Mixba'al Rev. Met. de Mat.
N$^{\underline{o}}$ 1, Vol. I, Junio 2010. 65 – 88

Espacios L_p no conmutativos

Gildardo Barrientos Roberto Quezada

Resumen

Siguiendo la referencia [2], presentamos una construcción elemental de los espacios L_p no conmutativos para matrices complejas $n \times n$.

Palabras clave: Valores singulares, normas de Schatten, normas de Ky Fan, normas Unitariamente Invariantes, desigualdades tipo Minkowski.

Clasificación de la AMS: 15A60, 81P45.

1. Introducción.

Los espacios L_p no conmutativos, también llamados ideales de Schatten, son una herramienta fundamental en Análisis Matemático. Sus aplicaciones recientes en Probabilidad Cuántica y Teoría Cuántica de la Información son muy interesantes, al respecto, consúltense por ejemplo las referencias [3], [4], [5] y [9]. En este trabajo presentamos una construcción elemental de estos espacios en dimensión finita, es decir para matrices complejas $n \times n$, siguiendo la referencia [2]. Incluiremos sólo aquellas demostraciones que son ilustrativas e interesantes, el lector puede consultar la demostración de cada resultado incluido en este trabajo en [1]. Para el caso de dimensión infinita se puede consultar la referencia [6].

Los conceptos de mayorización y matriz estocástica son centrales en este trabajo. Introducimos estos conceptos en la Sección 2 y discutimos algunas de sus consecuencias importantes. En la Sección 3 discutimos los conceptos de función simétrica y calibrada y norma unitariamente invariante, con ayuda de los cuales definimos las p-normas de Schatten. En la Sección 4 presentamos las desigualdades tipo Minkowski de E. Carlen y E. Lieb [3], [4], así como su aplicación para demostrar la subaditividad fuerte de la entropía de von Neumann.

2. Mayorización y matrices biestocásticas.

2.4. Mayorización.

Sea $x = (x_1, ..., x_n) \in \mathbb{R}^n$, denotaremos por $x^\downarrow$ y $x^\uparrow$ a los vectores obtenidos a partir de x reordenando las coordenadas en forma decreciente y en forma creciente respectivamente, es decir:

(i) $x^\downarrow = (x_1^\downarrow, ..., x_n^\downarrow)$ donde $x_1^\downarrow \geq x_2^\downarrow \geq ... \geq x_n^\downarrow$

(ii) $x^\uparrow = (x_1^\uparrow, ..., x_n^\uparrow)$ donde $x_1^\uparrow \leq x_2^\uparrow \leq ... \leq x_n^\uparrow$.

Obsérvese que

$$x_j^\uparrow = x_{n-j+1}^\downarrow \qquad\qquad j = 1, ..., n.$$

Definición 2.1. *Dados x,y* $\in \mathbb{R}^n$ *decimos que x es* ***mayorizado*** *por y, escribiremos* $x \prec y$*, si se cumplen las condiciones:*

$$\sum_{j=1}^{k} x_j^\downarrow \leq \sum_{j=1}^{k} y_j^\downarrow, \text{ para cada } 1 \leq k < n \text{ y} \tag{1}$$

$$\mathrm{Tr}\,(x) := \sum_{j=1}^{n} x_j^\downarrow = \sum_{j=1}^{n} y_j^\downarrow = \mathrm{Tr}\,(y). \tag{2}$$

A la última igualdad le llamaremos condición de la traza.

Ejemplo 2.2. *Si* $x_j \geq 0$ *y* $\sum_{j=1}^{n} x_j = 1$, *entonces*

$$\left(\frac{1}{n}, ..., \frac{1}{n}\right) \prec (x_1, ...x_n) \prec \mathbb{1} = (1, 0, ..., 0).$$

La noción de mayorización ocurre naturalmente en varios contextos de la vida real. Por ejemplo, en Física: supóngase que $\mathrm{Tr}\,(x) = \sum_{j=1}^{n} x_j = 1$ y que un sistema físico se describe mediante una cadena de Markov con espacio de estados $S = \{1, \cdots, n\}$, si $x_i \geq 0$ es la probabilidad de que el sistema se encuentre en el estado i; entonces dadas dos distribuciones de probabilidades $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots y_n)$, tenemos que

$$x \prec y \Longrightarrow H(x) \geq H(y)$$

donde $H(x) = -\sum_{i=1}^{n} x_i \log(x_i)$ es la entropía de Shannon, véase la Proposición 3.9. Esto significa que la distribución y es más desordenada (o caótica) que x.

Otro ejemplo ocurre en Economía, si $x_1, ..., x_n$, $y_1, ..., y_n$ denotan los ingresos de $n-$individuos, entonces $x \prec y$ significa que hay una distribución más equitativa en el estado x que en el estado y. El ejemplo anterior ilustra este hecho.

Denotaremos por $\mathbf{e}$ al vector $(1, 1, ..., 1)$ y para cualquier subconjunto I de $\{1, 2, ..., n\}$, $\mathbf{e}_I$ denotará la función indicadora de I y $|I|$ su cardinalidad. Nótese que podemos escribir $\mathrm{Tr}\,(x) := \sum_{j=i}^{n} \langle x, \mathbf{e} \rangle$, donde $\langle \cdot, \cdot \rangle$ es el producto interior en $\mathbb{R}^n$.

Las siguientes proposiciones se demuestran fácilmente a partir de la Definición 2.1.

Proposición 2.3. *Sea $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, para cada $1 \leq k \leq n$, se tiene que*

$$\sum_{j=1}^{k} x_j^{\downarrow} = \max_{|I|=k} \langle x, \mathbf{e}_I \rangle.$$

Proposición 2.4. *$x \prec y$ si y sólo si para cada subconjunto I de $\{1, 2, ..., n\}$ existe J tal que $|I| = |J|$,*

$$\langle x, \mathbf{e}_I \rangle = \langle y, \mathbf{e}_J \rangle \qquad y \qquad \mathrm{Tr}\,(x) = \mathrm{Tr}\,(y).$$

Definición 2.5. *Diremos que un vector x es* ***submayorizado débilmente*** *por un vector y si $\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}$, $k = 1, 2, ..., n$. Escribiremos $x \prec_{\omega} y$.*

De manera análoga diremos que un vector x es ***supermayorizado débilmente*** *por un vector y si $\sum_{j=1}^{k} x_j^{\uparrow} \geq \sum_{j=1}^{k} y_j^{\uparrow}$, $k = 1, 2, ..., n$. Y escribiremos $x \prec^{\omega} y$.*

No es difícil demostrar las siguientes propiedades.

Proposición 2.6. *(i) $x \prec y$ si y sólo si $x \prec_{\omega} y$, $x \prec^{\omega} y$.*

(ii) *Si α es un número positivo, entonces $x \prec_\omega y \Rightarrow \alpha x \prec_\omega \alpha y$, $x \prec^\omega y \Rightarrow \alpha x \prec^\omega \alpha y$.*

(iii) $x \prec_\omega y \Leftrightarrow -x \prec^\omega -y$.

(iv) $\forall \alpha \in \mathbb{R}$, $x \prec y \Rightarrow \alpha x \prec \alpha y$.

