Norm Inequalities Related to the Cartesian Decomposition of Matrices

Author
Pritam Chandra

Supervisor
Dr. Rajendra Bhatia

Department of Mathematics

Capstone Thesis
For
Post Graduate Diploma in
Advanced Studies and Research

May, 2022

Table of contents

1 Introduction

A basic beginning to our understanding of complex numbers lies in the Cartesian decomposition. That is, every complex number z is uniquely decomposable into its real and imaginary parts as $z = a + ib$, where a is real and ib is imaginary. The numbers a and b are given by

$$a = \frac{z + \bar{z}}{2} \text{ and } b = \frac{z - \bar{z}}{2i},$$

where $\bar{z}$ is the complex conjugate of z.

As numbers generalize to matrices, this decomposition generalizes as well. A Hermitian matrix replaces the real part, skew Hermitian replaces the imaginary, and the adjoint $(^*)$ replaces the conjugate. It is neat how the eigenvalues of matrices capture the connection to the basic case of complex numbers. That is, when we write a general matrix T as $T = A + iB$ with A Hermitian and iB skew Hermitian, the eigenvalues of A are all real, of B are all imaginary, while of T, generally, are complex numbers. The parts A and B imitating the complex numbers are called the ***real*** and ***imaginary*** parts of T and are given by an analogous formula

$$A = \frac{T + T^*}{2} \text{ and } B = \frac{T - T^*}{2i}.$$

For complex numbers the sizes of z, a and b are nicely related as $|z|^2 = a^2 + b^2$. It is therefore compelling to investigate the relationship shared by the sizes of T, A and B. The space of matrices provides ample liberty in choosing a notion of size. This makes the problem more difficult and interesting.

Taking the natural course of imitating the Euclidean distance provides a promising first step. For matrices the norm arising from this metric has a special name - the Frobenius norm. For $A = [a_{ij}]$, the ***Frobenius norm*** of A is denoted by $\|A\|_2$ and is defined as

$$\|A\|_2 = \sum_{i,j} |a_{ij}|^2. \tag{1}$$

Here, it is not difficult to see that

$$\|T\|_2^2 = \|A\|_2^2 + \|B\|_2^2. \tag{2}$$

This serves as the most natural analogue to the case of complex numbers (the result is also covered in the main results of the paper which cater to more general questions). One naturally enquires from here about what happens for other norms on matrices.

Of particular interest are the unitarily invariant norms. A norm $|||\cdot|||$ is ***unitarily invariant*** if $|||UAV||| = |||A|||$ for all unitaries U, V and all matrices A. A special class of these norms are the ***Schatten p-norms*** as defined below.

Every matrix A of dimension n admits the singular value decomposition, where the singular values $s_1 \geq \cdots \geq s_n \geq 0$ of A are eigenvalues of the matrix A^*A. Then, for $1 \leq p \leq \infty$ one denotes the Schatten p-norm of A as $\|A\|_p$ and defines it as the p-norm of the vector of its singular values $(s_1, \ldots, s_n)$. That is

$$\|A\|_p = (s_1^p + \cdots + s_n^p)^{\frac{1}{p}}.$$

It turns out that

$$\sum_{i,j} |a_{ij}|^2 = \sqrt{s_1^2 + \cdots + s_n^2} = \|A\|_2$$

and hence the notation used in (2) is justified. This paper motivates and establishes the relationships enjoyed by $\|T\|_p^2$, $\|A\|_p^2$ and $\|B\|_p^2$. We study several cases investigating whether enforcing special properties (like positivity) on A and (or) B enforces special improvements on this relationship. Apart from being significant for its own sake, this problem also finds importance in analysis of operators [1, 5], and mathematical physics [6].

Ideally, one first expects an equality, like in (2). However, we immediately see this to not be the case by considering $p = \infty$. For any matrix A, the norm $\|A\|_\infty$ is its highest singular value. It is also called the operator norm of A, and is equivalently calculated by

$$\|A\|_\infty = \sup_{\|x\|_2 = 1} \|Ax\|_2.$$

An example like the one below,

$$\underbrace{\begin{bmatrix} 1+2i & 0 \\ 0 & 2 \end{bmatrix}}_{T} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}}_{A} + i \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}}_{B},$$

where we have $\|T\|^2 = |1 + 2i|^2 = 5 < 2^2 + 2^2 = \|A\|^2 + \|B\|^2$, calls for a weaker demand than equality. We observe that the matrix T in the above example is normal. When T is normal its real and imaginary parts commute (see result (1) at the end in section **11**). For such matrices we quickly see that

$$\begin{aligned} \|T\|_\infty^2 = \|T^*T\|_\infty &= \|(A - iB)(A + iB)\|_\infty \\ &= \|A^2 + B^2 + i(AB - BA)\|_\infty \\ &\leq \|A\|_\infty^2 + \|B\|_\infty^2 + \|AB - BA\|_\infty. \end{aligned}$$

This leads to

$$\|T\|_\infty^2 \leq \|A\|_\infty^2 + \|B\|_\infty^2, \text{ when } T \text{ is normal.} \tag{3}$$

So, an inequality for general matrices is the next best expectation. Normal matrices along with the operator norm are able to provide this behavior. However, even this fails to hold generally. For example, for the decomposition given as

$$\underbrace{\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}}_{T} = \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{A} + i \underbrace{\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}}_{B}$$

we have $4 = \|T\|_\infty^2 > \|A\|_\infty^2 + \|B\|_\infty^2 = 1^2 + 1^2 = 2$. This shows that just for the ∞-norm and general matrices, an inequality between the left-hand side and the right-hand of (3) can go in either directions. This provides a testimony for possibly intricate properties existing between $\|T\|_p^2$, $\|A\|_p^2$ and $\|B\|_p^2$. These are revealed in the coming sections. We look at [3] to list satisfying and elegant relationships enjoyed by these quantities through cases that cover a variety of natural assumptions on A and B. Besides, we also present some surprising twists!

2 Acknowledgements

I would like to thank Professor Rajendra Bhatia for introducing me to mathematics in general, and these topics in particular; for nurturing my interests and abilities in the subject; for teaching and guiding me in my thesis and in several courses leading up to it.

3 Definitions, notations and resuts

Every matrix A on a Hilbert space $\mathcal{H}$ admits a singular value decomposition which allows it to be written as $A = USV^*$, where U and V are unitary, and S is a positive diagonal matrix. Adjust the columns of U and V such that the diagonal entries, say $s_1, \ldots, s_n$, are in nonincreasing order. Then s_j is called the ***j*-th singular value** of A. The columns of U are called the **left singular vectors**, and of V are called the **right singular vectors**.

It turns out that s_j^2 are eigenvalues of both A^*A and AA^* with eigenvectors v_j (j-th column of V) and u_j (j-th column of U) respectively.

We will quickly recall some types of matrices, and their relation to their eigenvalues. A matrix A is said to be **Hermitian** if it is its own adjoint. That is $A = A^*$. Hermitian matrices have real eigenvalues. These matrices admit the spectral decomposition, which allows $A = UDU^*$ where U is unitary and D is a real diagonal matrix with the eigenvalues of A as its diagonal entries. We call these eigenvalues $\lambda_1(A), \ldots,$ $\lambda_n(A)$ or $\lambda_1^\downarrow(A), \ldots, \lambda_n^\downarrow(A)$ where the second notation captures their arrangement in decreasing order. For any polynomial f, the matrix $f(A)$ has eigenvalues $f(\lambda_1), \ldots,$ $f(\lambda_n)$.

If a Hermitian matrix A has only nonnegative eigenvalues, we call it to be **positive semidefinite**. For the sake of brevity, we call them "positive" throughout this paper.

Let s_js and λ_js be the singular values and eigenvalues respectively of Hermitian A. In general, $\{s_1, \ldots, s_n\} = \{|\lambda_1|, \ldots, |\lambda_n|\}$, and $s_j = |\lambda|_j$ where $|\lambda|_j$ is the j-th largest eigenvalue of A in absolute value. When A is positive, $s_j = \lambda_j$.

Now, given a subspace S of $\mathcal{H}$, the space $\mathcal{H}$ can be decomposed as the direct sum $\mathcal{H} = S \oplus S^{\perp}$ where $S^{\perp}$ is the orthogonal complement of S. Then every vector $z \in \mathcal{H}$ can be uniquely decomposed as $z = x + y$ where $x \in S$, $y \in S^{\perp}$ and x and y are orthogonal to each other. We call the map P_S as the **orthoprojection** into S if for all z we have $P_S(z) = x$. Orthoprojections thus arise in this way out of subspaces of $\mathcal{H}$.

One can check that a map P is an orthoprojection if and only if it is Hermitian and idempotent. That is $P^* = P$ and $P^2 = P$ respectively. Here, $P = P_{\operatorname{ran} P}$, that is P is the orthoprojection into the subspace $\operatorname{ran} P$.

Finally, for a vector $x \in \mathbb{R}^n$, its trace is the sum of its coordinates. That is $\operatorname{tr} x = x_1 + \cdots + x_n$ where $x = (x_1, \ldots, x_n)$.

Finally, we list below some linear algebraic results which are going to be useful at several places in establishing the main results of this paper. The proofs for these results are listed at the end of the paper in section **11**.

Useful results from linear algebra

***Result* 1**. *Let T be normal with Cartesian decomposition $T = A + iB$. Then A and B must commute. That is $AB - BA = 0$.*

***Result* 2**. *If P is an orthoprojection then $P \leq I$. That is $I - P$ should be positive.*

***Result* 3**. *If for Hermitian A and B we have $A \leq B$, then $\operatorname{tr} A \leq \operatorname{tr} B$.*

***Result* 4**. *Let A be a positive matrix. Then $\|A\|_1 = \operatorname{tr} A$.*

***Result* 5**. *Let $A = [a_{ij}]$ be any $n \times n$ matrix. Then*

$$\|A\|_1 \geq \sum_{j=1}^{n} |a_{jj}|.$$

Result* 6**. ***(Equivalence of norms) *Let $\|\cdot\|_p$ denote the p-norm on $\mathbb{C}^n$. Then for $p \leq q$,*

$$\|x\|_p \leq n^{\frac{1}{p} - \frac{1}{q}} \|x\|_q.$$

After stating these we are all set to plunge into the primary ideas used in this paper.

4 Majorisation and convex functions

For any vector $x = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$, we define $x^{\downarrow}$ and $x^{\uparrow}$ as the vectors obtained by rearranging the coordinates of x in decreasing and increasing order respectively. That is if $x^{\downarrow} = (x_1^{\downarrow}, \ldots, x_n^{\downarrow})$ and $x^{\uparrow} = (x_1^{\uparrow}, \ldots, x_n^{\uparrow})$, then $x_1^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$ and $x_1^{\uparrow} \leq \cdots \leq x_n^{\uparrow}$, and the numbers $x_1^{\uparrow}, \ldots, x_n^{\uparrow}$ and $x_1^{\downarrow}, \ldots, x_n^{\downarrow}$ and $x_1, \ldots, x_n$ are the same as a collection.

