Mayorización de vectores y aplicaciones

Curso de la UMA 2018 - Pedro Massey*

Resumen

Las siguientes son notas para un curso de la UMA 2018 (La Plata) sobre mayorización de vectores y sus aplicaciones al análisis matricial.

Se sugiere imprimir este material para el curso, ya que las clases serán con proyector.

Índice

1. Introducción.

La mayorización de vectores y sus extensiones (mayorización de sucesiones finitas de vectores, de matrices autoadjuntas, de operadores autoajduntos en álgebras dotadas de una traza fiel, de funciones en espacios de probabilidad) juegan un papel en diversas ramas de la matemática. En estas notas desarrollamos la mayorización de vectores desde su inicio, y describimos algunas de las apariciones más elementales de esta relación en el análisis matricial. Hacia el final de las notas se mencionan una serie de textos para aquellos interesados en seguir investigando esta relación.

Es un gusto para mí agradecer a los profesores Jorge Antezana y Demetrio Stojanoff el haberme permitido usar su apunte sobre mayorización y su texto "Análisis matricial" para la escritura de estas notas.

Sin más, concluimos esta sección introduciendo algunas de las notaciones básicas que usaremos a continuación.

Notación

A lo largo de este trabajo consideraremos a $\mathbb{C}$ o $\mathbb{R}$ como cuerpo de escalares. Llamaremos $\mathbb{R}_+$ (resp. $\mathbb{R}_+^*$) al conjunto de números reales no negativos (resp. positivos). Dado $n \in \mathbb{N}$, usaremos el símbolo $\mathbb{I}_n$ para denotar al conjunto $\{1, 2, \ldots, n\} \subseteq \mathbb{N}$. Notaremos $\mathcal{M}_n(\mathbb{C}) = \mathbb{C}^{n\times n}$ al álgebra de

*Departamento de Matemática-FCE-UNLP & Instituto Argentino de Matemática "A.P. Calderón"-CONICET
e-mail: massey@mate.unlp.edu.ar

matrices cuadradas de $n \times n$ sobre $\mathbb{C}$. Análogamente, $\mathcal{M}_n(\mathbb{R}) = \mathbb{R}^{n\times n}$. Llamaremos $\mathcal{M}_{n,m}(\mathbb{C}) = \mathbb{C}^{n\times m}$, al espacio de matrices rectangulares. Si $A \in \mathcal{M}_n(\mathbb{C})$, la pensaremos también como un operador (o transformación lineal) sobre $\mathbb{C}^n$ por multiplicación: si $x \in \mathbb{C}^n$, entonces $A(x) = Ax$ (el producto usual de matrices). Para denotar las entradas de una matriz $A \in \mathcal{M}_{n,m}(\mathbb{C})$, usaremos indistintamente las notaciones $A = (A_{ij})_{i,j\in\mathbb{I}_n}$ ó $A = (a_{ij})_{i,j\in\mathbb{I}_n}$. Las columnas y filas de A se pueden pensar como vectores de $\mathbb{C}^n$. Se usará la notación $C_i(A)$ (respectivamente $F_i(A)$) para denotar a la i-ésima columna (resp. fila) de A. Notar que $C_i(A) = A\, e_i$ para $i \in \mathbb{I}_n$, donde $\{e_1, \dots, e_n\}$ denota la base canónica de $\mathbb{C}^n$.

Los vectores de $\mathbb{C}^n$ serán pensados como vectores columna, es decir que identificamos $\mathbb{C}^n$ con $\mathbb{C}^{n\times 1}$. Sin embargo, los describiremos como una fila (estilo $x = (x_1, \dots, x_n) \in \mathbb{C}^n$), para ahorrar espacio. Por ejemplo, si $A \in \mathcal{M}_n(\mathbb{C})$, $i \in \mathbb{I}_n$, entonces $C_i(A) = (a_{1i}, a_{2i}, \dots, a_{ni}) \in \mathbb{C}^n$.

Dado $x \in \mathbb{R}^n$, $x = (x_1, ..., x_n)$, notaremos $x^\downarrow$ al vector obtenido al reordenar las cordenadas de x en forma decreciente. Así, si $x^\downarrow = (x_1^\downarrow, ..., x_n^\downarrow)$, entonces $x_1^\downarrow \geq ... \geq x_n^\downarrow$. Es decir, por ejemplo, que

$$x_1^\downarrow = \max_i x_i \ , \quad x_1^\downarrow + x_2^\downarrow = \max_{i\neq j} x_i + x_j \ , \quad \text{etc.}$$

Otra notación: dado $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, escribiremos

$$\operatorname{tr} x = \sum_{i=1}^n x_i \ (= \operatorname{tr}(x^\downarrow) \ ) .$$

2. Mayorización de vectores

Comenzamos esta sección con la definición de mayorización y el desarrollo de sus propiedades elementales. Hacia el final, veremos la relación entre mayorización y familias de desigualdades en términos de funciones convexas.

2.1. Definición y primeras propiedades

Con las notaciones anteriores podemos dar la definición de mayorización:

Definición 2.1. Sean $x, y \in \mathbb{R}^n$. Se dice que x **es mayorizado por** y, notado $x \prec y$, si se verifica que $\operatorname{tr} y = \operatorname{tr} x$ y

$$\sum_{j=1}^k x_j^\downarrow \ \leq \ \sum_{j=1}^k y_j^\downarrow \quad , \quad 1 \leq k \leq n \ . \tag{1}$$

Si sólo se cumple la Eq. (1), se dice que x es **submayorizado** por y, notado $x \prec_w y$. △

Observación 2.2. La mayorización en $\mathbb{R}^n$ tiene las siguientes propiedades elementales:

1. Reflexividad: si $x \in \mathbb{R}^n$, $x \prec x$;
2. Transitividad: si $x,\, y,\, z \in \mathbb{R}^n$, $x \prec y$, $y \prec z \Rightarrow x \prec z$;
3. Antisimetría débil: si $x,\, y \in \mathbb{R}^n$, $x \prec y$, $y \prec x \Rightarrow x^\downarrow = y^\downarrow$.
4. Homogeneidad: si $\alpha \in \mathbb{R}$ y $x,\, y \in \mathbb{R}^n$ tales que $x \prec y \Rightarrow \alpha\, x \prec \alpha\, y$.
5. Si $x,\, y \in \mathbb{R}^n$, $x \prec y \Rightarrow$ usando la Eq. (1) junto con $-x \prec -y$ vemos que

   $$y_n^\downarrow = \min\{y_i : \ i \in \mathbb{I}_n\} \leq x_n^\downarrow = \min\{y_i : \ i \in \mathbb{I}_n\} \leq x_1^\downarrow = \max\{x_i : \ i \in \mathbb{I}_n\} \leq y_1^\downarrow = \max\{y_i : \ i \in \mathbb{I}_n\}$$

6. La mayorización no es un orden total: si $x = (1, 1, 0)$ y $y = (3/2, 1/4, 1/4)$ entonces $\operatorname{tr} x = \operatorname{tr} y$ pero $x \not\prec y$, $y \not\prec x$.

$\triangle$

Ejemplo 2.3. Sea $x \in \mathbb{R}^n$ tal que $x_i \geq 0$, $i \in \mathbb{I}_n$. Llamemos $\operatorname{tr} x = 1$. Entonces

$$\Big(\frac{1}{n} \, , \, \frac{1}{n} \, , \dots , \, \frac{1}{n}\Big) \prec (x_1, x_2, \dots, x_n) \prec (1, 0, \dots, 0). \tag{2}$$

La segunda relación es evidente. La primera es sencilla, pero la verificamos después. $\triangle$

Existe una relación muy estrecha entre las relaciones de mayorización y las matrices doblemente estocásticas que definimos a continuación.

Notación: Notaremos $A \in \mathcal{M}_n(\mathbb{R}_+)$ para indicar $A_{ij} \geq 0$ para $i, \, j \in \mathbb{I}_n$ es decir, si A tiene todas sus entradas no negativas.

Definición 2.4. Una matriz $A \in \mathcal{M}_n(\mathbb{R}_+)$ se denomina **doblemente estocástica** si para todo $i \in \mathbb{I}_n$, se verifica que

$$\operatorname{tr} F_i(A) = 1 \quad \text{ y } \quad \operatorname{tr} C_i(A) = 1.$$

Al conjunto de matrices doblemente estocásticas en $\mathcal{M}_n(\mathbb{C})$ lo denotaremos $\mathcal{DS}\,(n)$. $\triangle$

Observación 2.5. Sean $A, \, B \in \mathcal{DS}\,(n)$. Sea $\lambda \in [0, 1]$ y sea $C = (1 - \lambda)\, A + \lambda\, B \in \mathcal{M}_n(\mathbb{R}_+)$ entonces, para $i \in \mathbb{I}_n$:

$$F_i(C) = (1 - \lambda)\, F_i(A) + \lambda\, F_i(B) \;\Rightarrow\; \operatorname{tr}(F_i(C)) = (1 - \lambda)\, 1 + \lambda\, 1 = 1\, .$$

De forma similar se verifica $\operatorname{tr}(C_j(C)) = 1$ para $j \in \mathbb{I}_n$; así $C \in \mathcal{DS}\,(n)$.

Por otro lado, si $D = AB \in \mathcal{M}_n(\mathbb{R}_+)$ entonces, para $i \in \mathbb{I}_n$:

$$\operatorname{tr}(F_i(D)) = \sum_{j \in \mathbb{I}_n} D_{ij} = \sum_{j \in \mathbb{I}_n} \sum_{k \in \mathbb{I}_n} A_{ik}\, B_{kj} = \sum_{k \in \mathbb{I}_n} A_{ik} \sum_{j \in \mathbb{I}_n} B_{kj} = \sum_{k \in \mathbb{I}_n} A_{ik} = 1\, .$$

De forma similar se verifica $\operatorname{tr}(C_j(D)) = 1$ para $j \in \mathbb{I}_n$; así $D \in \mathcal{DS}\,(n)$.

Lo anterior muestra que $\mathcal{DS}\,(n)$ es un semigrupo convexo en $\mathcal{M}_n(\mathbb{C})$. $\triangle$

Una conjunto importante de matrices doblemente estocásticas es el llamado grupo de matrices de permutación, que describimos a continuación.

Observación 2.6. Sea $n \in \mathbb{N}$. Recordar que $\mathbb{I}_n = \{1, 2, \dots, n\}$.

1. Llamaremos S_n al n-ésimo grupo simétrico, es decir

   $$S_n = \{\sigma : \mathbb{I}_n \to \mathbb{I}_n \; : \; \sigma \;\text{ es biyectiva }\; \}\ ,$$

   con el producto dado por la composición de funciones.

2. Dada $\sigma \in S_n$, definimos $P_\sigma \in \mathcal{M}_n(\mathbb{R}_+)$ a la matrix dada por

   $$(P_\sigma)_{ij} = \delta_{\sigma(i)j} \in \{0, 1\}\ ,$$

   donde usamos la delta de Kronecker δ_{lk}.

3. Dados $\sigma \in S_n$ y $x \in \mathbb{C}^n$, llamaremos $x_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$; se verifica que

   $$P_\sigma \; x \; = \; x_\sigma \; , \quad x \in \mathbb{C}^n\ .$$

   Observar que, si $\tau \in S_n$, entonces $P_\sigma P_\tau = P_{\sigma\tau}$. Por otra parte, P_σ opera en la base canónica $\mathcal{B} = \{e_i\}_{i \in \mathbb{I}_n}$ por

   $$C_k(P_\sigma) = P_\sigma(e_k) = e_{\sigma^{-1}(k)}\, , \;\; k \in \mathbb{I}_n\, .$$

4. El grupo $\mathcal{U}_{\mathcal{P}}(n) = \{P_\sigma : \sigma \in S_n\}$ está incluido en $\mathcal{U}(n)$, dado que cada P_σ es claramente isométrico. Por lo tanto, para cada $\sigma \in S_n$, $P_{\sigma^{-1}} = P_\sigma^{-1} = P_\sigma^* = P_\sigma^T$.

5. Observar que $\mathcal{U}_{\mathcal{P}}(n) \subseteq \mathcal{DS}\,(n)$. En efecto, dada $\sigma \in S_n$,

$$C_k(P_\sigma) = P_\sigma(e_k) = e_{\sigma^{-1}(k)} \quad \text{y} \quad F_k(P_\sigma) = C_k(P_\sigma^T) = C_k(P_{\sigma^{-1}}) = e_{\sigma(k)} \ , \tag{3}$$

$\Rightarrow$ $\operatorname{tr}(C_k(P_\sigma)) = \operatorname{tr}(F_k(P_\sigma)) = 1$ para todo $k \in \mathbb{I}_n$.

6. En particular, dadas permutaciones $\sigma_1, \dots, \sigma_k \in \mathcal{U}_{\mathcal{P}}(n)$ y $\lambda_1, \dots, \lambda_k \geq 0$, $\sum_{i\in\mathbb{I}_k} \lambda_i = 1$ entonces $A = \sum_{i\in\mathbb{I}_k} \lambda_i \ P_{\sigma_i} \in \mathcal{DS}\,(n)$.

