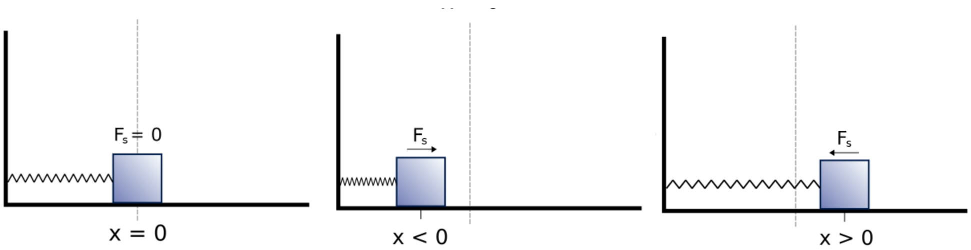
Índice

Capítulo 8

Autovalores y Autovectores

8.1. Autovalores y autovectores, conceptos básicos

Consideremos el problema del movimiento de una masa puntual de masa m sujeta a un resorte de constante k.



En este problema interesa obtener la posición $x(t)$ de la masa m en cualquier instante de tiempo t. La ecuación de movimiento puede obtenerse a partir de la segunda ley de Newton:

$$\begin{aligned} &F = ma, \\ &m\ddot{x} = -kx \\ &m\ddot{x} + kx = 0 \\ &\ddot{x} + \frac{k}{m}x = 0 \end{aligned} \tag{8.1}$$

$*$ En estas ecuaciones, cada punto sobre $x(t)$ indica una derivada respecto del tiempo. La última ecuación, en donde la incógnita es una función que aparece derivada dos veces respecto de la variable independiente, se denomina *Ecuación diferencial ordinaria de segundo orden a coeficientes constantes.* Para encontrar su solución podemos proceder como sigue. Sea

$$\begin{aligned} &U_1 = x, \\ &\dot{U}_1 = \dot{x} = U_2, \end{aligned}$$

de modo que (8.1) puede escribirse como

$$\begin{cases} \dot{U}_2 = -\dfrac{k}{m}U_1 \\ \dot{U}_1 = U_2 \end{cases} \qquad \text{ó} \qquad \begin{pmatrix} \dot{U}_1 \\ \dot{U}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k/m & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \tag{8.2}$$

con lo que convertimos nuestra ecuación original en un *sistema de ecuaciones diferenciales de primer orden.* Ecuaciones diferenciales de la forma $\dot{U}(t) = \lambda U(t)$ tienen soluciones $U(t) = \alpha e^{\lambda t}$, por lo tanto proponemos:

$$U_1 = \alpha_1 e^{\lambda t}, \qquad U_2 = \alpha_2 e^{\lambda t},$$

y reemplazando en 8.2, tenemos:

$$\begin{pmatrix} 0 & 1 \\ -k/m & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \lambda \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \tag{8.3}$$

que podemos escribir matricialmente como

$$A\vec{\alpha} = \lambda\vec{\alpha}. \tag{8.4}$$

Es decir, vamos a hallar las soluciones a la ecuación diferencial cuando encontremos la solución de (8.4), es decir, los valores de α_1, α_2 y λ que satisfacen el sistema. Es importante notar que $\vec{\alpha} = (0\ 0)^T$ es solución pero no sirve, puesto que no brinda información. Por lo tanto, lo que necesitamos son escalares λ y vectores $\vec{\alpha} \neq 0$ tales que

$$(A - \lambda I)\vec{\alpha} = 0,$$

que es equivalente a (8.4). En esta ecuación 0 es un vector, no un escalar. Así, queremos encontrar los $\vec{\alpha} \neq 0$ tales que

$$\vec{\alpha} \in \mathcal{N}(A - \lambda I). \tag{8.5}$$

Pero como

$$\mathcal{N}(A - \lambda I) \neq \vec{0} \quad \Leftrightarrow \quad (A - \lambda I) \text{ es singular}, \tag{8.6}$$

los λ que buscamos son los que hacen

$$\boxed{\det(A - \lambda I) = 0.} \tag{8.7}$$

* * *

Tenemos entonces las siguientes definiciones.

Definición. Para una matriz $A^{n\times n}$, los escalares λ y los vectores **no nulos** $x^{n\times 1}$ que satisfagan

$$Ax = \lambda x, \tag{8.8}$$

se llaman **autovalores y autovectores de A**, respectivamente.

* El conjunto $\sigma(A)$ de autovalores distintos se llama *espectro* de A.

* $\lambda \in \sigma(A) \iff A - \lambda I$ es singular $\iff \det(A - \lambda I) = 0$

* $\{x \neq 0 \ / \ x \in \mathcal{N}(A - \lambda I)\}$ es el conjunto de todos los autovectores asociados con λ.

* La unión de este conjunto con el vector nulo se llama espacio propio de A asociado al autovalor λ.

Definición. Se llama **polinomio característico** de $A^{n\times n}$ a

$$\det(A - \lambda I) = \mathcal{P}(\lambda) \tag{8.9}$$

y **ecuación característica** a

$$\mathcal{P}(\lambda) = 0. \tag{8.10}$$

El grado de $\mathcal{P}(\lambda)$ es n y el término de orden n es siempre $(-1)^n\lambda^n$.

Es posible probar que si $\mathcal{P}(\lambda) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_{n-1}\lambda + c_n$ es el polinomio característico de $A^{n\times n}$ entonces:

$$\begin{aligned} traza(A) &= \lambda_1 + \lambda_2 + \ldots + \lambda_n = -c_1 \\ det(A) &= \lambda_1\lambda_2\lambda_3 \ldots \lambda_n = -(1)^n c_n \end{aligned}$$

De acuerdo con esta definición, los autovalores son las raíces de la ecuación característica.

Vamos a continuar con el ejemplo del oscilador armónico, en el que habíamos llegado al sistema:

$$\begin{pmatrix} 0 & 1 \\ -k/m & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \lambda \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \tag{8.11}$$

que puede ser escrito de la forma

$$(A - \lambda I)\vec{\alpha} = \vec{0} \qquad \longrightarrow \qquad \begin{pmatrix} -\lambda & 1 \\ -k/m & -\lambda \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \vec{0}. \tag{8.12}$$

Buscamos los valores de λ que sean raíces del polinomio característico, es decir

$$\mathcal{P}(\lambda) = \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -k/m & -\lambda \end{vmatrix} = \lambda^2 + \frac{k}{m} = \left(\lambda - i\sqrt{\frac{k}{m}}\right)\left(\lambda + i\sqrt{\frac{k}{m}}\right) = 0$$

Por lo tanto, los autovalores de la matriz A serán

$$\boxed{\lambda_1 = i\sqrt{\frac{k}{m}},} \qquad \text{y} \qquad \boxed{\lambda_2 = -i\sqrt{\frac{k}{m}}.} \tag{8.13}$$

Los autovectores $\vec{\alpha}$ asociados a λ_1 serán aquéllos vectores no nulos que pertenezcan a

$$\mathcal{N}\left(A - i\sqrt{\frac{k}{m}}I\right), \tag{8.14}$$

mientras que los asociados a λ_2 pertenecerán a

$$\mathcal{N}\left(A + i\sqrt{\frac{k}{m}}I\right). \tag{8.15}$$

Para encontrar los autovectores correspondientes, por ejemplo, a λ_1 resolvemos el sistema

$$\left(A - i\sqrt{\frac{k}{m}}I\right)\vec{\alpha} = 0, \tag{8.16}$$

es decir, hay que obtener α_1 y α_2 tales que:

$$\begin{pmatrix} -i\sqrt{k/m} & 1 \\ -k/m & -i\sqrt{k/m} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = 0. \tag{8.17}$$

Si multiplicamos la primera fila de la matriz por $-i\sqrt{k/m}$ obtendremos la segunda fila. Esto es natural que suceda ya que los autovalores son aquellos escalares que hacen singular a la matriz. Por lo tanto, sólo una de las ecuaciones tiene información independiente y, utilizando la primera, podememos escribir

$$-i\sqrt{k/m}\,\alpha_1 + \alpha_2 = 0 \quad \Rightarrow \quad \alpha_2 = i\sqrt{k/m}\,\alpha_1 \quad \Rightarrow \quad \vec{\alpha} = \begin{pmatrix} 1 \\ i\sqrt{k/m} \end{pmatrix}. \tag{8.18}$$

Observen que si $\vec{\alpha}$ es solución, $c\vec{\alpha}$ también lo será (donde c es un escalar).