Cada una de las relaciones $\prec, \prec_\omega, \prec^\omega$ es reflexiva y transitiva, pero ninguna define un orden parcial. Por ejemplo, si $x \prec y$ y $y \prec x$, sólo podemos decir que $Py = x$, donde P es una matriz de permutación, es decir, si $P = (p_{ij})_{1 \leq i,j \leq n}$ entonces $p_{ij} = \delta_{\sigma(i)j}$, donde $\sigma \in S_n$. Escribiremos $P \in S_n$ si P es una matriz de permutación. Se demuestra fácilmente que tenemos una relación de equivalencia si definimos $x \sim y$ cuando $x = Py$ para alguna $P \in S_n$. Si denotamos por $\mathbb{R}^n_{sym} = \mathbb{R}^n / \sim$, entonces claramente $\prec$ define un orden parcial en $\mathbb{R}^n_{sym}$. Esta relación también es un orden parcial sobre el conjunto $\{x \in \mathbb{R}^n : x_1 \geq \cdots \geq x_n\}$. La misma afirmación se cumple para las relaciones $\prec_\omega$ y $\prec^\omega$.

Para $a, b \in \mathbb{R}$ escribiremos

$$a \vee b = \text{máx}(a, b), \quad a \wedge b = \text{mín}(a, b).$$

Y para $x, y \in \mathbb{R}^n$ definimos

$$x \vee y = (x_1 \vee y_1, x_2 \vee y_2, ..., x_n \vee y_n), \quad x \wedge y = (x_1 \wedge y_1, x_2 \wedge y_2, ..., x_n \wedge y_n).$$

Además, sean $x^+ = x \vee 0, \quad |x| = x \wedge (-x)$.

Nótese que x^+ es el vector obtenido al reemplazar las coordenadas negativas de x por ceros y $|x|$ es el resultado de tomar el valor absoluto en cada coordenada del vector.

Con las definiciones anteriores tenemos el siguiente resultado, que caracteriza la mayorización sin apelar al reordenamiento de las coordenadas.

Teorema 2.7. *Sean $x, y \in \mathbb{R}^n$, entonces se cumplen*

(i) $x \prec_\omega y$ *si y sólo si* $\forall t \in \mathbb{R}$, $\displaystyle\sum_{j=1}^{n} (x_j - t)^+ \leq \sum_{j=1}^{n} (y_j - t)^+$.

(ii) $x \prec^\omega y$ *si y sólo si para cualquier* $t \in \mathbb{R}$,

$$\sum_{j=1}^{n} (t - x_j)^+ \leq \sum_{j=1}^{n} (t - y_j)^+.$$

(iii) $x \prec y$ *si y sólo si* $\forall t \in \mathbb{R}$, $\displaystyle\sum_{j=1}^{n} |x_j - t| \leq \sum_{j=1}^{n} |y_j - t|$.

2.5. Matrices biestocásticas.

En la presente sección definiremos el concepto de matriz biestocástica y demostraremos algunos resultados que relacionan a estas matrices con la mayorización.

Definición 2.8. *Dada una matriz* $A \in \mathbf{M}_n$*, decimos que es una matriz* ***biestocástica*** *si cumple*

$$a_{ij} \geq 0, \quad \sum_{i=1}^{n} a_{ij} = 1, \quad y \ \sum_{j=1}^{n} a_{ij} = 1, \ \forall \, 1 \leq i, j \leq n.$$

La matriz A se llama **estocástica** si sólo satisface la primera y última condiciones. Las matrices estocásticas aparecen en Probabilidad como funciones de transición de cadenas de Markov.

Definición 2.9. *Diremos que*

(i) $A : \mathbb{C}^n \to \mathbb{C}^n$ *preserva la positividad si la imagen de cada vector con coordenadas no negativas es un vector con coordenadas no negativas.*

(ii) $A \in \mathbf{M}_n$ *preserva la traza si* $\operatorname{Tr}(Ax) = \operatorname{Tr}(x)$ *para cualquier* x*.*

(iii) $A \in \mathbf{M}_n$*, preserva la unidad si* $A\boldsymbol{e} = \boldsymbol{e}$*.*

Proposición 2.10. *Una matriz* $A \in \mathbf{M}_n$*, es biestocástica si y sólo si cumple con las condiciones de la Definición 2.9.*

Demostración. Supongamos que $A \in \mathbf{M}_n$ es matriz biestocástica y sea $x \in \mathbb{C}^n$ tal que $x_j \geq 0$, para toda $j = 1, ..., n$. Entonces es claro que $y_i = \displaystyle\sum_{j=1}^{n} a_{ij} x_j \geq 0$, pues $a_{ij} \geq 0$, para toda $1 \leq i, j \leq n$, por tanto, A preserva positividad. Ahora,

$$\operatorname{Tr}(Ax) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} a_{ij} x_j \Big) = \sum_{j=1}^{n} x_j \Big(\sum_{i=1}^{n} a_{ij} \Big) = \sum_{j=1}^{n} x_j = \operatorname{Tr}(x),$$

pues $\sum_{i=1}^{n} a_{ij} = 1$. Es decir A preserva traza. Finalmente,

$$A\mathbf{e} = A(1, \ldots, 1) = \Big(\sum_{j=1}^{n} a_{1j}, \ldots, \sum_{j=1}^{n} a_{nj} \Big) = (1, 1, ..., 1) = \mathbf{e}.$$

Recíprocamente, si A preserva la positividad, tomando e_j el j-ésimo elemento de la base canónica $1 \leq j \leq n$, obtenemos que $Ae_j = (a_{1j}, a_{2j}, \cdots, a_{nj})$ tiene coordenadas no negativas para cada j. Entonces $a_{ij} \geq 0$ para cada $1 \leq i, j \leq n$.

Si A preserva la traza, para cada vector $x \in \mathbb{C}^n$ se tiene la identidad $\sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right) = \sum_{j=1}^{n} x_j$, así tomando $x = e_j$ tenemos $\left(\sum_{i=1}^{n} a_{ij} - 1 \right) = 0$, para toda $j = 1, 2, \ldots, n$. En consecuencia $\sum_{i=1}^{n} a_{ij} = 1$ para toda $j = 1, \ldots, n$.

Finalmente, si A preserva la unidad tenemos que $\left(\sum_{j=1}^{n} a_{ij} \right)_i = (1, 1, ..., 1)$ de donde se obtiene que $\sum_{j=1}^{n} a_{ij} = 1$, para toda $i = 1, \ldots, n$. Esto demuestra que la matriz es biestocástica. $\square$

Teorema 2.11 (Birkhoff)**.** *Las matrices biestocásticas de tamaño $n \times n$ son un conjunto cerrado bajo la multiplicación de matrices y la toma de adjuntos, además este conjunto es convexo y sus puntos extremos son las matrices de permutación.*

Teorema 2.12. *Una matriz $A \in \mathbf{M}_n$ es biestocástica si y sólo si $Ax \prec x \ \forall x$.*

Proposición 2.13. *Sea A una matriz biestocástica, entonces, todos sus valores propios tienen módulo ≤ 1, 1 es valor propio de A y $\|A\| = 1$, donde $\|A\| = \sup_{\|x\|=1} \|Ax\|$.*

Demostración. Sea A matriz biestocástica, por la Proposición 2.10 preserva la unidad y, por lo tanto, 1 es valor propio de A con vector propio asociado $\mathbf{e} = (1, 1, ..., 1)$. Por el Teorema 2.12 tenemos que $Ax \prec x$, $\forall x$. Sea x vector propio de A con valor propio asociado λ, entonces $\lambda x \prec x$ y por (iii) de la Proposición 2.7, lo anterior se cumple si y sólo si $\forall t \in \mathbb{R}$

$$\sum_{j=1}^{k} |(\lambda x)_j - t| \leq \sum_{j=1}^{k} |x_j - t|, \;\; 1 \leq k \leq n.$$

Tomando $t = 0$ se obtiene que $|\lambda| \sum_{j=1}^{k} |x_j| \leq \sum_{j=1}^{k} |x_j|$, $1 \leq k \leq n$. Consecuentemente $|\lambda| \leq 1$.