$\triangle$

Proposición 2.7. *Sea* $t \in [0,1]^n$ *tal que* $\operatorname{tr}(t) = k \in \mathbb{I}_n$. *Si* $x \in \mathbb{R}^n$ *entonces*

$$\sum_{i\in\mathbb{I}_n} t_i\, x_i \ \leq \ \sum_{i\in\mathbb{I}_k} x_i^{\downarrow} .$$

Demostración. Sin pérdida de la generalidad, podemos asumir que $x = x^\downarrow$ (reemplazando t por $t_\sigma \in [0,1]^n$ para $\sigma \in S_n$ tal que $x^\downarrow = x_\sigma$). En este caso,

$$\begin{aligned}
\sum_{i=1}^n t_i x_i - \sum_{i=1}^k x_i &= \sum_{i=1}^n t_i x_i - \sum_{i=1}^k x_i + (k - \sum_{i=1}^n t_i) x_k \\
&= \sum_{i=1}^k t_i x_i + \sum_{i=k+1}^n t_i x_i - \sum_{i=1}^k x_i + k x_k - \sum_{i=1}^k t_i x_k - \sum_{i=k+1}^n t_i x_k \\
&= \sum_{i=1}^k (t_i - 1) x_i + \sum_{i=1}^k x_k - \sum_{i=1}^k t_i x_k + \sum_{i=k+1}^n t_i (x_i - x_k) \\
&= \sum_{i=1}^k (t_i - 1) x_i + \sum_{i=1}^k (1 - t_i) x_k + \sum_{i=k+1}^n t_i (x_i - x_k) \\
&= \sum_{i=1}^k (t_i - 1)(x_i - x_k) + \sum_{i=k+1}^n t_i (x_i - x_k) \leq 0,
\end{aligned}$$

pues los dos términos del último renglón son sumas de términos no positivos. ■

Observación 2.8. En lo que sigue vamos a usar repetidamente la siguiente observación: si $x \in \mathbb{R}^n$ entonces existe $\sigma \in S_n$ de forma que

$$x^\downarrow = x_\sigma = P_\sigma\, x$$

se decir, se verifica que $x_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$ con $x_{\sigma(1)} \geq \dots \geq x_{\sigma(n)}$.

Corolario 2.9. *Sea* $x \in \mathbb{R}^n$ *y sea* $k \in \mathbb{I}_n$, *entonces*

$$\sum_{i\in\mathbb{I}_k} x_i^\downarrow = \max_{K\subset\mathbb{I}_n\,,\,|K|=k} \ \sum_{i\in K} x_i \, .$$

□

Teorema 2.10. *Si* $A \in \mathcal{DS}\,(n)$ *entonces* $Ay \prec y$ *para todo* $y \in \mathbb{R}^n$.

Demostración. Supongamos que $A \in \mathcal{DS}(n)$ y llamemos $x = Ay$. Queremos probar que $x \prec y$. Se puede suponer que las cordenadas de x y de y están ordenadas en forma decreciente. En efecto, si $P, Q \in \mathcal{U}_{\mathcal{P}}(n)$ son matrices de permutación tales que $x^{\downarrow} = Px$, $y^{\downarrow} = Qy$; entonces $A' = PAQ^t \in \mathcal{DS}(n)$ por las observaciones anteriores,

$$A'y^{\downarrow} = PAy = Px = x^{\downarrow} \qquad \text{y} \qquad x \prec y \iff x^{\downarrow} \prec y^{\downarrow} .$$

Así, suponemos $x = x^{\downarrow}$, $y = y^{\downarrow}$. Como $x = Ay$,

$$\sum_{j=1}^{k} x_j = \sum_{j=1}^{k}\sum_{i=1}^{n} a_{ji} y_i = \sum_{i=1}^{n}\Big(\sum_{j=1}^{k} a_{ji}\Big) y_i \quad \text{para todo} \quad k \in \mathbb{I}_n . \tag{4}$$

Si llamamos $t_i = \sum\limits_{j=1}^{k} a_{ji}$, para $i \in \mathbb{I}_n$, entonces

$$0 \le t_i \le \operatorname{tr} C_i(A) = 1 \qquad \text{y} \qquad \sum_{i=1}^{n} t_i = \sum_{j=1}^{k} \operatorname{tr} F_j(A) = k .$$

Luego, aplicando la ecuación (4) y la Proposición 2.7,

$$\sum_{j=1}^{k} x_j = \sum_{j=1}^{k} t_j\, y_j \le \sum_{j=1}^{k} y_j$$

para todo $k \in \mathbb{I}_n$ y además, cuando $k = n$, obtenemos la igualdad, usando la ecuación (4) (para $k = n$, los $t_i = 1$). Así, $x \prec y$. ■

Ejemplo 2.11. Sea $y \in \mathbb{R}^n$, entonces

$$\Big(\frac{\operatorname{tr} y}{n}, \frac{\operatorname{tr} y}{n}, \dots, \frac{\operatorname{tr} y}{n}\Big) \prec (y_1, y_2, \dots, y_n).$$

Para verificar la primera mayorización, consideramos $E = (\frac{1}{n})_{ij} \in \mathcal{DS}(n)$. Por el teorema anterior

$$\Big(\frac{\operatorname{tr} y}{n}, \frac{\operatorname{tr} y}{n}, \dots, \frac{\operatorname{tr} y}{n}\Big) = Ey \prec y .$$

△

Teorema 2.12. *Sean $x, y \in \mathbb{R}^n$. Entonces, son equivalentes:*

1. *x es una combinación convexa de vectores que se obtienen de permutar las entradas de y.*
2. *Existe $A \in \mathcal{DS}(n)$ tal que $x = Ay$.*
3. *$x \prec y$.*
4. *$x = P_1 \cdots P_r\, y$, donde $P_j = \lambda_j I + (1-\lambda_j) P_{\sigma_j}$, con $\lambda_j \in [0,1]$ y $\sigma_j \in S_n$ es transposición, $j \in \mathbb{I}_r$.*

Demostración. Si vale 1, existen $\tau_1, \dots, \tau_m \in S_n$ y $\lambda_1, \dots, \lambda_m \in [0,1]$ tales que $\sum_{i=1}^{m} \lambda_i = 1$ tales que

$$x = \sum_{i \in \mathbb{I}_m} \lambda_i\, y_{\tau_i} = \sum_{i \in \mathbb{I}_m} \lambda_i\, P_{\tau_i}\, y = D\, y \qquad \text{con} \qquad D = \sum_{i \in \mathbb{I}_m} \lambda_i\, P_{\tau_i} \in \mathcal{DS}(n)$$

pues $\mathcal{DS}(n)$ es convexo y $\mathcal{U}_{\mathcal{P}}(n) \subseteq \mathcal{DS}(n)$. $2 \Rightarrow 3$ es el Teorema 2.10.

$3 \Rightarrow 4$. Lo haremos por inducción sobre la dimensión n. Para $n = 1$ es trivial. Sea $n \geq 2$; sin pérdida de generalidad podemos suponer que los vectores estan ordenados en forma decreciente. Luego, $y_n \leq x_n \leq x_1 \leq y_1$. Sea $k > 1$ tal que $y_k \leq x_1 \leq y_{k-1}$ y $\lambda \geq 0$ tal que $x_1 = \lambda y_1 + (1-\lambda) y_k$. Sea P_0 la matriz correspondiente a la transposición $(1\ k)$, que verifica

$$P_0(y_1, \dots, y_n) = (y_k, y_2, \dots, y_{k-1}, y_1, y_{k+1}, \dots, y_n).$$

Definamos $z = (\lambda\, I + (1-\lambda)\, P_0)(y)$: por construcción, $z \prec y$ y $z_1 = x_1$. Sean $x' = (x_2, \dots, x_n)$ y $z' = (z_2, \dots, z_n)$, como vectores indexados por el conjunto $\{2, \dots, n\}$. Vamos a probar que $x' \prec z'$: notemos que $x' = (x')^{\downarrow}$ pero en general $z' \neq (z')^{\downarrow}$ (ver la Observación 2.13 más abajo). Como $\operatorname{tr}(z') = \operatorname{tr}(z) - x_1 = \operatorname{tr}(y) - x_1$ y $\operatorname{tr}(x') = \operatorname{tr}(x) - x_1 = \operatorname{tr}(y) - x_1$, se tiene que $\operatorname{tr}(x') = \operatorname{tr}(z')$. Si $2 \leq m \leq k-1$, entonces, como $x_1 \leq y_{k-1}$,

$$\sum_{i=2}^{m} z_i = \sum_{i=2}^{m} y_i \geq (m-1) y_{k-1} \geq (m-1) x_1 \geq \sum_{i=2}^{m} x_i.$$

Por otro lado, si $m \geq k$.

$$\begin{aligned}
\sum_{i=2}^{m} z_i &= \sum_{i=2}^{k-1} y_i + (1-\lambda) y_1 + \lambda y_k + \sum_{i=k+1}^{m} y_i \\
&= \sum_{i=1}^{m} y_i + \lambda y_1 - (1-\lambda) y_k \\
&= \sum_{i=1}^{m} y_i - x_1 \geq \sum_{i=1}^{m} x_i - x_1 = \sum_{i=2}^{m} x_i.
\end{aligned}$$

Por hipótesis inductiva

$$x' = \sum_{i=1}^{s} \tilde{P}_1 \cdots \tilde{P}_s\, z'$$

para $\tilde{P}_i = \mu_i\, I + (1-\mu_i)\, P_{\tilde{\sigma}_i}$, para ciertas transposiciones $\tilde{\sigma}_i$ de $\{2, \dots, n\}$, $i \in \mathbb{I}_s$, donde consideramos la convención $P_{\tilde{\sigma}_s} z' = (z'_{\tilde{\sigma}_s(2)}, \dots, z'_{\tilde{\sigma}_s(n)})$ y continuando de esta forma con los demás factores.
Llamemos $\sigma_i \in S_n$ a la transposición que extiende a $\tilde{\sigma}_i$ a $\mathbb{I}_n$, poniendo $\sigma_i(1) = 1$ y sea $P_i = \mu_i\, I + (1-\mu_i)\, P_{\sigma_i}$. Luego, como $z_1 = x_1$, se tiene

$$x = P_1 \cdots P_s\, z \quad \Longrightarrow \quad x = P_1 \cdots P_s \cdot (\lambda\, I + (1-\lambda)\, P_0)\ y\,.$$

$4 \Rightarrow 1$. Supongamos que vale la representación en 2. Para cada $\alpha = (\alpha_i)_{i \in \mathbb{I}_r} \in \{0,1\}^r$ consideramos

$$\lambda_\alpha = \prod_{i=1}^{r} \lambda_j^{1-\alpha_i}\, (1-\lambda_j)^{\alpha_i} \qquad \text{y} \qquad P_\alpha = \prod_{i=1}^{r} (P_{\sigma_j})^{\alpha_i} \in \mathcal{U}_{\mathcal{P}}(n)\,.$$

Entonces, aplicando la propiedad distributiva se verifica que

$$\sum_{\alpha \in \{0,1\}^r} \lambda_\alpha = 1 \qquad \text{y} \qquad x = \sum_{\alpha \in \{0,1\}^r} \lambda_\alpha\, P_\alpha\ y\,,$$

y como cada $P_\alpha \in \mathcal{U}_{\mathcal{P}}(n)$ es una matriz de permutación $P_\alpha y \in \mathbb{R}^n$ es un vector que se obtiene de permutar las coordenadas de y. ■

Observación 2.13. Un error típico al tratar de demostrar mayorización entre dos vectores, es olvidarse de ordenarlos antes de sumar sus "primeras" k cordenadas. De hecho, esto sucede en la prueba anterior con el vector z'. Por suerte no es grave en este caso, porque z' está del lado de "los mayores", y lo grave es no reordenar del lado de "los menores". Más explícitamente, si $x, y \in \mathbb{R}^n$, como

$$\sum_{i=1}^{k} y_i \leq \sum_{i=1}^{k} y_i^{\downarrow},$$

es imprescindible ordenar a y para verificar que $y \prec x$, pero no para verificar que $x \prec y$. Notar que, en la prueba de la relación $x' \prec z'$, el vector x' ya venía ordenado correctamente. $\triangle$

En lo que sigue, dado $C \subset \mathbb{R}^n$ entonces $\operatorname{Conv} C$ denota la cápsula convexa de C, es decir

$$\text{Conv}C = \{ \sum_{i \in \mathbb{I}_k} \lambda_i \, c_i : \ \lambda_i \in [0,1], \, c_i \in C, i \in \mathbb{I}_k, \sum_{i \in \mathbb{I}_k} \lambda_i = 1 \, , \ k \in \mathbb{N} \ \} \subset \mathbb{R}^n \, .$$

Corolario 2.14. *Sea $y \in \mathbb{R}^n$ y definamos $\Omega(y) = \{x \in \mathbb{R}^n : \ x \prec y\} \subset \mathbb{R}^n$. Entonces*

$$\Omega(y) = \text{Conv}\{y_\sigma : \ \sigma \in S_n\} \, .$$

En particular los puntos extremales de $\Omega(y)$ son y_σ, $\sigma \in S_n$.

Demostración. La primer afirmación es la equivalencia entre los items 1 y 3 del teorema anterior. El segundo hecho es más técnico y se deja como ejercicio (no lo usaremos en estas notas). $\blacksquare$

Corolario 2.15. *Sean $w, z \in \mathbb{R}^m$ tales que $w \prec z$, y sean $x, y \in \mathbb{R}^k$ tales que $x \prec y$. Entonces los vectores $(x, w), (y, z) \in \mathbb{R}^{k+m}$ cumplen que $(x, w) \prec (y, z)$.*

Demostración. Por el Teorema 2.12, existen $A \in \mathcal{DS}\,(k)$ y $B \in \mathcal{DS}\,(m)$ tales que $Ay = x$ y $Bz = w$. Pero si consideramos

$$C = \begin{bmatrix} A & 0_{k\times m} \\ 0_{m\times k} & B \end{bmatrix} \in \mathcal{M}_{k+m}(\mathbb{C}),$$

es fácil ver que $C \in \mathcal{DS}\,(k+m)$ y que $C(y, z) = (x, w)$. $\blacksquare$

Observación 2.16. Sean $x, u \in \mathbb{R}^n$ tales que $x_i \leq u_i$, para todo $i \in \mathbb{I}_n$. En este caso notamos $x \leqslant u$.