* * *

Nota: La matriz $A^{n\times n}$ tiene n autovalores ya que el grado del polinomio característico es n. Algunos autovalores pueden ser números complejos, aún en el caso de ser $A \in \mathbb{R}^{n\times n}$ y pueden estar repetidos. Si $A \in \mathbb{R}^{n\times n}$ los autovalores complejos ocurren de a pares conjugados:

$$\text{si} \quad \lambda \in \sigma(A) \ \wedge \ \lambda \in \mathbb{C} \quad \Rightarrow \quad \overline{\lambda} \in \sigma(A).$$

Esto es una consecuencia del que las raíces complejas de polinomios con coeficientes reales ocurren de a pares conjugados (Teorema fundamental del Álgebra para polinomios).

* * *

Problema

Dada la matriz

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \tag{8.19}$$

probar que $\sigma(A) = \{1+i, 1-i\}$ y que los respectivos espacios propios son

$$\text{span}\left\{\begin{pmatrix} i \\ 1 \end{pmatrix}\right\} \quad \text{y} \quad \text{span}\left\{\begin{pmatrix} -i \\ 1 \end{pmatrix}\right\}.$$

Solución

La primera parte del problema nos pide calcular los autovalores, para lo cual debemos hallar los valores de λ que hacen que la matriz

$$A - \lambda I = \begin{pmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{pmatrix}, \tag{8.20}$$

sea singular. La ecuación característica es

$$P(\lambda) = (1-\lambda)^2 + 1 = \lambda^2 - 2\lambda + 2 = 0. \tag{8.21}$$

Esta ecuación cuadrática tiene por raíces $\lambda_1 = 1+i$ y $\lambda_2 = 1-i$, que son los autovalores de A. Para encontrar el autovector $x = (x_1\ x_2)^T$ asociado a λ_1, reemplazamos λ_1 en (8.20) y planteamos el sistema lineal

$$\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0. \tag{8.22}$$

Notamos que la segunda ecuación del sistema es equivalente a la primera (basta multiplicarla por $-i$ para obtener la primera), lo que no debería sorprendernos puesto que hallamos los λ exigiendo que el sistema fuera compatible indeterminado (infinitas soluciones). Por lo tanto, la única ecuación que nos da el sistema es

$$-ix_1 - x_2 = 0 \quad \Rightarrow \quad x_2 = -ix_1, \tag{8.23}$$

y, por lo tanto, los autovectores asociados a λ_1 tienen la forma

$$x = x_1\begin{pmatrix} 1 \\ -i \end{pmatrix} \qquad \text{o bien} \qquad x = \text{cte} \cdot \begin{pmatrix} i \\ 1 \end{pmatrix} \tag{8.24}$$

Naturalmente, la constante multiplicativa no puede ser cero, pero si sumamos esta posibilidad y consideramos todos los valores posibles, obtenemos el espacio propio asociado a λ_1:

$$\text{span}\left\{\begin{pmatrix} i \\ 1 \end{pmatrix}\right\}.$$

Procediendo de manera análoga se pueden obtener los autovectores asociados a λ_2.

Observen cómo aún matrices muy simples pueden tener autovalores y autovectores complejos.

* * *

Definición. Para una matriz A, el número

$$\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|, \tag{8.25}$$

se llama **radio espectral** de la matriz A.

En algunas ocasiones basta con conocer cotas de los autovalores y no sus valores exactos. Una cota "gruesa" (pero barata) de $\rho(A)$ es

$$\rho(A) \leq ||A||, \tag{8.26}$$

para cualquier norma matricial. Veamos cómo se demuestra esta relación. Sea (λ, x) un autovalor y su correspondiente autovector de $A^{n\times n}$. Sea, además, $X^{n\times n}$ una matriz cuya primera columna es x y las demás son nulas. Entonces, se cumplirá que

$$AX = \lambda X, \tag{8.27}$$

y, aplicando la norma a ambos miembros, tendremos que

$$||AX|| = ||\lambda X|| = |\lambda|\,||X||. \tag{8.28}$$

Una norma matricial debe cumplir con que

$$||AX|| \leq ||A||||X||, \tag{8.29}$$

por lo que, reemplazando ésta más arriba obtenemos

$$||A||||X|| \geq |\lambda|||X|| \quad \Rightarrow \quad ||A|| \geq |\lambda|, \qquad \lambda \in \sigma(A). \tag{8.30}$$

Es posible calcular cotas mejores para los autovalores λ usando el método de Gerschgorin:

Teorema. Los autovalores de la matriz $A \in \mathbb{C}^{n\times n}$ están contenidos en la unión G_r de los n círculos de Gerschgorin, definidos de la siguiente manera:

$$|z - a_{ii}| \leq r_i \qquad \text{donde} \qquad r_i = \sum_{\substack{j=1\\ i\neq j}}^{n} |a_{ij}|, \qquad i = 1, 2, \ldots, n.$$

Es decir, los autovalores están contenidos en una colección de círculos centrados en los a_{ii}, con radios dados por la suma de los valores absolutos de los elementos de la i-ésima fila, excepto el a_{ii}.

(*) Más aún, si la unión de los k círculos de Gerschgorin G_r no toca a ninguno de los otros $n-k$, entonces hay exactamente k autovalores (contando multiplicidades) en los círculos de la unión.

(*) Como $\sigma(A) = \sigma(A^T)$, las sumas en el cálculo de los círculos puede hacerse por columnas, es decir:

$$G_c : |z - a_{ii}| \leq r_i, \qquad \text{donde} \qquad r_i = \sum_{\substack{j=1\\ i\neq j}}^{n} |a_{ji}|.$$

Por lo tanto, los autovalores están contenidos en la intersección $G_r \cap G_c$.

Demostración

Lo probamos para G_r, para G_c es idéntico. Sea $\lambda \neq 0$ un autovalor de A y sea x su correspondiente autovector. Supongamos que el autovector está normalizado a uno ($||x||_\infty = 1$) lo que, sabemos, siempre se puede realizar. Si x_i es una componente tal que $|x_i| = 1$, como $Ax = \lambda x$, para la i-ésima componente tendremos:

$$\sum_{i\neq j} a_{ij}x_j + a_{ii}x_i = \lambda x_i \quad \Rightarrow \quad (a_{ii} - \lambda)x_i = \sum_{i\neq j} a_{ij}x_j.$$

Tomando módulo en ambos miembros

$$|a_{ii} - \lambda||x_i| = \left|\sum_{j\neq i} a_{ij}x_j\right| \leq \sum_{j\neq i} |a_{ij}||x_j| \leq \sum_{j\neq i} |a_{ij}||x_i| \quad \Rightarrow \quad |a_{ii} - \lambda| \leq \sum_{j\neq i} |a_{ij}|. \tag{8.31}$$

Por lo tanto, λ estará dentro de alguno de los círculos de Gerschgorin (no sabemos de cuál, porque hasta que no conozcamos los autovectores no sabremos cuál es la componente de mayor valor absoluto).