With this setting it is easy to see that

$$x_j^{\downarrow} = x_{n-j+1}^{\uparrow} \text{ for all } 1 \leq j \leq n. \tag{4}$$

Definition 7. *Let $x, y \in \mathbb{R}^n$. We say x is* ***majorised*** *by y, denoting it as* $\boldsymbol{x \prec y}$, *if*

$$\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow} \tag{5}$$

and

$$\sum_{j=1}^{n} x_j^{\downarrow} = \sum_{j=1}^{n} y_j^{\downarrow} \textit{ or } \operatorname{tr} x = \operatorname{tr} y. \tag{6}$$

Remark 8. In light of (4) and (6), equation (5) is redefined as follows.

$$\begin{aligned} \sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow} &\Longrightarrow \sum_{j=1}^{n} x_j^{\downarrow} - \sum_{j=k+1}^{n} x_j^{\downarrow} \leq \sum_{j=1}^{n} y_j^{\downarrow} - \sum_{j=k+1}^{n} y_j^{\downarrow} \\ \Longrightarrow \sum_{j=k+1}^{n} x_j^{\downarrow} \geq \sum_{j=k+1}^{n} y_j^{\downarrow} &\Longrightarrow \sum_{j=k+1}^{n} x_{n-j+1}^{\downarrow} \geq \sum_{j=k+1}^{n} y_{n-j+1}^{\downarrow} \\ &\Longrightarrow \sum_{j=1}^{n-k} x_j^{\downarrow} \geq \sum_{j=1}^{n-k} y_j^{\downarrow}. \end{aligned}$$

This gives us an alternate definiton of majorisation as

$$\sum_{j=1}^{k} x_j^{\downarrow} \geq \sum_{j=1}^{k} y_j^{\downarrow} \text{ for all } k, \text{ and } \sum_{j=1}^{n} x_j^{\uparrow} = \sum_{j=1}^{n} y_j^{\uparrow}. \tag{7}$$

We further note that if a vector x is rearranged into the vector Px for some permutation P, then the vectors $(Px)^{\downarrow} = x^{\downarrow}$, because they essentially have the same coordinates. This entails that if $x \prec y$, then $Px \prec Qy$ for all permutations P and Q.

Definition 9. *A matrix $A = [a_{ij}]$ is said to be doubly stochastic if all its entries are nonnegative and each of its rows and columns sum to 1. That is*

$$\sum_{j=1}^{n} a_{ij} = 1 \text{ for all } i, \ \sum_{i=1}^{n} a_{ij} = 1 \text{ for all } j \text{ and } a_{ij} \geq 0 \text{ for all } i, j.$$

After this definition we make the following remark.

Result 10. *For A, B doubly stochastic, for any $t \in [0, 1]$ their convex combination $(1-t)A + tB$ is doubly stochastic. Their product AB is also doubly stochastic.*

We provide a proof to this result at the end in section **11**.

Doubly stochastic matrices and majorisation are intricately connected (look at [1, chapter II]). This is captured in the next theorem. Though the implications of the theorem also work in the reverse direction, for the purpose of this paper we limit to the given direction only.

Theorem 11. *For $x, y \in \mathbb{R}^n$, if $x \prec y$ then $x = Ay$ for some $n \times n$ doubly stochastic matrix A.*

Proof of theorem. We proceed by induction on n. When $n = 1$, $x \prec y$ just means $x = y$. Hence the statement is trivially true.

Assume that for vectors in $\mathbb{R}^{n-1}$ the above theorem holds. Let $x, y \in \mathbb{R}^n$ such that $x \prec y$. Let us initiallly assume that $x_j = x_j^{\downarrow}$ and $y_j = y_j^{\downarrow}$ for all $1 \leq j \leq n$, that is, the entries of x and y are in nonincreasing order.

Then, from the definition of majorisation, we have $y_n \leq x_1 \leq y_1$. Choose k such that $y_k \leq x_1 \leq y_{k-1}$. Then as $x_1 \in (y_k, y_1)$ we can find t such that

$$x_1 = t y_1 + (1-t) y_k.$$

Then we construct two vectors $x', y' \in \mathbb{R}^{n-1}$ such that

$$\begin{aligned} x' &= (x'_2, \ldots, x'_n) = (x_2, \ldots, x_n) \\ y' &= (y'_1, \ldots, y'_n) = (y_2, \ldots, y_{k-1}, (1-t)y_1 + t y_k, y_{k+1}, \ldots, y_n) . \end{aligned}$$

We will show that $x' \prec y'$. By our choice we have $y_1 \geq \cdots \geq y_{k-1} \geq x_1 \geq \cdots \geq x_n$, so for $2 \leq m \leq k-1$ we have

$$\sum_{j=2}^{m} x_j \ \leq \ \sum_{j=2}^{m} y_j \ .$$

Further, for $k \le m \le n$ we have

$$\begin{aligned}
\sum_{j=2}^{m} y_j' &= \sum_{j=2}^{k-1} y_j + (1-t)y_1 + t y_k + \sum_{j=k+1}^{m} y_j \\
&= \sum_{j=1}^{m} y_j - [t y_1 + (1-t) y_k] \\
&= \sum_{j=1}^{m} y_j - x_1 \ge \sum_{j=1}^{m} x_j - x_1 \\
&= \sum_{j=2}^{m} x_j = \sum_{j=2}^{m} x_j',
\end{aligned}$$

where the step of inequality above is replaceable by an equality when $m = n$. This shows that $x' \prec y'$. Then, from our induction hypothesis, there exists an $(n-1) \times (n-1)$ matrix A' such that $x' = A' y'$.

Now let $P = [e_k\, e_2 \cdots e_{k-1}\, e_1\, e_{k+1} \ldots e_n]$ where e_j is the j-th standard basis (column) vector. The matrix P is basically a permutation that exchanges the first and the k-th columns of the matrix it is applied to. Then i-th entry of the vector $[tI + (1-t)P]y$ is given by

$$\sum_{j=1}^{n} t I_{ij} y_j + (1-t) P_{ij} y_j = \begin{cases} t y_i + (1-t) y_i = y_i & \text{if } i \neq 1, k \\ t y_1 + (1-t) y_k & \text{if } i = 1 \\ (1-t) y_1 + t y_k & \text{if } i = k \end{cases}.$$

Hence

$$[tI + (1-t)P]y = \begin{bmatrix} x_1 \\ y' \end{bmatrix}. \tag{8}$$

But we also define the doubly stochastic matrix B as below

$$B \begin{bmatrix} x_1 \\ y' \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & A' \end{bmatrix} \begin{bmatrix} x_1 \\ y' \end{bmatrix} = \begin{bmatrix} x_1 \\ A' y' \end{bmatrix} = \begin{bmatrix} x_1 \\ x' \end{bmatrix} = x. \tag{9}$$

Combining (8) and (9) we get

$$x = B[tI + (1-t)P]y.$$

Now B, I and P are all doubly stochastic, and hence by result 10 the matrix $B[tI + (1-t)P]$ is also doubly stochastic. This completes the proof of the theorem. □

Now, for any function $f: \mathbb{R} \to \mathbb{R}$, define the map induced by f on $\mathbb{R}^n$ as

$$f(x_1, \ldots, x_n) = (f(x_1), \ldots, f(x_n)).$$

We define the next theorem using the above notation.

Theorem 12. *For $x,y\in\mathbb{R}^n$ if $x\prec y$, then for every convex function $\varphi:\mathbb{R}\to\mathbb{R}$ we have*

$$\operatorname{tr}\varphi(x)\le\operatorname{tr}\varphi(y).$$

Proof of theorem. By theorem 11, if $x\prec y$ then there exists a doubly stochastic $A=[a_{ij}]$ for which $x=Ay$. Then for each $1\le i\le n$

$$\begin{aligned}
\varphi(x_i) &= \varphi\left(\sum_{j=1}^{n} a_{ij}\,y_j\right) \\
&\le \sum_{j=1}^{n} a_{ij}\,\varphi(y_j) \qquad \because \sum_{j=1}^{n} a_{ij} \text{ and } \varphi \text{ convex} \\
\implies \operatorname{tr}\varphi(x) &= \sum_{i=1}^{n} \varphi(x_i) \qquad \le \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\,\varphi(y_j) \\
&= \sum_{j=1}^{n} a_{ij}\,\varphi(y_j) \qquad = \operatorname{tr}\varphi(y).
\end{aligned}$$

This completes the proof of the above theorem. □

Corollary 13. *Let x,y be n-tuples such that $x\prec y$. Then*

$$\|x\|_p\le\|y\|_p \textit{ for } 1\le p\le\infty \textit{ and } \|x\|_p>\|y\|_p \textit{ for } 0\le p\le 1.$$

Proof of Corollary. For $1\le p\le\infty$ the map $\varphi(t)=|t|^p$ is convex. Hence, by theorem 12, we get

$$\|x\|_p^p = \sum_{j=1}^{n}|x_j|^p = \operatorname{tr}\varphi(x) \le \operatorname{tr}\varphi(y) = \sum_{j=1}^{n}|y_j|^p = \|y\|_p^p\ .$$

Now as the p-th root is a monotone function, we have $\|x\|_p\le\|y\|_p$.

When $0\le p\le 1$, then $\varphi(x)=-|t|^p$ is convex. Hence, in this case we get

$$-\|x\|_p^p = -\sum_{j=1}^{n}|x_j|^p = \operatorname{tr}\varphi(x) \le \operatorname{tr}\varphi(y) = -\sum_{j=1}^{n}|y_j|^p = -\|y\|_p^p.$$

Using monotonicity of the p-th root and taking negative on both sides we have $\|x\|_p\ge\|y\|_p$. This completes the proof of the theorem. □

For our convenience in this section we define the following notations. Given an n-tuple $\delta=(\delta_1,\dots,\delta_n)$ and an m-tuple $\sigma=(\sigma_1,\dots,\sigma_m)$, we write $\sigma\vee\delta$ as the $(n+m)$-tuple $(\delta_1,\dots,\delta_n,\sigma_1,\dots,\sigma_m)$. We write δ^2 for the tuple $\{\delta_1^2,\dots,\delta_n^2)$. Further, we identify $\delta\vee 0$ with δ.

Lemma 14. *Let s, α, β be three n-tuples such that $s^2 \prec \alpha^2 + \beta^2$. Then*

$$\begin{array}{rcll} \|s\|_p^2 & \leq & \|\alpha\|_p^2 + \|\beta\|_p^2 & \text{for } 2 \leq p \leq \infty \\ \|s\|_p^2 & \geq & \|\alpha\|_p^2 + \|\beta\|_p^2 & \text{for } 1 \leq p \leq 2 \end{array}.$$

Proof of lemma. Let $q = \frac{p}{2}$. Consider when $2 \leq p \leq \infty$, then $1 \leq q \leq \infty$. Applying corollary 13 with q we get

$$\begin{array}{rrcll} & \|s^2\|_q & \leq & \|\alpha^2 + \beta^2\|_q & \\ & & \leq & \|\alpha^2\|_q + \|\beta^2\|_q & \because \text{ Triangle inequality} \\ \Longrightarrow & (s_1^{2q} + \cdots + s_n^{2q})^{\frac{1}{q}} & \leq & (\alpha_1^{2q} + \cdots + \alpha_n^{2q})^{\frac{1}{q}} + (\beta_1^{2q} + \cdots + \beta_n^{2q})^{\frac{1}{q}} & \\ \Longrightarrow & (s_1^p + \cdots + s_n^p)^{\frac{2}{p}} & \leq & (\alpha_1^p + \cdots + \alpha_n^p)^{\frac{2}{p}} + (\beta_1^p + \cdots + \beta_n^p)^{\frac{2}{p}} & \\ \Longrightarrow & \|s\|_p^2 & \leq & \|\alpha\|_p^2 + \|\beta\|_p^2. & \end{array}$$

When $1 \leq p \leq 2$, then $1/2 \leq q \leq 1$. Applying corollary 13 and triangle inequality again in the corresponding steps we get all inequalities reversed. This completes the proof of the lemma. $\square$

Definition 15. *Consider the real tuple $\beta = (\beta_1, \ldots, \beta_n)$. We call the vector $\boldsymbol{p}(\beta)$ as **pairing** of β, and we define it as follows*

$$\boldsymbol{p}(\beta) = \begin{cases} \left(\beta_1 + \beta_2,\ \beta_3 + \beta_4,\ \ldots,\ \beta_{2m-1} + \beta_{2m},\ \underbrace{0, \ldots, 0}_{m \text{ times}}\right) & \text{if } n = 2m \\ \left(\beta_1 + \beta_2,\ \beta_3 + \beta_4^2,\ \ldots,\ \beta_{2m-1}^2 + \beta_{2m},\ \beta_{2m+1}, \underbrace{0, \ldots, 0}_{m \text{ times}}\right) & \text{if } n = 2m+1 \end{cases}.$$

Note that $\boldsymbol{p}(\beta)$ may have nonzero entries only until the $\tilde{n} := \lceil \frac{n}{2} \rceil$ component.