Finalmente, $\|A\| = \text{máx}\{|\lambda_j| : \lambda \text{ es valor propio de A}\} = 1$, pues 1 es valor propio de A. □

Proposición 2.14. *Si $A \in \mathbf{M}_n$ es matriz biestocástica, entonces*

$$|Ax| \leq A(|x|),$$

donde $|x| = (|x_1|, ..., |x_n|)$ y escribiremos $x \leq y$ siempre que $x_i \leq y_i$ para toda $i = 1, ..., n$.

Proposición 2.15. *(i) Dados $x, y \in \mathbb{R}^2$ y $0 \leq t \leq 1$ se tiene*

$$(x_1, x_2) = (ty_1 + (1-t)y_2, (1-t)y_1, +ty_2) \quad \textit{si y sólo si } x \prec y.$$

(ii) Sean $x, y \in \mathbb{R}^n$ tales que x se obtiene promediando cualesquiera dos coordenadas de y como en (i), dejando fijas a las demás coordenadas. Entonces $x \prec y$.

Esto motiva la siguiente definición.

Definición 2.16. *Diremos que una aplicación lineal T sobre $\mathbb{R}^n$, es una t-transformación si existen $0 \leq t \leq 1$ e índices i, j tales que:*

$$Ty = (y_1, ..., \underbrace{ty_i + (1-t)y_j}_{i}, ..., \underbrace{(1-t)y_i + ty_j}_{j}, ..., y_n). \tag{3}$$

Proposición 2.17. *Dada una $t-$transformación, entonces $Ty \prec y$ para toda $y \in \mathbb{R}^n$.*

Teorema 2.18. *Para toda $x, y \in \mathbb{R}^n$, las siguientes afirmaciones son equivalentes*

(i) $x \prec y$;

(ii) *x se obtiene a partir de y mediante un número finito de t-transformaciones;*

(iii) *x pertenece a la envolvente convexa de todos los vectores obtenidos al permutar las coordenadas de y;*

(iv) *$x = Ay$ para alguna matriz biestocástica A.*

Proposición 2.19. *(i) Si $U = (u_{ij})$ es una matriz unitaria, entonces la matriz $(|u_{ij}|^2)$ es biestocástica. Le llamaremos matriz uni-estocástica y se llamará ortoestocástica si U es ortogonal real.*

(ii) Si $x = Ay$ para alguna matriz biestocástica A, entonces existe una matriz ortoestocástica B tal que $x = By$

Teorema 2.20 (Lema de Schur)**.** *Sea $A_{n\times n}$ una matriz hermitiana ($A = A^*$); sea* $\operatorname{diag}(A)$ *el vector formado con los elementos en la diagonal de A y sea $\lambda(A)$ el vector cuyas coordenadas son los valores propios de A especificados en cualquier orden. Entonces* $\operatorname{diag}(A) \prec \lambda(A)$.

Demostración. Como A es matriz hermitiana, por el teorema espectral existe una matriz U unitaria tal que $A = U\operatorname{diag}(\lambda(A))U^*$. Es decir,

$$A = \begin{pmatrix} \sum_{j=1}^n u_{1j}\lambda_j\overline{u_{1j}} & \dots & \sum_{j=1}^n u_{1j}\lambda_j\overline{u_{nj}} \\ \vdots & & \vdots \\ \sum_{j=1}^n u_{nj}\lambda_j\overline{u_{1j}} & \dots & \sum_{j=1}^n u_{nj}\lambda_j\overline{u_{nj}} \end{pmatrix}. \tag{4}$$

De esta manera $\operatorname{diag}(A) = \left(\sum_{j=1}^n u_{1j}\lambda_j\overline{u_{1j}}, \dots, \sum_{j=1}^n u_{nj}\lambda_j\overline{u_{nj}}\right)$, entonces podemos escribir

$$a_{ii} = \sum_j |u_{ij}|^2\lambda_j.$$

Esto implica que $\operatorname{diag}(A) = S\lambda(A)$, donde S es la matriz $S = (|u_{ij}|^2)$, que es biestocástica por la parte (i) de la Proposición 2.19. Ahora, una aplicación de la parte (i) del Teorema 2.18 nos permite concluir que $\operatorname{diag}(A) \prec \lambda(A)$. □

Teorema 2.21 (Principio del máximo de Ky Fan)**.** *Si $\lambda(A)$ es el vector de valores propios de una matriz hermitiana A, entonces para* *toda* $k = 1, 2, ..., n$

$$\sum_{j=1}^{k} \lambda_j^{\downarrow} = \text{máx} \sum_{j=1}^{k} \langle x_j, Ax_j \rangle. \tag{5}$$

Donde el máximo es tomado sobre todas las $k-$*uplas de vectores ortonormales* $\{x_1, \cdots, x_k\}$ *en* $\mathbb{C}^n$.

Demostración. Sea $\{x_1, ..., x_k\}$ un conjunto ortonormal en $\mathbb{C}^n$, que podemos completar a una base ortonormal $\{x_1, ..., x_k, ..., x_n\}$ en $\mathbb{C}^n$. Si (a_{ij}) es la matriz de A respecto de esta base, entonces $a_{ij} = \langle x_i, Ax_j \rangle$, $1 \leq i, j \leq n$. Por el Lema de Schur, tenemos que $\sum_{j=1}^{m} \langle x_j, Ax_j \rangle^{\downarrow} \leq \sum_{j=1}^{m} \lambda_j^{\downarrow}$, $1 \leq m \leq n$. En particular, si $\{x_1, ..., x_k\}$ es un subconjunto ortonormal de vectores propios de A, entonces

$$\sum_{j=1}^{k} \langle x_j, Ax_j \rangle = \sum_{j=1}^{k} \langle x_j, \lambda_j x_j \rangle = \sum_{j=1}^{k} \lambda_j^{\downarrow}.$$

Es decir, la suma se satura alcanzando su máximo. Esto demuestra el Teorema. □

Proposición 2.22. *El principio del máximo de Ky Fan y el Lema de Schur son equivalentes.*

Demostración. En la demostración del resultado anterior se vió que el Lema de Schur implica el Principo del máximo de Ky Fan. Ahora, el Principio del máximo de Ky Fan implica que para cada $1 \leq k \leq n$ se tiene

$$\sum_{j=1}^{k} \lambda_j^{\downarrow} = \text{máx} \sum_{j=1}^{k} \langle x_j, Ax_j \rangle \geq \sum_{j=1}^{k} \langle x_j, Ax_j \rangle^{\downarrow} = \sum_{j=1}^{k} a_{ii}^{\downarrow}. \tag{6}$$

Además si $k = n$ el máximo del lado izquierdo se alcanza en la base ortonormal $\{x_1, \cdots, x_n\}$ de los vectores propios de A y la desigualdad anterior se convierte en igualdad. Es decir la condición de la traza también se cumple. Esto demuestra la proposición. □

Proposición 2.23. *Sean* A, B *matrices hermitianas, entonces para toda* $k = 1, ..., n$

$$\sum_{j=1}^{k} \lambda_j^{\downarrow}(A + B) \leq \sum_{j=1}^{k} \lambda_j^{\downarrow}(A) + \sum_{j=1}^{k} \lambda_j^{\downarrow}(B). \tag{7}$$

Demostración. Usando el Teorema 2.21

$$\sum_{j=1}^{k}\lambda_j^{\downarrow}(A+B)=\text{máx}\sum_{j=1}^{k}\langle x_j,(A+B)x_j\rangle\le\text{máx}\sum_{j=1}^{k}\langle x_j,Ax_j\rangle \tag{8}$$

$$+\,\text{máx}\sum_{j=1}^{k}\langle x_j,Bx_j\rangle=\sum_{j=1}^{k}\lambda_j^{\downarrow}(A)+\sum_{j=1}^{k}\lambda_j^{\downarrow}(B), \tag{9}$$

donde $\{x_1,\cdots,x_k\}$ es un subconjunto ortornormal en $\mathbb{C}^n$. □

Dada una matriz A de tamaño $n\times n$, decimos que B es la raíz cuadrada de la matriz A si $B\cdot B=A$ y el valor absoluto $|A|$ se define mediante la relación $|A|=\sqrt{A^*A}$. A los valores propios de la matriz $|A|$ tomando en cuenta sus multiplicidades se les llaman valores singulares de A.