Para demostrar lo que ocurre con k círculos de Gerschgorin que no se tocan con los demás $n-k$, debemos definir algunas matrices especiales. Sea $D = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$ y $B = A - D$. Sea

$$C(t) = D + tB = \begin{pmatrix} a_{11} & ta_{12} & ta_{13} & \ldots & ta_{1n} \\ ta_{21} & a_{22} & ta_{23} & \ldots & ta_{2n} \\ \vdots & & & & \vdots \\ ta_{n1} & ta_{n2} & \ldots & \ldots & a_{nn} \end{pmatrix}, \qquad t \in [0, 1]. \tag{8.32}$$

De lo que acabamos de probar, los autovalores $\lambda(t)$ de $C(t)$ están contenidos en la unión de los círculos de Gerschgorin $C_i(t)$, definidos por $|z - a_{ii}| \leq t\, r_i$, $i = 1, ..., n$. Estos círculos crecen con continuidad desde los puntos a_{ii},cuando $t = 0$, hasta los círculos de Gerschgorin de A, cuando $t = 1$. Luego, si los círculos de la unión aislada $\mathcal{U}$ están centrados en $(a_{i_1 i_1}, a_{i_2 i_2}, \ldots, a_{i_k i_k})$, entonces $\forall t \in [0, 1]$, la unión $\mathcal{U}(t) = C_{i_1}(t) \cup C_{i_2}(t) \cup \ldots \cup C_{i_k}(t)$ es disjunta de la unión $\overline{\mathcal{U}}(t)$ de los otros $n - k$ círculos de Gerschgorin de $C(t)$. Se puede demostrar que los autovalores $\lambda(t)$ varían también con continuidad, de manera que caen sobre una curva Γ_i que tiene su origen en $\lambda_i(0) = a_{ii}$ y termina en $\lambda_i(1) \in \sigma(A)$. Dado que $\mathcal{U}(t) \cap \overline{\mathcal{U}}(t) = \emptyset$, $\forall t \in [0, 1]$, las curvas Γ_{i1}, Γ_{i2}, ..., Γ_{ik} están enteramente en $\mathcal{U}$ y luego los puntos finales $\lambda_{i1}(1), \lambda_{i2}(1), \ldots, \lambda_{ik}(1)$ están en $\mathcal{U}$. De modo similar, los otros $n - k$ autovalores de A están en la unión del conjunto complementario de círculos.

■

Utilizando los círculos de Gerschgorin se puede inferir si una matriz A es no singular (invertible). Para ello, basta recordar que la matriz será no singular si $\det(A) \neq 0$, y por la propiedad de los autovalores que afirma que

$$\det(A) = \prod_i \lambda_i, \tag{8.33}$$

si ninguno de los círculos de Gerschgorin incluye al origen, los autovalores serán todos distintos de cero y, por ende, el determinante de A también.

$*$ El teorema de Gerschgorin puede ser utilizado para dar valores iniciales en el método de Müller para calcular ceros de polinomios, ya que dado un polinomio *mónico*

$$q(x) = x^n + c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \ldots + c_1 x + c_0,$$

podemos definir su *matriz acompañante* como

$$c(q) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -c_{n-1}, \end{pmatrix}$$

cuyo polinomio característico es precisamente $q(x)$. Por lo tanto, podemos usar Gerschgorin para acotar las posibles raíces complejas del polinomio $q(x)$.

8.2. Diagonalización

En Álgebra Lineal es muy importante, como ya vimos, lograr una representación lo más simple posible de un dado operador lineal φ sobre un espacio vectorial $\mathcal{V}$. Vimos ya que si B y B' son bases de $\mathcal{V}$,

$$[\varphi]_{BB'} = C_{BB'}[\varphi]_{BB} C_{BB'}^{-1}.$$

Por lo tanto, el problema se reduce, matricialmente hablando, a encontrar la matriz de cambio de base tal que $[\varphi]_{B'B'}$ sea tan simple como sea posible, preferentemente diagonal.

Recuerden que el proceso de convertir $[\varphi]_{BB}$ a $[\varphi]_{B'B'}$ se llama transformación de semejanza y que dos matrices relacionadas por esta transformación se dicen semejantes. Veamos un teorema con propiedades de matrices semejantes, que nos va a ser útil en lo sucesivo.

Teorema. Sean A y $B \in \mathbb{C}^{n\times n}$, con $A \sim B$, entonces

(a) $\det(A) = \det(B)$,

(b) A y B tienen el mismo polinomio característico.

Demostración

(a) Como A y B son semejantes, existe P invertible tal que

$$B = P^{-1}AP. \tag{8.34}$$

Entonces

$$\begin{aligned}
\det(B) &= \det(P^{-1}AP) \\
&= \det(P^{-1})\det(A)\det(P) \\
&= \frac{1}{\cancel{\det(P)}}\det(A)\cancel{\det(P)} \\
&= \det(A)
\end{aligned}$$

(b) El polinomio característico de B es

$$\begin{aligned}
\det(B - \lambda I) &= \det(P^{-1}AP - \lambda I) \\
&= \det(P^{-1}AP - \lambda P^{-1}IP) \\
&= \det(P^{-1}(A - \lambda I)P) \\
&= \det(P^{-1})\det(A - \lambda I)\det(P) \\
&= \frac{1}{\cancel{\det(P)}}\det(A - \lambda I)\cancel{\det(P)} \\
&= \det(A - \lambda I)
\end{aligned}$$

■

De la parte (b) del teorema se desprende que las matrices semejantes tienen los mismos autovalores.

Definición. Una matriz $A^{n\times n}$ se dice **diagonalizable** si es semejante a una matriz diagonal.

No todas las matrices cuadradas son diagonalizables. Un ejemplo de ésto son las matrices nilpotentes. Recordemos que una matriz $A^{n\times n}$ es nilpotente si existe $k \in \mathbb{N}$ tal que $A^k = 0$. Lo podemos ver en la práctica con la matriz

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad A^2 = 0. \tag{8.35}$$

Demostremos que una matriz nilpotente no es diagonalizable. Sea $A^{n\times n}$ nilpotente tal que $A \neq 0$ y $A^2 = 0$. Si esta matriz fuera diagonalizable, querría decir que existen una matriz D diagonal y otra matriz P, no singular, tales que

$$D = PAP^{-1}. \tag{8.36}$$

Por lo tanto,

$$D^2 = PAP^{-1}PAP^{-1} = PA^2P^{-1} = 0 \quad \Rightarrow \quad D = 0 \quad \Rightarrow \quad A = 0, \tag{8.37}$$

lo que está en contradicción con la hipótesis inicial. Concluimos que la contradicción se da porque las matrices nilpotentes no son diagonalizables.

* * *

Definición. Un **conjunto completo de autovectores** de $A^{n\times n}$ es cualquier conjunto de n autovectores de A linealmente independiente.

Ejemplo

Analicemos si la matriz

$$A = \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix} \tag{8.38}$$

tiene un conjunto completo de autovectores. Haciendo las cuentas correspondientes se encuentra que la ecuación característica de A es

$$\lambda^3 + 5\lambda^2 + 3\lambda - 9 = (\lambda - 1)(\lambda + 3)^2 = 0. \tag{8.39}$$

Por lo tanto, $\lambda_1 = 1$ es un autovalor simple mientra que $\lambda_2 = -3$ es un autovalor doble (multiplicidad algebráica igual a 2). Calculando los autovectores asociados a cada autovalor se obtiene que

$$N(A - \lambda_1 I) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} \right\} \qquad \text{y} \qquad N(A - \lambda_2 I) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}. \tag{8.40}$$

Es fácil probar que estos tres vectores forman un conjunto linealmente independiente.

No todas las matrices tienen conjuntos completos de autovectores.

Ejemplo

Sea

$$A = \begin{pmatrix} -3 & 1 & -3 \\ 20 & 3 & 10 \\ 2 & -2 & 4 \end{pmatrix}. \tag{8.41}$$

Haciendo las cuentas encontramos que el polinomio característico es

$$p(\lambda) = (\lambda - 3)^2(\lambda + 2). \tag{8.42}$$

Vemos que $\lambda = 3$ tiene multiplicidad algebráica igual a 2. Los espacios propios son

$$\mathcal{N}(A - 3I) = \operatorname{span}\left\{ \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \right\}, \qquad \text{(mult. geométrica 1)}$$

$$\mathcal{N}(A + 2I) = \operatorname{span}\left\{ \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

En consecuencia, encuentro sólo 2 autovectores linealmente independientes. Concluimos que A no posee un conjunto completo de autovectores.

* * *

El siguiente teorema demuestra que como consecuencia de no poseer un conjunto completo de autovectores la matriz no es diagonalizable.

Teorema. La matriz $A^{n\times n}$ es diagonalizable sí y sólo sí A posee un conjunto completo de autovectores o, más precisamente,

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_n), \tag{8.43}$$

sí y sólo sí las columnas de P constituyen un conjunto completo de autovectores y los λ_j son los autovalores asociados.