We make the following lemma based on the above defintion.

Lemma 16. *Let $\beta = (\beta_1, \ldots, \beta_n)$ be an n-tuple. Let $\gamma^2 = \boldsymbol{p}(\beta^2)$. Then*

$$\gamma^2 \vee \gamma^2 = (\gamma_1^2, \ldots, \gamma_{\tilde{n}}^2, \gamma_1^2, \ldots, \gamma_{\tilde{n}}^2) \prec 2\beta^2.$$

Proof of lemma. Let $\Gamma^2 = \gamma^2 \vee \gamma^2$. Without loss of generality assume β^2 and Γ^2 have coordinates arranged in decreasing order. We can do this because majorisation can be done up to permutations. Then we immediately note that

$$\sum_{j=1}^{2\tilde{n}} \Gamma_j^2 = 2\sum_{j=1}^{\tilde{n}} \gamma_j^2 = 2\sum_{j=1}^{\tilde{n}} (\beta_{2j-1}^2 + \beta_{2j}) = 2\sum_{j=1}^{2\tilde{n}} \beta_j^2$$

(after appropriate size adjustments by identifying $\beta \vee 0$ with β).
Further, for $1 \leq m \leq 2\tilde{n}$ we have

$$\begin{aligned}
\sum_{j=1}^{m} \Gamma_j^2 &= \sum_{j=1}^{2k} \Gamma_j^2 + \sum_{j=2k+1}^{m} \Gamma_j^2 \quad \text{where } \ 2k \leq m < 2(k+1) \\
&= 2\sum_{j=1}^{k} \gamma_j^2 + \sum_{j=2k+1}^{m} \gamma_{\frac{j+1}{2}}^2 \\
&= 2\sum_{j=1}^{k} (\beta_{2j-1}^2 + \beta_{2j}^2) + \sum_{j=2k+1}^{m} (\beta_j^2 + \beta_{j+1}^2) \\
&\leq 2\sum_{j=1}^{2k} \beta_j^2 + 2\sum_{j=2k+1}^{m} \beta_j^2 \\
&= 2\sum_{j=1}^{m} \beta_j^2.
\end{aligned}$$

This completes the proof of the lemma. $\square$

Lemma 17. *Let $\gamma^2 = \boldsymbol{p}(\beta^2)$ where β is an n-tuple. Then*

$$\begin{aligned}
\|\gamma\|_p^2 &\leq 2^{1-\frac{2}{p}} \|\beta\|_p^2 \quad \text{for } 2 \leq p \leq \infty \\
\|\gamma\|_p^2 &\geq 2^{1-\frac{2}{p}} \|\beta\|_p^2 \quad \text{for } 1 \leq p \leq 2
\end{aligned}$$

where elements of γ are positive square roots of elements of γ^2.

Proof of lemma. Assume β_j is such that $\beta_1^2 \geq \cdots \geq \beta_n^2$ (this can be assumed because norm is permutation invariant). Now for all $p \geq 2$ we have

$$\begin{aligned}
\|\gamma\|_p^2 &= (\gamma_1^p + \cdots + \gamma_n^p)^{\frac{2}{p}} \\
&= 2^{-\frac{2}{p}}(2[\gamma_1^p + \cdots + \gamma_n^p])^{\frac{2}{p}} \\
&= 2^{-\frac{2}{p}}\Big[(\gamma_1^2)^{\frac{p}{2}} + \cdots + (\gamma_n^2)^{\frac{p}{2}} + (\gamma_1^2)^{\frac{p}{2}} + \cdots + (\gamma_n^2)^{\frac{p}{2}}\Big]^{\frac{2}{p}} \\
&= 2^{-\frac{2}{p}}\|\gamma^2 \vee \gamma^2\|_{\frac{p}{2}} \\
&\leq 2^{-\frac{2}{p}} \cdot \|2\beta^2\|_{\frac{p}{2}} \qquad \text{(from lemma 16 and corollary 13)} \\
&= 2^{1-\frac{2}{p}}\Big[(\beta_1^2)^{\frac{p}{2}} + \cdots + (\beta_n^2)^{\frac{p}{2}}\Big]^{\frac{2}{p}} \\
&= 2^{1-\frac{2}{p}}\|\beta\|_p^2.
\end{aligned}$$

This concludes the proof of lemma for $2 \leq p \leq \infty$. All the inequalities are reveresed for $1 \leq p \leq 2$ (from corollary 13). $\square$

5 The minimax principle

In this section we state and prove the well known Minimax principle. To show this we go via Poincaré's inequality.

Theorem 18. *Let A be a Hermitian matrix on $\mathcal{H}$. Let $\mathcal{M}$ be a k-dimensional subspace of $\mathcal{H}$. Then there exists vectors x, y in $\mathcal{M}$ such that*

$$\langle x, Ax\rangle \leq \lambda_k^{\downarrow}(A) \text{ and } \langle x, Ay\rangle \geq \lambda_k^{\uparrow}(A).$$

Proof of theorem. Use the spectral theorem to find orthonormal vectors $u_1, \ldots, u_n$ such that

$$A u_j = \lambda_j^{\downarrow}(A)\, u_j.$$

Let $\mathcal{N} = [u_k, \ldots, u_n]$. Then $\dim \mathcal{N} = n - k + 1$ and $\dim \mathcal{M} = k$. In particular

$$\dim \mathcal{M} + \dim \mathcal{N} = n + 1.$$

This means that they must have a nontrivial intersection, or $\mathcal{M} \cap \mathcal{N} \neq \{0\}$. So choose a unit vector $x \in \mathcal{M} \cap \mathcal{N}$.

As $x \in \mathcal{M}$, it can be expanded as a linear combination of $u_k, \ldots, u_n$ as

$$x = \varsigma_k\, u_k + \cdots + \varsigma_n\, u_n \text{ where } |\varsigma_k|^2 + \cdots + |\varsigma_n|^2 = 1. \tag{10}$$

Then,

$$\begin{aligned}
\langle x, Ax\rangle &= \left\langle \sum_{j=k}^{n} \varsigma_j\, u_j, A\left(\sum_{j=k}^{n} \varsigma_j\, u_j\right)\right\rangle \\
&= \left\langle \sum_{j=k}^{n} \varsigma_j\, u_j, \left(\sum_{j=k}^{n} \varsigma_j\, A u_j\right)\right\rangle \\
&= \left\langle \sum_{j=k}^{n} \varsigma_j\, u_j, \left(\sum_{j=k}^{n} \varsigma_j\, \lambda_j^{\downarrow}(A)\, u_j\right)\right\rangle \\
&= \sum_{j=k}^{n} |\varsigma_j|^2\, \lambda_j^{\downarrow}(A) && \because u_j\text{s are orthonormal} \\
&\leq \sum_{j=k}^{n} |\varsigma_j|^2\, \lambda_k^{\downarrow}(A) && \because \lambda_j^{\downarrow}(A) \text{ are decreasing} \\
&= \lambda_k^{\downarrow}(A) && \because \text{from (10).}
\end{aligned}$$

But as $x \in \mathcal{M} \cap \mathcal{N}$, $x \in \mathcal{M}$, and the required vector has been found.

To find y, rename $u_1, \ldots, u_n$ as vectors such that $A u_j = \lambda_j^{\uparrow}(A)\, u_j$. Repeat the same argument using the increasing property of $\lambda_j^{\uparrow}(A)$ to get a lower bound. □

Theorem 19. *(The Minimax principle) Let A be a Hermitian matrix on $\mathcal{H}$. Then*

$$\begin{aligned}\lambda_k^{\downarrow}(A) &= \max_{\substack{\mathcal{M}\subset\mathcal{H}\\ \dim\mathcal{M}=k}} \min_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle \\ &= \min_{\substack{\mathcal{M}\subset\mathcal{H}\\ \dim\mathcal{M}=n-k+1}} \max_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle\end{aligned}.$$

Proof of theorem. Let $\mathcal{M}$ be any k-dimensional subspace of $\mathcal{H}$. Then by Poincaré's inequality (theorem 18) there exists $x_0\in\mathcal{M}$ such that

$$\langle x_0, Ax_0\rangle \le \lambda_k^{\downarrow}(A).$$

This means in particular

$$\min_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle \le \lambda_k^{\downarrow}(A). \tag{11}$$

But this is true for all k-dimensional subspaces $\mathcal{M}$, showing that

$$\max_{\substack{\mathcal{M}\subset\mathcal{H}\\ \dim\mathcal{M}=k}} \min_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle \ \le\ \lambda_k^{\downarrow}(A). \tag{12}$$

But now choose $\mathcal{M}_0 = [u_1, \ldots, u_k]$, where these vectors are the eigenvectors of A corresponding to the top k eigenvalues of A. Then for $u_k\in\mathcal{M}_0$, we have

$$\langle u_k, Au_k\rangle = \langle u_k, \lambda_k^{\downarrow}(A)\, u_k\rangle = \lambda_k^{\downarrow}(A).$$

Combining with (12) we get

$$\max_{\substack{\mathcal{M}\subset\mathcal{H}\\ \dim\mathcal{M}=k}} \min_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle \ =\ \lambda_k^{\downarrow}(A).$$

For the second principle we use the second part of Poincaré's inequality. This says that every $(n-k+1)$-dimensional subspace $\mathcal{M}$ has a vector y_0 for which

$$\langle y_0, Ay_0\rangle \ge \lambda_{n-k+1}^{\uparrow}(A) = \lambda_k^{\downarrow}(A).$$

This means for all such $\mathcal{M}$

$$\max_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle \ \ge\ \lambda_k^{\downarrow}(A) \implies \min_{\substack{\mathcal{M}\subset\mathcal{H}\\ \dim\mathcal{M}=n-k+1}} \max_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \langle x, Ax\rangle \ \ge\ \lambda_k^{\downarrow}(A).$$

The $\mathcal{M}_0$ for which equality is achieved is given by $[u_k, \ldots, u_n]$. This completes the proof of the theorem. $\square$

6 Cauchy's interlacing theorem

Theorem 20. *Let A be a Hermitian matrix on $\mathcal{H}$, and let $B = U^* AU$ be a compression of A to an $(n-k)$-dimensional subspace $\mathcal{N}$. That is $U = [u_1 \ldots u_{n-k}]$ for some $u_1, \ldots, u_{n-k}$ orthonormal. Then*

$$\lambda_j^\downarrow(A) \geq \lambda_j^\downarrow(B) \geq \lambda_{j+k}^\downarrow(A).$$