1. Entonces se cumple que $x^{\downarrow} \leqslant u^{\downarrow}$.

2. Lo anterior permite deducir que $x \prec_w u$.

La verificación de estas afirmaciones queda a cargo del lector. $\triangle$

Proposición 2.17. *Sean $x, y \in \mathbb{R}^n$. Entonces $x \prec_w y$ si y solo si existe $u \in \mathbb{R}^n$ tal que*

$$x \leqslant u \prec y \ .$$

Demostración. Antes que nada, es claro que si el tal u existe, entonces $x \prec_w y$ (por la Observación 2.16 y transitividad de $\prec_w$). Para probar la recíproca, podemos asumir que $\operatorname{tr} x < \operatorname{tr} y$, porque sinó basta tomar $u = x$. Asumiremos también que x e y están ordenados en forma decreciente, dado que una ves encontrado el u para este caso, luego se lo puede reordenar igual que a x, lo que preserva la relación $x \leqslant u$ y no afecta la relación $u \prec y$.

Se hará inducción sobre n. Si $n = 1$, el resultado es trivial (en ese caso $\prec$ significa igualdad). Si $n > 1$, cosideramos dos casos:

Caso 1: Si existe $k \in \mathbb{I}_{n-1}$ tal que $\displaystyle\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} y_i$, llamamos $a = (x_1, \dots, x_k)$ y $b = (y_1, \dots, y_k) \in \mathbb{R}^k$. Es claro que $a \prec b$. Por otra parte, si llamamos $w = (x_{k+1}, \dots, x_n)$ y $z = (y_{k+1}, \dots, y_n) \in \mathbb{R}^{n-k}$, es también claro que $w \prec_w z$, porque están bien ordenados y, si $r \in \mathbb{I}_{n-k}$, entonces

$$\sum_{i=1}^{r} z_i - \sum_{i=1}^{r} w_i = \sum_{i=1}^{k+r} y_i - \sum_{i=1}^{k+r} x_i \geq 0.$$

Ahora bien, por hipótesis inductiva, debe existir $v \in \mathbb{R}^{n-k}$ tal que $w \leqslant v \prec z$. Entonces basta tomar $u = (a, v) \in \mathbb{R}^n$ que cumple lo pedido, porque $w \leqslant v$ implica que $x = (a, w) \leqslant (a, v) = u$. Por otra parte, como $a \prec b$ y $v \prec z$, el Corolario 2.15 dice que $u = (a, v) \prec (b, z) = y$.

Caso 2: Supongamos que $d = \min\limits_{k \in \mathbb{I}_n} \left\{ \displaystyle\sum_{i=1}^{k} y_i - \sum_{i=1}^{k} x_i \right\} > 0$, y que se realiza en cierto $k_0 \in \mathbb{I}_n$. Tomemos $v = x + d e_1$, es decir que agrandamos en d la primera cordenada de x. Observar que v está ordenado decrecientemente, por estarlo x. Por ser d quien es, es claro que $x \leqslant v$ y que $v \prec_w y$. Pero claramente v cae en el Caso 1 (sumando hasta k_0, y si k_0 era n, bingo). Entonces existe $u \in \mathbb{R}^n$ tal que $x \leqslant v \leqslant u \prec y$. ■

2.2. Mayorización y funciones convexas

Definición 2.18. *Si $I \subset \mathbb{R}$ es un intervalo (semirrecta o todo $\mathbb{R}$), decimos que f es convexa si, dados $m \in \mathbb{N}$, $x_1, \dots, x_m \in I$ y $\lambda_1, \dots, \lambda_m \in [0, 1]$ tales que $\sum_{i \in \mathbb{I}_m} \lambda_i = 1$, se cumple que*

$$f\Big(\sum_{i \in \mathbb{I}_m} \lambda_i \, x_i \Big) \leq \sum_{i \in \mathbb{I}_m} \lambda_i \, f(x_i) \, .$$

△

En general, dado $\mathrm{I} \subseteq \mathbb{R}$, una función $f : \mathrm{I} \to \mathbb{R}$, y un vector $y \in \mathrm{I}^n$, notaremos por

$$f(y) = (f(y_1), \dots, f(y_n)) \in \mathbb{R}^n \quad \Big(\Rightarrow \operatorname{tr}(f(y)) = \sum_{i \in \mathbb{I}_n} f(y_i) \ \Big) \, .$$

Si I es un intervalo, $f : \mathrm{I} \to \mathbb{R}$ es función convexa, $y_i = (y_{ij})_{j \in \mathbb{I}_n} \in \mathrm{I}^n$, para $j \in \mathbb{I}_m$, y $\lambda_1, \dots, \lambda_m \in [0, 1]$ son tales que $\sum_{i \in \mathbb{I}_m} \lambda_i = 1$ entonces, usando la notación anterior,

$$f\Big(\sum_{i \in \mathbb{I}_m} \lambda_i \, y_i \Big) = \Big(f\Big(\sum_{i \in \mathbb{I}_m} \lambda_i \, y_{ij} \Big) \Big)_{j \in \mathbb{I}_n} \leqslant \Big(\sum_{i \in \mathbb{I}_m} \lambda_i \, f(y_{ij}) \Big)_{j \in \mathbb{I}_n} = \sum_{i \in \mathbb{I}_m} \lambda_i \, f(y_i) \, ,$$

donde la desigualdad entre los vectores es coordenada a coordenada.

Teorema 2.19. *Sean $x, y \in \mathbb{R}^n$. Sea $I \subseteq \mathbb{R}$ un intervalo (semirrecta o todo $\mathbb{R}$) tal que $x, y \in I^n$. Entonces, son equivalentes:*

1. $x \prec y$
2. $\operatorname{tr} f(x) \leq \operatorname{tr} f(y)$ *para toda función convexa* $f : I \to \mathbb{R}$.
3. $\displaystyle\sum_{i=1}^{n} |x_i - t| \leq \sum_{i=1}^{n} |y_i - t|$ *para todo* $t \in \mathbb{R}$.

Análogamente, son equivalentes

1'. $x \prec_w y$ *(submayorización)*

2'. $\operatorname{tr} f(x) \leq \operatorname{tr} f(y)$ *para toda función* $f: \mathbb{R} \to \mathbb{R}$ *convexa y* **no decreciente**.

3'. $\displaystyle\sum_{i=1}^{n} (x_i - t)^+ \leq \sum_{i=1}^{n} (y_i - t)^+$ *para todo* $t \in \mathbb{R}$.

Demostración. Sólo probaremos la primer parte, puesto que los argumentos para probar la segunda son similares (para 1' → 2', se aplica 1 → 2 y la Proposición 2.17, que será útil para funciones no decrecientes). Supongamos que $x \prec y$. Entonces, por el Teorema 2.12,

$$x = \sum_{i=1}^{s} \lambda_i \, P_{\sigma_i}(y)$$

para ciertos $\lambda_i \geq 0$ que suman uno, y para ciertas $\sigma_i \in S_n$. Luego $f\big(\sum_{i=1}^{s} \lambda_i \, P_{\sigma_i} y\big) \leqslant \sum_{i=1}^{s} \lambda_i \, f\big(P_{\sigma_i} y\big)$ (i.e., se tiene desigualdad en cada cordenada), y

$$\operatorname{tr} f(x) = \operatorname{tr} f\Big(\sum_{i=1}^{s} \lambda_i \, P_{\sigma_i} y\Big) \leq \operatorname{tr} \sum_{i=1}^{s} \lambda_i \, f\big(P_{\sigma_i} y\big) = \sum_{i=1}^{s} \lambda_i \, \operatorname{tr} P_{\sigma_i}\big(f(y)\big) = \sum_{i=1}^{s} \lambda_i \, \operatorname{tr}\big(f(y)\big) = \operatorname{tr} f(y)\ .$$

La implicación $2 \to 3$ (respectivamente, $2' \to 3'$) se deduce de que la función $x \mapsto |x - t|$ (resp. $x \mapsto (x-t)^+$) es convexa (resp. convexa no decreciente) para todo $t \in \mathbb{R}$. Probemos $3 \to 1$. Supongamos que los vectores x e y están ordenados de forma decreciente (ni 3 ni 1 depende del orden de las cordenadas). Sean $M = \max\{x_1, y_1\}$ y $m = \min\{x_n, y_n\}$. Tomando $t > M$, se tiene que

$$\sum_{i=1}^{n} |y_i - t| = \sum_{i=1}^{n} t - y_i = nt - \sum_{i=1}^{n} y_i\ ,$$

y lo mismo para x. Luego la desigualdad del item 3 para estos valores de t implica que $\operatorname{tr} y \leq \operatorname{tr} x$. Análogamente, la desigualdad 3 para valores de t tales que $t < m$ implica que $\operatorname{tr} x \leq \operatorname{tr} y$. Luego $\operatorname{tr} x = \operatorname{tr} y$. Por otro lado, dado $z \in \mathbb{R}$, se tiene que $2z^+ = z + |z|$. Luego

$$\begin{aligned}
2\sum_{i=1}^{n} (x_i - t)^+ &= \sum_{i=1}^{n} (x_i - t) + \sum_{i=1}^{n} |x_i - t| = \operatorname{tr} x - nt + \sum_{i=1}^{n} |x_i - t| \\
&\leq \operatorname{tr} y - nt + \sum_{i=1}^{n} |x_i - t| = \sum_{i=1}^{n} (y_i - t) + \sum_{i=1}^{n} |y_i - t| \\
&= 2\sum_{i=1}^{n} (y_i - t)^+.
\end{aligned}$$

Fijemos $k \in \mathbb{I}_n$. Tomando $t = y_k$, resulta que

$$\sum_{i=1}^{n} (y_i - t)^+ = \sum_{i=1}^{k} (y_i - t)^+ = \sum_{i=1}^{k} y_i - kt\ .$$

Por lo tanto

$$\begin{aligned}
\sum_{i=1}^{k} x_i - kt = \sum_{i=1}^{k} (x_i - t) &\leq \sum_{i=1}^{k} (x_i - t)^+ \leq \sum_{i=1}^{n} (x_i - t)^+ \\
&\leq \sum_{i=1}^{n} (y_i - t)^+ = \sum_{i=1}^{k} y_i - kt,
\end{aligned}$$

lo cual muestra que $\displaystyle\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$. ∎

Corolario 2.20. *Sea $I \subseteq \mathbb{R}$ un intervalo. Sea $g : I \to \mathbb{R}$ una función convexa (resp. convexa creciente). Entonces, dados $x, y \in I^n$,*

$$x \prec y \quad (\textit{resp. } x \prec_w y) \quad \Longrightarrow \quad g(x) \prec_w g(y).$$

En particular $x \prec y \implies |x| \prec_w |y|$, y también $x \prec_w y \implies x^+ \prec_w y^+$.

Demostración. Sea $f : \mathbb{R} \to \mathbb{R}$ convexa no decreciente. Es fácil ver, entonces, que $f \circ g$ es una función convexa. Por el Teorema 2.19, si $x \prec y$, entonces

$$\operatorname{tr} f(g(x)) = \operatorname{tr} f \circ g\ (x) \leq \operatorname{tr} f \circ g\ (y) = \operatorname{tr} f(g(y)).$$

Pero por el mismo teorema (en su segunda parte), como lo anterior vale para toda $f : \mathbb{R} \to \mathbb{R}$ convexa no decreciente, deducimos que $g(x) \prec_w g(y)$.

Si g es creciente y $x \prec_w y$, por el Corolario 2.17 existe $u \in \mathbb{R}^n$ tal que $x \leqslant u \prec y$. Luego, por el caso anterior, $g(x) \leqslant g(u) \prec_w g(y)$. Para concluir que $g(x) \prec_w g(y)$ basta aplicar la Observación 2.16. ■

Corolario 2.21. *Sean $x, y \in \mathbb{R}^n$, tales que $x > 0$ e $y > 0$. Entonces, se tiene que*

$$x \prec y \quad \Longrightarrow \quad \prod_{i=1}^n x_i \geq \prod_{i=1}^n y_i\ .$$

Demostración. Sea $g(t) = -\log t$, que es una función convexa (pero decreciente), definida en $\mathrm{I} = (0, +\infty)$. Por el Corolario 2.20, si $x \prec y$, entonces $g(x) \prec_w g(y)$. En particular,

$$-\log \prod_{i=1}^n x_i = -\sum_{i=1}^n \log x_i \leq -\sum_{i=1}^n \log y_i = -\log \prod_{i=1}^n y_i\ ,$$

de lo que se concluye que $\prod_{i=1}^n x_i \geq \prod_{i=1}^n y_i$. ■

Corolario 2.22. *Sean $x, y \in \mathbb{R}^n$, tales que $x > 0$ e $y > 0$. Si se cumple que*

$$\prod_{i=1}^k x_i \leq \prod_{i=1}^k y_i\ , \qquad \textit{para todo} \quad k \in \mathbb{I}_n\ , \tag{5}$$

entonces $x \prec_w y$.

Demostración. Notar que la ecuación (5) equivale a decir que $\log(x) \prec_w \log(y)$. Como la función $t \mapsto \exp t := e^t$ es convexa y creciente, podemos aplicar el Corolario 2.20 y deducir que $x = \exp \log x \prec_w \exp \log y = y$. ■

Observación 2.23. Sean x, $y \in \mathbb{R}^n$ tales que $x \prec y$. Notemos que el item 3. del Teorema 2.19 indica entonces que la dispersión de las entradas del vector x alrededor de t es menor ó igual que la dispersión de las entradas del vector y alrededor de t, con respecto a cualquier $t \in \mathbb{R}$. En este caso, la dispersión es con respecto a la distancia usual en $\mathbb{R}$ dada por $d(a, b) = |a - b|$ para a, $b \in \mathbb{R}$. Más generalmente, si $p \in [1, \infty)$ entonces podemos considerar la dispersión con respecto a la distancia $d_p(a, b) = |a - b|^p$, a, $b \in \mathbb{R}$. Notemos que como $g_t(x) = |x - t|$ ($t \in \mathbb{R}$) es convexa en $\mathbb{R}$ y la función $f_p(x) = x^p$ ($p \in [1, \infty)$) es convexa y creciente en $\mathbb{R}_+$ entonces, el Teorema 2.19 muestra que $g(x) \prec_w g(y) \ \Rightarrow f_p(g(x)) \prec_w f_p(g(y))$. En particular

$$\sum_{i \in \mathbb{I}_n} |x_i - t|^p \leq \sum_{i \in \mathbb{I}_n} |y_i - t|^p$$

de forma que la dispersión de las entradas de x es menor ó igual que la dispersión de las entradas de y con respecto a toda una familia de medidas de dispersión. △

3. Aplicaciones: matrices autoadjuntas

Observación 3.1. En lo que sigue vamos a usar los siguientes hechos y nociones:

1. Si $A \in \mathcal{M}_n(\mathbb{C})$ entonces $A^* \in \mathcal{M}_n(\mathbb{C})$ denota $A^* = \overline{A^t}$ la transpuesta conjugada de A:
$$(A^*)_{jk} = \overline{A_{kj}} \quad \text{para} \quad j,\, k \in \mathbb{I}_n\,.$$
Al conjunto de todas las matrices $A \in \mathcal{M}_n(\mathbb{C})$ autoadjuntas, i.e. tales que $A^* = A$, lo notamos $\mathcal{H}(n)$.