Proposición 2.24.

(a) Para cualquier matriz A, sea $\widetilde{A}$ la matriz:

$$\widetilde{A}=\begin{pmatrix}0 & A\\ A^* & 0\end{pmatrix}.$$

Entonces $\widetilde{A}$ es hermitiana y sus valores propios son los valores singulares de A junto con sus negativos.

(b) Sean $s_1(A),\dots,s_n(A)$ los valores singulares de A ordenados en forma decreciente, entonces para cualesquiera matrices A,B y todo $k=1,2,\dots,n$ se cumple

$$\sum_{j=1}^{k}s_j(A+B)\le\sum_{j=1}^{k}s_j(A)+\sum_{j=1}^{k}s_j(B).$$

Demostración. Es fácil ver que $\widetilde{A}$ es hermitiana. Sea λ un valor propio de $\widetilde{A}$, con vector propio (X_1,X_2). Entonces tenemos que $AX_2=\lambda X_1$ y $A^*X_1=\lambda X_2$, como $\widetilde{A}$ es hermitiana $\lambda\in\mathbb{R}$, consecuentemente $|A|^2X_2=A^*AX_2=\lambda A^*X_1=\lambda^2X_2$. Por lo tanto $|\lambda|=\sqrt{\lambda^2}$ es un valor singular de A. Reciprocamente, si $W^*AQ=S$ es la descomposición de valores singulares de A, es decir, $S=\text{diag}(s_1(A),...,s_n(A))$, W es la matriz unitaria cuyas columnas W_j son los vectores propios de

A^*A y Q es la matriz unitaria cuyas columnas Q_j son los vectores propios de AA^* correspondientes a los valores propios $s_j^2(A)$, $1 \leq j \leq n$. Entonces para cada j tenemos que

$$AQ_j = s_j(A)W_j, \quad \text{y} \quad A^*W_j = s_j(A)Q_j,$$

es decir, $s_j(A)$ y $-s_j(A)$ son valores propios de $\tilde{A}$ con vectores propios asociados (W_j, Q_j) y $(-W_j, Q_j)$, respectivamente. Esto demuestra (a).

Sean $\widetilde{A}$ y $\widetilde{B}$ como en (a). Por la Proposición 2.23 tenemos que

$$\sum_{j=1}^{k} \lambda_j^{\downarrow}(\widetilde{A} + \widetilde{B}) \leq \sum_{j=1}^{k} \lambda_j^{\downarrow}(\widetilde{A}) + \sum_{j=1}^{k} \lambda_j^{\downarrow}(\widetilde{B}).$$

Pero por el inciso anterior, para $1 \leq k \leq n$ podemos escribir

$$\sum_{j=1}^{k} s_j(A + B) \leq \sum_{j=1}^{k} s_j(A) + \sum_{j=1}^{k} s_j(B),$$

pues los negativos de los valores propios de $|A|$ aparecen después de la posición n. □

Si en la proposición anterior, tomamos $k = 1$ obtenemos que $s_1(A+B) \leq s_1(A) + s_1(B)$, que es la desigualdad del triángulo para la norma de operadores, i.e., $||A + B|| \leq ||A|| + ||B||$.

3. Normas simétricas sobre $\mathbb{C}^n$

3.6. Normas en $\mathbb{C}^n$

Para cada $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$ y cada $1 \leq p \leq \infty$

$$||x||_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}$$
$$||x||_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Es bien conocido que estas relaciones definen normas en $\mathbb{C}^n$.

Las siguientes propiedades se demuestran fácilmente.

Proposición 3.1. *Para toda $x \in \mathbb{C}^n$ y toda $1 \leq p \leq \infty$ se cumplen* *(i) Propiedad del módulo o calibración.* $\|x\|_p = \| \, |x| \, \|_p$ *, donde el vector* $|x|$ *se define mediante* $|x| = (|x_1|, \dots, |x_n|)$.

(ii) Propiedad de monotonía. $\|x\|_p \leq \|y\|_p$ *siempre que* $|x| \leq |y|$ *donde ésta última desigualdad significa que* $|x_j| \leq |y_j|$ *para toda* $1 \leq j \leq n$.

(iii) Propiedad de simetría. $\|x\|_p = \|Px\|_p \; \forall P \in S_n$.

Observación 3.2. *Supongamos que el espacio es de dimensión infinita y está dotado de la p-norma, es decir se trata del espacio* ℓ_p. *Las primeras dos condiciones en la proposición 3.1 se cumplen claramente pues las sucesiones en* ℓ_p *son absolutamente convergentes. Y por el teorema de reordenamiento de series la última condición en la Proposición 3.1 también se cumple. Entonces la proposición también es válida en los espacios* ℓ_p.

Proposición 3.3. *Una norma en* $\mathbb{C}^n$ *es calibrada si y sólo si es monótona.*

3.7. Normas simétricas y calibradas

Por la propiedad de calibración vista en la sección anterior, las normas simétricas y calibradas en $\mathbb{C}^n$ están determinadas por aquellas en $\mathbb{R}^n$.

Definición 3.4. *Una función* $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ *es simétrica y calibrada si*

(i) Φ *es una norma.*

(ii) $\Phi(Px) = \Phi(x)$ *para toda* $x \in \mathbb{R}^n$ *y toda* $P \in S_n$.

(iii) $\Phi(\varepsilon_1 x_1, \dots, \varepsilon_n x_n) = \Phi(x_1, \dots, x_n)$, *donde* $\varepsilon_j = \pm 1$, $j = 1, 2, ..., n$.

(iv) $\Phi(\mathbf{1}) = \Phi(1, 0, \dots, 0) = 1$, *decimos que* Φ *esta normalizada.*

Por la condición *(iii)* una función simétrica y calibrada está determinada por sus valores en $\mathbb{R}^n_+$

Definición 3.5. *Sea* $x \in \mathbb{R}^n$ *y supóngase que sus coordenadas son ordenadas de tal forma que* $|x_1| \geq |x_2| \geq \cdots \geq |x_n|$. *Para* $k = 1, 2, ..., n$ *la norma de Ky Fan se define como*

$$\Phi_{(k)}(x) = \sum_{j=1}^{k} |x_j|. \tag{10}$$

Por las propiedades del módulo, la función $\Phi_{(k)}$ es claramente una función simétrica y calibrada. Además $\Phi_{(1)}(x) = |x_1| = \|x\|_\infty$ y $\Phi_{(n)}(x) = \|x\|_1$. Usaremos también $\|\cdot\|_{(k)}$ para referirnos a la k-ésima norma de Ky Fan, con paréntesis en el subíndice para no confundirla con la p-norma.

Proposición 3.6.

(i) Toda función simétrica y calibrada es continua.