Proof. Fix any j. If A is Hermitian, then so is B. By spectral theorem, let $v_1, \ldots, v_j$ be the eigenvectors of B corresponding to $\lambda_1^\downarrow(B), \ldots, \lambda_j^\downarrow(B)$, for $1 \leq j \leq n-k$. Let $\mathcal{M} = [v_1, \ldots, v_{n-k}]$. Then for any $x \in \mathcal{M}$,

$$\langle x, Bx\rangle = \langle x, U^* AUx\rangle = \langle Ux, AUx\rangle = \langle x, Ax\rangle.$$

Hence, using Minimax principle, we obtain,

$$\begin{aligned} \lambda_j^\downarrow(B) &= \min_{\substack{x \in \mathcal{M} \\ \|x\|=1}} \langle x, Bx\rangle = \min_{\substack{x \in \mathcal{M} \\ \|x\|=1}} \langle x, Ax\rangle \\ &\leq \max_{\dim \mathcal{M}=j} \min_{\substack{x \in \mathcal{M} \\ \|x\|=1}} \langle x, Ax\rangle \qquad (13) \\ &= \lambda_j^\downarrow(A) \end{aligned}$$

Now we apply the same to $-A$ and its compression $-B$, noting that

$$-\lambda_i^\downarrow(A) = \lambda_i^\uparrow(-A) = \lambda_{n-i+1}^\downarrow(-A) \text{ for } 1 \leq i \leq n$$

and

$$-\lambda_j^\downarrow(B) = \lambda_j^\uparrow(-B) = \lambda_{(n-k)-j+1}^\downarrow(-B) \text{ for } 1 \leq j \leq n-k.$$

Observe that $n-(k+j)+1 = (n-k)-j+1 \leq j$. So putting $i = k+j$ and applying (13) we get

$$\begin{aligned} \lambda_{(n-k)-j+1}^\downarrow(-B) &\leq \lambda_{(n-k)-j+1}^\downarrow(-A) \\ \Longrightarrow \quad -\lambda_j^\downarrow(B) &\leq -\lambda_{j+k}^\downarrow(A) \\ \Longrightarrow \quad \lambda_j^\downarrow(B) &\geq \lambda_{j+k}^\downarrow(A). \end{aligned}$$

The above along with (13) completes the proof of the theorem. $\square$

7 Antisymmetric tensor products

Definition 21. *Given two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, define the Hilbert space $\mathcal{H} \otimes \mathcal{K}$ as follows.*

1. *Let $\mathcal{H} \otimes \mathcal{K}$ is the collection of all linear combinations of vectors of the form $x \otimes y$, where $x \in \mathcal{H}$ and $y \in \mathcal{K}$.*

2. *The vector $x \otimes y$ is defined such that it interacts with the inner product on $\mathcal{H} \otimes \mathcal{K}$ in the following way,*

$$\langle u \otimes v, x \otimes y \rangle = \langle u, x \rangle \langle v, y \rangle \tag{14}$$

   for all $u, x \in \mathcal{H}$ and $v, y \in \mathcal{K}$. Naturally, the definition of this inner product is extended to all members of $\mathcal{H} \otimes \mathcal{K}$ using linearity.

The vector $x \otimes y$ is called the ***tensor product*** *of x and y. Such tensor products are called* ***elementary tensors*** *in $\mathcal{H} \otimes \mathcal{K}$. Therefore, the members of $\mathcal{H} \otimes \mathcal{K}$ are linear combinations of elementary tensors.*

If $\{e_1, \ldots, e_n\}$ and $\{f_1, \ldots, f_m\}$ are bases for $\mathcal{H}$ and $\mathcal{K}$ respetively, then set $\{e_i f_j\}$ is a basis for $\mathcal{H} \otimes \mathcal{K}$. Moreover, if e_is and f_js are chosen to be orthonormal, then so is $e_i f_j$. This follows easily as a consequence of the defintions in (14). As a consequence of this we have

$$\dim \mathcal{H} \otimes \mathcal{K} = \dim \mathcal{H} \cdot \dim \mathcal{K}.$$

Following this, given matrices A and B on $\mathcal{H}$, we define the matrix $A \otimes B$ on elementary tensors as

$$(A \otimes B)(x \otimes y) := (Ax) \otimes (By) \tag{15}$$

and extend linearly to all members of $\mathcal{H} \otimes \mathcal{K}$.

Note. requiring the tensor product to be associative, we can define these products for multiple vectors, Hilbert spaces and matrices.

Now, from the definition above we see that

$$\begin{aligned} (A \otimes B) \cdot (C \otimes D)(x \otimes y) &= (A \otimes B)(Cx \otimes Dy) \\ &= (AC)x \otimes (BD)\, y \\ &= (AC \otimes BD)(x \otimes y). \end{aligned}$$

This implies

$$(A \otimes B) \cdot (C \otimes D) = AC \otimes BD. \tag{16}$$

And, for all $u \otimes v, x \otimes y \in \mathcal{H} \otimes \mathcal{K}$ we have

$$\begin{aligned} \langle u \otimes v, (A \otimes B)(x \otimes y) \rangle &= \langle u \otimes v, Ax \otimes By \rangle \\ &= \langle u, Ax \rangle \langle v, By \rangle \\ &= \langle A^* u, x \rangle \langle B^* v, y \rangle \\ &= \langle A^* u \otimes B^* v, x \otimes y \rangle \\ &= \langle (A^* \otimes B^*)(u \otimes v), x \otimes y \rangle. \end{aligned}$$

Extending linearly to all tensors tells us that

$$(A \otimes B)^* = A^* \otimes B^*. \tag{17}$$

As a consequence of the above fact we get $A \otimes B$ is Hermitian if A and B are Hermitian.

Now note that from the definition of tensor product, the order of operands is significant. That is $x \otimes y \neq y \otimes x$. However, for several purposes we might want to call these the same. For such purposes we define equivalence relations which equate tensor products whose operands appear in different orders. One way of doing this is through defining antisymmetric tensors.

For this we recall permutation groups and signatures. Let S_k be the group of permutations on k elements. Then for $\sigma \in S_k$ its signature, denoted by ε_σ, is given by

$$\varepsilon_\sigma = \begin{cases} 1 & \text{if } \sigma \text{ is a product of even number of permutations} \\ -1 & \text{otherwise} \end{cases}.$$

In this notation we recall that the determinant of the matrix $A = [a_{ij}]$ is given by

$$\det A = \sum_{\sigma \in S_n} \left(\varepsilon_\sigma \prod_{i=1}^{n} a_{i\sigma(i)} \right).$$

Definition 22. *For $x_1, \ldots, x_k \in \mathcal{H}$ denote their **antisymmetric tensor product** by $x_1 \wedge x_2 \wedge \cdots \wedge x_k$ and define it as*

$$x_1 \wedge \cdots \wedge x_k = \frac{1}{\sqrt{k!}} \sum_{\sigma \in S_k} \varepsilon_\sigma x_{\sigma(1)} \otimes x_{\sigma(2)} \otimes \cdots \otimes x_{\sigma(k)}.$$

One can use the same techniques used earlier to get from elementary tensors to tensor products of Hilbert spaces and matrices for antisymmetric tensors.

It is clear from the above definition that it is blind to the ordering of the vectors upto a factor of -1. That is $x_1 \wedge \cdots \wedge x_k = \varepsilon_\sigma x_{\sigma(1)} \wedge \cdots \wedge x_{\sigma(k)}$ for all permutations σ. As every transposition has a negative signature, if any two operands of an antisymmetric tensor product are equal, the product must be 0.

Similarly, when e_is are basis for $\mathcal{H}$, we have $\{e_{i_1} \wedge \cdots \wedge e_{i_k}\}$ as basis of $\wedge^k \mathcal{H}$, where $1 \leq i_1 < \cdots < i_k \leq n$. This entails that

$$\dim \wedge^k \mathcal{H} = \binom{n}{k}.$$

Theorem 23. $\langle x_1 \wedge \cdots \wedge x_k, y_1 \wedge \cdots \wedge y_k \rangle = \det [\langle x_i, y_j \rangle]$.

Proof of theorem.

$$\begin{aligned}
& \langle x_1 \wedge \cdots \wedge x_k, y_1 \wedge \cdots \wedge y_k \rangle \\
= & \left\langle \frac{1}{\sqrt{k!}} \sum_{\sigma \in S_k} x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(k)}, \frac{1}{\sqrt{k!}} \sum_{\tau \in S_k} y_{\tau(1)} \otimes \cdots \otimes y_{\tau(k)} \right\rangle \\
= & \frac{1}{k!} \sum_{\sigma \in S_k} \sum_{\tau \in S_k} \langle x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(k)}, y_{\tau(1)} \otimes \cdots \otimes y_{\tau(k)} \rangle \\
= & \frac{1}{k!} \sum_{\sigma \in S_k} \sum_{\tau \in S_k} \langle x_{\sigma(1)}, y_{\tau(1)} \rangle \cdots \langle x_{\sigma(k)}, y_{\tau(k)} \rangle \\
= & \frac{1}{k!} \sum_{\sigma \in S_k} \det [\langle x_i, y_j \rangle] \\
= & \det [\langle x_i, y_j \rangle]. \quad \square
\end{aligned}$$

Let $Q_{k,n}$ denote the set of all k-length strictly increasing tuples from $\{1, \ldots, n\}$. That is

$$Q_{k,n} = \{(i_1, \ldots, i_k) : 1 \leq i_1 < \cdots < i_k \leq n\}.$$

For each $\mathcal{I} \in Q_{k,n}$ define the tensor $e_{\mathcal{I}}$ as

$$e_{\mathcal{I}} = e_{i_1} \wedge \cdots \wedge e_{i_k} \text{ where } \mathcal{I} = (i_1, \ldots, i_k).$$

Let the elements of $Q_{k,n}$ be ordered lexicographically. Then the set $\{e_{\mathcal{I}}\}_{\mathcal{I} \in Q_{k,n}}$ is an ordered basis of $\wedge^k \mathcal{H}$. Now, for every matrix A on $\mathcal{H}$, the matrix $\wedge^k A$ is a matrix on $\wedge^k \mathcal{H}$ of order $\binom{n}{k} \times \binom{n}{k}$. With respect to the basis $\{e_{\mathcal{I}}\}$ we describe the $(\mathcal{I}, \mathcal{J})$-th entry of $\wedge^k \mathcal{H}$.

