ldots, n-1$, podemos desarrollar x_n en la base formada por éstos y obtener

$$x_n = \sum_{i=1}^{n-1} \frac{\langle x_n, y_i\rangle}{\|y_i\|^2}\, y_i + y_n,$$

de donde y_n es la "parte" de x_n ortogonal a los vectores anteriores. También se puede obtener un sistema ortonormal $\{z_n\}$ dado por

$$z_n = \frac{y_n}{\|y_n\|}.$$

3.3 Matrices

Una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{K})$, donde a veces escribiremos $A \in \mathbb{K}^{m\times n}$, está formada por una tabla de $m\cdot n$ escalares de un cuerpo $\mathbb{K}$ dispuestos en m filas y n columnas, para las que se han definido

[17] Desarrollada por el danés Ludvik Henrik Ferdinand Oppermann (1817–1883) en 1863, y utilizado por el también danés Jorgen Pedersen Gram (1850–1916), en su tesis doctoral sobre ecuaciones integrales (1879). Más tarde fue referenciado por el alemán (ahora sería estonio) Erhard Schmidt (1876–1959), en un artículo escrito en 1905 y publicado en 1907, que lo hizo popular. Sin embargo, el procedimiento era ya conocido por el francés Pierre-Simon Laplace (1749–1827), y fue utilizado en 1836 por el también francés Augustin-Louis Cauchy (1789–1857).

las operaciones de suma, producto, y producto por escalar[18].

Denotaremos al elemento de la i-ésima fila y de la j-ésima columna como a_{ij}, con el nombre de la matriz en minúsculas, o directamente como $(A)_{ij}$ indicando explícitamente éste. Lo habitual es definir una matriz por sus elementos y escribir $A \equiv (a_{ij})$ donde los elementos vienen dados por una tabla como la siguiente

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Definimos las siguientes operaciones binarias sobre matrices: suma

$$C = A + B, \qquad c_{ij} = a_{ij} + b_{ij}, \qquad A, B, C \in \mathcal{M}_{m\times n}(\mathbb{K}),$$

producto por escalar

$$C = \alpha\, A, \qquad c_{ij} = \alpha\, a_{ij}, \qquad A, C \in \mathcal{M}_{m\times n}(\mathbb{K}), \quad \alpha \in \mathbb{K}$$

y producto de matrices

$$C = A\, B, \qquad c_{ij} = \sum_{k=1}^{n} a_{ik}\, b_{kj}, \qquad A \in \mathcal{M}_{m\times n}(\mathbb{K}), \quad B \in \mathcal{M}_{n\times p}(\mathbb{K}), \quad C \in \mathcal{M}_{m\times p}(\mathbb{K}).$$

Se denomina 0 a la matriz nula, la que tiene todos sus elementos nulos. El espacio $\mathcal{M}_{m\times n}(\mathbb{K})$ es un espacio vectorial sobre el cuerpo $\mathbb{K}$. Se denomina I a la matriz identidad, la que tiene todos sus elementos diagonales iguales a uno (unitarios) y sus elementos no diagonales nulos, lo que denotaremos $(I)_{ij} = \delta_{ij}$, la delta de Kronecker.

El producto de matrices no es conmutativo $A\, B \neq B\, A$, aunque sí es distributivo respecto de la suma, por lo que el espacio de las matrices cuadradas $\mathcal{M}_{n\times n}(\mathbb{K})$ está dotado de una estructura de álgebra no conmutativa (con I, la matriz identidad y 0, la matriz nula).

En este curso nos limitaremos a matrices sobre el cuerpo de los reales ($\mathbb{R}$) o de los complejos ($\mathbb{C}$), que denotaremos por $\mathcal{M}_{m\times n}$ (sin especificar directamente el cuerpo de escalares).

[18]La teoría de las matrices fue descubierta en 1857 por el inglés Arthur Cayley (1821–1895) con objeto de simplificar la solución de sistemas de ecuaciones lineales simultáneos. Sin embargo, el nombre "matriz" fue utilizado por primera vez por el también inglés James Joseph Sylvester (1814–1897) en 1850.

3.3.1 Los vectores como matrices fila o columna

Los vectores de $\mathbb{R}^n$ (o $\mathbb{C}^n$) se pueden interpretar como matrices de dos formas diferentes, como vectores columna, $\mathcal{M}_{n\times 1}$, que es lo habitual, o como vectores fila, $\mathcal{M}_{1\times n}$. De esta forma se puede utilizar el producto matricial para (pre-) multiplicar una matriz por un vector columna,

$$w = A\,v, \qquad w \in \mathcal{M}_{m\times 1}, \quad A \in \mathcal{M}_{m\times n}, \quad v \in \mathcal{M}_{n\times 1},$$

o para (post-) multiplicar una matriz por un vector fila

$$w = v\,A, \qquad w \in \mathcal{M}_{1\times n}, \quad A \in \mathcal{M}_{m\times n}, \quad v \in \mathcal{M}_{1\times m}.$$

Para convertir vectores fila en vectores columna se utiliza el operador unario de trasposición de matrices, definido como

$$C = A^\top, \qquad c_{ij} = a_{ji}, \qquad A \in \mathcal{M}_{m\times n}, \quad C \in \mathcal{M}_{n\times m}.$$

El lector puede comprobar fácilmente que este operador se comporta, ante las operaciones sobre matrices, de la siguiente forma

$$(A+B)^\top = A^\top + B^\top, \qquad (A\,B)^\top = B^\top\,A^\top, \qquad (\alpha\,A)^\top = \alpha\,A^\top, \qquad \left(A^\top\right)^\top = A,$$

y que además podemos escribir

$$w = A\,v, \qquad w^\top = v^\top\,A^\top.$$

En este curso siempre representaremos los vectores como vectores columna, $v \in \mathcal{M}_{n\times 1}$, de forma que cuando queramos escribir un vector fila escribiremos $v^\top$. De esta forma, el producto interior (euclídeo) entre dos vectores (columna) en $\mathbb{R}^n$ se puede escribir como

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i\,y_i = x^\top\,y.$$

Para poder escribir el producto interior en $\mathbb{C}^n$ hemos de introducir otro operador unario para matrices, la traspuesta hermítica, también llamada conjugada, definida por

$$C = A^* = \bar{A}^\top, \qquad c_{ij} = \bar{a}_{ji}, \qquad A \in \mathcal{M}_{m\times n}, \quad C \in \mathcal{M}_{n\times m}.$$

donde la barra $\bar{A}$ indica la matriz conjugada, cuyos elementos son los complejos conjugados de los de A. Esta operación unaria se comporta de la siguiente forma respecto a las operaciones matriciales

$$(A+B)^* = A^* + B^*, \qquad (A\,B)^* = B^*\,A^*, \qquad (\alpha\,A)^* = \overline{\alpha}\,A^\top, \qquad (A^*)^* = A,$$

y permite escribir el producto interior en $\mathbb{C}^n$ de la forma

$$\langle x, y\rangle = \sum_{i=1}^{n} \overline{x_i}\, y_i = \overline{x}^\top y = x^*\, y,$$

donde hemos utilizado x^* como traspuesta conjugada del vector columna x.