(ii) Sea Φ una función simétrica y calibrada. Entonces para toda $x \in \mathbb{R}^n$, $\|x\|_\infty \leq \Phi(x) \leq \|x\|_1$.

(iii) Sea Φ función simétrica y calibrada y sean $0 \leq t_j \leq 1$, entonces

$$\Phi(t_1x_1, ..., t_nx_n) \leq \Phi(x).$$

(iv) Toda función Φ simétrica y calibrada es monótona en $\mathbb{R}^n_+$.

(v) Sean $x, y \in \mathbb{R}^n$ tales que $|x| \leq |y|$, entonces $\Phi(x) \leq \Phi(y)$.

Consideraremos funciones $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, cuyo dominio es todo $\mathbb{R}^n$ o bien un subconjunto de este espacio que es convexo e invariante bajo la permutación de sus coordenadas. Dados dos vectores en $\mathbb{R}^n$, $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$, escribiremos $x \leq y$ si $x_j \leq y_j,\ 1 \leq j \leq n$. Retomando el concepto mayorización presentado en la sección anterior introducimos ahora la siguiente.

Definición 3.7. *Diremos que una función $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ es*

(i) ***isotona*** *si*

$$x \prec y \Rightarrow \Phi(x) \prec_\omega \Phi(y) \tag{11}$$

(ii) ***fuertemente isotona*** *si*

$$x \prec_\omega y \Rightarrow \Phi(x) \prec_\omega \Phi(y); \tag{12}$$

(iii) ***monótona creciente*** *si $x \leq y$ implica $\Phi(x) \leq \Phi(y)$;*

(iv) ***convexa*** *si*

$$\Phi(tx + (1-t)y) \leq t\Phi(x) + (1-t)\Phi(y),\ \ 0 \leq t \leq 1,\ \ x, y \in \mathbb{R}^n.$$

Si $m = 1$ la submayorización $\prec_\omega$ en el lado derecho de (11) se convierte en desigualdad. En este caso las funciones isotonas también se llaman ***Schur-convexas*** *o bien* ***S-convexas****.*

Teorema 3.8. *Sea $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ una aplicación convexa. Supóngase que para toda $P \in S_n$ existe $P' \in S_m$ tal que*

$$\Phi(Px) = P'\Phi(x) \qquad \forall x \in \mathbb{R}^n. \tag{13}$$

Entonces Φ es isotona. Si además Φ es monótona creciente entonces es fuertemente isotona.

Proposición 3.9. *La función $-H$, donde H es la entropía de Shannon, es una función isotona.*

Demostración. Para $x \geq 0$ sea $\varphi(x) = x\log(x)$, con la convención de que $0\log(0) = 0$. Como φ es una función convexa se tiene que $-H(x) = \sum_{j=1}^{n} \varphi(x_j)$ es convexa. Además es claro que $-H(x)$ es invariante bajo permutaciones de las coordenadas de x.

Supóngase que $x \prec y$, entonces por el inciso (iii) del Teorema 2.18, para $1 \leq j \leq n$, existen $P_j \in S_n$ y $0 \leq t_j \leq 1$ con $\sum_{j=1}^{n} t_j = 1$ tales que $x = \sum_{j=1}^{n} t_j P_j y$. Por la convexidad de $-H$ y su invarianza bajo permutaciones se obtiene que

$$-H(x) \leq -\sum_j t_j H(P_j y) = -H(y).$$

□

Proposición 3.10. *Sean $x, y \in \mathbb{R}^n_+$. Si $x \prec_\omega y$ entonces $\Phi(x) \leq \Phi(y)$ para toda función Φ simétrica y calibrada.*

Demostración. Por ser norma Φ es convexa, por la simetría se cumple la identidad (13) para toda $P \in S_n$, con $P' = Id$. Entonces por el Teorema 3.8, Φ es fuertemente isotona, es decir $\Phi(x) \leq \Phi(y)$. □

Proposición 3.11. *Sea $\Phi_{(k)}(\cdot)$ la k-ésima norma de Ky Fan, para $k = 1, 2, ..., n$. Si $\Phi_{(k)}(x) \leq \Phi_{(k)}(y)$ para toda $k = 1, 2, ..., n$ y para toda $x, y \in \mathbb{R}^n$, entonces $\Phi(x) \leq \Phi(y)$ para toda función Φ simétrica y calibrada y para toda $x, y \in \mathbb{R}^n$.*

Demostración. Supóngase que $\Phi_{(k)}(x) \leq \Phi_{(k)}(y)$ para toda $k = 1, 2, \dots, n$ por la Definición 3.5 tenemos que $|x| \prec_\omega |y|$, entonces de la Proposición 3.10 se tiene que para toda Φ función simétrica y calibrada $\Phi(|x|) \leq \Phi(|y|)$. En consecuencia $\Phi(x) \leq \Phi(y)$ por la propiedad de calibración. $\square$

Proposición 3.12. *(Desigualdad de Young) Sean $p, q, r \in \mathbb{R}_+$ tales que $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Entonces para cada función simétrica y calibrada Φ y cada x, $y \in \mathbb{R}^n$ se cumple*

$$[\Phi(|x \cdot y|^r)]^{1/r} \leq [\Phi(|x|^p)]^{1/p} [\Phi(|y|^q)]^{1/q}. \tag{14}$$

La desigualdad que se obtiene tomando $r = 1$ en (14) es la desigualdad de Hölder para funciones simétricas y calibradas. Si tomamos $r = 1$ y $p = 2$ en (14) obtenemos la desigualdad de Cauchy-Schwarz para funciones simétricas y calibradas.

Si $\Phi = \Phi_{(n)}$ y $r = 1$ la desigualdad (14) se reduce a la bien conocida desigualdad de Hölder para las p-normas en $\mathbb{C}^n$.

Teorema 3.13 (Desigualdad de Minkowski)**.** *Sea Φ una función simétrica y calibrada, sea $p \geq 1$, entonces para toda $x, y \in \mathbb{R}^n$ se cumple*

$$[\Phi(|x + y|^p)]^{1/p} \leq [\Phi(|x|^p)]^{1/p} + [\Phi(|y|^p)]^{1/p}. \tag{15}$$

Nótese que si $\Phi = \Phi_{(n)}$ tenemos

$$\left(\sum_{j=1}^{n} |x_j + y_j|^p\right)^{1/p} \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} + \left(\sum_{j=1}^{n} |y_j|^p\right)^{1/p} \tag{16}$$

que es la desigualdad de Minkowski para la p-norma.

3.8. Normas en $\mathbb{C}^n$ unitariamente invariantes

En esta sección $\mathbb{C}^n$ denotará el espacio de Hilbert $\mathbb{C}^n$ dotado del producto interno $\langle \cdot, \cdot \rangle$ y con la norma asociada $\| \cdot \|$. Si $A \in M_n$, $\|A\|$ denotará su norma como operador, definida como

$$\|A\| = \sup_{\|x\|=1} \|Ax\|_{\mathbb{C}^n}. \tag{17}$$

Si $A \in M_n$, su módulo $|A|$ se define por $|A| = (A^*A)^{1/2}$, donde $(A^*A)^{1/2}$ es la raíz cuadrada positiva del operador positivo A^*A.

Denotaremos por $s(A) = (s_1(A), s_2(A), \dots, s_n(A))$ al vector de valores singulares de A ordenados de manera que $s_1(A) \geq s_2(A) \geq \dots \geq s_n(A)$.