Lemma 24. *The $(\mathcal{I}, \mathcal{J})$-th entry of $\wedge^k \mathcal{H}$ is the det $A[\mathcal{I}, \mathcal{J}]$, where $A[\mathcal{I}, \mathcal{J}]$ is the submatrix $[a_{ij}]_{i \in \mathcal{I}, j \in \mathcal{J}}$. That is $A[\mathcal{I}, \mathcal{J}]$ is the intesection of rows with indices picked from $\mathcal{I}$, and columns with indices picked from $\mathcal{J}$.*

Proof of lemma. We use the definition of $Q_{k,n}$ and theorem 23. Let $\mathcal{I} = (i_1, \ldots, i_k)$ and $\mathcal{J} = (j_1, \ldots, j_k)$. Then $(\mathcal{I}, \mathcal{J})$-th entry of $\wedge^k A$ is given by

$$\begin{aligned}
& \langle e_{\mathcal{I}}, (\wedge^k A) e_{\mathcal{J}} \rangle \\
= & \langle e_{i_1} \wedge \cdots \wedge e_{i_k}, (\wedge^k A) e_{j_1} \wedge \cdots \wedge e_{j_k} \rangle \\
= & \langle e_{i_1} \wedge \cdots \wedge e_{i_k}, A e_{j_1} \wedge \cdots \wedge A e_{j_k} \rangle \\
= & \det \langle e_{i_r}, A e_{i_s} \rangle && \because \text{theorem 23} \\
= & \det [a_{i_r i_s}]_{i_r \in \mathcal{I}, i_s \in \mathcal{J}} \\
= & \det A[\mathcal{I} | \mathcal{J}]
\end{aligned}$$

This completes the proof of the theorem. $\square$

Theorem 25. *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Then the eigenvalues of $\wedge^k A$ are $\{\lambda_{\mathcal{I}}\}_{\mathcal{I} \in Q_{k,n}}$ where $\lambda_{\mathcal{I}} = \lambda_{i_1} \cdots \lambda_{i_k}$ for $\mathcal{I} = (i_1, \ldots, i_k)$.*

Proof of theorem. Assume that the λ_js are distinct. Choose eigenvectors x_j of A such that $Ax_j = \lambda_j$. For $\mathcal{I} \in Q_{k,n}$ construct $x_{\mathcal{I}} = x_{i_1} \wedge \cdots \wedge x_{i_k}$ where $\mathcal{I} = (i_1, \ldots, i_k)$. Then

$$\begin{aligned} (\wedge^k A)\, x_{\mathcal{I}} &= (\wedge^k A)\, x_{i_1} \wedge \cdots \wedge x_{i_k} &&= (A x_{i_1}) \wedge \cdots \wedge (A x_{i_k}) \\ &= (\lambda_{i_1} \cdots \lambda_{i_k})\, x_{i_1} \wedge \cdots \wedge x_{i_k} &&= \lambda_{\mathcal{I}} x_{\mathcal{I}}\,. \end{aligned}$$

So for each $\mathcal{I} \in Q_{k,n}$, we have $\lambda_{\mathcal{I}}$ is an eigenvalue of $\wedge^k A$. They are all different as λ_js are distinct. But as $\wedge^k A$ is of dimension $\binom{n}{k}$, these have to all.

When λ_js are not distinct, then more has to be argued. For the sake of brevity, we only see an outline of the proof. By density of matrices with distinct eigenvalues, we can find a matrix A' which is arbitrarily close to A and has distinct eigenvalues. We know the eigenvalues of $\wedge^k A'$ through the previous argument. Using continuity of eigenvalues we obtain the result for such cases. □

Corollary 26. *If A is positive, so is $\wedge^k A$.*

Proof of corollary. One can derive from (17) with a little work that if $A = A^*$ then $\wedge^k A = (\wedge^k A)^*$. Theorem 25 says that the eigenvalues of $\wedge^k A$ are all positive, as they are products of positive eigenvalues of A. Hence, $\wedge^k A$ is positive. □

Finally, in the spirit of submatrices, we state the following small result.

Result 27. *Let T have the Cartesian decomposition $T = A + iB$. For $\mathcal{I} \subset [1, \ldots, n]$, let $T_{\mathcal{I},\mathcal{I}}$ be a principal submatrix of T. Then*

$$T_{\mathcal{I},\mathcal{I}} = A_{\mathcal{I},\mathcal{I}} + i B_{\mathcal{I},\mathcal{I}}$$

is the Cartesian decomposition of $T_{\mathcal{I},\mathcal{I}}$.

We see a proof of this at the end in section **11**.

8 Some useful results

In this section we prove some results which are particular to our discussion. They will straightforwardly lead to our main result in the coming section.

Lemma 28. *Let k, n be positive integers such that $k < \frac{n}{2}$. Let $\mu_1 \geq \cdots \geq \mu_n$ and $\lambda_1 \geq \cdots \geq \lambda_{n-k}$ be two sets of real numbers obeying the interlacing relationships. That is,*

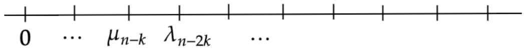
$$\mu_i \geq \lambda_i \geq \mu_{i+k} \text{ for } 1 \leq i \leq n-k.$$

Then there exists a subset $\{i_1, \ldots, i_{n-2k}\}$ of $\{1, \ldots, n-k\}$ such that

$$|\lambda_{i_s}| \geq |\mu_{k+s}| \text{ for } 1 \leq s \leq n-2k.$$

Proof of lemma. For convenience we put the μs on a number and argue in 3 cases.

<u>Case 1</u> $\mu_{n-k} \geq 0$.

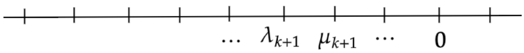


This implies that $\lambda_{n-2k} \geq \mu_{n-k} \geq 0$. Therefore the number $\lambda_1, \ldots, \lambda_{n-2k}$ and $\mu_{k+1}, \ldots, \mu_{n-k}$ are all positive. The interlacing relations imply

$$\lambda_i \geq \mu_{i+k} \Longrightarrow |\lambda_i| \geq |\mu_{i+k}| \text{ for } 1 \leq i \leq n-2k.$$

Therefore, the required set of indices are $\{1, \ldots, n-2k\}$.

<u>Case 2</u> $\mu_{k+1} < 0$.

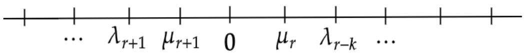


This implies that $\lambda_{k+1} \leq \mu_{k+1} \leq 0$. Therefore the numbers $\lambda_{k+1}, \ldots, \lambda_{n-k}$ and $\mu_{k+1}, \ldots, \mu_{n-k}$ are all negative. The interlacing relations say

$$\lambda_{i+k} \leq \mu_{i+k} \Longrightarrow |\lambda_{i+k}| \geq |\mu_{i+k}| \text{ for } 1 \leq i \leq n-2k.$$

Therefore, the required set of indices are $\{k+1, \ldots, n-k\}$.

<u>Case 3</u> $\mu_r \geq 0 > \mu_{r+1}$ for some $k < r < n-k$.



In this case we split our indices following case 1 and case 2. We note that $\lambda_{r-k} \geq \mu_r \geq 0$ and $\lambda_{r+1} \leq \mu_{r+1} < 0$. Therefore the numbers $\lambda_1, \ldots, \lambda_{r-k}$ and $\mu_{k+1}, \ldots, \mu_r$ are all positive, and the numbers $\lambda_{r+1}, \ldots, \lambda_{n-k}$ and $\mu_{r+1}, \ldots, \mu_{n-k}$ are all negative. Hence, using the interlacing relations, we have,

$$\begin{aligned} \lambda_i \geq \mu_{i+k} \quad &\Longrightarrow \quad |\lambda_i| \geq |\mu_{i+k}| \quad \text{for } 1 \leq i \leq r-k \text{ and,} \\ \lambda_{r+i} \leq \mu_{r+i} \quad &\Longrightarrow \quad |\lambda_{r+i}| \geq |\mu_{r+i}| \quad \text{for } 1 \leq i \leq n-r-k. \end{aligned}$$

Thus the required indices are $\{1, \ldots, r-k, r+1, \ldots, n-k\}$. This completes the proof of the lemma. ☐

Lemma 29. *Let U be an $n \times (n-k)$ isometry, that is the columns columns of U are orthonormal. Then for any $T = A + iB$ we have*

$$\operatorname{tr} U^* T^* T U \geq \operatorname{tr} (U^* A U)^2 + \operatorname{tr}(U^* B U)^2.$$

Proof of lemma. Look at the matrix UU^*. It is positive. Moreover,

$$(UU^*)^2 = UU^* UU^* = U I_k U^* = UU^*,$$

or UU^* is idempotent. This shows that UU^* is an orthoprojection. From result 2 we therefore have $UU^* \leq I$. So,

$$\begin{aligned} U^*T^*TU &= U^*T^* \cdot I \cdot TU \\ &\geq U^*T^* \cdot UU^* \cdot TU \\ &= U^*(A - iB)U \cdot U^*(A + iB)U \\ &= (U^*AU)^2 + (U^*BU)^2 + i(U^*AU \cdot U^*BU - U^*BU \cdot U^*AU) \\ &= (U^*AU)^2 + (U^*BU)^2 + i[U^*AU, U^*BU] \end{aligned}$$

where $[X, Y]$ is the commutator $XY - YX$. From here we use result 3 to see that

$$\operatorname{tr} U^*T^*TU \geq \operatorname{tr}(U^*AU)^2 + \operatorname{tr}(U^*BU)^2. \quad \square$$

Lemma 30. *Let A be positive with eigenvalues $\alpha_1 \geq \cdots \geq \alpha_n$. Then for any k-dimensional compression U^*AU of A, we have*

$$\operatorname{tr}(U^*AU)^2 \geq \sum_{j=k+1}^{n} \alpha_j^2.$$

Proof of lemma. As U^*AU is a compression, its eigenvalues interlace with the eigenvalues of A. In particular, from theorem 20 we have

$$\lambda_j^{\downarrow}(U^*AU) \geq \lambda_{j+k}^{\downarrow}(A) = \alpha_{j+k} \text{ for } 1 \leq j \leq n - k,$$

where $\lambda_{j+k}^{\downarrow}(A) = \alpha_{j+k}$ follows from the fact that A is positive.

Therefore,

$$\operatorname{tr}(U^*AU)^2 = \sum_{j=1}^{n-k} [\lambda_j^{\downarrow}(U^*AU)]^2 \geq \sum_{j=1}^{n-k} \alpha_{j+k}^2 = \sum_{j=k+1}^{n} \alpha_j^2. \quad \square$$

Lemma 31. *Let B be Hermitian with eigenvalues $\mu_1 \geq \cdots \geq \mu_n$. Further, let the eigenvalues of B^2 be $\beta_1^2 \geq \cdots \geq \beta_n^2$. Then for any k-dimensional compression U^*BU of B with $k < \frac{n}{2}$, we have*

$$\operatorname{tr}(U^*BU)^2 \geq \sum_{j=2k+1}^{n} \beta_j^2.$$

Proof of lemma. Let $\lambda_1 \geq \cdots \geq \lambda_{n-k}$ be the eigenvalues of U^*BU. Then theorem 20 says

$$\mu_j \geq \lambda_j \geq \mu_{j+k} \text{ for } 1 \leq j \leq n - k.$$

The conditions of lemma 28 are satisfied, which says that there are indices $\{i_1, \ldots, i_{n-2k}\} \subset \{1, \ldots, n-k\}$ such that

$$\lambda_{i_s}^2 \geq \mu_{s+k}^2 \text{ for } 1 \leq s \leq n - 2k.$$

So,

$$\begin{aligned}\operatorname{tr}(U^* BU)^2 &= \sum_{j=1}^{n-k} \lambda_j^2 \;\geq\; \sum_{s=1}^{n-2k} \lambda_{i_s}^2 \;\geq\; \sum_{s=1}^{n-2k} \mu_{s+k}^2\\ &= \sum_{j=k+1}^{n-k} \mu_j^2 \;\geq\; \sum_{j=2k+1}^{n} \beta_j^2 .\end{aligned}$$

Where the last inequality follows from the containment $\{\mu_{k+1}^2, \ldots, \mu_{n-k}^2\} \subset \{\beta_1^2, \ldots, \beta_n^2\}$ and the fact that β_j^2 are in decreasing order. □

Note. The equality $\lambda_{j+k}^{\downarrow}(A) = \alpha_{j+k}$ is a crucial part of the proof. This allows the positive matrix A to appear without a constant in the main inequality. The matrix B does not enjoy this privilege, and hence this case has to be argued more carefully. When positivity is dropped, we need pairing.