3.3.2 Matrices como representación de aplicaciones lineales

Las matrices tienen una interpretación geométrica como aplicaciones lineales, funciones que, por ejemplo en el plano, transforman cada paralelogramo colocado en el origen en otro paralelogramo también colocado en el origen; en $\mathbb{R}^n$ el comportamiento es similar respecto a un hiper-paralepípedo. Estas transformaciones del paralelogramo incluyen rotaciones, simetrías especulares, escalados, cizalladuras (*shear maps*), y combinaciones de las anteriores.

Las transformaciones (funciones o aplicaciones) lineales entre espacios vectoriales de dimensión finita se representan mediante matrices[19]. Sean V y W dos espacios vectoriales con bases $\mathcal{B}_V = \{v_1, \ldots, v_n\}$ y $\mathcal{B}_W = \{w_1, \ldots, w_m\}$, respectivamente, y una aplicación lineal $f : V \longrightarrow W$, que transforma un vector $v \in V$ en un vector $w \in W$ de la forma

$$v = \sum_{j=1}^{n} \alpha_j\, v_j, \qquad f(v) = w = \sum_{i=1}^{m} \beta_i\, w_i.$$

La matriz $A \equiv (a_{ij})$ que se dice que representa a f se define como

$$f(v_j) = \sum_{i=1}^{m} a_{ij}\, w_i, \qquad 1 \le j \le n,$$

y permite calcular los coeficientes β_i de w en la base $\mathcal{B}_W$ a partir de los coeficientes α_j de v en $\mathcal{B}_V$; aprovechando la linealidad de f obtenemos

$$f(v) = \sum_{j=1}^{n} \alpha_j\, f(v_j) = \sum_{j=1}^{n} \alpha_j \sum_{i=1}^{m} a_{ij}\, w_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_j\, a_{ij} \right) w_i = \sum_{i=1}^{m} \beta_i\, w_i,$$

luego

$$\beta_i = \sum_{j=1}^{n} a_{ij}\, \alpha_j.$$

Llamando α y β a los vectores (columna) cuyos elementos son los coeficientes α_j y β_i, de los vectores v y w, respectivamente, podemos escribir, $f(v) = w$ como

$$\beta = A\, \alpha.$$

[19] La diferenciación entre operadores lineales y matrices, que los representan, se debe a Carvallo en 1891.

Dadas dos bases distintas de un mismo espacio vectorial, existe una aplicación lineal que realiza el cambio de base entre ellas. La matriz que representa este cambio de base puede ser aplicada directamente a las coordenadas de los vectores considerados éstos como vectores columna, como ya hemos visto, o directamente a los vectores base, utilizando la traspuesta de la matriz y considerando éstos como vectores fila. Sean $\{v_i\}$ y $\{w_i\}$ dos bases de V, de forma que un vector cualquiera puede representarse como

$$u = \sum_{j=1}^{n} \alpha_j \, v_j = \sum_{j=1}^{n} \beta_j \, w_j.$$

La representación de la aplicación cambio de base mediante una matriz $A = (a_{ij})$ nos indica, como ya hemos visto, que podemos transformar la coordenadas α_j y $beta_i$ de v en las dos bases de la forma

$$\beta_i = \sum_{j=1}^{n} a_{ij} \, \alpha_j, \qquad \beta = A \, \alpha.$$

Igualmente podemos considerar que se transforman los vectores base sin cambiar las coordenadas. Aplicando la transformación anterior

$$u = \sum_{i=1}^{n} \beta_i \, w_i = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \, \alpha_j \, w_i = \sum_{j=1}^{n} \alpha_j \sum_{i=1}^{n} a_{ij} \, w_i = \sum_{j=1}^{n} \alpha_j \, v_j,$$

con lo que obtenemos finalmente

$$v_j = \sum_{i=1}^{n} w_i \, a_{ij}, \qquad [v]^\top = [w]^\top A^\top,$$

donde $[v]$ y $[w]$ son matrices cuyas filas están definidas por las componentes de los vectores base $\{v_i\}$ y $\{w_i\}$, respectivamente.

3.3.3 Algunos tipos de matrices

Hay una serie de tipos de matrices que se obtienen cuando se encuentra algún tipo de patrón en la distribución de sus elementos no nulos.

Se denominan matriz cuadrada a la que tiene el mismo número de filas que de columnas, $A \in \mathcal{M}_{n\times n}$, y se dice que es una matriz de orden n. En caso contrario se dice que la matriz es rectangular.

Sean n vectores columna, $\{a_1, a_2, \ldots, a_n\}$, $a_i \in \mathbb{R}^n$. La matriz A cuyos vectores columna son los a_i se escribe como $A \equiv [a_1, a_2, \ldots, a_n]$. Dados n vectores fila, $\{a_1^\top, a_2^\top, \ldots, a_n^\top\}$, escribimos la matriz cuyos vectores fila son los $a_i^\top$ como $A \equiv [a_1^\top, a_2^\top, \ldots, a_n^\top]$.

Dada la base ortonormal canónica $\{e_i\}$ de $\mathbb{R}^n$, donde $(e_i)_j = \delta_{ij}$, y $\{i_1, i_2, \ldots, i_n\}$ una permutación de los números $1, 2, \ldots, n$, la matriz cuyas columnas son $P = [e_{i_1}, e_{i_2}, \ldots, e_{i_n}]$, se denomina matriz de perturbación. Esta matriz es la matriz identidad con sus columnas permutadas.