Demostración

$\Rightarrow$)

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \quad \Rightarrow \quad AP = P \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \tag{8.44}$$

lo que podemos escribir también de la siguiente manera

$$A(P_1|P_2|\ldots|P_n) = (P_1|P_2|\ldots|P_n) \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \tag{8.45}$$

donde los P_i son las columnas de P. Entonces encontramos que

$$A(P_1|P_2|\ldots|P_n) = (\lambda_1 P_1|\lambda_2 P_2|\ldots|\lambda_n P_n), \tag{8.46}$$

y, finalmente,

$$(AP_1|AP_2|\ldots|AP_n) = (\lambda_1 P_1|\lambda_2 P_2|\ldots|\lambda_n P_n). \tag{8.47}$$

Por lo tanto, debe verificarse que

$$AP_j = \lambda_j P_j, \qquad j = 1, ..., n. \tag{8.48}$$

Vemos que los λ_j son los autovalor de A y P_j su respectivo autovector. Como P es invertible, las columnas deben ser base del espacio columna, por lo que los P_j forman un conjunto linealmente independiente.

$\Leftarrow$) Si tengo un conjunto completo de autovectores, construyo P no singular y si D es la matriz diagonal, tal que los elementos de la diagonal son los autovalores, voy de atrás para adelante en la demostración anterior y llego a que

$$P^{-1}AP = D. \tag{8.49}$$

■

Vamos a ver un par de conceptos que nos van a permitir caracterizar la posibilidad de diagonalizar a una matriz sin tener que computar los autovectores.

Definición. Para $\lambda_i \in \sigma(A) = \{\lambda_1, \ldots, \lambda_s\}$ decimos que

(a) la multiplicidad algebráica de λ_i es el número de veces (a_i) que se repite como raíz de la ecuación característica y lo denotamos por

$$\mathrm{MA}_A(\lambda_i) = a_i, \tag{8.50}$$

(b) si $\mathrm{MA}_A(\lambda_i) = 1$, el autovalor λ_i se llama simple,

(c) la multiplicidad geométrica del autovalor λ_i es la dimensión del espacio propio asociado, es decir,

$$\mathrm{MG}_A(\lambda_i) = \dim(\mathcal{N}(A - \lambda_i I)), \tag{8.51}$$

(d) los autovalores tales que

$$\mathrm{MG}_A(\lambda_i) = \mathrm{MA}_A(\lambda_i), \tag{8.52}$$

se llaman semisimples.

Teorema (S/D). Para todo matriz $A \in \mathbb{C}^{n\times n}$ y para cada $\lambda \in \sigma(A)$

$$\mathrm{MG}_A(\lambda) \leq \mathrm{MA}_A(\lambda). \tag{8.53}$$

Corolario: Si un autovalor es simple, entonces es semisimple.

No toda matriz cuadrada es diagonalizable, pero sí ”triangularizable”.

Teorema (S/D). **Teorema de triangularización de Schur**. Toda matriz cuadrada es unitariamente semejante a una matriz triangular superior. Esto es, para toda matriz $A^{n\times n}$ existe U unitaria ($U^* = U^{-1}$ y una matriz triangular superior T, tales que

$$T = U^* AU. \tag{8.54}$$

Corolario: los autovalores de A son los elementos de la diagonal de T.

Demostracion

$$T \sim A \quad \Rightarrow \quad A \text{ y } T \text{ tienen los mismos autovalores.} \tag{8.55}$$

y, por lo tanto, A y T tendrán el mismo polinomio característico

$$\det(A - \lambda I) = \det(T - \lambda I) = \Pi_{i=1}^{n}(t_{ii} - \lambda). \tag{8.56}$$

De esta última concluimos que los t_{ii} son los autovalores de A.

Teorema (S/D). Sean $\{\lambda_1, ..., \lambda_k\}$ un conjunto de autovalores distintos de la matriz $A^{n\times n}$, entonces

(a) el conjunto $S = \{x_1, ..., x_k\}$ de autovectores asociados respectivamente a los autovalores es linealmente independiente,

(b) si B_i es una base del subespacio propio asociado a λ_i, es decir, si B_i es una base de $\mathcal{N}(A - \lambda_i I)$, entonces

$$B = B_1 \cup B_2 \cup ... \cup B_k, \tag{8.57}$$

es un conjunto linealmente independiente.

Teorema. Sea $A^{n\times n}$ con autovalores $\lambda_1, ..., \lambda_k$ distintos. Las siguientes afirmaciones son equivalentes:

(a) A es diagonalizable

(b) la unión B de las bases de los espacios propios de los autovalores de A tiene n vectores

(c) la multiplicidad algebráica de cada autovalor es igual a su multiplicidad geométrica.

Demostración

(a) $\Rightarrow$ (b)

Recordemos que

$$A \text{ diagonalizable} \quad \Rightarrow \quad A \text{ tiene un conjunto completo de autovectores.}$$

Si n_i de los n autovectores corresponden al autovalor λ_i, entonces B_i tiene al menos n_i vectores (B_i: base del espacio propio de λ_i). Sabemos que estos n_i vectores forman un conjunto linealmente independiente, así que, a lo sumo puede pasar que falten vectores para formar una base B_i del espacio propio correspondiente a λ_i. En consecuencia,

$$B = B_1 \cup B_2 ... \cup B_i ... \cup B_k \tag{8.58}$$

tendrá AL MENOS n vectores ($\sum_{i=1}^{k} n_i = n$). Por el teorema anterior sabemos que B es un conjunto linealmente independiente. Como B es un conjunto de n vectores de $\mathbb{R}^n$, linealmente independiente, no es posible agregar otro vector sin que éste sea combinación lineal de los anteriores. Por lo tanto, B tiene EXACTAMENTE n vectores.

(b) $\Rightarrow$ (c)

Por un teorema anterior sabemos que

$$\mathrm{MG}_A(\lambda_i) \leq \mathrm{MA}_A(\lambda_i), \qquad i = 1, ..., k. \tag{8.59}$$

Como A es de $n \times n$, el grado del polinomio característico es n. Luego

$$\sum_{i=1}^{k} \mathrm{MG}_A(\lambda) \leq \sum_{i=1}^{k} \mathrm{MA}_A(\lambda_i) = n. \tag{8.60}$$

Por hipótesis,

$$\sum_{i=1}^{k} \mathrm{MG}_A(\lambda_i) = n, \tag{8.61}$$

por lo que resulta que que

$$\sum_{i=1}^{k} \mathrm{MG}_A(\lambda_i) = \sum_{i=1}^{k} \mathrm{MA}_A(\lambda_i) \quad \Rightarrow \quad \sum_{i=1}^{k} [\mathrm{MG}_A(\lambda_i) - \mathrm{MA}_A(\lambda_i)] = 0. \tag{8.62}$$

Teniendo presente (8.59) concluimos que

$$\mathrm{MG}_A(\lambda_i) = \mathrm{MA}_A(\lambda_i), \qquad i = 1, ..., k. \tag{8.63}$$

(c) $\Rightarrow$ (a)

Ahora, por hipótesis tenemos que

$$\mathrm{MG}_A(\lambda_i) = \mathrm{MA}_A(\lambda_i) = a_i, \qquad i = 1, ..., k. \tag{8.64}$$

Sea B_i base de $\mathcal{N}(A - \lambda_i I)$. Entonces

$$B = \bigcup_{i=1}^{k} B_i \quad \text{tiene} \quad \sum_{i=1}^{k} a_i = n \quad \text{vectores.} \tag{8.65}$$

Luego B es un conjunto completo de autovectores de A y, por lo tanto, A es diagonalizable.

■

Corolario. Si $A^{n\times n}$ tiene n autovalores distintos, entonces es diagonalizable.

Demostración Sabemos que si A posee n autovalores distintos entonces tiene n autovectores asociados que forman un conjunto linealmente independiente. Esto implica que la matriz tiene un conjunto completo de autovectores y, por lo tanto, que es diagonalizable.