Proposición 3.14. *Sea A operador lineal en $\mathbb{C}^n$, entonces*

$$\|A\| = \|\,|A|\,\| = s_1(A). \tag{18}$$

Proposición 3.15. *Si U, V son operadores unitarios sobre $\mathbb{C}^n$ entonces*

$$|U\,A\,V| = V^*|A|V \tag{19}$$

y además

$$\|A\| = \|U\,A\,V\|. \tag{20}$$

Demostración. Sean U *y* V operadores unitarios, entonces

$$\begin{aligned}|U\,A\,V|^2 &= (U\,A\,V)^*(U\,A\,V) = (V^*\,A^*\,U^*)(U\,A\,V)\\ &= V^*\,A^*\,Id\,A\,V = V^*\,|A|^2\,V = V^*\,|A|\,VV^*\,|A|\,V\\ &= (V^*\,|A|\,V)^2.\end{aligned}$$

Así, extrayendo la raíz cuadrada tenemos $|U\,A\,V| = V^*|A|\,V$. Ahora usaremos esta relación para demostrar la segunda identidad de la proposición. Sea $x \in \mathbb{C}^n$

$$\begin{aligned}\|U\,A\,Vx\|^2 = \|V^*|A|\,Vx\|^2 &= \langle V^*|A|\,Vx, V^*|A|\,Vx\rangle\\ &= \langle V\,V^*|A|\,Vx, |A|\,Vx\rangle = \langle |A|\,Vx, |A|Vx\rangle.\end{aligned}$$

Como V es unitario se tiene que $\|Vx\| = \|x\|$; además $\|\,|A|\,\| = \|A\|$, por lo tanto $\|A\| = \|U\,A\,V\|$. □

Una norma en los operadores sobre $\mathbb{C}^n$ que satisface la propiedad (20) para todo U, V operadores unitarios, se llama unitariamente invariante. Denotaremos a tales normas mediante $|||\cdot|||$ y las normalizaremos de manera que $|||\mathrm{diag}(1, 0, \cdots, 0)||| = 1$.

El siguiente teorema relaciona las normas unitariamente invariantes con las funciones simétricas y calibradas, mediante los valores singulares.

Teorema 3.16. *(i) Sea Φ una función simétrica y calibrada sobre $\mathbb{R}^n$. Definimos la función $|||\cdot|||_\Phi$ sobre $\mathbf{M}_{n\times n}$ como sigue*

$$|||A|||_\Phi = \Phi(s(A)). \tag{21}$$

Entonces $|||\cdot|||_\Phi$ *es una norma.*

(ii) Sea $|||\cdot|||$ *una norma unitariamente invariante. Definimos la función* $\Phi_{|||\cdot|||}$ *sobre* $\mathbb{R}^n$ *mediante*

$$\Phi_{|||\cdot|||}(x) = |||\mathrm{diag}\,(x)|||. \tag{22}$$

Donde $\mathrm{diag}\,(x)$ *es la matriz diagonal con coordenadas* $x_1, x_2, \ldots, x_n$. *Entonces* $\Phi_{|||\cdot|||}$ *es una función simétrica y calibrada.*

Las funciones simétricas y calibradas conducen a varios ejemplos de normas unitariamente invariantes en el espacio de las matrices complejas $n \times n$. Dos clases de tales normas son:

(1) Las p-normas de Schatten, para $1 \le p \le \infty$,

$$\|A\|_p = \Phi_p(s(A)) = \left[\sum_{j=1}^{n} (s_j(A))^p\right]^{1/p} \tag{23}$$

$$\|A\|_\infty = \Phi_\infty(s(A)) = s_1(A) = \|A\|. \tag{24}$$

(2) Las k-normas de Ky Fan

$$\|A\|_{(k)} = \sum_{j=1}^{k} s_j(A) = \Phi_{(k)}(s(A)) \qquad \text{para } 1 \le k \le n. \tag{25}$$

Hemos visto que $\|A\|_{(1)} = \|A\|_\infty = \|A\|$ y $\|A\|_{(n)} = \|A\|_1 = \|\,|A|\,\|_1 = \mathrm{Tr}\,(|A|)$. En la base que diagonaliza a $|A|$ tenemos que

$$(\mathrm{Tr}|A|^p)^{\frac{1}{p}} = \Big(\sum_{j=1}^{n} s_j(A)^p\Big)^{\frac{1}{p}} = \|A\|_p. \tag{26}$$

La traza es un funcional sobre el espacio de matrices complejas $n \times n$ con propiedades similares a las de la integral. Debido a esta analogía y en vista de la identidad (26), al espacio $\mathbf{M}_{n\times n}(\mathbb{C})$ provisto con la norma $\|\cdot\|_p$ le llamaremos **espacio L_p no conmutativo** y lo denotaremos por $L_p(\mathbb{C}^n)$.

Igual que en el caso conmutativo el espacio $L_2(\mathbb{C}^n)$ tiene un estructura de espacio de Hilbert, con el producto interno de Hilbert-Schmidt que se define mediante

$$\langle A, B\rangle_{HS} = \mathrm{Tr}\big(A^*B\big).$$

La norma inducida por este producto interno es la norma Hilbert-Schmidt,

$$\|A\|_{HS} = \mathrm{Tr}\big(A^*A\big)^{\frac{1}{2}} = \mathrm{Tr}(|A|^2)^{\frac{1}{2}} = \Big(\sum_j (s_j(A))^2\Big)^{\frac{1}{2}} = \|A\|_2,$$

que coincide con la $2-$norma de Schatten. A esta norma también se le llama norma de Frobenius. Si $A = (a_{ij})_{1\leq i,j\leq n}$ entonces después de un cálculo simple se ve que $\|A\|_2 = \Big(\sum_{ij} |a_{ij}|^2\Big)^{1/2}$. Es decir, la norma de Hilbert-Schmidt es la norma euclidiana de la matriz A pensada como vector en $\mathbb{C}^{n^2}$.

Proposición 3.17. *Si $1 \leq p < \infty$ y $\frac{1}{p} + \frac{1}{q} = 1$, entonces se cumple la desigualdad de Hölder para las p-normas de Schatten.*

Demostración. Se sigue de (14) con $r = 1$, $\Phi = \Phi_p$ y la identidad (23). □

El siguiente Teorema muestra la relevancia de las normas de Ky Fan.

Teorema 3.18 (Dominación de Ky Fan)**.** *Sean $A, B \in \mathbf{M}_{n\times n}$; si $\|A\|_{(k)} \leq \|B\|_{(k)}$ para $k = 1, 2, \ldots, n$. Entonces*

$$|||A||| \leq |||B|||,$$

para todas las normas unitariamente invariantes.

Demostración. Sea $|||\cdot|||$ una norma unitariamente invariante y $\Phi := \Phi_{|||\cdot|||}$ la función simétrica y calibrada asociada. Por la descomposición de valores singulares, para cualquier matriz A se tiene que $\mathrm{diag}(s(A)) = W^*AQ$ con W y Q matrices unitarias, consecuentemente $|||A||| = |||W^*AQ||| = |||\mathrm{diag}(s(A))|||$. Por otra parte, del Teorema 3.16 obtenemos que $|||A|||_\Phi = \Phi_{|||\cdot|||}(s(A)) = |||\mathrm{diag}(s(A))|||$, entonces

$$|||A||| = |||A|||_\Phi = \Phi_{|||\cdot|||}(s(A)).$$

Sean A, B matrices de tamaño $n \times n$. Si $\|A\|_{(k)} \leq \|B\|_{(k)}$ para $k = 1, 2, \ldots, n$; por la ecuación (25) tenemos que

$$\sum_{j=1}^{k} s_j(A) \leq \sum_{j=1}^{k} s_j(B) \qquad \text{para toda } 1 \leq k \leq n.$$

Entonces $s(A) \prec_\omega s(B)$. Por la Proposición 3.10 se tiene que $|||A||| = \Phi_{|||\cdot|||}(s(A)) \leq \Phi_{|||\cdot|||}(s(B)) = |||B|||$.