Una matriz D es diagonal si son nulos sus elementos no diagonales, $d_{ij} = 0$, $i \neq j$. Se denomina diagonal principal al vector formado por los elementos d_{ii}.

Una matriz es triangular superior (o inferior) si sus elementos por debajo (o encima) de la diagonal principal son nulos. Para una matriz U triangular superior, $u_{ij} = 0$, $i > j$, y para una L triangular inferior, $l_{ij} = 0$, $i < j$. Una matriz triangular (superior o inferior) se denomina unitaria si todos los elementos de su diagonal principal son iguales a 1.

Los elementos de la diagonal inmediatamente inferior (o superior) a la diagonal principal se llaman elementos de la subdiagonal (superdiagonal) principal. Se denomina matriz bidiagonal inferior (superior) a la matriz cuyas únicos elementos no nulos están en la diagonal y subdiagonal (superdiagonal) principales. Una matriz es tridiagonal si todos sus elementos no nulos están en la diagonal, subdiagonal y superdiagonal principales.

Una matriz es de Hessenberg superior (o inferior) si sus elementos por debajo de la subdiagonal (o encima de la superdiagonal) son nulos, es decir, $a_{ij} = 0$ si $i > j + 1$ (o $j > i + 1$).

Se denomina producto exterior de dos vectores $x, y \in \mathbb{C}^n$ a la matriz cuadrada $A = x\, y^*$, cuyos elementos son $a_{ij} = x_i\, \overline{y_j}$. En $\mathbb{R}^n$ se tiene $A = x\, y^\top$.

3.3.4 La traza, el determinante y la inversa

Se define la traza $\operatorname{tr}(A)$ de una matriz A de orden n como la suma de sus elementos de la diagonal principal

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

El determinante de una matriz cuadrada $\det(A)$ es una forma multilineal alternada asociada a dicha matriz que se puede definir de forma recursiva: para $n = 1$, sea $[a]$ la matriz de $\mathcal{M}_{1\times 1}$ cuyo único elemento es el escalar a, entonces definimos $\det([a]) = a$; para $n > 1$, sea A una matriz de $\mathcal{M}_{n\times n}$, definimos $\det(A) = \sum_{j=1}^{n} (-1)^{j+1}\, a_{1j}\, \det(A_{1j})$, donde A_{1j} es la matriz de $\mathcal{M}_{(n-1)\times(n-1)}$ que se obtiene suprimiendo la primera fila y la j-ésima columna de A.

A partir de la definición de determinante se pueden demostrar las siguientes propiedades:

1. $\det(A\, B) = \det(A)\, \det(B), \quad A, B \in \mathcal{M}_{n\times n}$,

2. $\det(\alpha\, A) = \alpha^n \det(A), \quad A \in \mathcal{M}_{n\times n}, \quad \alpha \in \mathbb{R}$ (o $\alpha \in \mathbb{C}$),
3. $\det(A^\top) = \det(A), \quad A \in \mathcal{M}_{n\times n}$,
4. $\det(A^*) = \overline{\det(A)}, \quad A \in \mathcal{M}_{n\times n}$.

La relación entre el determinante de la suma $\det(A+B)$ y los de los sumandos $\det(A)$ y $\det(B)$ no se conoce en el caso general[20].

Se dice que una matriz (cuadrada) A es no singular si $\det(A) \neq 0$. Dada una matriz no singular A, siempre es posible encontrar su matriz inversa, que es única y se denota por A^{-1}, que cumple que $A\, A^{-1} = A^{-1}\, A = I$.

Utilizando la definición de inversa se pueden demostrar las siguientes propiedades:

1. $(A\, B)^{-1} = A^{-1}\, B^{-1}, \quad A, B \in \mathcal{M}_{n\times n}$,
2. $(\alpha\, A)^{-1} = \frac{1}{\alpha}\, A^{-1}, \quad A \in \mathcal{M}_{n\times n}, \quad \alpha \in \mathbb{R}$ (o $\alpha \in \mathbb{C}$,
3. $(A^\top)^{-1} = (A^{-1})^\top, \quad A \in \mathcal{M}_{n\times n}$,
4. $(A^*)^{-1} = (A^{-1})^*, \quad A \in \mathcal{M}_{n\times n}$,
5. $\det(A^{-1}) = 1/\det(A)$.

3.3.5 Sistemas de ecuaciones lineales

Gracias a la definición del producto de matrices, un sistema de ecuaciones lineales se puede escribir como $A\, x = b$, donde $A \in \mathcal{M}_{n\times n}$ se denomina matriz de coeficientes, $x \in \mathcal{M}_{n\times 1}$ es el vector de incógnitas y $b \in \mathcal{M}_{n\times 1}$ es el término no homogéneo. La existencia y unicidad de su solución viene dada por el teorema del rango o de Rouché-Frobenius, que se demuestra en los cursos de álgebra lineal [1].

Se define el rango de una matriz A, rango (A), como el número de filas (columnas) linealmente independientes de A, y coincide con la dimensión del subespacio vectorial $\{y = A\, x, \forall x \in \mathbb{R}^n\}$. Se puede demostrar que

[20] Más aún, es un problema muy difícil. El problema matemático aún sin resolver más importante del siglo XXI es la hipótesis (o conjetura) de Riemann. Este problema es equivalente a mostrar que $\det(D_n + C_n) = O\left(n^{1/2+\epsilon}\right)$, para todo $\epsilon > 0$, donde $D_n \in \mathcal{M}_{n\times n}$ es la matriz de divisores, cuyo elemento (i,j) es 1 si i es múltiplo de j, y 0 en otro caso, y $C_n \in \mathcal{M}_{n\times n}$ es la matriz cuyos elementos $(2,1)$, $(3,1)$, $\ldots$, $(n,1)$ son iguales a 1 y todos los demás nulos. El Instituto de Matemáticas Clay anunció en el 2000 que premiará con un millón de dólares a quien resuelva este problema.