In particular, let $\lambda_1, \ldots, \lambda_{n-k}$ be the eigenvalues of U^*AU in decreasing order. Then the eigenvalues of $(U^* AU)^2$ are $\lambda_1^2, \ldots, \lambda_{n-k}^2$. And they satisfying the following interlacing equations.

$$\alpha_j \geq \lambda_j \geq \alpha_{j+k}.$$

If $A \geq 0$ then all the above terms are positive. And hence

$$\lambda_j^2 \geq \alpha_{j+k}^2 .$$

This ensures the good behavior. We cannot do this generally. However, lemma 28 comes to be crucial here which shows that we can still recover an $(n-2k)$ size subset of $\{\lambda_j\}$ such that

$$\lambda_{i_j}^2 \geq \alpha_{j+k}^2 .$$

9 Main results

The results in this section were first introduced and proved in the papers [2, 3, 4].

9.1 Norm inequalities in different cases

We present the main results in two parts. The first part posits the majorisation relations enjoyed by the singular values of a matrix with the eigenvalues of its real and imaginary parts under three different circumtances. This immediately leads to the p-norm inequalities which we are in search of.

Theorem 32. *Let $T = A + iB$, and say $s^2 = (s_1^2, \ldots, s_n^2)$, $\alpha^2 = (\alpha_1^2, \ldots, \alpha_n^2)$ and $\beta^2 = (\beta_1^2, \ldots, \beta_n^2)$ be the eigenvalues of T^*T, A and B respectively arranged in decreasing order.*

Let $\gamma^2 = \boldsymbol{p}(\beta^2)$ and $\delta^2 = \boldsymbol{p}(\alpha^2)$. Then the following majorisations hold.

i. $s^2 \prec \delta^2 + \gamma^2$.

ii. If A is positive then $s^2 \prec \alpha^2 + \gamma^2$.

iii. If both A and B are positive then $s^2 \prec \alpha^2 + \beta^2$.

Proof of theorem. The equation (2) states that $\|T\|_2^2 = \|A\|_2^2 + \|B\|_2^2$. By definition this precisely means

$$\sum_{j=1}^{n} s_j^2 = \sum_{j=1}^{n} \alpha_j^2 + \sum_{j=1}^{n} \beta_j^2 = \sum_{j=1}^{n} \alpha_j^2 + \sum_{j=1}^{n} \gamma_j^2 = \sum_{j=1}^{n} \delta_j^2 + \sum_{j=1}^{n} \gamma_j^2, \tag{18}$$

where we see that

$$\sum_{j=1}^{n} \beta_j^2 = \sum_{j=1}^{n} \gamma_j^2 \text{ and } \sum_{j=1}^{n} \alpha_j^2 = \sum_{j=1}^{n} \delta_j^2 \tag{19}$$

by definition of the γ^2 and δ^2.

This fulfils the one of inequalities required for each majorisation in *(i)*, *(ii)* and *(iii)*. Now, from spectral theorem we know that there exists orthonormal vectors $u_1, \ldots, u_n$ such that

$$(T^*T)u_j = s_j^2\, u_j\,.$$

Collecting the bottom $n-k$ eigenvectors to into the matrix $U = [u_{k+1}, \ldots, u_n]$ we get

$$U^*(T^*T)\,U = \begin{bmatrix} s_{k+1}^2 & & \\ & \ddots & \\ & & s_n^2 \end{bmatrix}.$$

So this leaves us

$$\sum_{j=k+1}^{n} s_j^2 = \operatorname{tr} U^*(T^*T)\,U \quad \text{for the above choice of } U.$$

From here, lemma 29 tells us that

$$\operatorname{tr} U^*T^*TU \geq \operatorname{tr}(U^*AU)^2 + \operatorname{tr}(U^*BU)^2. \tag{20}$$

We now show *(i)*, *(ii)* and *(iii)* one by one.

For *(iii)* Note that both A and B are positive. So by lemma 30 we have

$$\operatorname{tr}(U^*AU)^2 \geq \sum_{j=k+1}^{n} \alpha_j^2 \quad \text{and} \quad \operatorname{tr}(U^*BU)^2 \geq \sum_{j=k+1}^{n} \beta_j^2.$$

Combining with (20) gives us

$$\sum_{j=k+1}^{n} s_j^2 \;\geq\; \sum_{j=k+1}^{n} \alpha_j^2 \;+\; \sum_{j=k+1}^{n} \beta_j^2\,.$$

This along with (18) completes the proof of *(iii)*.

For *(ii)* Here, only A is positive. So by lemma 30 we have

$$\operatorname{tr}(U^*AU)^2 \geq \sum_{j=k+1}^{n} \alpha_j^2. \tag{21}$$

Now consider the case when $k<\frac{n}{2}$. Then lemma 31 tells us that

$$\operatorname{tr}(U^* BU)^2 \geq \sum_{j=2k+1}^{n} \beta_j^2.$$

Using (20) and (21) we get

$$\begin{aligned}
\sum_{j=k+1}^{n} s_j^2 &\geq \sum_{j=k+1}^{n} \alpha_j^2 + \sum_{j=2k+1}^{n} \beta_j^2 \\
&= \sum_{j=k+1}^{n} \alpha_j^2 + \sum_{j=1}^{n} \beta_j^2 - \sum_{j=1}^{2k} \beta_j^2 \\
&= \sum_{j=k+1}^{n} \alpha_j^2 + \sum_{j=1}^{n} \gamma_j^2 - \sum_{j=1}^{k} \gamma_j^2 \\
&= \sum_{j=k+1}^{n} \alpha_j^2 + \sum_{j=k+1}^{n} \gamma_j^2 .
\end{aligned}$$

When $k\geq\frac{n}{2}$, then from (20) and (21), noting that $\operatorname{tr}(U^*BU)^2\geq 0$ (as $U^* BU$ is Hermitian), we get

$$\begin{aligned}
\sum_{j=k+1}^{n} s_j^2 &= \operatorname{tr} U^* T^* TU \geq \operatorname{tr}(U^* AU)^2 \geq \sum_{j=k+1}^{n} \alpha_j^2 \\
&= \sum_{j=k+1}^{n} \alpha_j^2 + \sum_{j=k+1}^{n} \gamma_j^2
\end{aligned}$$

where the last equality comes from the fact that $\gamma_j=0$ for $j\geq\frac{n}{2}+1$. Combinig the cases when $k\leq\frac{n}{2}$ and when $k>\frac{n}{2}$, along with (18) gives us the majorisation of *(ii)*.

For *(i)* In this case when $k<\frac{n}{2}$, lemma 31 tells us that

$$\operatorname{tr}(U^* AU)^2 \geq \sum_{j=2k+1}^{n} \alpha_j^2 \quad \text{and} \quad \operatorname{tr}(U^* BU)^2 \geq \sum_{j=2k+1}^{n} \beta_j^2.$$

Equation (20) gives us

$$\begin{aligned}
\sum_{j=k+1}^{n} s_j^2 &\geq \sum_{j=2k+1}^{n} \alpha_j^2 + \sum_{j=2k+1}^{n} \beta_j^2 \\
&= \sum_{j=1}^{n} \delta_j^2 - \sum_{j=1}^{k} \delta_j^2 + \sum_{j=1}^{n} \gamma_j^2 - \sum_{j=1}^{k} \gamma_j^2 \\
&= \sum_{j=k+1}^{n} \delta_j^2 + \sum_{j=k+1}^{n} \gamma_j^2 .
\end{aligned}$$

Finally, when $k \geq \frac{n}{2}$ then $\delta_j = \gamma_j = 0$ for all $j \geq k+1$. So the inequality

$$\sum_{j=k+1}^{n} s_j^2 \;\geq\; 0 \;=\; \sum_{j=k+1}^{n} \delta_j^2 + \sum_{j=k+1}^{n} \gamma_j^2$$

is trivial. Both these cases combined with (18) completes the majorisation in *(i)*. $\square$

Finally, we are set to present best known norm inequalities related to the Cartesian decomposition of a matrix. Theorem 33 displays the power that theorem 32 holds.

Theorem 33. *Let $T = A + iB$. Then for $2 \leq p \leq \infty$*

i. In general $\|T\|_p^2 \leq 2^{1-\frac{2}{p}}(\|A\|_p^2 + \|B\|_p^2)$.

ii. If A is positive, then $\|T\|_p^2 \leq \|A\|_p^2 + 2^{1-\frac{2}{p}}\|B\|_p^2$.

iii. If A and B are both positive, then $\|T\|_p^2 \leq \|A\|_p^2 + \|B\|_p^2$.

All inequalities are reversed for $1 \leq p \leq 2$.

Proof of theorem. We prove the three parts separately.

For *(iii)* In this case, theorem 32 part *(iii)* tells us that $s^2 \prec \alpha^2 + \beta^2$. We use lemma 14 on the above majorisation to get

$$\|s\|_p^2 \leq \|\alpha\|_p^2 + \|\beta\|_p^2 \quad \text{for } 2 \leq p \leq \infty.$$

But as the set s, α and β are singular values of T, A and B, the above precisely means that

$$\|T\|_p^2 \leq \|A\|_p^2 + \|B\|_p^2.$$

For *(ii)* In this case, theorem 32 part *(ii)* says $s^2 \prec \alpha^2 + \gamma^2$. Then, for $2 \leq p \leq \infty$,

$$\begin{aligned} s^2 \prec \alpha^2 + \gamma^2 \implies & \|s\|_p^2 \leq \|\alpha\|_p^2 + \|\gamma\|_p^2 && \because \text{lemma 14} \\ \implies & \|s\|_p^2 \leq \|\alpha\|_p^2 + 2^{1-\frac{2}{p}}\|\beta\|_p^2 && \because \text{lemma 17} \end{aligned}.$$

Now we note that the values $|\beta_1|, \ldots, |\beta_n|$ are singular values of B. Therefore, the above inequality precisely means

$$\|T\|_p^2 \leq \|A\|_p^2 + 2^{1-\frac{2}{p}}\|B\|_p^2.$$

For *(iii)* Similarly, here theorem 32 part *(i)* gives $s^2 \prec \delta^2 + \gamma^2$. Then again, for $2 \le p \le \infty$, we can argue

$$\begin{aligned} s^2 \prec \delta^2 + \gamma^2 &\implies \|s\|_p^2 \le \|\delta\|_p^2 + \|\gamma\|_p^2 && \because \text{lemma 14} \\ &\implies \|s\|_p^2 \le 2^{1-\frac{2}{p}} \|\alpha\|_p^2 + 2^{1-\frac{2}{p}} \|\beta\|_p^2 && \because \text{lemma 17} \end{aligned}.$$

Now $|\alpha_j|$ and $|\beta_j|$ are singular values of A and B, and thus we recover the required inequality

$$\|T\|_p^2 \le 2^{1-\frac{2}{p}}(\|A\|_p^2 + \|B\|_p^2).$$

When $1 \le p \le \infty$, the inequalities in lemma 14 and lemma 17 are reversed, and hence we get the result stated in the theorem.

Remark 34. Note that when $T = A + iB$ with only B positive and A Hermitian, we do the following. We write $-iT = B + i(-A)$. Applying theorem 33 part *(ii)* for $2 \le p \le \infty$ to $-iT$ we get

$$\begin{aligned} \|-iT\|_p^2 &\le \|B\|_p^2 + 2^{1-\frac{2}{p}} \|-A\|_p^2 \\ \implies \quad \|T\|_p^2 &\le 2^{1-\frac{2}{p}} \|A\|_p^2 + \|B\|_p^2 \end{aligned}$$

The inequalities reverse for $1 \le p \le 2$. $\square$

9.2 A discussion on sharpness and raising a question

Section **9.1** gave us the inequalities summarised in the following table for the Cartesian decomposition of $T = A + iB$.