$\square$

Recordemos que de la Proposición 3.6 se cumple que $\Phi_{(1)}(x) \leq \Phi(x) \leq \Phi_{(n)}(x)$ para toda $x \in \mathbb{C}^n$ y toda función simétrica y calibrada Φ. Entonces por el Teorema 3.16 tenemos que $\|A\|_\infty = \|A\| \leq |||A||| \leq \|A\|_{(n)} = \|A\|_1$, para toda $A \in \mathbf{M}_{n\times n}$ y toda norma unitariamente invariante $|||\cdot|||$.

4. Desigualdades tipo Minkowski.

La noción de traza parcial es central en esta parte, para una discusión más amplia de este operador el lector puede consultar, por ejemplo, las referencias [7], [8].

Definición 4.1. *Sean $\mathcal{H}_1$ y $\mathcal{H}_2$ espacios de Hilbert de dimensión finita. Sea $\{e_i\}_i$ una base ortonormal de $\mathcal{H}_2$. Sea ρ un operador sobre $\mathcal{H}_1 \otimes \mathcal{H}_2$. La traza parcial de ρ es el operador $\mathrm{Tr}_2\rho \in \mathcal{B}(\mathcal{H}_1)$ definido para todos u, $v \in \mathcal{H}_1$ mediante*

$$\langle u, \mathrm{Tr}_2 \rho v\rangle_{\mathcal{H}_1} = \sum_i \langle u \otimes e_i, \rho(v \otimes e_i)\rangle_{\mathcal{H}_1\otimes\mathcal{H}_2}. \tag{27}$$

Se puede demostrar que el operador $\mathrm{Tr}_2\rho$ no depende de la base ortonormal utilizada. Además, la traza parcial es un operador $\mathrm{Tr}_2 : \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1)$, que preserva la traza y la positividad. De hecho, es un operador completamente positivo, veáse por ejemplo, [7] y [8].

Denotaremos mediante $\rho_1 = \mathrm{Tr}_2\rho$, a la traza parcial del operador $\rho \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ y de manera análoga $\rho_2 = \mathrm{Tr}_1\rho$.

La siguientes desigualdades tipo Minkowski se deben a E. A. Carlen y E. Lieb, [3], [4]. Para las condiciones de igualdad consúltese [9]. Demostraremos sólo la primera desigualdad, la demostración de la segunda es más difícil.

Teorema 4.2 (Desigualdad Tipo Minkowski I)**.** *Sea $A \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ positivo. Entonces, para toda $p \geq 1$*

$$\left(\mathrm{Tr}_2(\mathrm{Tr}_1 A)^p\right)^{1/p} \leq \mathrm{Tr}_1(\mathrm{Tr}_2 A^p)^{1/p}. \tag{28}$$

Si $0 < p \leq 1$, la desigualdad se invierte.

Demostración. Por la desigualdad de Hölder, Proposición 3.17, para $\frac{1}{p}+\frac{1}{q}=1$ tenemos que

$$\operatorname{Tr}(B\operatorname{Tr}_1 A) \leq (\operatorname{Tr} B^q)^{\frac{1}{q}} \operatorname{Tr}((\operatorname{Tr}_1 A)^p)^{\frac{1}{p}} \tag{29}$$

para $B \in \mathcal{B}(\mathcal{H}_2)$ positivo. Tomando $B = \sum_j b_j |\beta_j\rangle\langle\beta_j|$, donde $b_j = \alpha\lambda_j$ con $\alpha = \left(\sum_j \lambda_j^q\right)^{\frac{1}{q}}$ y $\{\lambda_j\}$, $\{\beta_j\}$ los valores propios y la base ortonormal que diagonaliza a $\operatorname{Tr}_1 A$, respectivamente, es fácil ver que se cumple la igualdad en (29) y que $\operatorname{Tr} B^q = 1$. Entonces tenemos

$$\begin{aligned}
\operatorname{Tr}\left((\operatorname{Tr}_1 A)^p\right)^{\frac{1}{p}} &= \operatorname{Tr}\left(B\operatorname{Tr}_1(A)\right) = \operatorname{Tr}\left((I\otimes B)A\right) \\
&= \sum_{i,j} \langle e_i \otimes \beta_j, (I\otimes B)A(e_i\otimes\beta_j)\rangle \\
&= \sum_{i,j} \langle e_i \otimes B\beta_j, A(e_i\otimes\beta_j)\rangle,
\end{aligned}$$

donde $\{e_i\}$ es base ortonormal de $\mathcal{H}_1$.

Podemos estimar el lado derecho como sigue

$$\begin{aligned}
\sum_{i,j} \langle e_i \otimes B\beta_j, A(e_i\otimes\beta_j)\rangle &\leq \sum_i \left(\sum_j \lambda_j^q\right)^{\frac{1}{q}} \left(\sum_j (\langle e_i\otimes\beta_j, A(e_i\otimes\beta_j)\rangle)^p\right)^{\frac{1}{p}} \\
&= \sum_i \left(\sum_j (\langle e_i\otimes\beta_j, A\,(e_i\otimes\beta_j)\rangle)^p\right)^{\frac{1}{p}}.
\end{aligned}$$

Una aplicación del Teorema espectral en dimensión finita permite demostrar que

$$\langle e_i\otimes\beta_j, A(e_i\otimes\beta_j)\rangle \leq \Big(\langle e_i\otimes\beta_j, A^p(e_i\otimes\beta_j)\rangle\Big)^{\frac{1}{p}},$$

de donde se obtiene

$$\begin{aligned}
(\operatorname{Tr}(\operatorname{Tr}_1 A^p))^{\frac{1}{p}} &\leq \sum_i \left(\sum_j \Big(\langle e_i\otimes\beta_j, A^p(e_i\otimes\beta_j)\rangle\Big)^{\frac{1}{p}}\right) \\
&= \sum_i \Big(\langle e_i, \operatorname{Tr}_2 A^p e_i\rangle\Big)^{\frac{1}{p}}.
\end{aligned}$$

Si se elige la base $\{e_j\}_j$ como los valores propios de $\mathrm{Tr}_2 A^p$. Entonces el lado derecho de la desigualdad anterior es $\mathrm{Tr}\,(\mathrm{Tr}_2 A^p)^{\frac{1}{p}}$. Con esto se concluye la demostración. □

Teorema 4.3 (Desigualdad Tipo Minkowski II)**.** *Para $1 \leq q \leq p \leq 2$ y todo A operador positivo sobre $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$,*

$$\mathrm{Tr}_3 \left(\mathrm{Tr}_2 \left((\mathrm{Tr}_1 A^q)^{\frac{p}{q}} \right) \right)^{\frac{q}{p}} \leq \mathrm{Tr}_3 \left(\mathrm{Tr}_1 [(\mathrm{Tr}_2 A^p)^{\frac{q}{p}}] \right). \tag{30}$$

Para $0 \leq p \leq 1$, y cualquier $q \geq p$ la desigualdad se invierte.