1. rango $(A+B) \leq$ rango $(A)+$ rango $(B), \quad A, B \in \mathcal{M}_{n\times n}$,
2. rango $(\alpha A)=$ rango $(A), \quad \mathbb{R} \ni \alpha \neq 0, \quad A \in \mathcal{M}_{n\times n}$,
3. rango $(A B) \leq \min\{$rango (A), rango $(B)\}, \quad A, B \in \mathcal{M}_{n\times n}$,

El teorema del rango o de Rouché-Frobenius asegura que son equivalentes las siguientes sentencias:

1. El sistema de n ecuaciones lineales $A\,x = b$ tiene solución única.
2. El sistema lineal homogéneo $A\,x = 0$ tiene como única solución $x = 0$.
3. Existe la inversa A^{-1} de A.
4. $\det(A) \neq 0$, es decir, A es no singular.
5. rango $(A) = n$.

Como corolario de este teorema observamos que el sistema lineal de ecuaciones homogéneo, $A\,x = 0$, tiene una solución no trivial ($x \neq 0$) si y solamente si $\det(A) = 0$.

3.3.6 Tipos fundamentales de matrices

En este curso usaremos fundamentalmente matrices cuadradas. Una matriz cuadrada se dice simétrica si $A = A^\top$, hermítica[21] si $A = A^*$, antisimétrica si $A^\top = -A$ y antihermítica si $A = -A^*$. Las matrices hermíticas (simétricas) son muy importantes debido a su actuación dentro de un producto interior complejo de la forma

$$\langle A\,x, y\rangle = (A\,x)^*\,y = x^*\,A^*\,y = \langle x, A^*\,y\rangle, \qquad A \in \mathcal{M}_{n\times n}(\mathbb{C}),$$

(o en el caso real,

$$\langle A\,x, y\rangle = (A\,x)^T\,y = x^T\,A^T\,y = \langle x, A^T\,y\rangle, \qquad A \in \mathcal{M}_{n\times n}(\mathbb{R})$$

donde se ha sustituido la traspuesta hermítica por la traspuesta). En un contexto más amplio, se dice que la matriz A^* (A^T) representa la aplicación dual o adjunta de la aplicación representada por A con respecto al producto interior complejo (real). El concepto de operador (diferencial) dual es muy importante ya que sustenta la teoría de Sturm-Liouville que el alumno ha estudiado en cursos anteriores y que revisaremos brevemente en un tema posterior.

[21] A veces se dice hermitiana, del inglés *Hermitian*.

Una matriz real se dice ortogonal si su inversa es igual a su traspuesta,

$$O^\top = O^{-1}, \qquad O^\top O = O\,O^\top = I.$$

De forma similar, una matriz compleja es unitaria si su inversa es igual a su traspuesta hermítica,

$$U^* = U^{-1}, \qquad U^*\,U = U\,U^* = I.$$

Los vectores filas y columnas de una matriz ortogonal (o unitaria) son ortonormales entre sí y su determinante es de módulo unitario.

Demostración en $\mathbb{C}$: Sea una matriz unitaria U con vectores columna u_i, y U^* su traspuesta hermítica, cuyos vectores fila son $u_i^* \equiv \overline{u}_i^\top$, es decir,

$$U = [u_1\; u_2\; \cdots\; u_n], \qquad U^* = \begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix},$$

donde hemos utilizado corchetes cuadrados para indicar una matriz definida por sus vectores. Esta notación nos permite escribir el producto $U^*\,U = I$ explícitamente multiplicando vectores fila por vectores columna de la forma

$$U^*\,U = \begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix} \cdot [u_1\; \cdots\; u_n] = I,$$

de donde se obtiene la condición de ortonormalidad $u_i^*\,u_j = \langle u_i, u_j\rangle = \delta_{ij}$, la delta de Kronecker. Además, aplicando determinantes a la definición de matriz unitaria,

$$1 = \det(I) = \det(U^*\,U) = \det(U^*)\,\det(U) = \overline{\det(U)}\,\det(U) = |\det(U)|^2,$$

luego el módulo del determinante es 1.

3.3.7 Autovalores y autovectores

Se dice que λ es un autovalor de una matriz A y que $x \neq 0$ es uno de sus autovectores asociados, si $A\,x = \lambda\,x$. Para determinar los autovalores se puede tener en cuenta el teorema del rango que afirma que el sistema $(A - \lambda\,I)\,x = 0$ tiene solución no trivial (distinta del vector 0), si y sólo si su determinante es nulo, $|A - \lambda I| = 0$. Este determinante es un polinomio de grado n en λ y se denomina polinomio característico; sus raíces son los autovalores de A. Los autovectores asociados a un autovalor dado forman un subespacio vectorial que se denomina autoespacio $S(\lambda)$,

$$S(\lambda) = \{x : A\,x = \lambda\,x\},$$

y cuya dimensión es $m_g(\lambda) = n - \text{rango}\,(A - \lambda I)$.

Los autovalores tienen una interpretación geométrica sencilla. Consideremos, para simplificar, el plano, y una matriz de 2×2, que representa una transformación geométrica en él. Aplicando dicha matriz a un círculo unidad centrado en el origen, éste se transforma en una elipse (si la matriz tiene rango 2). Los vectores que definen los ejes mayor y menor de la elipse corresponden a los dos autovectores de la matriz. Cada uno de estos autovectores corta al círculo en un punto. La matriz ha actuado sobre el vector que pasa por el origen y ese punto multiplicándolo por un número, éste es el autovalor correspondiente. Si las autovectores se usan para definir una base, la matriz de la transformación geométrica toma en dicha base una forma diagonal, la forma más simple posible.

Se denomina multiplicidad algebraica ($m_a(\lambda)$) de un autovalor λ a su multiplicidad como raíz del polinomio característico, es decir, al número de veces que está repetido como raíz de dicho polinomio. Se denomina multiplicidad geométrica ($m_g(\lambda)$) de un autovalor a la dimensión del autoespacio $S(\lambda)$ que tiene asociado, es decir, al número de autovectores linealmente independientes que están asociados a dicho autovalor. Se cumple siempre que $n \geq m_a \geq m_g$.

Los autovalores indican el "tamaño" de una matriz ya que

$$A\,x = \lambda\,x, \quad \Rightarrow \quad \langle A\,x, A\,x\rangle = |\lambda|^2\,\langle x, x\rangle \geq 0$$

por lo que

$$|\lambda|^2 = \frac{\|A\,x\|_2^2}{\|x\|_2^2},$$

y de este modo, el autovalor mide lo grande o pequeña que es una matriz cuando actúa sobre un vector de su autoespacio asociado.