	$A \ge 0$?	$B \ge 0$?	Inequality	Sharpness
			When $2 \le p \le \infty$	
$(e1)$	✓	✓	$\|T\|_p^2 \le \|A\|_p^2 + \|B\|_p^2$	✓
$(e2)$	×	×	$\|T\|_p^2 \le 2^{1-\frac{2}{p}}(\|A\|_p^2 + \|B\|_p^2)$	✓
$(e3)$	✓	×	$\|T\|_p^2 \le \|A\|_p^2 + 2^{1-\frac{2}{p}} \|B\|_p^2$	✓
			When $1 \le p \le 2$	
$(e4)$	✓	✓	$\|T\|_p^2 \ge \|A\|_p^2 + \|B\|_p^2$	✓
$(e5)$	×	×	$\|T\|_p^2 \ge 2^{1-\frac{2}{p}}(\|A\|_p^2 + \|B\|_p^2)$	✓
$(e6)$	✓	×	$\|T\|_p^2 \ge \|A\|_p^2 + 2^{1-\frac{2}{p}} \|B\|_p^2$	×

Table 1.

We will now see nontrivial matrices T which attain sharpness for the inequalities (e1)-(e5).

Consider the following T for which A and B are both positive.

$$\underbrace{\begin{bmatrix} 1+i & 0 \\ 0 & 0 \end{bmatrix}}_{T} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{A} + i \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{B}.$$

Then $\|T\|_p^2 = |1+i|^2 = 2$, and $\|A\|_p^2 + \|B\|_p^2 = 1+1 = 2$, for all $1 \leq p \leq \infty$. This shows the sharpness of the inequalities (e1) and (e4).

Consider the following T now for which neither A nor B is positive.

$$\underbrace{\begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}}_{T} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}_{A} + i \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{B}.$$

The singular values of T, A and B are the tuples $(2,0), (1,1)$ and $(1,1)$. Then it is easy verify that the right-hand side (2.) and (5.) equals

$$2^{1-\frac{2}{p}} (\|(1,1)\|_p^2 + \|(1,1)\|_p^2) \;=\; 2^{1-\frac{2}{p}} \cdot 2 \cdot 2^{\frac{2}{p}} \;=\; 4 \;=\; \|(2,0)\|_p^2$$

which is the left-hand side of (e2) and (e5). This shows that the inequalities in (e2) and (e5) are also sharp.

Now construct the following matrix $T = T(t)$ such that A is positive and B is not.

$$\underbrace{\begin{bmatrix} 1 & it \\ it & 0 \end{bmatrix}}_{T} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{A} + i \underbrace{\begin{bmatrix} 0 & t \\ t & 0 \end{bmatrix}}_{B}.$$

Note that for $t \neq 0$ the matrix B is not positive. The highest singular value s of T is given by

$$s^2 = \|T\|_\infty^2 = \frac{(1+2t^2) + \sqrt{1+4t^2}}{2},$$

whereas the singular values of A and B are $(1,0)$ and (t,t). Then the right-hand side of (e3) equals

$$1^2 + 2^{1-\frac{2}{p}} (t^p + t^p)^{\frac{2}{p}} = 1 + 2t^2.$$

Using (e3) we write

$$s^2 = \|T\|_\infty^2 \leq \|T\|_p^2 \leq 1 + 2t^2 \quad \text{for all } 2 \leq p \leq \infty.$$

But $1+2t^2-s^2=\frac{(1+2t^2)-\sqrt{1+4t^2}}{2}$ can be chosen to be arbitrarily small by choosing t small. This shows that for any ε there exists $t\neq 0$, and hence matrices $T=T(t)$, for which

$$1+2t^2-\varepsilon\leq\|T\|_p^2\leq 1+2t^2.$$

This shows that the constant appearing in the right-hand side of (e3) is the best possible one.

In light of all this it is rather surprising that (e6) is not sharp. To illustrate this we take the case when $p=1$. The inequality (e6), in this case, says

$$\|T\|_1^2\geq\|A\|_1^2+\frac{1}{2}\|B\|_1^2. \tag{22}$$

However, we prove the following theorem which shows that the constant $\frac{1}{2}$ appearing in the right-hand side of (22) can be improved to 1.

Theorem 35. *Let $T=A+iB$ where A is positive. Then*

$$\|T\|_1^2\geq\|A\|_1^2+\|B\|_1^2.$$

Proof of theorem. As $\|\cdot\|_1$ is unitarily invariant, we can rename T,A and B as U^*TU, U^*AU and U^*BU. For a suitable matrix U, this can lead us to assume B to be diagonal, such that $T=A+iB$ holds and the norms are unchanged. So let us suppose B is a diagonal matrix.

Let $s_1,\ldots,s_n$ be the singular values of T, and $\alpha_1,\ldots,\alpha_n$ and $|\beta_1|,\ldots,|\beta_n|$ be the eigenvalues of A and B respectively.

Then equation (2) says

$$s_1^2+\cdots+s_n^2=\alpha_1^2+\cdots+\alpha_n^2+\beta_1^2+\cdots+\beta_n^2. \tag{23}$$

Theorem 35, in this case, in light of (23), requires us to show

$$\begin{aligned}
&(s_1+\cdots+s_n)^2 &\geq\ &(\alpha_1+\cdots+\alpha_n)^2+(|\beta_1|+\cdots+|\beta_n|)^2\\
\Longleftrightarrow\quad &s_1^2+\cdots+s_n^2+2\sum_{i<j}s_i\,s_j &\geq\ &\alpha_1^2+\cdots+\alpha_n^2+2\sum_{i<j}\alpha_i\,\alpha_j+\\
&&&|\beta_1|^2+\cdots+|\beta_n|^2+2\sum_{i<j}|\beta_i|\,|\beta_j|\\
\Longleftrightarrow\quad &\sum_{i<j}s_i\,s_j &\geq\ &\sum_{i<j}\alpha_i\,\alpha_j+\sum_{i<j}|\beta_i|\,|\beta_j| \qquad (24)
\end{aligned}$$

First, consider $n=2$. Then (24) is reread as

$$|\det T|\ \geq\ \det A+|\det B|\ . \tag{25}$$

Now suppose $A = \begin{bmatrix} a & c \\ \bar{c} & b \end{bmatrix}$ and $B = \begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix}$.

Then, showing (25) requires showing (after renaming $|c|$ as c)

$$\begin{aligned} |(a+is)(b+it) - c^2| &\geq ab - c^2 + |st| \\ \Longrightarrow |ab - st - c^2 + i(at+bs)| &\geq ab - c^2 + |st|. \end{aligned}$$

If $st \leq 0$ then the right-hand side is $ab - st - c^2$, which is clearly smaller than the left-hand side. Let $st > 0$. Then

$$\begin{aligned} (ab - c^2 + st)^2 - (ab - c^2 - st)^2 &= 4(ab - c^2)st \\ &\leq 4abst \qquad \because st > 0 \\ &\leq (at+bs)^2 \qquad \because \text{AM-GM inequality}. \end{aligned}$$

In this case therefore,

$$\begin{aligned} |ab - st - c^2 + i(at+bs)|^2 &= (ab - st - c^2) + (at+bs)^2 \\ &\geq (ab - st - c^2) + (ab - c^2 + st)^2 - (ab - c^2 - st)^2 \\ &= (ab - c^2 + st)^2. \end{aligned}$$

This proves (24) when T is 2-dimensional.

For general n, we proceed with tools from section **7**. Let us consider the operator $\Lambda^2 T, \Lambda^2 A$ and $\Lambda^2 B$. Then theorem states that they have $s_i s_j, \alpha_i \alpha_j$ and $|\beta_i|\,|\beta_j|$ for $1 \leq i < j \leq n$ as their respective singular values. From this (24) is restated as

$$\|\Lambda^2 T\|_1 \geq \|\Lambda^2 A\|_1 + \|\Lambda^2 B\|_1.$$

But, lemma 24 from section **7** tells us that for any matrix $n \times n$ X, the matrix of $\Lambda^2 X$ is $\binom{n}{2} \times \binom{n}{2}$ with diagonal entries enumerated as

$$\mathcal{C}_{ij}(X) = \det M_{ij}(X) = \begin{vmatrix} X_{ii} & X_{ij} \\ X_{ji} & X_{jj} \end{vmatrix} \text{ for } 1 \leq i < j \leq n.$$

So, using result 5 on $\Lambda^2 T$ we get

$$\|\Lambda^2 T\|_1 \geq \sum_{i<j} |\mathcal{C}_{ij}(T)|. \tag{26}$$

Now, lemma result 27 tells us that $M_{ij}(T)$ has cartesian decomposition

$$M_{ij}(T) = M_{ij}(A) + i\,M_{ij}(B).$$

As these are 2×2 matrices, using (25) on the above gives us

$$\begin{aligned} |\det \mathcal{M}_{ij}(T)| &\geq \det \mathcal{M}_{ij}(A) + |\det \mathcal{M}_{ij}(B)| \\ \Longrightarrow \qquad |\mathcal{C}_{ij}(T)| &\geq \mathcal{C}_{ij}(A) + |\mathcal{C}_{ij}(B)|. \end{aligned} \tag{27}$$

But again as B is Hermitian and diagonal, so is $\Lambda^2 B$ (lemma 24). Therefore $|\mathcal{C}_{ij}(B)|$ are the diagonal entries of the positive matrix $|\Lambda^2 B|$. This tells us by result 4,

$$\|\Lambda^2 B\|_1 = \operatorname{tr}|\Lambda^2 B| = \sum_{i<j} |\mathcal{C}_{ij}(B)|. \tag{28}$$

Similarly, as A is positive, so is $\Lambda^2 A$. And again,

$$\|\Lambda^2 A\|_1 = \operatorname{tr}\Lambda^2 A = \sum_{i<j} \mathcal{C}_{ij}(A). \tag{29}$$

One now uses (26), (27), (28) and (29) to conclude that

$$\begin{aligned} \|\Lambda^2 T\|_1 \;\geq\; \sum_{i<j} |\mathcal{C}_{ij}(T)| \;&\geq\; \sum_{i<j} \mathcal{C}_{ij}(A) + \sum_{i<j} |\mathcal{C}_{ij}(B)| \\ &=\; \|\Lambda^2 A\|_1 + \|\Lambda^2 B\|_1 . \end{aligned}$$

This concludes the proof of the theorem. □

Remark 36. The above inequality is (obviously) sharp. Any skew symmetric matrix T will have a Cartesian decomposition $T = \mathbf{0} + iB$. Then $\|T\|_p = \|B\|_p$, and therefore the right-hand side and the left-hand side are equal.

We see the well behavior of positive matrices through the inequalities (e1) and (e4). General matrices which do not display such conditioning like of positive matrices appear with an adjustment factor of magnitude $2^{1-2/p}$. This same constant appearing in (e2), (e3), (e5) and (e6) is quite elegant and satisfying. Therefore, the anomaly of sharpness in (e6) is a surprise.