2. Si $A \in \mathcal{H}(n)$ entonces existen reales $\lambda_1 \geq \ldots \geq \lambda_n$ y una base ortonormal (BO) $\{v_i\}_{i\in\mathbb{I}_n}$ de $\mathbb{C}^n$ tales que
$$A\, v_i = \lambda_i\, v_i \quad \text{para} \quad i \in \mathbb{I}_n$$
y en este caso notamos $\lambda(A) = (\lambda_i)_{i\in\mathbb{I}_n}$ (los autovalores de A contando multiplicidades y ordenados en forma decreciente).
Otra forma equivalente de lo anterior es: existen reales $\lambda_1 \geq \ldots \geq \lambda_n$ y una base ortonormal (BO) $\{v_i\}_{i\in\mathbb{I}_n}$ de $\mathbb{C}^n$ tales que
$$A = \sum_{i\in\mathbb{I}_n} \lambda_i\ v_i\, v_i^* \tag{6}$$
(recordemos que $\mathbb{C}^n = \mathbb{C}^{n\times 1}$ y luego $v_i^* \in \mathbb{C}^{1\times n}$ y entonces $v_i\, v_i^* \in \mathcal{H}(n)$).

3. Si $A \in \mathcal{H}(n)$ y consideramos una representación como en Eq. (6) entonces, dada una función $f : I \to \mathbb{C}$ con $\lambda(A) \in I^n$, definimos
$$f(A) = \sum_{i\in\mathbb{I}_n} f(\lambda_i)\ v_i\, v_i^* \tag{7}$$
que resulta una matriz normal en general. Notemos que los autovalores de $f(A)$ contando multiplicidades son: $f(\lambda_1), \ldots, f(\lambda_n) \in \mathbb{C}$. Si $f(\lambda_i) \in \mathbb{R}$, $i \in \mathbb{I}_n$, entonces $f(A) \in \mathcal{H}(n)$.

4. Una matriz $A \in \mathcal{H}(n)$ es semidefinida positiva (SDP) si $\lambda(A) \in \mathbb{R}^n_+$. Equivalentemente, $A \in \mathcal{H}(n)$ es SDP si solo si
$$\langle A\, v, v\rangle \geq 0 \quad \text{para todo} \quad v \in \mathbb{C}^n\,, \tag{8}$$
donde $\langle v\,,\, w\rangle = \sum_{i\in\mathbb{I}_n} v_i,\, \overline{w_i}$ denota el producto interno usual de $\mathbb{C}^n$. Al conjunto de todas las matrices semidefinidas positivas lo notamos $\mathcal{M}_n(\mathbb{C})^+$.

5. Dada $A \in \mathcal{M}_n(\mathbb{C})^+$, entonces $\lambda(A) \in \mathbb{R}^n_+$. Si $f : \mathbb{R}_+ \to \mathbb{R}_+$, $f(x) = x^{1/2}$, llamaremos $A^{1/2} = f(A) \in \mathcal{M}_n(\mathbb{C})^+$. Por construcción $(A^{1/2})^2 = A$.

6. Dada $B \in \mathcal{M}_n(\mathbb{C})$ y una BO arbitraria $\{w_i\}_{i\in\mathbb{I}_n}$ entonces la traza de B está dada por
$$\operatorname{tr}(B) = \sum_{i\in\mathbb{I}_n} \langle B\, w_i,\, w_i\rangle \in \mathbb{C}\,.$$
$\operatorname{tr}(B)$ no depende de la BO considerada. Recordemos las propiedades fundamentales de tr: es una funcional lineal que satisface $\operatorname{tr}(CD) = \operatorname{tr}(DC)$ para todas $C,\, D \in \mathcal{M}_n(\mathbb{C})$.

7. Si $A \in \mathcal{H}(n)$ entonces con autovalores $\lambda(A) \in \mathbb{R}^n$ entonces $\operatorname{tr}(A) = \operatorname{tr}(\lambda(A)) = \sum_{i\in\mathbb{I}_n} \lambda_i(A)$: basta tomar la BO de autovectores de A para calcular $\operatorname{tr}(A)$. (Vale mencionar que la fórmula vale para matrices arbitrarias, considerando los autovalores - posiblemente complejos - contando multiplicidades algebraicas).

$\triangle$

3.1. El teorema de Schur

La primer aplicación de la mayorización es la descripción de la relación que existe entre la diagonal de una matriz autoadjunta A y sus autovalores $\lambda(A)$.

En lo que sigue, dada $A \in \mathcal{H}(n)$ y $x \in \mathbb{R}^n$ notamos $\mathrm{d}\,(A) = (a_{11}, \dots, a_{nn}) \in \mathbb{R}^n$ a la diagonal principal de A (que resulta real, porqué?) y $\mathrm{diag}\,(x) \in \mathcal{H}(n)$ a la matriz diagonal tal que $\mathrm{d}\,(\mathrm{diag}\,(x)) = x$.

Teorema 3.2 (Teorema de mayorización de Schur)**.** *Sea $A \in \mathcal{H}(n)$ y $\lambda(A) \in \mathbb{R}^n$ el vector de los autovalores de A. Entonces,*

$$d\,(A) \prec \lambda(A).$$

Demostración. Para demostrar que $\mathrm{d}\,(A) \prec \lambda(A)$ vamos a probar que $\mathrm{d}\,(A) = B \cdot \lambda(A)$, para cierta $B \in \mathcal{DS}\,(n)$. Como $A \in \mathcal{H}(n)$, si $D = \mathrm{diag}\,(\lambda(A))$, existe $U \in \mathcal{U}(n)$ tal que $A = U^*DU$ (de hecho, U^* es la matriz cuyas columnas son los vectores de una BO de autovectores de A como en la Observación 3.1, item 2). Mediante cuentas elementales de matrices, se puede verificar que cada entrada de A tiene la forma: dados $i, j \in \mathbb{I}_n$,

$$a_{ij} = \sum_{k=1}^{n} \overline{u}_{ki} \lambda_k u_{kj} \ , \qquad \text{en particular,} \qquad a_{ii} = \sum_{k=1}^{n} \lambda_k |u_{ki}|^2.$$

Consideremos ahora la matriz $B = (|u_{ji}|^2)_{ij}$ que, por ser U unitaria, cumple

$$\mathrm{tr}(C_j(B)) = \|F_j(U)\|^2 = 1 \qquad \text{y} \qquad \mathrm{tr}(F_j(B)) = \|C_j(U)\|^2 = 1 \qquad \text{para} \qquad j \in \mathbb{I}_n \, .$$

Así, $B \in \mathcal{DS}\,(n)$. Además

$$B \cdot \lambda(A) = \begin{bmatrix} |u_{11}|^2 & \cdots & |u_{n1}|^2 \\ \vdots & \ddots & \vdots \\ |u_{1n}|^2 & \cdots & |u_{nn}|^2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{k=1}^{n} |u_{k1}|^2 \lambda_k \\ \vdots \\ \sum\limits_{k=1}^{n} |u_{kn}|^2 \lambda_k \end{bmatrix} = \mathrm{d}\,(A)\, .$$

Luego, el Teorema 2.12 completa la demostración. ∎

Proposición 3.3 (Principio del Máximo de Ky Fan)**.** *Sea $A \in \mathcal{H}(n)$. Entonces para todo $k \in \mathbb{I}_n$, se tiene que*

$$\sum_{j=1}^{k} \lambda_j(A) = \max \sum_{j=1}^{k} \langle A w_j, w_j \rangle,$$

donde el máximo se toma sobre todas las k-uplas ortonormales $\{w_1, ..., w_k\}$ en $\mathbb{C}^n$.

Demostración. Fijemos $k \in \mathbb{I}_n$. Sea $\{w_1, ..., w_k\}$ una k-upla cualquiera de vectores ortonormales. Sea $U \in \mathcal{U}(n)$ tal que sus primeras k columnas sean los vectores dados. Notemos $B = U^*AU$, que verifica $\lambda(B) = \lambda(A)$ y, además, si $\{e_i\}_{i \in \mathbb{I}_n}$ es la BO canónica

$$\sum_{j=1}^{k} \langle A w_j, w_j \rangle = \sum_{j=1}^{k} \langle AU\, e_j, U\, e_j \rangle = \sum_{j=1}^{k} \langle U^* A\, U e_j, e_j \rangle = \sum_{j=1}^{k} b_{jj} \, .$$

Por el Teorema de mayorización de Schur 3.2,

$$\sum_{j=1}^{k} \langle A w_j, w_j \rangle = \sum_{j=1}^{k} b_{jj} \leq \sum_{j=1}^{k} \lambda_j(A).$$

Si consideramos en particular una k-upla ortonormal $\{v_1, ..., v_k\}$ compuesta por autovectores de A correspondientes a los autovalores $\lambda_1(A), \dots, \lambda_k(A)$, obtenemos

$$\sum_{j=1}^{k} \langle A\, v_j\, ,\, v_j \rangle = \sum_{j=1}^{k} \langle \lambda_j(A)\, v_j\, ,\, v_j \rangle = \sum_{j=1}^{k} \lambda_j(A).$$

De esta manera vemos que se alcanza la igualdad cuando se toma el máximo sobre todas las k-uplas de vectores ortonormales. ■

Teorema 3.4. *Sean A y $B \in \mathcal{H}(n)$. Entonces*

$$\lambda(A+B) \prec \lambda(A) + \lambda(B).$$

Demostración. Sea $k \in \mathbb{I}_n$ y $\{w_1, \dots, w_k\}$ un conjunto ortonormal en $\mathbb{C}^n$ tal que

$$\sum_{j=1}^{k} \lambda_j(A+B) = \sum_{j=1}^{k} \langle (A+B)\, w_j\, ,\, w_j \rangle\, .$$

Entonces, por la Proposición 3.3 tenemos que

$$\begin{aligned} \sum_{j=1}^{k} \lambda_j(A+B) &= \sum_{j=1}^{k} \langle (A+B)\, w_j\, ,\, w_j \rangle \\ &= \sum_{j=1}^{k} \langle A\, w_j\, ,\, w_j \rangle + \sum_{j=1}^{k} \langle B\, w_j\, ,\, w_j \rangle \\ &\le \sum_{j=1}^{k} \lambda_j(A) + \sum_{j=1}^{k} \lambda_j(B) = \sum_{j=1}^{k} (\lambda(A) + \lambda(B))_j\, . \end{aligned}$$

Finalmente, para $k = n$ hay igualdad, porque $\operatorname{tr}(A+B) = \operatorname{tr}\,(A) + \operatorname{tr}(B)$. ■

3.2. El teorema de Schur-Horn

Para motivar el próximo resultado fundamental sobre mayorización, consideramos el siguiente problema de representación de matrices SDP: dada $A \in \mathcal{M}_n(\mathbb{C})^+$ entonces $\lambda(A) = (\lambda_i)_{i \in \mathbb{I}_n} \in \mathbb{R}_+^{*}$ y existe una BO $\{v_i\}_{i \in \mathbb{I}_n}$ de $\mathbb{C}^n$ tal que

$$A = \sum_{i \in \mathbb{I}_n} \lambda_i \; v_i\, v_i^*\, . \tag{9}$$

El hecho de que los v_i's en la Eq. (9) sean mutuamente ortogonales nos permite recuperar el hecho de que son autovectores de A y que $\lambda_1, \dots, \lambda_n$ son los autovalores de A. En efecto, como $v^* w = \langle w, v \rangle$ para $v,\, w \in \mathbb{C}^n$ vemos que

$$A v_j = \sum_{i \in \mathbb{I}_n} \lambda_i \; v_i\, v_i^*\, v_j = \sum_{i \in \mathbb{I}_n} \lambda_i \; \langle v_j\, ,\, v_i \rangle\, v_i = \lambda_j\, v_j\, .$$

Sin embargo, si consideramos vectores unitarios $w_1, \dots, w_n \in \mathbb{C}^n$ (no necesariamente ortogonales entre sí) y coeficientes $c_1, \dots, c_n \in \mathbb{R}_+$ entonces la matriz

$$\sum_{i \in \mathbb{I}_n} c_i \; w_i\, w_i^* \in \mathcal{M}_n(\mathbb{C})^+ \qquad \text{(porqué?)}$$

pero en general sus autovalores y autovectores no tienen porqué coincidir con los c_i's y w_i's. En este sentido, dada $A \in \mathcal{M}_n(\mathbb{C})^+$ y coeficientes $c_1, \ldots, c_m \in \mathbb{R}_+$ (con $n \neq m$ posiblemente) podemos preguntar bajo qué condiciones existen vectores unitarios $w_1, \ldots, w_m \in \mathbb{C}^n$ para los que se verifique

$$A = \sum_{i \in \mathbb{I}_m} c_i \; w_i \, w_i^* \, . \tag{10}$$

Una condición obviamente necesaria es que la dimensión del rango de A, notada $\text{rk}(A)$, verifique $\text{rk}(A) \leq m$. Para justificar el estudio de este problema consideramos el siguiente caso particular: si $m > n$ y $c_i = \frac{n}{m}$ para $i \in \mathbb{I}_m$ entonces veremos más adelante que existen vectores unitarios $w_1, \ldots, w_m \in \mathbb{C}^m$ tales que

$$I = \sum_{i \in \mathbb{I}_m} \frac{n}{m} \; w_i \, w_i^* \implies z = \frac{n}{m} \sum_{i \in \mathbb{I}_m} \langle z, w_i \rangle \, w_i \quad \text{para todo} \quad z \in \mathbb{C}^n \, . \tag{11}$$

Salvo el factor n/m, la representación de los vectores de $\mathbb{C}^n$ arriba es formalmente análoga a la representación con respecto a una BO de $\mathbb{C}^n$; sin embargo, el conjunto $\{w_1, \ldots, w_m\}$ es linealmente dependiente ($m > n$). La dependencia lineal de este sistema de generadores puede ser una ventaja en ciertas situaciones que surjen en las aplicaciones.