Al conjunto de todos los autovalores de una matriz se le denomina espectro de dicha matriz, y al mayor de sus autovalores en valor absoluto radio espectral, que se escribe como

$$\rho(A) = \max\{|\lambda| : |A - \lambda I| = 0\}.$$

Se dice que dos matrices A y B son semejantes si existe una matriz P no singular ($|P| \neq 0$) tal que $B = P^{-1}\,A\,P$. En ese caso los determinantes de las dos matrices son iguales, $|A| = |B|$.

Las matrices semejantes tienen los mismos autovalores, ya que

$$A\,x = \lambda_A\,x, \quad \Rightarrow \quad P^{-1}\,A\,P\,P^{-1}\,x = \lambda_A\,P^{-1}\,x, \quad \Rightarrow \quad B\,P^{-1}\,x = \lambda_A\,P^{-1}\,x,$$

cuyo autovector es $y = P^{-1}\,x$. La traza, el determinante y el rango de una matriz son invariantes ante transformaciones de semejanza.

Se pueden calcular el determinante y la traza si se conocen todos sus autovalores, en concreto

$$\det(A) = \prod_{i=1}^{n} \lambda_i, \qquad \operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i,$$

donde $A\,x_i = \lambda_i\,x_i$.

Los autovalores de una matriz hermítica $A = A^*$ (en $\mathbb{C}$), o simétrica $A = A^\top$ (en $\mathbb{R}$), son reales, ya que en ese caso

$$\langle A\,x, x\rangle = \langle x, A^*\,x\rangle = \langle x, A\,x\rangle,$$

por lo que con x un autovector, $A\,x = \lambda\,x$,

$$\langle A\,x, x\rangle = \langle \lambda\,x, x\rangle = \overline{\lambda}\langle x, x\rangle = \langle x, A\,x\rangle = \langle x, \lambda\,x\rangle = \lambda\langle x, x\rangle,$$

por lo que $\lambda = \overline{\lambda}$ y $\lambda \in \mathbb{R}$.

Los autovectores de una matriz hermítica (o simétrica) correspondientes a dos autovalores distintos son ortogonales entre sí. Sean $A\,x_i = \lambda_i\,x_i$ y $A\,x_j = \lambda_j\,x_j$. Se tiene que para

$$0 = \langle A\,x_i, x_j\rangle - \langle x_i, A\,x_j\rangle = \lambda_i\langle x_i, x_j\rangle - \lambda_j\langle x_i, x_j\rangle = (\lambda_i - \lambda_j)\,\langle x_i, x_j\rangle,$$

ya que los autovalores son reales. Si los autovalores son distintos ($i \neq j$), es necesario que los autovectores sean ortogonales

$$\langle x_i, x_j\rangle = 0.$$

De esta forma los autovectores de una matriz hermítica definen un conjunto de n vectores ortogonales entre sí, es decir, cada matriz hermítica define una base ortonormal. **Cqd**.

Finalmente, recordaremos el teorema de Cayley-Hamilton[22], que dice que toda matriz A satisface su ecuación característica $|A - \lambda\,I| = p_A(\lambda) = 0$, es decir, $p_A(A) = 0$ donde $A^n = A^{n-1}\,A$. Demostrar este resultado no es fácil. Seguidamente, por completitud, presentamos dos demostraciones interesantes del mismo.

Demostración. Recordemos que un determinante se puede calcular utilizando su desarrollo respecto a cualquiera de sus filas de la forma

$$|B| = \sum_{j=1}^{n} (-1)^{j+1}\,b_{ij}\,|B_{ij}|.$$

[22]Probado por el inglés Arthur Cayley (1821–1895) en 1857.

Llamemos matriz de adjuntos de B, $\mathrm{Adj}(B)$, a la que tiene por elementos

$$(\mathrm{Adj}(B))_{ij} = (-1)^{j+1}\,|B_{ij}|,$$

de forma que el determinante se puede calcular como

$$|B| = \sum_{j=1}^{n} b_{ij}\,(\mathrm{Adj}(B))_{ij},$$

para cualquier fila i. Podemos escribir esta identidad de forma matricial

$$\det(B)\,I = |B|\,I = B\,\mathrm{Adj}(B).$$

Tomemos $B = A - \lambda\,I$, lo que da

$$p_A(\lambda)\,I = (A - \lambda\,I)\,\mathrm{Adj}(A - \lambda\,I). \tag{3.1}$$

Los elementos $(\mathrm{Adj}(A-\lambda\,I))_{ij}$ son polinomios en λ de grado a lo sumo $n-1$, por ser determinantes de matrices de $(n-1)\times(n-1)$. Podemos escribir estos polinomios en forma matricial, sea

$$\mathrm{Adj}(A - \lambda\,I) = B_0 + B_1\,\lambda + \cdots + B_n\,\lambda^{n-1},$$

donde B_j son matrices de números reales (independientes de λ). Si ahora escribimos

$$p_A(\lambda) = c_0 + c_1\,\lambda + \cdots + c_n\,\lambda^n,$$

la ecuación (3.1) nos da

$$c_0\,I + c_1\,\lambda\,I + \cdots + c_n\,\lambda^n\,I = (A - \lambda\,I)\,(B_0 + B_1\,\lambda + \cdots + B_n\,\lambda^{n-1}),$$

que igualando coeficientes en λ conduce a las identidades matriciales

$$\begin{aligned}
c_0\,I &= A\,B_0,\\
c_1\,I &= A\,B_1 - B_0,\\
c_2\,I &= A\,B_2 - B_1,\\
&\vdots\\
c_{n-1}\,I &= A\,B_{n-1} - B_{n-2},\\
c_n\,I &= -B_{n-1},
\end{aligned}$$

que buscando potencias de A nos dan

$$\begin{aligned} c_0\, I &= A\, B_0, \\ c_1\, A &= A^2\, B_1 - A\, B_0, \\ c_2\, A^2 &= A^3\, B_2 - A^2\, B_1, \\ &\vdots \\ c_{n-1}\, A^{n-1} &= A^n\, B_{n-1} - A^{n-1}\, B_{n-2}, \\ c_n\, A^n &= -A^n B_{n-1}, \end{aligned}$$

que sumadas en columnas nos da la expresión (matricial) deseada

$$p_A(A) = c_0\, I + c_1\, A + c_2\, A^2 + \cdots + c_n\, A^n = 0.$$

Otra demostración. Podemos utilizar el “truco del determinante”. El polinomio característico $p_A(\lambda) = |A - \lambda\, I| \equiv |B(\lambda)|$ se puede calcular desarrollando su primera fila de la forma,

$$p_A(\lambda) = |B(\lambda)| = \sum_{j=1}^{n} (-1)^{j+1}\, (a_{1j} - \lambda\, \delta_{1j})\, |B_{1j}(\lambda)|,$$

donde el determinante del menor $(-1)^{j+1}\, |B_{1j}(\lambda)| \equiv p_j(\lambda)$ es un polinomio de grado (a lo sumo) $n-1$. De esta forma

$$p_A(\lambda) = p_1(\lambda)(a_{11} - \lambda) + p_2(\lambda)(a_{12}) + \cdots + p_n(\lambda)(a_{1n}).$$