■

Nota: la recíproca NO es cierta, **no es necesario que los autovalores de A sean todos distintos para que sea diagonalizable**.

Teorema espectral. Si $A^{n\times n}$ es hermítica, entonces A es diagonalizable por una matriz unitaria.

Demostración

Como $A \in \mathbb{C}^{n\times n}$, vale el teorema de triangularización de Schur, luego

$$T = U^* A U, \tag{8.66}$$

con U matriz unitaria ($U^* = U^{-1}$), donde T es triangular. Entonces

$$T^* = (U^* A U)^* = U^* A^* U = U^* A U = T, \tag{8.67}$$

donde hemos usado que A es hermítica ($A = A^*$). T^*=T sólo puede darse si T es diagonal. Probamos así que A es diagonalizable por medio de una matriz unitaria.

■

*Observen que en el caso de ser A hermítica, los autovectores que forman las columnas de U son ortonormales, cosa que en general no es cierto (sólo son linealmente independientes).

*La recíproca no es cierta, es decir, que A sea diagonalizable por una matriz unitaria NO implica que A sea hermítica. Por ejemplo, la matriz

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{8.68}$$

puede ser diagonalizada haciendo

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = U^* AU, \qquad \text{con} \quad U = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \tag{8.69}$$

y, sin embargo, A no es hermítica.

*En el caso real puede demostrarse un teorema más fuerte ($\Leftrightarrow$).

> *Teorema (S/D).* Sea $A \in \mathbb{R}^{n\times n}$. Entonces vale que
>
> $$A \text{ es simétrica} \quad \Leftrightarrow \quad A \text{ es ortogonalmente diagonalizable.} \tag{8.70}$$

Veamos algunos resultados que aplican a matrices especiales.

> *Teorema.* Sea $A \in \mathbb{C}^{n\times n}$ hermítica, entonces A tiene todos sus autovalores reales.

Demostración

Sea λ un autovalor de A y x su correspondiente autovector. Dado que $x \neq 0$,

$$||x||^2 = x^* x \neq 0. \tag{8.71}$$

Calculamos

$$\begin{aligned} x^* x(\lambda - \overline{\lambda}) &= x^*(\lambda - \overline{\lambda})x \\ &= x^* \lambda x - x^* \overline{\lambda} x \\ &= x^* Ax - \overline{\lambda} x^* x. \end{aligned} \tag{8.72}$$

Pero

$$Ax = \lambda x \quad \Rightarrow \quad (Ax)^* = (\lambda x)^* \quad \Rightarrow \quad x^* A^* = \overline{\lambda} x^* \quad \Rightarrow \quad x^* A = \overline{\lambda} x^*, \tag{8.73}$$

donde hemos usado que A es hermítica. Por lo tanto, reemplazando en (8.72)

$$\begin{aligned} x^* x(\lambda - \overline{\lambda}) &= x^* Ax - x^* A^* x \\ &= x^* Ax - x^* Ax \\ &= 0, \end{aligned} \tag{8.74}$$

donde nuevamente hemos usado que A es hermítica. Dado que $x^* x \neq 0$, concluimos que

$$\boxed{\lambda = \overline{\lambda},} \tag{8.75}$$

es decir, que los autovalores son reales.

■

Corolario. Sea $A \in \mathbb{R}^{n\times n}$ simétrica, entonces A tiene todos sus autovalores reales.

Demostración

Repetimos los pasos del teorema anterior pero usando la simetría y que la matriz es real. Si λ es un autovalor de A y x su correspondiente autovector:

$$\begin{aligned} x^*x(\lambda-\overline{\lambda}) &= x^*(\lambda-\overline{\lambda})x = x^*\lambda x - x^*\overline{\lambda}x = \\ &= x^*Ax - (\overline{\lambda}x^*)x = x^*Ax - (\lambda x)^*x = \\ &= x^*Ax - (Ax)^*x = x^*Ax - x^*A^Tx = \\ &= x^*Ax - x^*Ax = 0. \end{aligned} \tag{8.76}$$

■

Teorema. Sea A una matriz hermítica, entonces los autovectores asociados a autovalores distintos son ortogonales.

Demostración

Sean λ y μ autovalores asociados a autovectores u y v. Entonces

$$\mu u^*v = u^*(\mu v) = u^*Av = u^*A^*v = (Au)^*v = (\lambda u)^*v = \overline{\lambda}u^*v = \lambda u^*v, \tag{8.77}$$

donde hemos utilizado el teorema anterior y la hermiticidad de la matriz A. Por lo tanto,

$$\mu u^*v = \lambda u^*v \quad \Rightarrow \quad (\mu-\lambda)u^*v = 0, \tag{8.78}$$

y, en consecuencia, si $\mu \neq \lambda$

$$u^*v = 0 \quad \Rightarrow \quad \boxed{u \perp v.} \tag{8.79}$$

■

Corolario 1: Si A es hermítica, $\mathcal{N}(A-\lambda_i I) \perp \mathcal{N}(A-\lambda_j I), \forall i \neq j$.

Corolario 2: Lo anterior también vale para matrices reales simétricas.

8.2.1. Introducción (informal) a las formas cuadráticas

Una expresión de la forma

$$\alpha x^2 + \beta y^2 + \gamma xy, \tag{8.80}$$

se denomina forma cuadrática en x y en y. Similarmente

$$\alpha x^2 + \beta y^2 + \gamma z^2 + \delta xy + \varepsilon xz + \zeta yz, \tag{8.81}$$

es una forma cuadrática en x, y, z. O sea, una forma cuadrática es una suma de términos, cada uno de los cuales tiene grado DOS en todas sus variables.

Las podemos expresar usando matrices:

$$\alpha x^2 + \beta y^2 + \gamma xy = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \alpha & \gamma/2 \\ \gamma/2 & \beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{8.82}$$

y

$$\alpha x^2 + \beta y^2 + \gamma z^2 + \delta xy + \varepsilon xz + \zeta yz = \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} \alpha & \delta/2 & \epsilon/2 \\ \delta/2 & \beta & \zeta/2 \\ \varepsilon/2 & \zeta/2 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{8.83}$$

(tarea: probar que son ciertas)

Notar que en ambos casos las expresiones tienen la forma

$$v^T A v, \qquad \text{(con } A \text{ simétrica)}. \tag{8.84}$$

Ahora sí vamos a definir las formas cuadráticas de manera más formal.

Definición. Se llama **forma cuadrática** $f : \mathbb{R}^n \to \mathbb{R}$ a una función de la forma

$$f(x) = x^T A x = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{i<j} 2a_{ij} x_i x_j, \tag{8.85}$$

donde $A \in \mathbb{R}^{n \times n}$ es simétrica. A se llama "matriz asociada con f".

De acuerdo con la definición, la matriz asociada a

$$f(x_1, x_2, x_3) = 2x_1^2 - x_2^2 + 5x_3^2 + 6x_1x_2 - 3x_1x_3, \tag{8.86}$$

es

$$A = \begin{pmatrix} 2 & 3 & -3/2 \\ 3 & -1 & 0 \\ -3/2 & 0 & 5 \end{pmatrix}, \tag{8.87}$$

y podemos escribir

$$f = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{8.88}$$

Teorema de los ejes principales. Toda forma cuadrática puede diagonalizarse. Específicamente, si $A^{n \times n}$ es la matriz asociada a la forma cuadrática $x^T A x$ y Q es la matriz ortogonal tal que

$$D = Q^T A Q, \tag{8.89}$$

con D diagonal, entonces $x = Qy$ transforma la forma cuadrática original en otra de la forma $y^T D y$, que no tiene términos cruzados. Si los autovalores de A son $\{\lambda_1, ...\lambda_n\}$, entonces

$$x^T A x = y^T D y = \sum_{i=1}^{n} \lambda_i y_i^2. \tag{8.90}$$

Demostración

Que la matriz A (simétrica) puede diagonalizarse ortogonalmente ya fue mencionado antes, aunque no lo demostramos. Por lo tanto, existen D y Q ortogonal tales que