Para poder desarrollar una primera caracterización de la existencia de solución del problema anterior consideramos la siguiente construcción estándar en $\mathcal{M}_n(\mathbb{C})$.

Definición 3.5. Sea $A \in \mathcal{M}_n(\mathbb{C})$. Llamaremos "módulo de A" a la matriz

$$|A| = (A^*A)^{1/2} \in \mathcal{M}_n(\mathbb{C})^+.$$

$\triangle$

Teorema 3.6 (Descomposición polar). *Sea $A \in \mathcal{M}_n(\mathbb{C})$. Entonces*

1. *Para todo $x \in \mathbb{C}^n$, se verifica que $\|Ax\| = \|\,|A|x\|$.*

2. *Si definimos $\tilde{U} : R(|A|) \to R(A)$ dado por $\tilde{U}(|A|x) = Ax$, para $x \in \mathbb{C}^n$ entonces $\tilde{U}$ es una isometría entre los subespacios que son los rangos de $|A|$ y A respectivamente. Si $U \in \mathcal{U}(n)$ es cualquier extensión de $\tilde{U}$ a una isometría de $\mathbb{C}^n$ entonces*

$$A = U|A|,$$

que es llamada una **descomposición polar** *(DP) de A; en general no es única.*

3. *Cualquier $U \in \mathcal{U}(n)$ que cumpla $A = U|A|$, verifica que $UA^*AU^* = AA^*$, y por lo tanto*

$$U|A|U^* = |A^*| \quad y \quad A = |A^*|U.$$

Esto dice que U^* es un unitario admisible para la DP de A^*.

Demostración. Ver el apéndice. ■

Observación 3.7. Sea $W \in \mathcal{M}_n(\mathbb{C})$ y sea $W = U\,|W|$ una DP de W. Entonces, el Teorema 3.6 implica que

$$U\,W^*W\,U^* = W\,W^* \, ,$$

es decir, WW^* y W^*W son conjugados mediante un unitario (y en particular, son similares). En particular, concluimos que WW^* y W^*W tienen la misma ecuación característica de forma que $\lambda(WW^*) = \lambda(W^*W)$, dado que WW^*, $W^*W \in \mathcal{M}_n(\mathbb{C})^+ \subset \mathcal{H}(n)$. $\triangle$

Proposición 3.8. *Sean $c = (c_i)_{i \in \mathbb{I}_n} \in \mathbb{R}^n_+$ y $A \in \mathcal{M}_n(\mathbb{C})^+$. Son equivalentes:*

1. *Existen vectores unitarios* $w_1, \dots, w_n \in \mathbb{C}^n$ *tales que* $A = \sum_{j=1}^{n} c_j \ w_j \ w_j^*$.
2. Existe $B \in \mathcal{M}_n(\mathbb{C})^+$ tal que $\mathrm{d}\,(B) = c$ y $\lambda(B) = \lambda(A)$.

Demostración. Si se verifica 1, sea $W \in \mathcal{M}_n(\mathbb{C})$ definida por $We_k = C_k(W) = c_k^{1/2} w_k$, $k \in \mathbb{I}_n$. Veamos que $WW^* = A$. En efecto, notemos que si llamamos $W_k \in \mathcal{M}_n(\mathbb{C})$ a la matriz cuya columna k-ésima es $C_k(W)$, pero las demás son nulas, se tiene que

$$W = \sum_k W_k \ , \quad W_k W_j^* = 0 \quad \text{si} \quad j \neq k \quad \text{y} \quad W_k W_k^* = c_k \ w_k w_k^*.$$

Es claro que todo esto implica que $WW^* = A$. Por lo tanto $\lambda(A) = \lambda(WW^*) = \lambda(W^*W)$, por la Observación 3.7. Si $B = W^*W$, es fácil ver, además, que $B_{ii} = c_i \|w_i\|^2 = c_i$, $i \in \mathbb{I}_n$, lo que prueba 2. Recíprocamente, si $B \in \mathcal{M}_n(\mathbb{C})^+ \subset \mathcal{H}(n)$ cumple $\lambda(B) = \lambda(A)$ y $\mathrm{d}\,(B) = c$, sean $V, W \in \mathcal{U}(n)$ tales que

$$V^* A\, V = \operatorname{diag}(\lambda(A)) \quad \text{y} \quad W^* B\, W = \operatorname{diag}(\lambda(A)) \implies U^* A\, U = B \quad \text{para} \quad U = VW^* \in \mathcal{U}(n)\,.$$

Consideremos la matriz $W = A^{1/2}U$. Entonces

$$W^*W = U^* \,(A^{1/2})^2\, U = U^* A\, U = B \quad \text{y} \quad WW^* = A^{1/2} U \ U^* A^{1/2} = A,$$

mientras que $c_i = B_{ii} = \|C_i(W)\|^2$. Basta ahora definir $w_i = \dfrac{C_i(W)}{\|C_i(W)\|}$, si $c_i > 0$ ó $w_i = e_1$ si $c_i = 0$: entonces $C_i(W) = c_i^{1/2}\, w_i$ y se verifica como al comienzo de esta prueba que

$$A = WW^* = \sum_{j=1}^{n} c_j \ w_j \ w_j^* \ ,$$

que prueba 1. ■

La proposición anterior, junto con el Teorema de Schur 3.2 muestran que la relación de mayorización $c \prec \lambda(A)$ es una condición necesaria para que existan vectores unitarios $x_1, \dots, x_n \in \mathbb{C}^n$ tales que $A = \sum_{j=1}^{n} c_j \ x_j \ x_j^*$. El siguiente resultado debido a Alfred Horn (que complementa el Teorma de Schur) muestra que la relación de mayorización es también suficiente.

Teorema 3.9 (Schur-Horn)**.** *Sean* $b, c \in \mathbb{R}^n$. *Entonces son equivalentes:*

1. $c \prec b$.
2. *Existe* $B \in \mathcal{H}(n)$ *tal que* $\mathrm{d}\,(B) = c$ *y* $\lambda(B) = b^{\downarrow}$.

Si, además, b y c tienen entradas no negativas, lo anterior equivale a

3. *Existen vectores unitarios* $w_1, \dots, w_n \in \mathbb{C}^n$ *tales que*

$$\operatorname{diag}(b) = \sum_{j=1}^{n} c_j \ w_j \ w_j^*.$$

Para probar el teorema de Schur-Horn consideramos el siguiente lema:

Lema 3.10. *Sea* $A = (a_{ij})_{i,j \in \mathbb{I}_n} \in \mathcal{H}(n)$ *con* $d(A) = y$. *Sea* $\sigma = (k\ l) \in S_n$ *una transposición y*

$$x = (\lambda\, I + (1 - \lambda)\, P_\sigma)\, y \in \mathbb{R}^d\,.$$

Entonces, existe $U \in \mathcal{U}(n)$ *tal que* $d(U^*AU) = x$.

Demostración. Primero notamos que podemos asumir (sin pérdida de generalidad) que $\sigma = (1\ 2)$ y que $a_{11} \geq a_{22}$. Esta reducción es posible gracias a la siguiente observación (los detalles se dejan al lector): notemos que si $\tau \in S_n$ entonces $P_\tau \in \mathcal{U}_{\mathcal{P}}(n) \subseteq \mathcal{U}(n)$ y $P_\tau^* = P_{\tau^{-1}}$; además, $\langle P_\tau^* \, A \, P_\tau \, e_i \, , e_i\rangle = \langle A \, e_{\tau(i)}, e_{\tau(i)}\rangle$ de forma que

$$\mathrm{d}\,(P_\tau^* \, A \, P_\tau) = P_\tau(\mathrm{d}\,(A)) = (a_{\tau(1)\tau(1)}, \ldots, a_{\tau(n)\tau(n)}) \ .$$

Así, asumimos que $\sigma = (1\ 2)$ y $a_{11} \geq a_{22}$; entonces $y = (a_{11}, \ldots, a_{nn})$ y

$$\begin{aligned} x &= (\lambda \, a_{11} + (1-\lambda) \, a_{22}, (1-\lambda) \, a_{11} + \lambda \, a_{22}, a_{33}, \ldots, a_{nn}) \\ &= (x_1, x_2, a_{33}, \ldots, a_{nn}) \end{aligned}$$

En particular $a_{22} \leq x_1$, $x_2 \leq a_{11}$ y $a_1 + a_2 = x_1 + x_2$. Consideremos la ecuación

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix}^* \begin{pmatrix} a_{11} & a_{12} \\ \overline{a_{12}} & a_{22} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} = \begin{pmatrix} x_1 & z \\ \overline{z} & w \end{pmatrix} \qquad (\ z \in \mathbb{C} \ w \in \mathbb{R})$$

con c, $s \in \mathbb{R}$, $c^2 + s^2 = 1$, que garantiza que el (primer y) tercer factor es matriz unitaria. La igualdad de las entradas (1,1) equivale a

$$c^2 \, a_{11} - 2\, sc \operatorname{Re} a_{21} + s^2 \, a_{22} = x_1 \ .$$

La ecuación se transforma en cuadrática en la variable $t := s/c$,

$$(a_{22} - x_1) \, t^2 - 2 \, t \operatorname{Re} a_{21} + (a_{11} - x_1) = 0$$

con discriminante $4 \cdot [(\operatorname{Re} a_{21})^2 - (a_{11} - x_1)(a_{22} - x_1)] \geq 0$ dado que $(a_{11} - x_1)(a_{22} - x_1) \leq 0$. Entonces

$$\Rightarrow \quad c = \frac{1}{\sqrt{1+t^2}} \qquad \text{y} \qquad s = ct \ \text{ satisfacen la igualdad de las entradas (1,1)} \, .$$

Pero notemos que por construcción

$$a_{11} + a_{22} = \operatorname{tr}(\begin{pmatrix} c & s \\ -s & c \end{pmatrix}^* \begin{pmatrix} a_{11} & a_{12} \\ \overline{a_{12}} & a_{22} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix}) = x_1 + w \implies w = x_2$$

Así si c, $s \in \mathbb{R}$ son como antes (en particular $c^2 + s^2 = 1$)

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix}^* \begin{pmatrix} a_{11} & a_{12} \\ \overline{a_{12}} & a_{22} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} = \begin{pmatrix} x_1 & z \\ \overline{z} & x_2 \end{pmatrix} \qquad (\ z \in \mathbb{C}\ ) \, .$$

Definimos

$$U = \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \oplus I_{n-2} \quad \Rightarrow \quad U \in \mathcal{U}(n) \, .$$

Más aún,

$$U^*AU = \begin{pmatrix} x_1 & z & * & \cdots & * \\ \overline{z} & x_2 & * & \cdots & * \\ * & * & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

$$\Rightarrow \quad \mathrm{d}\,(U^*AU) = (x_1, x_2, a_{33}, \ldots, a_{nn}) = x \ .$$

$\blacksquare$

Ya podemos probar el teorema de Schur-Horn:

Demostración del Teorema 3.9. Notemos que 2. $\Rightarrow$ 1. es el Teorema de Schur, mientras que la equivalencia de 2. y 3. (bajo la hipótesis de que $b,\, c \in \mathbb{R}^n_+$) es la Proposición 3.8. De esta forma, solo resta probar la implicación 1. $\Rightarrow$ 2. Supongamos que $c \prec b$; entonces existen transposiciones de $\mathbb{I}_n$, σ_j and $\lambda_j \in [0,1]$, $1 \le j \le r$ tales que si $P_j = \lambda_j\, I + (1-\lambda_j)\, P_{\sigma_j}$ entonces

$$c = P_r \cdots P_1\, b$$

Por el lema previo existe $U_1 \in \mathcal{U}(n)$ tal que

$$\mathrm{d}\,(U_1^* \operatorname{diag}(b)\, U_1) = P_1\, \mathrm{d}\,(\operatorname{diag}(b)) = P_1\, b\ .$$

Aplicando el lema anterior nuevamente, a la matriz $U_1^*\operatorname{diag}(b)\, U_1 \in \mathcal{H}(n)$ con diagonal principal $P_1\, b$ y P_2, concluimos que existe $U_2 \in \mathcal{U}(n)$ tal que

$$\mathrm{d}\,(U_2^*\ (U_1^* \operatorname{diag}(b)\, U_1)\ U_2) = P_2\, P_1\, b\ .$$

Repitiendo el procedimiento anterior concluimos que existen $U_1, \ldots, U_r \in \mathcal{U}(n)$ tales que

$$\mathrm{d}\,(U_r^* \cdots U_1^* \operatorname{diag}(b)\, U_1 \cdots U_r) = P_r \cdots P_1\, b = c$$

$$\Rightarrow \quad \mathrm{d}\,((U_1 \cdots U_r)^* \operatorname{diag}(b)\ (U_1 \cdots U_r)) = c$$

de forma que basta tomar $B = \operatorname{diag}(b)$ y $U = U_1 \cdots U_r \in \mathcal{U}(n)$. ∎

En lo que sigue usaremos el siguiente hecho: dadas $A,\, B \in \mathcal{H}(n)$ entonces existe $U \in \mathcal{U}(n)$ tal que si $U\,AU^* = B$ si y solo si $\lambda(A) = \lambda(B)$ (para verificar este hecho basta recordar que toda matriz autoadjunta tiene asociada una BO de autovectores de $\mathbb{C}^n$).