El “truco del determinante” consiste en recordar que en un primer curso de Álgebra Lineal se demostró que si desarrollamos un determinante a partir de una fila dada, pero utilizando los menores de otra, éste es automáticamente cero. De esta forma obtenemos las expresiones

$$\begin{aligned} 0 &= p_1(\lambda)(a_{21}) + p_2(\lambda)(a_{22} - \lambda) + \cdots + p_n(\lambda)(a_{2n}), \\ &\vdots \\ 0 &= p_1(\lambda)(a_{n1}) + p_2(\lambda)(a_{n2}) + \cdots + p_n(\lambda)(a_{nn} - \lambda). \end{aligned}$$

Ahora bien, si evaluamos formalmente esta serie de expresiones en A,

$$\begin{aligned}
p_A(A) &= p_1(A)(a_{11}\, I - A) + p_2(A)(a_{12}\, I) + \cdots + p_n(A)(a_{1n}\, I). \\
0 &= p_1(A)(a_{21}\, I) + p_2(A)(a_{22}\, I - A) + \cdots + p_n(A)(a_{2n}\, I), \\
&\vdots \\
0 &= p_1(A)(a_{n1}\, I) + p_2(A)(a_{n2}\, I) + \cdots + p_n(A)(a_{nn}\, I - A),
\end{aligned}$$

multiplicamos cada una de estas expresiones matriciales por cada uno de los vectores unitarios (en columnas) de la base canónica de $\mathbb{R}^n$, sea $\{e^i\}$, obtenemos

$$\begin{aligned}
p_A(A)\, e^1 &= p_1(A)(a_{11}\, e^1 - A\, e^1) + p_2(A)(a_{12}\, e^1) + \cdots + p_n(A)(a_{1n}\, e^1). \\
0\, e^2 = 0 &= p_1(A)(a_{21}\, e^2) + p_2(A)(a_{22}\, e^2 - A\, e^2) + \cdots + p_n(A)(a_{2n}\, e^2), \\
&\vdots \\
0\, e^n = 0 &= p_1(A)(a_{n1}\, e^n) + p_2(A)(a_{n2}\, e^n) + \cdots + p_n(A)(a_{nn}\, e^n - A\, e^n),
\end{aligned}$$

y sumamos todas estos ecuaciones por columnas, recordando que $A\, e^1$ es la primera columna de A, obtenemos que son nulos los factores que multiplican a los $p_j(A) = 0$, $1 \le j \le n$. Por ello $p_A(A)\, e^1 = 0$. Permutando cíclicamente los vectores de la base canónica es obvio que $p_A(A)\, e^2 = 0, \ldots, p_A(A)\, e^n = 0$, con lo que $p_A(A) = 0$ idénticamente.

3.3.8 Formas canónicas de matrices

La semejanza entre matrices permite definir una serie de formas canónicas para la escritura de las mismas.

1. **Forma normal de Schur**. Para toda matriz A existe una matriz unitaria U, con $U^{-1} = U^*$, tal que $T = U^*\, A\, U$ es una matriz triangular. El producto de los elementos de la diagonal de T es igual al determinante de A, es decir, $|A| = |T| = \prod_k t_{kk}$. El teorema de la forma normal de Schur será demostrado en los ejercicios resueltos.

2. **Descomposición en ejes principales**. La aplicación de la forma normal de Schur a una matriz hermítica A, nos indica que existe una matriz unitaria U tal que $U^*\, A\, U = \Lambda$ es una matriz diagonal, ya que la forma normal de Schur garantiza que $U^*\, A\, U = \Lambda$ es triangular superior, y que $(U^*\, A\, U)^* = \Lambda^*$ es triangular inferior, pero como A es hermítica,

$(U^* A U)^* = U^* A U$, luego Λ es diagonal. Los vectores columna (o fila) de U se denominan ejes principales de la matriz hermítica A, y permiten diagonalizarla.

Los elementos en la diagonal de la matriz Λ son los autovalores de A. También esto es fácil de demostrar. Podemos escribir la matriz U mediante sus vectores columna, sean u_i, $U = [u_1, u_2, \ldots, u_n]$. Como U es unitaria, sus vectores columna son ortonormales, y como $A U = U U^* A U = U \Lambda$ se puede escribir

$$A\,[u_1, \ldots, u_n] = [u_1, \ldots, u_n] \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

o en la notación de columnas,

$$[A\,u_1, A\,u_2, \ldots, A\,u_n] = [\lambda_1\,u_1, \lambda_2\,u_2, \ldots, \lambda_n\,u_n],$$

es decir, $A\,u_i = \lambda_i\,u_i$.

3. **Descomposición en valores singulares**. Toda matriz $A \in \mathcal{M}_{n\times m}$ se puede factorizar en la forma

$$A = V\,D\,U,$$

donde $V \in \mathcal{M}_{m\times m}$ y $U \in \mathcal{M}_{n\times n}$ son matrices unitarias y $D \in \mathcal{M}_{m\times n}$ es una matriz diagonal. Los $r = \text{rango}\,(A)$ valores $\mu_i > 0$ en la diagonal de la matriz D se denominan valores singulares, que coinciden con las raíces cuadradas de los valores propios de la matriz $A^* A$ (que es hermítica), es decir, $A^* A\,x_i = \mu_i^2\,x_i$. La demostración del teorema de la descomposición de en valores singulares de una matriz (rectangular) aparece en los ejercicios resueltos de este tema.