$$D = Q^T A Q. \tag{8.91}$$

Sea ahora un vector definido por $y = Q^T x$. Por ser Q ortogonal podemos escribir $x = Qy$ y reemplazando en la forma cuadrática obtenemos

$$x^T A x = y^T Q^T A Q y = y^T D y. \tag{8.92}$$

Como $D = \text{diag}(\lambda_1, ..., \lambda_n)$, donde los λ_i son los autovalores de A, encontramos finalmente que

$$y^T D y = \sum_{i=1}^{n} \lambda_i y_i^2. \tag{8.93}$$

■

Ejemplo. En $\mathbb{R}^2$ una forma cuadrática representa a una cónica que, si tiene términos no diagonales, está rotada. Consideremos la función

$$\begin{aligned} f(x) &= 5x_1^2 + 4x_1x_2 + 2x_2^2 - 6 \\ &= x^T Ax - 6, \end{aligned}$$

donde

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}. \tag{8.94}$$

A tiene autovalores $\lambda_1 = 6$ y $\lambda_2 = 1$. Haciendo las cuentas se puede hallar la matriz

$$Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}. \tag{8.95}$$

Haciendo $x' = Q^T x$ obtenemos

$$f(x') = (x')^T Dx' - 6 = (x_1')^2 + 6(x_2')^2 - 6. \tag{8.96}$$

En forma más general, las formas cuadráticas suelen aparecer en problemas de optimización (hallar máximos y mínimos, con y sin restricciones. En estos casos es conveniente poder clasificar a las formas cuadráticas como sigue.

Definición. Sea $f(x) = x^T Ax$ una forma cuadrática, entonces $f(x)$ es

(a) **definida positiva** si $f(x) > 0, \forall x \neq 0$

(b) **semi-definida positiva** si $f(x) \geq 0, \forall x$

(c) **definida negativa** si $f(x) < 0, \forall x \neq 0$

(d) **semi-definida negativa** si $f(x) \leq 0, \forall x$

(e) **indefinida** si $f(x)$ toma tanto valores positivos como negativos.

La matriz simétrica A asociada a la forma cuadrática sigue la clasificación de esta última.

El teorema de los ejes principales permite clasificar una forma cuadrática usando los autovalores de A.

Teorema (S/D). Sea $A \in \mathbb{R}^{n \times n}$ simétrica. La forma cuadrática $x^T Ax$ es

(a) **definida positiva** $\iff$ todos los autovalores de A son positivos

(b) **semi-definida positiva** $\iff$ todos los autovalores de A son no negativos

(c) **definida negativa** $\iff$ todos los autovalores de A son negativos

(d) **semi-definida negativa** $\iff$ todos los autovalores de A son no positivos

(e) **indefinida** $\iff$ los autovalores de A son positivos y negativos.

Para finalizar con las formas cuadráticas damos el siguiente teorema.

Teorema. Sea $f(x) = x^T Ax$ una forma cuadrática y sean $\lambda_1 \geq \lambda_2 \geq ...\lambda_n$ los autovalores de A (recordar que como A es simétrica, todos sus autovalores son reales). Entonces, para x tal que $||x|| = 1$ vale

(a) $\lambda_1 \geq f(x) \geq \lambda_n$

(b) el valor máximo de $f(x)$ es λ_1 y se produce cuando x es el autovector asociado a λ_1 y $||x|| = 1$

(c) el valor mínimo de $f(x)$ es λ_n y se produce cuando x es el autovector asociado a λ_n y $||x|| = 1$.

Demostración

a) Como A es simétrica sabemos que es ortogonalmente diagonalizable

$$D = Q^T AQ. \tag{8.97}$$

Definiendo $y = Q^T x$ tenemos que

$$||y||^2 = y^T y = (Q^T x)^T Q^T x = x^T QQ^T x = x^T x = ||x||^2 = 1. \tag{8.98}$$

Trabajando ahora sobre la forma cuadrática

$$\begin{aligned} f(x) &= x^T Ax = (Qy)^T AQy = y^T Q^T AQy = y^T Dy \\ &= \lambda_1 y_1^2 + ... + \lambda_n y_n^2 \leq \lambda_1 \sum_{i=1}^{n} y_i^2 = \lambda_1 ||y||^2 = \lambda_1. \end{aligned}$$

De forma análoga se demuestra que $f(x) \geq \lambda_n$.

b) Sea x autovector, $||x|| = 1$, asociado a λ_1, entonces

$$f(x) = x^T Ax = x^T \lambda_1 x = \lambda_1 x^T x = \lambda_1. \tag{8.99}$$

c) Se demuestra igual que b) pero considerando el autovector correspondiente al autovalor λ_n.

■

Vamos a ver ahora algunos métodos para el cálculo aproximado de autovalores y autovectores. Si bien hay muchos, vamos a centrarnos especialmente en 2, el método de potencias y el método QR.

Definición. Sean $\lambda_1, ..., \lambda_n$ los autovalores de $A \in \mathbb{R}^{n\times n}$. Si

$$|\lambda_j| > |\lambda_i|, \qquad i = 1, ..., n\,(i \neq j), \tag{8.100}$$

entonces λ_j se llama **autovalor dominante**. El autovector x_j asociado a λ_j se denomina **autovector dominante**.

8.2.2. Método de potencias

Se aplica a la matriz $A^{n\times n}$ que tiene un autovalor dominante λ_1. El método de potencias es ITERATIVO, produce una sucesión de escalares que converge hacia λ_1 y una sucesión de vectores que converge hacia el autovector correspondiente.

Teorema. Sea $A^{n\times n}$ una matriz diagonalizable con autovalor dominante λ_1, entonces existe $x_0 \neq 0$ tal que la secuencia

$$x_1 = Ax_0, \qquad x_2 = Ax_1, \qquad ... \qquad x_k = Ax_{k-1}, \tag{8.101}$$

se aproxima a un autovector dominante de A.

Demostración

Sabemos que

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| ... \geq |\lambda_n|. \tag{8.102}$$

Sean v_1, v_2,..., v_n los autovectores correspondientes. Por ser A diagonalizable sabemos que $\{v_1, v_2, ..., v_n\}$ forma un conjunto completo (es decir, un conjunto linealmente independiente). Esto permite utilizar al conjunto como base de $\mathbb{C}^n$, y así escribir a cualquier vector como

$$x_0 = \sum_{i=1}^{n} c_i v_i, \qquad \forall x_0 \in \mathbb{C}^n. \tag{8.103}$$

La secuencia

$$x_1 = Ax_0, \qquad x_2 = Ax_1, \qquad ... \qquad x_k = Ax_{k-1}, \tag{8.104}$$

se puede escribir como

$$x_k = A^k x_0, \tag{8.105}$$

y si usamos que $\lambda_1 \neq 0$ podemos hace lo siguiente

$$x_k = A^k x_0 = A^k \left(\sum_{i=1}^{n} c_i v_i \right) = \sum_{i=1}^{n} c_i \left(A^k v_i \right) = \sum_{i=1}^{n} c_i \lambda_i^k v_i =$$

$$= \lambda_1^k \sum_{i=1}^{n} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i = \lambda_1^k c_1 v_1 + \lambda_1^k \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i$$

Por ser λ_1 autovalor dominante

$$k \to \infty \quad \Rightarrow \quad x_k \to \lambda_1^k c_1 v_1. \tag{8.106}$$

Como $\lambda_1 \neq 0$ y $v_1 \neq 0$, si $c_1 \neq 0$, x_k se irá aproximando a un múltiplo de v_1, es decir, a un autovector dominante.

■

Este resultado vale sólo si x_0 tiene una componente (c_1) distinta de cero en la dirección del autovector dominante.

¿Cómo se procede en la práctica?

Dada una matriz $A^{n\times n}$, se elige un x_0 arbitrario y se empieza a aplicar repetidamente A a x_0

$$x_k = A^k x_0. \tag{8.107}$$

x_k debería tender a un autovector dominante. ¿Cómo se obtiene λ_1? Teniendo en cuenta que

$$x_{k+1} = Ax_k \approx \lambda_1 x_k, \tag{8.108}$$

podemos hacer el cociente entre componentes de x_{k+1} y x_k y así hallar λ_1 (se puede tomar un promedio).