Teorema 3.11. *Sea $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ y $A \in \mathcal{H}(n)$ tal que $\lambda(A) = a^\downarrow$. Entonces,*

$$\Big\{\ \mathrm{d}\,(\ UAU^*) : U \in \mathcal{U}(n)\Big\} = \{x \in \mathbb{R}^n : x \prec a\}.$$

Demostración. Si $B = UAU^*$, con $U \in \mathcal{U}(n)$, entoces $\lambda(A) = \lambda(B) = a^\downarrow$. Luego por el Teorema de mayorización de Schur 3.2, se tiene que $\mathrm{d}\,(B) \prec a$, que prueba la inclusión del conjunto izquierdo en el derecho.

Recíprocamente, si $x \in \mathbb{R}^n$ cumple $x \prec a$, por el Teorema 3.9 existe $B \in \mathcal{H}(n)$ tal que $\mathrm{d}\,(B) = x$ y $\lambda(B) = a^\downarrow$. Por lo tanto, usando el hecho mencionado previamente, debe existir $U \in \mathcal{U}(n)$ tal que $B = UAU^*$. Luego $x \in \{\mathrm{d}\,(UAU^*) : U \in \mathcal{U}(n)\}$, lo que prueba la otra inclusión. ∎

Corolario 3.12. *Sean $A \in \mathcal{M}_n(\mathbb{C})^+$ y $c \in \mathbb{R}^m_+$, con $m > n$. Entonces existen vectores unitarios $w_1, \ldots, w_m \in \mathbb{C}^n$ tales que*

$$A = \sum_{k=1}^{m} c_k\, w_k\, w_k^* \qquad \Longleftrightarrow \qquad c \prec (\lambda(A), 0, \ldots, 0) := \tilde{\lambda}(A) \in \mathbb{R}^m_+.$$

Demostración. Sea

$$A_1 = \begin{bmatrix} A & 0_{n,\, m-n} \\ 0_{m-n,\, n} & 0_{m-n,\, m-n} \end{bmatrix} \in \mathcal{M}_m(\mathbb{C})^+\ .$$

Entonces $\lambda(A_1) = \tilde{\lambda}(A) \in \mathbb{R}^m_+$.

Claim: la existencia de vectores unitarios $w_1, \ldots, w_m \in \mathbb{C}^n$ tales que $A = \sum_{k=1}^m c_k\, w_k\, w_k^*$ equivale a la existencia de vectores unitarios $\tilde{w}_1, \ldots, \tilde{w}_m \in \mathbb{C}^m$ tales que $\tilde{A} = \sum_{k=1}^m c_k\, \tilde{w}_k\, \tilde{w}_k^*$.

En efecto, si asumimos la existencia de los $w_1, \dots, w_m \in \mathbb{C}^n$ como arriba, definimos $\tilde{w}_i = (w_i, 0_{m-n})$ para $i \in \mathbb{I}_m$. Ahora es un ejercicio sencillo verificar que los $\tilde{w}_1, \dots, \tilde{w}_m \in \mathbb{C}^m$ tienen la propiedad deseada. Recíprocamente, si asumimos la existencia de los $\tilde{w}_1, \dots, \tilde{w}_m \in \mathbb{C}^m$ como arriba, entonces necesariamente se tiene que $\tilde{w}_i = (w_i, 0_{m-n})$ para vectores unitarios $w_i \in \mathbb{C}^n$, para $i \in \mathbb{I}_m$: esto se mirando la diagonal principal de A_1 pues, si $n+1 \le i \le m$

$$0 = (A_1)_{ii} = \sum_{k \in \mathbb{I}_m} (\tilde{w}_k \, \tilde{w}_k^*)_{ii} = \sum_{k \in \mathbb{I}_m} |(\tilde{w}_k)_i|^2 \quad \text{(verificarlo !)} \quad .$$

El claim anterior junto con el Teorema 3.9 prueban el resultado. ■

Observación 3.13. El resultado anterior resuelve el problema planteado al comienzo de esta sección, al menos para el caso $m \ge n$. Es fácil ver, usando el Teorema 3.9, que si $A \in \mathcal{M}_n(\mathbb{C})^+$, $\operatorname{rk}(A) \le m < n$ y $c \in \mathbb{R}_+^m$, entonces la condición necesaria y suficiente para que A pueda ser representado $A = \sum_{k=1}^m c_k \, w_k \, w_k^*$ para ciertos vectores unitarios $w_1, \dots, w_m \in \mathbb{C}^n$, es que $\lambda(A) \succ (c, 0, \dots, 0) \in \mathbb{R}^n$. En particular, la representación en la Eq. (11) es una consecuencia de la mayorización $(n/m)_{i \in \mathbb{I}_m} \prec (1, \dots, 1, 0_{m-n}) \in \mathbb{R}^m$ (ver Ejemplo 2.11). △

3.3. Desigualdades de Lidskii para autovalores

Sean $A, B \in \mathcal{H}(n)$. En general, el vector de autovalores de $A + B \in \mathcal{H}(n)$ no tiene una expresión explícita en función de $\lambda(A)$ y $\lambda(B)$. En este sentido, la relación sencilla relación de mayorización $\lambda(A+B) \prec \lambda(A)+\lambda(B)$ da una información importante sobre $\lambda(A+B)$ en términos de $\lambda(A)$ y $\lambda(B)$. La desigualdad de Lidskii da otra relación de mayorización fundamental entre los vectores $\lambda(A)$, $\lambda(B)$ y $\lambda(A+B)$.
Para poder desarrollar esta desigualdades, consideramos los siguientes hechos y resultados previos. En lo que sigue usamos el siguiente hecho fundamental de álgebra lineal: si $\mathcal{S}$, $\mathcal{M}$ son subespacios de $\mathbb{C}^n$ entonces $n \ge \dim(\mathcal{S} + \mathcal{M}) = \dim \mathcal{S} + \dim \mathcal{M} - \dim(\mathcal{S} \cap \mathcal{M})$.

Teorema 3.14 (Courant-Fisher). *Sea $A \in \mathcal{H}(n)$ y sea $k \in \mathbb{I}_n$. La letra $\mathcal{S}$ denotará subespacios de $\mathbb{C}^n$, y $\mathcal{S}_1$ denotará al conjunto de los vectores de $\mathcal{S}$ de norma uno. Entonces,*

$$\lambda_k(A) = \max_{\dim \mathcal{S} = k} \; \min_{x \in \mathcal{S}_1} \langle Ax, x \rangle .$$

Demostración. Sea $\{v_i\}_{i \in \mathbb{I}_n}$ una BO de $\mathbb{C}^n$ tal que $Av_i = \lambda_i(A)\, v_i$ para $i \in \mathbb{I}_n$. Sea $\mathcal{S}$ un subespacio de $\mathbb{C}^n$ tal que $\dim \mathcal{S} = k$ y sea $\mathcal{M}$ = el subespacio generado por $\{v_k, \dots, v_n\}$. Entonces $\dim \mathcal{S} + \dim \mathcal{M} = k + (n - k + 1) > n$ de forma que $\mathcal{S} \cap \mathcal{M} \ne \{0\}$ (ver el comentario antes del enunciado de este resultado). Si $z \in \mathcal{S} \cap \mathcal{M}$, $\|z\| = 1$ entonces (como $z \in \mathcal{M}$)

$$\langle Az, z \rangle = \sum_{i=k}^{n} \lambda_i(A)\, |\langle z, v_i \rangle|^2 \le \sum_{i=k}^{n} \lambda_k(A)\, |\langle z, v_i \rangle|^2 = \lambda_k(A)\, \|z\|^2 = \lambda_k(A) \, .$$

Como además $z \in \mathcal{S}$ entonces $\min_{x \in \mathcal{S}_1} \langle Ax, a \rangle \le \langle Az, z \rangle \le \lambda_k(A)$. Así

$$\max_{\dim \mathcal{S} = k} \; \min_{x \in \mathcal{S}_1} \langle Ax, x \rangle \le \lambda_k(A) \, .$$

Notar que si elegimos $\mathcal{S}'$ como el subespacio de $\mathbb{C}^n$ generado por $\{v_1, \dots, v_k\}$ entonces $\dim \mathcal{S}' = k$ y $\min_{x \in (\mathcal{S}')_1} \langle Ax, x \rangle = \langle A\, v_k, \, v_k \rangle = \lambda_k(A)$, donde usamos que $v_k \in \mathcal{S}'$ y las estimaciones anteriores. ■

Definición 3.15. *Sean $A, B \in \mathcal{H}(n)$. Decimos que B es más grande que A en el orden de operadores, notado $B \ge A$ ó $A \le B$, si $B - A \in \mathcal{M}_n(\mathbb{C})^+$. Equivalentemente, $A \le B$ si y solo si*

$$\langle Ax, \, x \rangle \le \langle Bx, \, x \rangle \quad \textit{para todo} \quad x \in \mathbb{C}^n \, .$$

△

Corolario 3.16 (Monotonía de Weyl)**.** *Sean A, $B \in \mathcal{H}(n)$ tales que $A \leq B$. Entonces $\lambda_k(A) \leq \lambda_k(B)$ para $k \in \mathbb{I}_n$.*

Demostración. Por hipótesis $\langle Ax, x\rangle \leq \langle Bx, x\rangle$ para $x \in \mathbb{C}^n$. Si $\mathcal{S}$ es un subespacio de $\mathbb{C}^n$ tal que $\dim \mathcal{S} = k$ entonces

$$\min_{x\in\mathcal{S}_1} \langle Ax, x\rangle \leq \min_{x\in\mathcal{S}_1} \langle Bx, x\rangle \implies \lambda_k(A) = \max_{\dim\mathcal{S}=k} \min_{x\in\mathcal{S}_1} \langle Ax, x\rangle \leq \max_{\dim\mathcal{S}=k} \min_{x\in\mathcal{S}_1} \langle Bx, x\rangle = \lambda_k(B).$$

■

Observación 3.17. Sea $B \in \mathcal{H}(n)$ y consideremos $f, g : \mathbb{R} \to \mathbb{R}_+$ dadas por $f(x) = x^+ = \text{máx}\{0, x\}$, $g(x) = x^- = \text{máx}\{0, -x\}$, para $x \in \mathbb{R}$. Entonces definimos

$$B^+ = f(B) \in \mathcal{M}_n(\mathbb{C})^+ \quad \text{ y } \quad B^- = g(B) \in \mathcal{M}_n(\mathbb{C})^+ ,$$

que son la **parte positiva** (PP) y **parte negativa** (PN) de B. Si $\{v_i\}_{i\in\mathbb{I}_n}$ es un BO de $\mathbb{C}^n$ tal que

$$B = \sum_{i\in\mathbb{I}_n} \lambda_i(B)\ v_i\, v_i^* \ \Rightarrow\ B^+ = \sum_{i\in\mathbb{I}_n,\, \lambda_i(B)\geq 0} \lambda_i(B)\ v_i\, v_i^* \quad \text{ y } \quad B^- = \sum_{i\in\mathbb{I}_n,\, \lambda_i(B)<0} -\lambda_i(B)\ v_i\, v_i^* \,.$$

Lo anterior muestra bien que $B = B^+ - B^-$. Por otro lado, $B \leq B^+$ pues si $x \in \mathbb{C}^n$ entonces

$$\langle Bx, x, \rangle = \sum_{i\in\mathbb{I}_n} \lambda_i(B)\ |\langle x, v_i\rangle|^2 \leq \sum_{i\in\mathbb{I}_n,\, \lambda_i(B)\geq 0} \lambda_i(B)\ |\langle x, v_i\rangle|^2 = \langle B^+x, x\rangle \,.$$

△

Comenzamos con una versión técnica de la desigualdad de Lidskii.

Teorema 3.18 (Lidskii - versión técnica)**.** *Sean $A, B \in \mathcal{H}(n)$, $k \in \mathbb{I}_n$ y $J \subseteq \mathbb{I}_n$ con $|J| = k$. Entonces*

$$\sum_{j\in J} \lambda_j(A + B) - \sum_{j\in J} \lambda_j(A) \leq \sum_{i=1}^{k} \lambda_i(B). \tag{12}$$

Demostración. Supongamos primero que $B \in \mathcal{H}(n)$ además verifica que $\lambda_k(B) = 0$. Sea $B = B_+ - B_-$ la descomposición de B en partes positiva y negativa de B. Como $\lambda_k(B) = 0 \geq \lambda_i(B)$ con $k \leq i \leq n$, se tiene que:

$$\text{tr}(B_+) = \sum_{j=1}^{k} \lambda_j(B).$$

Como $B \leq B^+$ entonces $A + B \leq A + B^+$; por el teorema de Weyl deducimos que $\lambda_j(A + B) \leq \lambda_j(A + B_+)$ para todo $j \in \mathbb{I}_n$. En consecuencia,

$$\sum_{j\in J} \Big(\lambda_j(A + B) - \lambda_j(A)\Big) \leq \sum_{j\in J} \Big(\lambda_j(A + B_+) - \lambda_j(A)\Big).$$

Finalmente, como $B^+ \geq 0$ entonces $A + B^+ \geq A$ y usando nuevamente el teorema de Weyl, resulta que $\lambda_j(A + B_+) \geq \lambda_j(A)$, para todo $j \in \mathbb{I}_n$. Por lo tanto

$$\begin{aligned}\sum_{j\in J} \Big(\lambda_j(A + B_+) - \lambda_j(A)\Big) &\leq \sum_{j=1}^{n} \Big(\lambda_j(A + B_+) - \lambda_j(A)\Big) \\ &= \text{tr}(A + B_+) - \text{tr}(A) = \text{tr}(B_+) \\ &= \sum_{j=1}^{k} \lambda_j(B)\ .\end{aligned}$$