Los valores singulares se pueden interpretar geométricamente. Si A es una transformación lineal de un espacio euclídeo, entonces convierte la esfera unidad en ese espacio en un elipsoide de otro espacio euclídeo. Los valores singulares son las longitudes de los semi-ejes de este elipsoide. Las matrices V y U nos dan información sobre la posición de los ejes y sobre los vectores del primer espacio que se transforman en éstos, respectivamente.

La aplicación más importante de la descomposición en valores singulares es la selección de los términos dominantes (cuyos valores singulares son los más grandes) en una aplicación lineal o matricial. Es decir, permite calcular aproximaciones en un subespacio vectorial de una aplicación lineal en el correspondiente espacio vectorial.

4. **Forma canónica de Jordan**. Toda matriz cuadrada A de orden $n\times n$ es semejante a una matriz diagonal por bloques que tiene bloques de Jordan $J(\lambda_i, n_i)$ sobre la diagonal. Sean $\lambda_1, \ldots, \lambda_k$ los k autovalores distintos de A con multiplicidades geométricas y algebraicas m_{gi} y m_{ai}, i=1, $\ldots$, k, respectivamente. Entonces para cada autovalor λ_i existen m_{gi} números naturales $n_j^{(i)}$, j= 1, 2, $\ldots m_{gi}$, (únicos salvo por su orden), tales que

$$m_{ai} = \sum_{j=1}^{m_{gi}} n_j^{(i)},$$

y una matriz no singular P (que en general no es única) tal que, $J = P^{-1}\,A\,P$ toma la forma (canónica de Jordan)

$$J = \begin{pmatrix} J(\lambda_1, n_1^{(1)}) & 0 & & \cdots & & & 0 \\ 0 & \ddots & & & & & \\ & & J(\lambda_1, n_{m_{g1}}^{(1)}) & & & & \\ \vdots & & & \ddots & & & \vdots \\ & & & & J(\lambda_k, n_1^{(k)}) & & \\ & & & & & \ddots & 0 \\ 0 & & & \cdots & & 0 & J(\lambda_k, n_{m_{gk}}^{(k)}) \end{pmatrix},$$

que se puede escribir como $J = D + N$, donde D es una matriz diagonal y N es una matriz nilpotente $N^n = 0$. De hecho cada bloque de Jordan toma la forma

$$J(\lambda_i, n_j^{(i)}) = \lambda_i\, I_{n_j^{(i)}} + N_{n_j^{(i)}},$$

donde $I_{n_j^{(i)}}$ es la matriz identidad en $\mathcal{M}_{n_j^{(i)}\times n_j^{(i)}}$ y $N_{n_j^{(i)}}$ es nilpotente, $(N_{n_j^{(i)}})^{n_j^{(i)}}$. Escrito en forma matricial se tiene

$$J(\lambda_i, n_j^{(i)}) = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}.$$

Los polinomios característicos de cada uno de los bloques de Jordan $J(\lambda_i, n_j^{(i)})$ son

$$(\lambda_i - \lambda)^{n_j^{(i)}} = |J(\lambda_i, n_j^{(i)}) - \lambda\, I|,$$

y se denominan divisores elementales.

3.3.9 Normas de matrices

Podemos definir una norma matricial de la misma forma que se define una vectorial, dado que las matrices forman un espacio vectorial. Sin embargo, en ese caso, no sabemos como se comporta la norma ante un producto de matrices. Por ello, conviene utilizar normas matriciales que sean submultiplicativas, es decir, tales que $\|A\,B\| \leq \|A\|\,\|B\|$. Aún así, no hay posibilidad de relacionar directamente estas normas matriciales con alguna norma vectorial, algo muy útil en muchos casos. Es por ello que usualmente se utilizan normas matriciales subordinadas a una norma vectorial, que están ligadas directamente con ésta y permiten la desigualdad $\|A\,x\| \leq \|A\|\,\|x\|$.

Dada una norma vectorial $\|\cdot\|$, se define su norma matricial asociada o subordinada como

$$\|A\| = \sup_{\|x\|\neq 0} \frac{\|A\,x\|}{\|x\|}.$$

Haciendo $u = x/\|x\|$, se obtiene esta otra definición equivalente a la anterior

$$\|A\| = \sup_{\|u\|=1} \|A\,u\|.$$

Presentaremos seguidamente varios ejemplos de normas matriciales que utilizaremos a lo largo de este curso.

- **Norma matricial de Frobenius (o de Schur)**: Es un ejemplo de norma matricial, que se puede demostrar que no está asociada o subordinada a ninguna norma vectorial, y se define como

  $$F(A) = \|A\|_F = \left(\sum_i \sum_j |a_{ij}|^2\right)^{1/2}.$$

  Aunque no es una norma subordinada, sí es submultiplicativa y está relacionada con la norma vectorial euclídea, ya que se tiene que

  $$\|A\,x\|_2^2 = \sum_i \left|\sum_j a_{ij}\,x_j\right|^2,$$

que aplicando la desigualdad de Cauchy-Schwarz, siendo $a_{i\cdot}$ el i-ésimo vector fila de A, se deduce

$$\left|\sum_j a_{ij}\, x_j\right|^2 = |\langle a_{i\cdot}, x\rangle|^2 \leq \|a_{i\cdot}\|_2^2\, \|x\|_2^2 = \left(\sum_j |a_{ij}|^2\right) \left(\sum_j |x_j|^2\right),$$

y por tanto

$$\begin{aligned}\|A\,x\|_2^2 &\leq \sum_i \left(\sum_j |a_{ij}|^2\right) \left(\sum_j |x_j|^2\right) = \left(\sum_j |x_j|^2\right) \left(\sum_i \sum_j |a_{ij}|^2\right)\\ &= F(A)^2\, \|x\|_2^2.\end{aligned}$$

Sin embargo, la norma de Frobenius es muy poco utilizada en las aplicaciones prácticas.