Un problema que se puede dar con este método (de las potencias) es que las componentes de los x_k se vayan haciendo muy grandes (si $|\lambda_1| > 1$) o muy chicos (si $|\lambda_1| < 1$) y pueden causar errores de redondeo significativos (o incluso overflow o underflow). Podemos evitar estos problemas si multiplicamos por un escalar para normalizar a 1, ya que eso no cambiará la dirección del vector.

Observaciones

- Si x_0 tiene $c_1 = 0$ (componente nula en la dirección del autovector dominante) el método no debería converger. Sin embargo, los errores de redondeo muy probablemente provoquen que aparezca una componente $c_1 \neq 0$ en alguna iteración. Curiosamente, encontramos un caso en el que los errores de redondeo nos ayudan...
- La velocidad de convergencia dependerá esencialmente del valor de $|\lambda_2/\lambda_1|$. A menor valor de este cociente, más rápida será la convergencia.

- Otra manera de calcular el autovalor dominante λ_1 de A en conjunto con el método de potencias es usando el cociente de Rayleigh

$$R(x) = \frac{(Ax) \cdot x}{x \cdot x} = \frac{(\lambda_1 x) \cdot x}{x \cdot x} = \frac{\lambda_1 (x \cdot x)}{x \cdot x} = \lambda_1. \tag{8.109}$$

  Entonces, a medida que calculamos x_k, los cocientes $R(x_k)$ deberían aproximarse a λ_1.

Hasta acá sólo tenemos un método para calcular autovalores/autovectores dominates. ¿Qué hacemos si queremos calcular los demás? Aplicamos una técnica llamada deflación. Así, una vez calculado λ_1, transformamos A en una matriz con los mismos autovalores de antes salvo el autovalor λ_1, que es reemplazado por uno nulo. Así podremos seguir aplicando el método de potencias.

8.2.3. Deflación de Hotelling para matrices simétricas

Se utiliza el reemplazo

$$\widetilde{A} = A - \lambda_1 x_1 x_1^T, \tag{8.110}$$

donde λ_1 es el autovalor dominante y x_1 su autovector asociado (autovector dominante) normalizado a 1.
Apliquemos $\widetilde{A}$ a un autovector x_i de A, de norma 1:

$$\widetilde{A} x_i = (A - \lambda_1 x_1 x_1^T) x_i = \lambda_i x_i - \lambda_1 x_1 x_1^T x_i. \tag{8.111}$$

Recordando que los autovectores de una matriz simétrica son ortogonales tenemos que

$$x_1^T x_i = 0, \quad \forall i \neq 1. \tag{8.112}$$

Por lo tanto,

$$\widetilde{A} x_i = \begin{cases} 0 & \text{si } i = 1 \\ \lambda_i x_i & \text{si } i \neq 1. \end{cases} \tag{8.113}$$

En consecuencia, los autovalores de $\widetilde{A}$ son 0, $\lambda_2, \lambda_3, ..., \lambda_n$.

* * *

Para matrices no simétricas se puede usar una estrategia parecida, pero primero tenemos que definir algunos conceptos adicionales.

Definición. $y \neq 0$ es una **autovector por izquierda de A**, asociado al autovalor λ, si

$$y^T A = \lambda y^T. \tag{8.114}$$

Definición. $x \neq 0$ es una **autovector por derecha de A**, asociado al autovalor λ, si

$$Ax = \lambda x. \tag{8.115}$$

Observen que

$$y^T A = \lambda y^T \quad \Rightarrow \quad y^T A - y^T \lambda I = 0 \quad \Rightarrow \quad y^T (A - \lambda I) = 0 \quad \Rightarrow \quad (A - \lambda I)^T y = 0. \tag{8.116}$$

Dado que para toda matriz de $n \times n$

$$\det(B) = \det(B^T), \tag{8.117}$$

concluimos que los autovalores siguen siendo los mismos, pero no así los autovectores, dado que para hallarlos se necesita resolver un sistema lineal distinto

$$(A - \lambda I)^T y = 0. \tag{8.118}$$

Teorema. Sea $A \in \mathbb{R}^{n\times n}$. Los autovectores por derecha e izquierda asociados a autovalores distintos son ortogonales respecto del producto escalar canónico.

Demostración

Sea x un autovector por derecha, con autovalor λ y sea y un autovector por izquierda, con autovalor μ. Entonces

$$\lambda y^T x = y^T \lambda x = y^T A x = \mu y^T x \quad \Rightarrow \quad (\lambda - \mu) y^T x = 0. \tag{8.119}$$

Por lo tanto,

$$\lambda \neq \mu \quad \Rightarrow \quad y^T x = 0 \quad \Rightarrow \quad x \perp y. \tag{8.120}$$

8.2.4. Deflación de Hotelling para matrices no simétricas

Sea λ_1 el autovalor dominante y sus autovectores asociados por derecha (x_1) y por izquierda (y_1), de forma tal que

$$Ax_1 = \lambda_1 x_1 \tag{8.121}$$

$$y_1^T A = \lambda_1 y_1^T. \tag{8.122}$$

Como $x_1 \neq 0$ e $y_1 \neq 0$, siempre será posible normalizar de la siguiente forma

$$y_1^T x_1 = 1. \tag{8.123}$$

Definimos ahora a una nueva matriz

$$\widetilde{A} = A - \lambda_1 x_1 y_1^T. \tag{8.124}$$

Veamos qué pasa si la aplicamos a los autovectores (x_i, $i = 1, ..., n$) de A

$$\widetilde{A} x_i = (A - \lambda_1 x_1 y_1^T) x_i = \lambda_i x_i - \lambda_1 x_1 (y_1^T x_i) = \begin{cases} 0 & \text{si } i = 1 \\ \lambda_i x_i & \text{si } i \neq 1 \end{cases} \tag{8.125}$$

En consecuencia, los autovalores de $\widetilde{A}$ son: $0, \lambda_2, \lambda_3, ..., \lambda_n$. Aplicando el método de potencias sobre $\widetilde{A}$ podemos obtener el autovalor dominante y, repitiendo el proceso, se encuentran los demás autovalores.

8.2.5. Deflación de Wielandt

Sea (λ_1, x_1) el par autovalor/autovector dominante, respectivamente. Los suponemos conocidos, por aplicación del método de potencias.
Definimos la matriz

$$\widetilde{A} = A - x_1 u_1^T \tag{8.126}$$

donde u_1 es un vector cualquiera que verifica la igualdad $u_1^T x_1 = \lambda_1$. Si aplicamos esta matriz sobre el autovector dominante de A tendremos

$$\widetilde{A} x_1 = (A - x_1 u_1^T) x_1 = \lambda_1 x_1 - x_1 \underbrace{u_1^T x_1}_{=\lambda_1} = 0. \tag{8.127}$$

En consecuencia, x_1 es también autovector de $\widetilde{A}$ con autovalor igual a cero.
Tomemos ahora un conjunto de escalares $\alpha_i \neq 0$, $i = 2, ..., n$, y calculemos para x_i (autovector de A) lo siguiente

$$\widetilde{A}(x_i - \alpha_i x_1) = \widetilde{A} x_i - \alpha_i \underbrace{\widetilde{A} x_1}_{=0} \tag{8.128}$$

$$= (A - x_1 u_1^T) x_i \tag{8.129}$$

$$= \lambda_i x_i - x_1 u_1^T x_i \tag{8.130}$$

$$= \lambda_i \left[x_i - \left(\frac{u_1^T x_i}{\lambda_i} \right) x_1 \right] \tag{8.131}$$

donde $i = 2, ..., n$. Por lo tanto, si hacemos

$$\alpha_i = \frac{u_1^T x_i}{\lambda_i} \tag{8.132}$$

tendremos que

$$\widetilde{A}(x_i - \alpha_i x_1) = \lambda_i (x_i - \alpha_i x_1) \tag{8.133}$$

con lo que queda demostrado que $\widetilde{A}$ tiene los mismos autovalores λ_i que A, para $i = 2, ..., n$, y que el autovalor dominante de A ha sido reemplazado por el autovalor nulo. Aplicando el método de potencias sobre $\widetilde{A}$ obtendremos su autovalor dominante, y repitiendo el proceso se pueden calcular los autovalores restantes.