4.9. Subaditividad Fuerte de la entropía de von Neumann

Consideremos una función $f(p) = \rho^p$, ρ un estado, i.e., un operador positivo de traza unitaria. Por cálculo funcional podemos escribir $\rho^p = \sum_j \lambda_j^p |e_j\rangle\langle e_j|$, entonces

$$\begin{aligned} \frac{d}{dp}(Tr\rho^p)\Big|_{p=1} &= \frac{d}{dp}\Big[Tr\Big(\sum_j \lambda_j^p |e_j\rangle\langle e_j|\Big)\Big]_{p=1} \\ &= \frac{d}{dp}\Big[\sum_j \lambda_j^p Tr\big(|e_j\rangle\langle e_j|\big)\Big]_{p=1} \\ &= \frac{d}{dp}\Big[\sum_j \lambda_j^p\Big]_{p=1} = \sum_j \lambda_j \log \lambda_j = -S(\rho). \end{aligned}$$

La serie de McLaurin de $f(x) = a^{x+1}$ alrededor de 0 tiene la forma $f(x) = a + xa \ln a + a(\ln a)^2 \frac{x^2}{2!} + \dots$. Entonces para una matriz definida positiva A y $\varepsilon > 0$ suficientemente pequeño, tenemos

$$A^{1\pm\varepsilon} = A \pm \varepsilon\, A \log\, A \pm \varepsilon^2 \frac{A(\log\, A)^2}{2!} \pm \dots \tag{31}$$

Teorema 4.4. *(Subaditividad fuerte de la entropía de von Neumann) Sea ρ_{123} un estado sobre el espacio de Hilbert $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. Entonces*

$$S(\rho_{1,3}) + S(\rho_{2,3}) \geq S(\rho_{1,2,3}) + S(\rho_3),$$

donde S es la entropía de von Neumann.

Demostración. Sea $\rho = \rho_{1,2,3} \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3)$ un estado. Usaremos las siguientes notaciones: $\rho_{2,3} = \mathrm{Tr}_1 \rho$, $\rho_3 = \mathrm{Tr}_1 \mathrm{Tr}_2 \rho$, etc. Aplicamos traza parcial sobre el segundo factor en (31),

$$Tr_2(\rho^{1+\varepsilon}) = Tr_2\rho + \varepsilon\, Tr_2 \rho \log\, \rho + \mathcal{O}(\varepsilon^2) = \rho_{1,3} + \varepsilon\, Tr_2 \rho \log\, \rho + \mathcal{O}(\varepsilon^2).$$

Por otro lado, sabemos que $\frac{1}{1+\varepsilon} = \sum_{n\geq 0}(-\varepsilon)^n = 1 - \varepsilon + \mathcal{O}(\varepsilon^2)$. Así

$$
\begin{aligned}
&\Big[Tr_2(\rho^{1+\varepsilon})\Big]^{\frac{1}{1+\varepsilon}} = \Big(\rho_{1,3} + \varepsilon\, Tr_2\rho \log\, \rho + \mathcal{O}(\varepsilon^2)\Big)^{\frac{1}{1+\epsilon}} \\
=&\rho_{1,3} + \varepsilon\, Tr_2\rho \log\, \rho - \varepsilon\big(\rho_{1,3} + \varepsilon\, Tr_2\rho \log\, \rho\big) \log(\rho_{1,3} + \varepsilon\, Tr_2\rho \log\, \rho\big) \\
&\quad + \mathcal{O}(\varepsilon^2) \\
=&\rho_{1,3} + \varepsilon\, Tr_2\rho \log\, \rho - \epsilon\rho_{1,3} \log \rho_{1,3} + \mathcal{O}(\varepsilon^2).
\end{aligned}
$$

Y tomando Tr_1 y Tr_3 obtenemos que

$$
Tr_3\, Tr_1 \Big[Tr_2(\rho^{1+\varepsilon})\Big]^{\frac{1}{1+\varepsilon}} = 1 - \varepsilon S(\rho) + \varepsilon S(\rho_{1,3}) + \mathcal{O}(\varepsilon^2). \tag{32}
$$

De manera análoga se obtiene que

$$
Tr_3 \Big[Tr_2\big(\rho_{2,3}^{1+\varepsilon}\big)\Big]^{\frac{1}{1+\varepsilon}} = 1 - \varepsilon S(\rho_{2,3}) + \varepsilon S(\rho_3) + \mathcal{O}(\varepsilon^2) \tag{33}
$$

Ahora, aplicando el Teorema 4.2, con $A = \rho$, $p = 1 + \varepsilon$, y usando que la traza parcial es una transformación completamente positiva, de (32) y (33) obtenemos,

$$
1 - \varepsilon S(\rho) + \varepsilon S(\rho_{1,3}) + \mathcal{O}(\varepsilon^2) \leq 1 - \varepsilon S(\rho_{2,3}) + \varepsilon S(\rho_3) + \mathcal{O}(\varepsilon^2)
$$

Equivalentemente

$$
1 - S(\rho) + S(\rho_{1,3}) + \frac{\mathcal{O}(\varepsilon^2)}{\varepsilon} \leq 1 - S(\rho_{2,3}) + S(\rho_3) + \frac{\mathcal{O}(\varepsilon^2)}{\varepsilon}.
$$

Haciendo $\varepsilon \to 0$, obtenemos la subaditividad fuerte para la entropía de von Neumann. □

5. Agradecimiento

Agradecemos al árbitro por varias observaciones y sugerencias que nos permitieron mejorar este artículo.

Referencias

[1] Barrientos Sánchez, G., *Espacios L_p no conmutativos y aplicaciones*, Tesis de Maestría, Posgrado en Matemáticas, UAM-I, México, 2010.

[2] Bhatia, R., *Matrix Analysis*, Springer-Verlag, New York U.S.A., 1997.

[3] Carlen, E.A. and Lieb, E., *A Minkowski type trace inequality and strong subadditivity of quantum entropy*, Advances in the Mathematical Science, AMS Transl., **189**, Series 2, 1999, 59-68.

[4] Carlen, E.A. and Lieb, E., *A Minkowski type trace inequality and strong subadditivity of quantum entropy II: convexity and concavity*, Lett. math. Phys., **83**, 2008, 107-126.

[5] Gibilisco, P., Imparato, D. and Isola, T., *Schrödinger Equation, L^p-Duality and the Geometry of Wigner-Yanase-Dyson Information*, in Quantum Probability and Related Topics, Proceedings of the 28th Conference, 2008, p. 157-164.

[6] McCarthy, C. A., $\mathcal{C}_p$, Israel J. Math., **5**, 1967, p. 249-271.

[7] Pantaleón, L. y Quezada, R., *Una introducción a la Teoría cuántica de la información*, Texto para el *Primer Coloquio del Departamento de Matemáticas*, Departamento de Matemáticas UAM-I, México, 2008.

[8] Petz, D.,*Quantum Information Theory and Quantum Statistics*, Springer, Berlin Heidelberg, 2008.

[9] Ruskai, M. B. and Jenčová, A., *A Unified Treatment of Convexity of Relative Entropy and Related Trace Functions, with Conditions for Equality*, arxiv:0903.2895v1 [quant-ph] 17 Mar 2009.

Dirección de los autores

Gildardo Barrientos

Universidad Autónoma Metropolitana,

Unidad Iztapalapa,

División de Ciencias Básicas e Ingeniería,

Departamento de Matemáticas.

Av. San Rafael Atlixco 186, Col. Vicentina

Del. Iztapalapa, C.P. 09340 México, D.F.

e-mail: `gil_barrientos@hotmail.com`

Roberto Quezada
Universidad Autónoma Metropolitana,
Unidad Iztapalapa,
División de Ciencias Básicas e Ingeniería,
Departamento de Matemáticas.
Av. San Rafael Atlixco 186, Col. Vicentina
Del. Iztapalapa, C.P. 09340 México, D.F.
e-mail: `roqb@xanum.uam.mx`