- **Norma matricial uno**: La norma matricial asociada a la norma vectorial uno, se deduce fácilmente

$$\begin{aligned}\|A\,x\|_1 &= \sum_i \left|\sum_j a_{ij}\, x_j\right| \leq \sum_i \sum_j |a_{ij}|\, |x_j|\\ &= \sum_j \sum_i |a_{ij}|\, |x_j| = \sum_j \left(|x_j| \sum_i |a_{ij}|\right) \leq \max_j \left(\sum_i |a_{ij}|\right) \sum_j |x_j|,\end{aligned}$$

por lo que

$$\|A\|_1 \leq \max_j \sum_i |a_{ij}|.$$

Hemos obtenido una cota superior, ahora tenemos que demostrar que es óptima y coincide con el supremo, para lo que bastará demostrar la igualdad para algunos vectores convenientemente elegidos. Sean éstos los vectores base e_j para los que se verifica

$$\|Ae_j\|_1 = \sum_i |a_{ij}| = \sum_i |a_{ij}| \|e_j\|_1,$$

por lo que la norma matricial uno de A es el máximo de la sumas por columnas de los valores absolutos de sus elementos. Por esta razón también se conoce como norma del máximo por columnas[23].

- **Norma matricial infinito**: La norma matricial asociada a la norma vectorial del máximo o infinito, se obtiene de forma del todo similar. Operando

$$\|A\,x\|_\infty = \max_i \left|\sum_j a_{ij}\, x_j\right| \leq \max_i \sum_j |a_{ij}|\, |x_j|$$

[23]Una regla sencilla permite recordar esta definición, 1 corresponde a vertical |, es decir, a suma por columnas.

$$\leq \max_i \left(\sum_j |a_{ij}| \max_j |x_j| \right) = \max_j |x_j| \max_i \sum_j |a_{ij}|,$$

por lo que

$$\|A\|_\infty \leq \max_i \sum_j |a_{ij}|.$$

Para demostrar la igualdad basta considerar el vector $x = (\pm 1, ..., \pm 1)$, con $\|x\|_\infty = 1$, y observar que

$$\|Ax\|_\infty = max_i(\sum_j \pm a_{ij}) = max_i(\sum_j \pm a_{ij})\|x\|_\infty,$$

por lo que eligiendo de forma adecuada el signo + ó - en cada componente de x se obtiene el resultado buscado, y por lo tanto la norma infinito de A coincide con el máximo de la sumas por filas de los valores absolutos de sus elementos. A esta norma también se la llama norma del máximo por filas[24].

- **Norma matricial dos**: La norma matricial asociada a la norma vectorial euclídea o norma dos no es la norma de Frobenius, como ya hemos indicado, pero se puede determinar fácilmente siguiendo el siguiente proceso. Como

$$\|A\,x\|_2^2 = \langle A\,x, A\,x\rangle = \langle x, A^*\,A\,x\rangle,$$

es necesario estudiar el producto $A^*\,A$ que es una matriz hermítica ($(A^*\,A)^* = A^*\,A$) y, por tanto, sus autovalores λ_j son reales y sus n autovectores definen una base ortonormal, $\{u_i\}$. Estos autovectores coinciden con las columnas de la matriz unitaria que la transforma en su forma canónica de Schur, que es diagonal. Todo vector x se puede escribir en dicha base de autovectores como

$$x = \sum_j \langle x, u_j\rangle\, u_j = \sum_j x_j\, u_j,$$

y de esta forma

$$\|A\,x\|_2^2 = \langle x, A^*\,A\,x\rangle = \langle \sum_i x_i\,u_i, \sum_j x_j\,\lambda_j\,u_j\rangle = \sum_i \sum_j \lambda_j\,\overline{x_i}\,x_j\,\langle u_i, u_j\rangle,$$

que aplicando $\langle u_i, u_j\rangle = \delta_{ij}$ nos da

$$\|A\,x\|_2^2 = \sum_j \lambda_j\,|x_j|^2 \leq \sum_j |\lambda_j|\,|x_j|^2 \leq \lambda_{max}\,\|x\|_2^2,$$

[24]Una regla sencilla permite recordar esta definición, ∞ corresponde a horizontal $-$, es decir, a suma por filas.

donde

$$\lambda_{max} = \max_j |\lambda_j| = \rho(A^*\, A).$$

Por lo tanto,

$$\|A\|_2^2 \leq \rho(A^*\, A).$$

Ahora bien tomando el vector x igual a un autovector asociado al mayor autovalor λ_{max}, se obtiene la igualdad

$$\|A\|_2 = \sqrt{\rho(A^*\, A)}.$$

Es interesante notar que para cualquier norma matricial asociada a una norma vectorial se cumple

$$\rho(A) \leq \|A\|,$$

ya que para todos los autovalores λ de A se tiene

$$|\lambda| \leq \|A\|.$$

Para verificarlo basta tomar un autovector x, para el que

$$\|A\, x\| = \|\lambda\, x\| = |\lambda|\, \|x\| \leq \|A\|\, \|x\|.$$

De hecho se puede probar que el radio espectral satisface la ecuación

$$\rho(A) = \inf_{\|\cdot\|} \|A\|,$$

en la que se toma el ínfimo sobre todas las normas matriciales subordinadas. De esta forma, si $\rho(A) < 1$ entonces sabemos que existe alguna norma matricial subordinada tal que $\|A\|_* = \rho(A) < 1$, aunque determinar dicha norma en la práctica es muy difícil.

Se puede demostrar que para toda matriz y para todo $\epsilon > 0$, existe una norma vectorial $\|\cdot\|_K$ tal que la norma matricial $\|\cdot\|_K$ inducida por esta norma vectorial es tal que

$$\|A\|_K \leq \rho(A) + \epsilon.$$

Así, el radio espectral es el ínfimo de todas las normas matriciales subordinadas de A. La demostración se encuentra en la sección 4.6 de [3]

BIBLIOGRAFÍA

[1] Artin, Michael, “*Algebra*,” Prentice-Hall (1991).

[2] Schechter, Eric, “*The Axiom of Choice. Introduction and Links*,” Vanderbilt University (2001). Disponible en `http://www.math.vanderbilt.edu/~schectex/ccc/choice.html`

[3] Kincaid, David and Cheney, Ward, “*Análisis Numérico*,” Addison-Wesley Iberoamericana (1994).

[4] Burgos Román, Juan de, “*Álgebra Lineal*,” McGraw-Hill Interamericana de España (1993).

[5] Hernández Rodríguez, Eugenio, “*Álgebra y geometría*,” (2a. edición), Addison-Wesley Iberoamericana España (1998).

[6] Granero Rodríguez, Francisco, “*Álgebra y geometría analítica*,” McGraw-Hill Interamericana de España (1985).