8.2.6. Método de potencias inverso

Este método aprovecha el que si λ es autovalor de una matriz invertible A, entonces λ^{-1} es autovalor de A^{-1}. Lo demostramos

$$Ax = \lambda x \quad \Rightarrow \quad \underbrace{A^{-1}A}_{=I} x = \lambda A^{-1} x \quad \Rightarrow \quad \boxed{A^{-1}x = \lambda^{-1} x} \tag{8.134}$$

De este modo, el autovalor dominante de A^{-1} es el más pequeño (en valor absoluto) de A. El método de potencias inverso no se aplica calculando explícitamente A^{-1}, sino utilizando que

$$\omega_{k+1} = A^{-1}\omega_k \quad \Rightarrow \quad A\,\omega_{k+1} = \omega_k \tag{8.135}$$

que se resuelve numéricamente para obtener ω_{k+1}.

RECORDAR: El médoto de potencias inverso sólo es aplicable si A es invertible.

8.2.7. Método de potencias inverso modificado

Teorema. Sea λ autovalor de una matriz A y sea $\alpha \neq 0$, $\alpha \notin \sigma(A)$. Entonces $\lambda - \alpha$ es un autovalor de $A - \alpha I$ y $(\lambda - \alpha)^{-1}$ autovalor de $(A - \alpha I)^{-1}$.

Demostración: Si

$$Ax = \lambda x \tag{8.136}$$

entonces

$$(A - \alpha I)x = Ax - \alpha x = \lambda x - \alpha x = (\lambda - \alpha)x \tag{8.137}$$

y queda demostrado que $\lambda - \alpha$ es autovalor de $A - \alpha I$.
Por ser $\alpha \notin \sigma(A)$, los autovalores de $A - \alpha I$ son todos distintos de cero. En consecuencia, $\det(A - \alpha I) \neq 0$ y $A - \alpha I$ es invertible. Así, podemos aplicar el método de potencias inverso a $A - \alpha I$

$$\text{si } (\lambda - \alpha) \text{ es autovalor de } (A - \alpha I) \quad \Rightarrow \quad (\lambda - \alpha)^{-1} \text{ es autovalor de } (A - \alpha I)^{-1}. \tag{8.138}$$

■

* * *

La ventaja de este método es que si puedo estimar α de modo tal que sea una buena aproximación de λ, entonces $(\lambda - \alpha)^{-1}$ será un número grande y, por consiguiente, el método de potencias va a converger rápidamente.

Para estimar α se puede usar el siguiente resultado relacionado con el teorema de Gerschgorin: *Si la unión de k círculos de Gerschgorin no toca a ninguno de los otros $n - k$ círculos, entonces hay exactamente k*

autovalores (contando multiplicidades) en los círculos de la unión. En particular, si un disco está separado de los demás, entonces debe contener exactamente UN autovalor de la matriz.

En la práctica, procedemos como en el caso del método de potencias inverso. En lugar de calcular la matriz inversa y aplicar la iteración

$$\omega_{k+1} = (A - \alpha I)^{-1}\omega_k \tag{8.139}$$

se resuelve, en cada paso, el sistema lineal

$$(A - \alpha I)\omega_{k+1} = \omega_k, \qquad k = 0, 1, ... \tag{8.140}$$

8.2.8. Algoritmo iterativo QR

Finalmente damos los lineamientos generales de este método que permite la obtención simultánea de todos los autovalores de una matriz $A \in \mathbb{R}^{n\times n}$. La idea de este algoritmo es la siguiente. Sea

$$A_1 = A \tag{8.141}$$

y apliquemos la factorización QR

$$A_1 = Q_1 R_1. \tag{8.142}$$

Con Q_1 y R_1 conocidas, definimos ahora una nueva matriz

$$A_2 = R_1 Q_1. \tag{8.143}$$

Factorizamos a esta matriz, como lo hicimos con A_1

$$A_2 = Q_2 R_2 \tag{8.144}$$

y definimos

$$A_3 = R_2 Q_2. \tag{8.145}$$

Continuando con este proceso, tendremos

$$A_k = Q_k R_k, \qquad\qquad A_{k+1} = R_k Q_k. \tag{8.146}$$

Definamos ahora a la matriz P_k tal que

$$P_k = Q_1 Q_2 ... Q_k. \tag{8.147}$$

Es fácil ver que P_k es una matriz ortogonal

$$P_k^T P_k = (Q_1 Q_2 ... Q_k)^T Q_1 Q_2 ... Q_k \tag{8.148}$$

$$= Q_k^T Q_{k-1}^T ... Q_2^T \underbrace{Q_1^T Q_1}_{=I} Q_2 ... Q_k \tag{8.149}$$

$$= Q_k^T Q_{k-1}^T ... Q_3^T \underbrace{Q_2^T Q_2}_{=I} Q_3 ... Q_k \tag{8.150}$$

$$= ... \tag{8.151}$$

$$= I \tag{8.152}$$

y se verifican las siguientes igualdades

$$P_1^T A P_1 = Q_1^T A Q_1 = Q_1^T (Q_1 R_1) Q_1 = A_2 \tag{8.153}$$

$$\begin{aligned} P_2^T A P_2 &= (Q_1 Q_2)^T A (Q_1 Q_2) && (8.154)\\ &= Q_2^T Q_1^T A Q_1 Q_2 && (8.155)\\ &= Q_2^T A_2 Q_2 && (8.156)\\ &= Q_2^T (Q_2 R_2) Q_2 && (8.157) \end{aligned}$$

$$= R_2Q_2 \tag{8.158}$$
$$= A_3 \tag{8.159}$$

y, en general,

$$P_k^T A P_k = A_{k+1}. \tag{8.160}$$

En palabras, las matrices A_2, A_3,... A_{k+1} son ortogonalmente semejantes a A. En consecuencia, tienen el mismo espectro

$$\sigma(A_i) = \sigma(A), \qquad i = 2, ..., k+1 \tag{8.161}$$

De las expresiones anteriores, se ve que

$$P_k = Q_1Q_2...Q_{k-1}Q_k = P_{k-1}Q_k \quad \Rightarrow \quad P_{k-1}^T P_k = Q_k \tag{8.162}$$

Si el proceso iterativo converge ($P_k \longrightarrow P$) tendremos que para k grande

$$P_{k-1}^T P_k = Q_k \quad \longrightarrow \quad P^T P = I \tag{8.163}$$

y, además,

$$A_{k+1} = R_k Q_k \quad \longrightarrow \quad RI \tag{8.164}$$

En consecuencia

$$\lim_{k\to\infty} A_{k+1} = R \tag{8.165}$$

que es una matriz triangular superior cuyos elementos diagonales son los autovalores de A.

NOTAS

- El método QR no puede implementarse basándose sólo en estas ideas, ya que como vimos en la factorización QR, los elementos de la diagonal son reales y positivos, lo que excluye a matrices con autovalores complejos y/o reales negativos.
- Es además necesario, en general, introducir transformaciones a la matriz A que permitan una convergencia rápida. Una de estas transformaciones consiste en convertir a la matriz A en una matriz de la forma de Hessenberg.

$$H = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix} \tag{8.166}$$

  Esto debe hacerse mediante transformaciones de semajanza, con matrices unitarias u ortogonales.
- Una posibilidad es hacerlo mediante transformaciones de Householder, que en $\mathbb{R}^n$ pueden pensarse como reflexiones alrededor de un vector dado.
- Otra es hacerlo mediante transformaciones de Givens, que pueden pensarse como rotaciones alrededor de un eje.
- Una vez obtenida la matriz H, en vez de factorizarla $H_k = Q_k R_k$ suelen utilizarse desplazamientos

$$H_k - \alpha_k I = Q_k R_k \tag{8.167}$$

  y

$$H_{k+1} = R_k Q_k + \alpha_k I \tag{8.168}$$

  donde α_k es una aproximación para algún autovalor real. Se puede demostrar que esto acelera la convergencia del método.