Consideremos ahora $B \in \mathcal{H}(n)$ arbitraria. En este caso, si definimos $\tilde{B} = B - \lambda_k(B)I \in \mathcal{H}(n)$ entonces $\lambda(\tilde{B}) = (\lambda_i(B) - \lambda_k(B))_{i\in\mathbb{I}_n}$. En particular, $\lambda_k(\tilde{B}) = 0$. Por la primera parte de la prueba, y el hecho de que $A + \tilde{B} = (A + B) - \lambda_k(B)\, I$

$$\sum_{j\in J} \lambda_j(A+B) - k\,\lambda_k(B) = \sum_{j\in J} \lambda_j(A+\tilde{B}) \le \sum_{j\in J} \lambda_j(A) + \sum_{i=1}^{k} \lambda_i(\tilde{B}) = \sum_{j\in J} \lambda_j(A) + \sum_{i=1}^{k} \lambda_i(B) - k\,\lambda_k(B)$$

que prueba la desigualdad en Eq. (12) en el caso general. ∎

Teorema 3.19 (Desigualdad de Lidskii - caso autoadjunto)**.** *Sean $C, D \in \mathcal{H}(n)$. Entonces*

$$\lambda(C) - \lambda(D) \prec \lambda(C - D)\,.$$

Demostración. Reemplazando en el Teorema 3.18 (Lidskii versión técnica) $A = D$ por $B = C - D$ (de forma que $A + B = C$), obtenemos

$$\sum_{j=1}^{k} \big(\lambda(C) - \lambda(D)\big)^{\downarrow}_j = \max_{J\subseteq\mathbb{I}_n\ |J|=k} \left\{ \sum_{j\in J} \lambda_j(C) - \lambda_j(D) \right\} \le \sum_{j=1}^{k} \lambda_j(C - D), \tag{13}$$

para todo $k \in \mathbb{I}_n$. Además,

$$\operatorname{tr}(\lambda(C) - \lambda(D)) = \operatorname{tr}(C) - \operatorname{tr}(D) = \operatorname{tr}(C - D) = \operatorname{tr}(\lambda(C - D))\,.$$

Entonces, $\lambda(C) - \lambda(D) \prec \lambda(C - D)$. ∎

3.4. Valores singulares y normas unitariamente invariantes.

Valores singulares. Comenzamos recordando la noción de valor singular de una matriz y su descomposición en valores singulares.

Definición 3.20. Sea $A \in \mathcal{M}_n(\mathbb{C})$,

1. Recordemos que el módulo de A está definido por $|A| = (A^*A)^{1/2} \in \mathcal{M}_n(\mathbb{C})^+$.
2. Llamaremos **valores singulares** de A a los autovalores de $|A|$ ordenados en forma decreciente, notándolos $s_1(A) \ge \cdots \ge s_n(A) \ge 0$ es decir, $s_i(A) = \lambda_i(|A|) = \lambda_i(A^*A)^{1/2}$, $i \in \mathbb{I}_n$.
3. Llamaremos $s(A) = (s_1(A), \ldots, s_n(A)) = \lambda(|A|)$ y $\Sigma(A)$ a la matriz diagonal

$$\Sigma(A) = \operatorname{diag}(s(A)) = \begin{pmatrix} s_1(A) & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & s_n(A) \end{pmatrix}.$$

△

Teorema 3.21 (Descomposición en valores singulares)**.** *Sea $A \in \mathcal{M}_n(\mathbb{C})$. Entonces existen dos matrices unitarias $V, W \in \mathcal{M}_n(\mathbb{C})$ tales que*

$$A = W\Sigma(A)V^*.$$

En este caso, las columnas $C_i(V)$ forman una B.O. de autovectores de $|A|$ (y A^*A), y las columnas $C_i(W)$ forman una B.O. de autovectores de $|A^*|$ (y AA^*).

Demostración. Veamos la existencia de la descomposición en valores singulares; el resto de las propiedades enunciadas queda como ejercicio para el lector. Primero consideramos una descomposición polar $A = U\,|A|$, donde $U \in \mathcal{U}(n)$. Como $|A| \in \mathcal{M}_n(\mathbb{C})^+$ podemos diagonalizar esta matriz mediante un unitario $V \in \mathcal{U}(n)$ de forma que $V^*\,|A|\,V = \operatorname{diag}(\lambda(|A|)) = \Sigma(A)$, puesto que $\lambda(|A|) = s(A)$. Así

$$U\,V\,\Sigma(A)\,V^* = U\,|A| = A\,,$$

con lo que basta definir $W = UV \in \mathcal{U}(n)$. La prueba dl resto de las afirmaciones es un ejercicio para el lector. ■

El próximo resultado se lo conoce como el truco del sombrero.

Teorema 3.22. *dada $A \in \mathcal{M}_n(\mathbb{C})$, si definimos*

$$\widehat{A} = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \in \mathcal{H}(2n),$$

entonces se tiene que

$$\lambda(\widehat{A}) = (s_1(A), \cdots, s_n(A), -s_n(A), \cdots, -s_1(A)) \in (\mathbb{R}^{2n})^{\downarrow}\,, \tag{14}$$

donde $s(A) = (s_1(A), \ldots, s_n(A)) \in (\mathbb{R}^n_+)^{\downarrow}$ denota el vector de valores singulares de A.

Demostración. Consideremos matrices unitarias V, $W \in \mathcal{U}(n)$ para una descomposición en valores singulares de A, de forma que $A = W\,\Sigma(A)\,V^*$. Entonces construimos la matriz unitaria

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} W & V \\ -W & V \end{pmatrix} \in \mathcal{M}_{2n}(\mathbb{C})\,.$$

La verificación de que U es matriz unitaria se puede hacer notando que, en general, si tenemos matrices $X, Y \in \mathcal{M}_{2n}(\mathbb{C})$ por bloques cuadrados de tamaño n es decir tales que

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \quad \text{y} \quad Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}$$

entonces se verifica que (ejercicio para el lector)

$$X^* = \begin{pmatrix} X_{11}^* & X_{21}^* \\ X_{12}^* & X_{22}^* \end{pmatrix} \quad \text{y} \quad X\,Y = \begin{pmatrix} X_{11}\,Y_{11} + X_{12}\,Y_{21} & X_{11}\,Y_{12} + X_{12}\,Y_{22} \\ X_{21}\,Y_{11} + X_{22}\,Y_{21} & X_{21}\,Y_{12} + X_{22}\,Y_{22} \end{pmatrix}.$$

De hecho, operando con bloques también podemos verificar que

$$\frac{1}{\sqrt{2}}\begin{pmatrix} W & V \\ -W & V \end{pmatrix}\begin{pmatrix} \Sigma(A) & 0 \\ 0 & -\Sigma(A) \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} W^* & -W^* \\ V^* & V^* \end{pmatrix} = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$$

ó, de forma equivalente, que $U^*\widehat{A}\,U$ es la matriz diagonal con diagonal principal $(s(A), -s(A)) \in \mathbb{R}^{2n}$. De esta forma, el vector de autovalores (ordenado en forma decreciente y contando multiplicidades) de $\widehat{A}$ resulta $(s_1(A), \cdots, s_n(A), -s_n(A), \cdots, -s_1(A))$.

■

Corolario 3.23. *Sean A, $B \in \mathcal{M}_n(\mathbb{C})$. Entonces $s(A+B) \prec_w s(A) + s(B)$.*

Demostración. Dadas A, $B \in \mathcal{M}_n(\mathbb{C})$ entonces

$$\widehat{A} + \widehat{B} = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} + \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} = \begin{pmatrix} 0 & A+B \\ (A+B)^* & 0 \end{pmatrix} = \widehat{A+B} \in \mathcal{H}(2n)\,.$$

Por el Corolario 3.4 deducimos que

$$\lambda(\widehat{A+B}) = \lambda(\widehat{A} + \widehat{B}) \prec \lambda(\widehat{A}) + \lambda(\widehat{B}) \, .$$

Por el truco del sombrero (Teorema 3.22) y la relación de mayorización anterior vemos que si $k \in \mathbb{I}_n$ entonces

$$\sum_{i \in \mathbb{I}_k} s_i(A+B) = \sum_{i \in \mathbb{I}_k} \lambda_i(\widehat{A+B}) \le \sum_{i \in \mathbb{I}_k} \lambda_i(\widehat{A}) + \lambda_i(\widehat{B}) = \sum_{i \in \mathbb{I}_k} s_i(A) + s_i(B) \, .$$

■

Corolario 3.24 (Desigualdad de Lidskii para valores singulares)**.** *Sean* A, $B \in \mathcal{M}_n(\mathbb{C})$. *Entonces* $|s(A) - s(B)| \prec_w s(A-B)$.

Demostración. Argumentando como en la prueba anterior, notemos que dadas A, $B \in \mathcal{M}_n(\mathbb{C})$ entonces $\widehat{A-B} = \widehat{A} - \widehat{B} \in \mathcal{H}(2n)$. Si aplicamos la desigualdad de Lidskii a las matrices autoadjuntas $\widehat{A}$, $\widehat{B}$ vemos que

$$\lambda(\widehat{A}) - \lambda(\widehat{B}) \prec \lambda(\widehat{A} - \widehat{B}) = \lambda(\widehat{A-B}) \, . \tag{15}$$

Notemos que

$$\lambda(\widehat{A}) - \lambda(\widehat{B}) = (s_1(A)-s_1(B), \ldots, s_n(A)-s_n(B), -(s_n(A)-s_n(B)), \ldots, -(s_1(A)-s_1(B))) \in \mathbb{R}^{2n} \, .$$

Por otro lado, $|s(A) - s(B)| = (|s_i(A) - s_i(B)|)_{i \in \mathbb{I}_n}$. Sea $\sigma \in S_n$ tal que

$$|s(A) - s(B)|^{\downarrow} = |s(A) - s(B)|_{\sigma} \quad \text{i.e.} \quad |s_{\sigma(1)}(A) - s_{\sigma(1)}(B)| \ge \ldots \ge |s_{\sigma(n)}(A) - s_{\sigma(n)}(B)| \ge 0 \, .$$

Dado $k \in \mathbb{I}_n$ definimos $J \subset \mathbb{I}_{2n}$ de forma que: $j \in J$ si se verifica alguna de las siguiente condiciones:

1. $j \in \sigma(\mathbb{I}_k)$, $s_j(A) - s_j(B) \ge 0$;
2. $j \in \{\sigma(1) + n, \ldots \sigma(k) + n\} = \sigma(\mathbb{I}_k) + n$, $-(s_j(A) - s_j(B)) > 0$.

Entonces $|J| = k$ y

$$\sum_{j \in \mathbb{I}_k} |s(A) - s(B)|^{\downarrow}_j = \sum_{j \in J} (\lambda(\widehat{A}) - \lambda(\widehat{B}))_j \le \sum_{j \in \mathbb{I}_k} \lambda(\widehat{A-B})^{\downarrow}_j = \sum_{j \in \mathbb{I}_k} s_j(A-B) \, ,$$

donde hemos usado la definición de J, el Corolario 2.9, la relación de mayorización en la Eq. (15) y el truco del sombrero (Teorema 3.22). ■

Normas unitariamente invariantes. En lo que sigue consideramos la siguiente clase de normas fundamentales del análisis matricial.

Definición 3.25. Una norma N en $\mathcal{M}_n(\mathbb{C})$ se dice que es una norma unitariamente invariante (NUI) , si cumple que

$$N(UAV) = N(A)$$

para toda $A \in \mathcal{M}_n(\mathbb{C})$ y $U, V \in \mathcal{U}(n)$. Notar que, en tal caso, por el Teorema 3.6 se tiene que $N(A) = N(\Sigma(A))$. △

Ejemplos 3.26. A modo de ejemplo, consideramos las siguientes familias de nuis en $\mathcal{M}_n(\mathbb{C})$.

1. Si $k \in \mathbb{I}_n$ definimos la k-norma KyFan, dada por $\|A\|_{(k)} = \sum_{j \in \mathbb{I}_k} s_j(A)$ para $A \in \mathcal{M}_n(\mathbb{C})$.

2. Si $p \in (1, \infty)$ definimos la norma p, dada por $\|A\|_p = (\sum_{j \in \mathbb{I}_n} s_j(A)^p)^{1/p}$ para $A \in \mathcal{M}_n(\mathbb{C})$.

El hecho de que estas sean nuis en $\mathcal{M}_n(\mathbb{C})$ es una consecuencia de que los valores singulares no cambian por multiplicar por unitarios (de cualquier lado) y del Corolario 3.23. Más adelante veremos una caracterización de las nuis que permite verificar que las expresiones de arriba son nuis.
Notemos que estas familias de nuis están relacionadas "por los extremos", en el siguiente sentido

$$\lim_{p \to 1^+} \|A\|_p = \|A\|_{(n)} \qquad \text{y} \qquad \lim_{p \to \infty} \|A\|_p = \|A\|_{(1)} .$$

A raíz de las identidades anteriores también se nota $\|A\|_1$ para la n-norma KyFan, y $\|A\|_\infty$ para la 1-norma KyFan, que se denominan norma traza y norma espectral respectivamente. $\triangle$

Definición 3.27. Dada una norma $N(\cdot)$ unitariamente invariante, se define la función $g_N : \mathbb{R}^n \to \mathbb{R}_+$ como

$$g_N(x_1, \ldots, x_n) = N\left(\operatorname{diag}(x_1, \ldots, x_n)\right) .$$

$\triangle$

Proposición 3.28. *Sea N una NUI. Entonces:*

1. *g_N es una norma en $\mathbb{R}^n$.*
2. *$g_N(x_1, \ldots, x_n) = g_N(|x_1|, \ldots, |x_n|)$.*
3. *$g_N(x_1, \ldots, x_n) = g_N(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$, para toda $\sigma \in S_n$.*

Observación 3.29. Una función $f : \mathbb{R}^n \to \mathbb{R}$ que cumple los ítems 1, 2 y 3 de la Proposición anterior se denomina **gauge simétrica**.

Demostración.

1. Se deduce de que la aplicación $\mathbb{C}^n \ni x \mapsto \operatorname{diag}(x) \in \mathcal{M}_n(\mathbb{C})$ es lineal e inyectiva y que N es una norma en $\mathcal{M}_n(\mathbb{C})$.

2. Sea $x_j = \omega_j |x_j|$ donde $w_j = e^{i\theta_j}$. Luego, como $\operatorname{diag}(\omega_1, \ldots, \omega_n) \in \mathcal{U}(n)$,

$$\begin{aligned} g_N(|x_1|, \ldots, |x_n|) &= N(\operatorname{diag}(|x_1|, \ldots, |x_n|)) \\ &= N(\operatorname{diag}(\omega_1, \ldots, \omega_n) \operatorname{diag}(|x_1|, \ldots, |x_n|)) \\ &= N(\operatorname{diag}(x_1, \ldots, x_n)) \\ &= g_N(x_1, \ldots, x_n). \end{aligned}$$

3. Si $\sigma \in S_n$ es una permutación, sea $P_\sigma \in \mathcal{U}_{\mathcal{P}}(n)$ la matriz de permutación asociada, que resulta unitaria. Además, se verifica que

$$P_\sigma \operatorname{diag}(x_1, \ldots, x_n) P_\sigma^{-1} = \operatorname{diag}\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right) .$$

Entonces,

$$\begin{aligned} g_N(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) &= N(\operatorname{diag}\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right)) \\ &= N(P_\sigma \operatorname{diag}(x_1, \ldots, x_n) P_\sigma^{-1}) \\ &= N(\operatorname{diag}(x_1, \ldots, x_n)) \\ &= g_N(x_1, \ldots, x_n). \end{aligned}$$

■

Lema 3.30. *Si g es una función gauge simétrica, entonces, g es monótona, es decir, si $|x_i| \le |y_i|$ para todo $i \in \{1, \dots, n\}$, entonces, $g(x) \le g(y)$.*

Demostración. Sin pérdida de generalidad, podemos suponer que $x, y \in \mathbb{R}^n_+$. Por un argumento inductivo, es suficiente verificar que si $t \in [0, 1]$, entonces

$$g(y_1, \dots, ty_k, \dots, y_n) \le g(y_1, \dots, y_k, \dots, y_n).$$

Verifiquemos esto último:

$$\begin{aligned}
g(y_1, \dots, ty_k, \dots, y_n) &= g\left(\left(\frac{1+t}{2}y_1, \dots, \frac{1+t}{2}y_k, \dots, \frac{1+t}{2}y_n\right)\right. \\
&\quad + \left.\left(\frac{1-t}{2}y_1, \dots, \frac{1-t}{2}(-y_k), \dots, \frac{1-t}{2}y_n\right)\right) \\
&\le \frac{1+t}{2}\, g(y) + \frac{1-t}{2}\, g(y_1, \dots, -y_k, \dots, y_n) \\
&= \frac{1+t}{2}\, g(y) + \frac{1-t}{2}\, g(y) = g(y).
\end{aligned}$$

■

Teorema 3.31. *Sea g es una función gauge simétrica y $x, y \in \mathbb{R}^n_+$ tales que $x \prec_w y$. Entonces, $g(x) \le g(y)$.*

Demostración. Como $x \prec_w y$, por la Proposición 2.17, existe $u \in \mathbb{R}^n$ tal que $x \leqslant u \prec y$. Por el lema anterior, $g(x) \le g(u)$. Dado $\sigma \in S_n$, notemos $y_\sigma = (y_{\sigma(1)}, \dots, y_{\sigma(n)})$. Por el Teorema 2.12, existe $A \subseteq S_n$ tal que $u = \sum_{\sigma \in A} \lambda_\sigma y_\sigma$ para ciertos λ_σ tales que $\lambda_\sigma \in [0, 1]$ y $\sum_{\sigma \in A} \lambda_\sigma = 1$. Luego

$$g(u) = g\left(\sum_{\sigma \in A} \lambda_\sigma\, y_\sigma\right) \le \sum_{\sigma \in A} \lambda_\sigma\, g(y_\sigma) = \sum_{\sigma \in A} \lambda_\sigma\, g(y) = g(y).$$

■

Teorema 3.32.

1. *Si N es una NUI, entonces, g_N es una función gauge simétrica.*

2. *Si g es una función gauge simétrica, entonces,*

$$\|A\|_g = g(s_1(A), \dots, s_n(A)) \ , \quad A \in \mathcal{M}_n(\mathbb{C})$$

   es una NUI en $\mathcal{M}_n(\mathbb{C})$.

Demostración. 1. Esto es la Proposición 3.28.

2. Sólo demostraremos la desigualdad triangular. Las demás propiedades quedan como ejercicio para el lector. Sean $A, B \in \mathcal{M}_n(\mathbb{C})$. Por el Corolario 3.23 tenemos que $s(A + B) \prec_w s(A) + s(B)$, y entonces

$$\begin{aligned}
\|A + B\|_g &= g(s(A + B)) \le g(s(A) + s(B)) \\
&\le g(s(A)) + g(s(B)) = \|A\|_g + \|B\|_g.
\end{aligned}$$

■

El Teorema 3.32 permite ahora verificar que las expresiones en los Ejemplos 3.26 corresponden a nuis en $\mathcal{M}_n(\mathbb{C})$.

Teorema 3.33. *Sean $A, B \in \mathcal{M}_n(\mathbb{C})$. Entonces son equivalentes:*

1. $N(A) \leq N(B)$ *para toda norma unitariamente invariante* N.
2. $\|A\|_{(k)} \leq \|B\|_{(k)}$ *para todo* $k \in \mathbb{I}_n$.
3. $s(A) \prec_w s(B)$.

Demostración. Es consecuencia del Teorema 3.31 para funciones gauge simétricas, y del Teorema 3.32. En efecto, si $\|A\|_{(k)} \leq \|B\|_{(k)}$ para todo $k \in \mathbb{I}_n$, entonces $s(A) \prec_w s(B)$. Por lo tanto $g(s(A)) \leq g(s(B))$ para toda función gauge simétrica. La recíproca es evidente. ■

Corolario 3.34. *Sean $A, B \in \mathcal{M}_n(\mathbb{C})^+$ tales que $A \leq B$. Entonces, $N(A) \leq N(B)$ para toda norma unitariamente invariante N.*

Demostración. Aplicando el Corolario 3.16, obtenemos que

$$s_k(A) = \lambda_k(A) \leq \lambda_k(B) = s_k(B) \ , \quad \text{para todo} \quad k \in \mathbb{I}_n \ .$$

Luego basta aplicar el Teorema 3.33. ■

Corolario 3.35. *Dadas $A, B \in \mathcal{H}(n)$. Entonces*

$$\lambda(A) \prec \lambda(B) \quad \Longrightarrow \quad s(A) \prec_w s(B).$$

Demostración. Sea $f(t) = |t|$, $t \in \mathbb{R}$. Como f es una función convexa, por el Teorema 2.19 tenemos que

$$\lambda(A) \prec \lambda(B) \implies (|\lambda_i(A)|)_{i\in\mathbb{I}_n} \prec_w (|\lambda_i(B)|)_{i\in\mathbb{I}_n} \ .$$

Como A, $B \in \mathcal{H}(n)$, se tiene que $s(A) = (|\lambda_i(A)|)_{i\in\mathbb{I}_n}{}^{\downarrow}$ y $s(B) = (|\lambda_i(B)|)_{i\in\mathbb{I}_n}{}^{\downarrow}$. Por lo tanto, $s(A) \prec_w s(B)$.

■

Corolario 3.36. *Dadas $A, B \in \mathcal{H}(n)$, si $\lambda(A) \prec \lambda(B)$ entonces $N(A) \leq N(B)$ para toda norma unitariamente invariante N.*

Demostración. Se deduce del Corolario 3.35 y del Teorema 3.32. ■

4. Comentarios finales

La teoría de mayorización de vectores es elemental y a su vez profunda, en términos de los resultados que se obtienen a partir de ella. En estas notas se ha cubierto de forma más que modesta una serie de tópicos sesgados hacia el análisis matricial. Para una lectura más detallada se recomienda el texto [5]. Como hemos mencionado, esta exposición está basada en el texto [1]. Otras referencias clásicas en donde se desarrolla la mayorización son los libros [2, 3, 4]. La influencia de la relación de mayorización en el desarrollo de desigualdades en ambientes no conmutativos puede apreciarse en los textos de análisis matricial previamente mencionados y también en el texto [6]. La mayorización como herramienta para probar desigualdades traciales así como los teoremas de tipo Schur-Horn (de matrices y en en el contexto más general de álgebras de operadores) han atraído la atención del expositor y están relacionados con parte de su trabajo de investigación, lo que ha motivado la elección de este tópico para el curso.

Referencias

[1] J. Antezana y D. Stojanoff, *Análisis Matricial.* Cursos y Seminarios de Matemática - Serie B. Depto. de Matemática - FCEyN - UBA (297 pags.) 2008.

[2] R. Bhatia; *Matrix Analysis*, Springer, New York, 1997.

[3] R. Horn, C. Johnson, *Matrix analysis.* Second edition. Cambridge University Press, Cambridge, (643 pp) 2013.

[4] R. Horn y C. Johnson; *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[5] A. Marshall, I. Olkin, B. Arnold, Inequalities: theory of majorization and its applications. Second edition. Springer Series in Statistics. Springer, New York, (909 pp) 2011.

[6] B. Simon, *Trace ideals and their applications.* Second edition. Mathematical Surveys and Monographs, 120. American Mathematical Society, Providence, RI, (150 pp) 2005.

5. Apéndice: descomposición polar de matrices

En esta sección probamos el Teorema 3.6 sobre la descomposición polar. Recordemos su enunciado:

Teorema. *Sea $A \in \mathcal{M}_n(\mathbb{C})$. Entonces*

1. *Para todo $x \in \mathbb{C}^n$, se verifica que $\|Ax\| = \|\,|A|x\|$.*

2. *Si definimos $\tilde{U} : R(|A|) \to R(A)$ dado por $\tilde{U}(|A|x) = Ax$, para $x \in \mathbb{C}^n$ entonces $\tilde{U}$ es una isometría entre los subespacios que son los rangos de $|A|$ y A respectivamente. Si $U \in \mathcal{U}(n)$ es cualquier extensión de $\tilde{U}$ a una isometría de $\mathbb{C}^n$ entonces*

$$A = U|A|,$$

que es la llamada **descomposición polar** *(DP) de A, aunque no es siempre única.*

3. *Cualquier $U \in \mathcal{U}(n)$ que cumpla $A = U|A|$, verifica que $UA^*AU^* = AA^*$, y por lo tanto*

$$U|A|U^* = |A^*| \quad y \quad A = |A^*|U.$$

Esto dice que U^* es un unitario admisible para la DP de A^*.

Demostración. Si $A = 0$ el enunciado es trivialmente cierto ($|A| = 0$ y se puede tomar cualquier unitario $U \in \mathcal{U}(n)$). Suponemos entonces que $A \in \mathcal{M}_n(\mathbb{C})$, $A \neq 0$. Si $x \in \mathbb{C}^n$ entonces

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

donde hemos usado que $A^*A = |A|^2$. Si consideramos los dos subespacios de $\mathbb{C}^n$ dados por los rangos $R(|A|)$ y $R(A)$ podemos definir la función

$$\tilde{U} : R(|A|) \to R(A) \quad \text{dada por} \quad \tilde{U}(|A|x) = Ax\ , \quad x \in \mathbb{C}^n\,.$$

Notemos que $\tilde{U}$ está bien definida: si $x,\, y \in \mathbb{C}^n$ son tales que $|A|x = |A|y$ entonces $|A|(x-y) = 0$ entonces $\||A|(x-y)\,\| = 0$ de forma que $\|A(x-y)\| = \||A|(x-y)\,\| = 0$ y entonces $Ax = Ay$. Además, es sencillo verificar que $\tilde{U}$ es lineal. Como $\|\tilde{U}w\| = \|w\|$ por construcción y por el item 1. ($w = |A|x$ para algún $x \in \mathbb{C}^n$ de forma que $\tilde{U}w = Ax$) entonces $\tilde{U}$ es una isometría entre

$R(|A|)$ y $R(A)$. En particular, $\dim R(|A|) = \dim R(A)$ de forma que $\dim R(|A|)^\perp = \dim R(A)^\perp$. Sea $V : \dim R(|A|)^\perp \to \dim R(A)^\perp$ una isometría (arbitraria) y definamos

$$U : \mathbb{C}^n = R(|A|) \oplus R(|A|)^\perp \to \mathbb{C}^n = R(A) \oplus R(A)^\perp$$

mediante

$$U(x_1 \oplus x_2) = \tilde{U} x_1 \oplus V x_2 \quad \text{donde} \quad x_1 \in R(|A|) \quad \text{y} \quad x_2 \in R(|A|)^\perp .$$

Es sencillo verificar que la matriz de U con respecto a la base canónica (que seguimos denotando U) satisface $U \in \mathcal{U}(n)$ (pues U es una isometría de $\mathbb{C}^n$ en $\mathbb{C}^n$) y es tal que

$$U|A|x = U(|A|x \oplus 0) = \tilde{U}(|A|x) \oplus V0 = Ax \quad \text{para} \quad x \in \mathbb{C}^n .$$

Para verificar el item 3 notamos que si $A = U|A|$ entonces

$$UA^*AU^* = U|A|^2U^* = (U|A|)(|A|U^*) = AA^* .$$

En este caso, si $f : \mathbb{R}_+ \to \mathbb{C}$ entonces $f(AA^*) = f(U(A^*A)U^*) = Uf(A^*A)U^*$, por construcción del cálculo funcional. En particular, si $f(x) = x^{1/2}$ para $x \in \mathbb{R}_+$ entonces

$$|A^*| = f(AA^*) = Uf(A^*A)U^* = U|A|U^* .$$

La prueba de las afirmaciones restantes se deja a cargo del lector. ■