For p lying in the interval $[1,2]$, let the best constant for (e6) be C_p. That is, say the inequality (for T with positive real part)

$$\|T\|_p^2 \geq \|A\|_p^2 + C_p \|B\|_p^2 \quad \text{for } 1 \leq p \leq 2$$

be sharp. Then (e6) tells us $C_p \geq 2^{1-2/p}$. The inequality (e6) tells us that $C_1 = 1$, and (2) tells us that $C_2 = 1$. Now note that $2^{1-2/p}$ is a monotone function of p. That is p-norms pertaining to higher values of p require a higher adjustment factor. So, is C_p also a monontone function of p? If it were so, as $C_1 = C_2 = 1$, this would demand that $C_p = 1$ for all $p \in (1,2)$. So, can the adjustment factor $2^{1-2/p}$ in (e6) be improved to 1? That is

$$\|T\|_p^2 \geq_{?} \|A\|_p^2 + \|B\|_p^2 \quad \text{for } 1 \leq p \leq 2.$$

This is the question Bhatia and Zhan raised in [3]. We find a negative answer to this question in the coming section. It shows us that the interval $(1,2)$, for matrices with positive real part, does not enjoy a monotone optimum constant unlike all the other cases.

9.3 Answering the question

Theorem 37. *There exists a 2×2 matrix T with a positive real part such that*

$$\|T\|_p^2 < \|A\|_p^2 + \|B\|_p^2 \text{ for all } 1<p<2.$$

Proof of theorem. Consider the matrix T with the following Cartesian decomposition.

$$\underbrace{\begin{bmatrix} 1+i & -1 \\ -1 & 1-i \end{bmatrix}}_{T} = \underbrace{\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}}_{A} + i \underbrace{\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}}_{B}.$$

One can calculate the singular values of A to be 2 and 0, and of B to be 1 and 1. We further note that the characteristic polynomial of T^*T is $(x-3)^2-8$, and therefore, the singular values of T are $\sqrt{2}+1$ and $\sqrt{2}-1$. We denote these two values as a and ${}^1/_a$. Thus, we are required to show that for $1<p<2$,

$$\left\|\begin{pmatrix} a \\ {}^1/_a \end{pmatrix}\right\|_p^2 < \left\|\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right\|_p^2 + \left\|\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\|_p^2 = 2^2 + 2^{\frac{2}{p}}$$

or equivalently,

$$\left\|\begin{pmatrix} a \\ {}^1/_a \end{pmatrix}\right\|_p < \sqrt{2^2+2^{\frac{2}{p}}} = \left\|\begin{pmatrix} 2 \\ 2^{\frac{1}{p}} \end{pmatrix}\right\|_2. \tag{30}$$

Now, we use the following inequality for the equivalence of the p-norm and the 2-norm from lemma result 6,

$$\left\|\begin{pmatrix} 2 \\ 2^{\frac{1}{p}} \end{pmatrix}\right\|_2 \geq 2^{\frac{1}{2}-\frac{1}{p}} \left\|\begin{pmatrix} 2 \\ 2^{\frac{1}{p}} \end{pmatrix}\right\|_p.$$

Using the above, to show the main inequality in (30) it is sufficient to show that

$$\begin{aligned} & \left\|\begin{pmatrix} a \\ {}^1/_a \end{pmatrix}\right\|_p < 2^{\frac{1}{2}-\frac{1}{p}} \left\|\begin{pmatrix} 2 \\ 2^{\frac{1}{p}} \end{pmatrix}\right\|_p \\ \iff & \left\|\begin{pmatrix} a \\ {}^1/_a \end{pmatrix}\right\|_p^p < 2^{\frac{p}{2}-1} \left\|\begin{pmatrix} 2 \\ 2^{\frac{1}{p}} \end{pmatrix}\right\|_p^p \end{aligned}$$

or equivalently, we need to show that

$$a^p + \frac{1}{a^p} < 2^{\frac{3p}{2}-1} + 2^{\frac{p}{2}} \text{ for } 1<p<2. \tag{31}$$

We first note that at $p=1$ and $p=2$, both the RHS and the LHS are equal, taking values $2\sqrt{2}$ and 6 respectively. Thus, LHS − RHS as a function of p equals 0 at $p=1$ and $p=2$. If we can show that LHS − RHS is convex in the interval $(1,2)$, then we will have the required inequality in (31).

So we first obtain the second derivative of LHS − RHS as.

$$\begin{aligned}
& a^p(\ln a)^2+\frac{1}{a^p}\left(\ln\frac{1}{a}\right)^2-\left(\frac{3}{2}\right)^2 2^{\frac{3p}{2}-1}(\ln 2)^2-\left(\frac{1}{2}\right)^2 2^{\frac{p}{2}}(\ln 2)^2 && (32)\\
=\; & (\ln 2)^2\left[\left(a^p+\frac{1}{a^p}\right)\left(\frac{\ln a}{\ln 2}\right)^2-\frac{1}{8}\left(9\cdot 8^{\frac{p}{2}}+2\cdot 2^{\frac{p}{2}}\right)\right]\\
=\; & (\ln 2)^2\left[\left(a^p+\frac{1}{a^p}\right)\log_2^2 a-\frac{11}{8}\left(\frac{9}{11}\cdot 8^{\frac{p}{2}}+\frac{2}{11}\cdot 2^{\frac{p}{2}}\right)\right]\\
>\; & (\ln 2)^2\left[\left(a^p+\frac{1}{a^p}\right)\log_2^2 a-\frac{11}{8}\left(\frac{76}{11}\right)^{\frac{p}{2}}\right]\\
=\; & \frac{11}{8}(\ln 2)^2\left(a^p+\frac{1}{a^p}\right)\left[\frac{8}{11}\log_2^2 a-\frac{\left(\frac{76}{11}\right)^{\frac{p}{2}}}{\left(a^p+\frac{1}{a^p}\right)}\right] && (33)
\end{aligned}$$

Note that when $p\in(1,2)$, the function $x^{\frac{p}{2}}$ is concave for $x>0$. Therefore, the following inequality

$$\frac{9}{11}\cdot 8^{\frac{p}{2}}+\frac{2}{11}\cdot 2^{\frac{p}{2}} \;<\; \left(\frac{9}{11}\cdot 8+\frac{2}{11}\cdot 2\right)^{\frac{p}{2}} \;=\; \left(\frac{76}{11}\right)^{\frac{p}{2}}$$

is used in the penultimate step above.

Now, to show that (32) is positive, it is sufficient to show that the function in (33) is positive, which happens if and only if for $p\in(1,2)$

$$\frac{8}{11}\log_2^2 a \;>\; \frac{\sqrt{\frac{76}{11}}^{\,p}}{a^p+\frac{1}{a^p}}\,. \tag{34}$$

Now as $\sqrt{76/11}$ is strictly greater than both a and $\frac{1}{a}$, the reciprocal of the RHS of (34) is strictly decreasing, and hence the RHS strictly increasing. Therefore, the RHS is upper bounded by its value at $p=2$ in the interval $(1,2)$. Or, to show (34) it is sufficient to show that

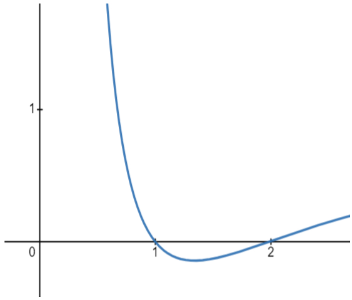
$$\frac{8}{11}\log_2^2 a>\frac{\frac{76}{11}}{\left(a^2+\frac{1}{a^2}\right)}=\frac{76}{66} \quad \text{or equivalently} \quad \log_2 a>\sqrt{\frac{19}{12}}\,.$$

One can easily check the last inequality using rational approximations. This completes the required proof. □

10 Conclusion

The optimum constant in (e6) is yet to be found. However, we know now that the optimum constant for $p \in (1, 2)$ is not monotone in p (and hence not 1). This in itself is a surprising result.

Even considering the difference LHS − RHS of the above example as a function of p, one finds the following relationship.



One now expects the the optimum possible constants to have convex properties in $(1, 2)$. While both cases (e3) and (e6) follow from theorem 33, it is intriguing that tightness is compromised in one and not the other. One still has to find out why the elegant pattern of the inequalities in (e1) to (e5) is broken in (e6). Perhaps something more elegant is lying underneath!

References

[1] R. Bhatia, Matrix Analysis, Springer, Berlin, 1997.

[2] R. Bhatia, F. Kittaneh, Cartesian decompositions and Schatten norms, Linear Algebra Appl. 318(2000), 109–116.

[3] R. Bhatia, X. Zhan, Norm inequalities for operators with positive real part, J. Operator Theory 50(2003), 67–76.

[4] R. Bhatia, X. Zhan, Compact operators whose real and imaginary parts are positive, Proc. Amer. Math. Soc. 129(2001), 2277–2281.

[4] A.W. Marshall, I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, 1979.

[5] B. Simon, Trace Ideals and their Applications, Cambridge Univ. Press, Cambridge– New York, 1979.

11 Proofs of used results from linear algebra

Proof of result 1.

If T is normal then by equating $TT^* = T^*T$ we get

$$\begin{aligned} & (A - iB)(A + iB) &&= (A + iB)(A - iB) \\ \Longrightarrow\ & A^2 + B^2 + i(AB - BA) &&= A^2 + B^2 + i(BA - AB) \\ \Longrightarrow\ & AB - BA &&= BA - AB. \end{aligned} \qquad \square$$

Proof of result 2.

For any x we have

$$\|x\|^2 = \|Px + (I - P)x\|^2 = \|Px\|^2 + \|(I - P)x\|^2$$

where the last equality is the Pythagoras theorem. This shows that

$$\|Px\|^2 \leq \|x\|^2.$$

But,

$$\|Px\|^2 = \langle Px, Px\rangle = \langle x, P^* Px\rangle = \langle x, P^2 x\rangle = \langle x, Px\rangle$$

So for all x we have,

$$\|Px\|^2 \leq \|x\|^2 \Longrightarrow \langle x, Px\rangle \leq \langle x, x\rangle \Longrightarrow \langle x, (I - P)x\rangle \geq 0.$$

This shows that $I - P \geq 0$ or $I \geq P$. $\square$

Proof of result 3.

If $A \leq B$ then $B - A \geq 0$. And hence $B - A$ has positive eigenvalues. Therefore the sum of its eigenvalues $\operatorname{tr}(B - A) \geq 0$. This shows $\operatorname{tr} A \leq \operatorname{tr} B$. $\square$

Proof of result 4.

By the spectral theorem write $A = UDU^*$, where diagonals of A are positive real numbers. By unitary invariance of trace we have $\operatorname{tr} A = \operatorname{tr} D$. But again, by unitary invariance of $\|\cdot\|_1$ we have $\|A\|_1 = \|D\|_1 = \operatorname{tr} D$ as D is a positive diagonal matrix. This proves the required result. $\square$

Proof of result 5.

Let $U = \begin{bmatrix} 1 & & & \\ & \omega & & \\ & & \ddots & \\ & & & \omega^{n-1} \end{bmatrix}$ where $\omega = e^{\frac{2\pi i}{n}}$ be the n-th root of unity.

Let us calculate the (i, j)-th entry of the matrix D given by

$$D = \frac{1}{n} \sum_{r=0}^{n-1} U^r A (U^*)^r.$$

So,

$$\begin{aligned}
D_{ij} &= \frac{1}{n}\sum_{r=0}^{n-1} [U^r A (U^*)^r]_{ij} \\
&= \frac{1}{n}\sum_{r=0}^{n-1} \omega^{r(i-j)} a_{ij} \\
&= \frac{a_{ij}}{n}\sum_{r=0}^{n-1} \omega^{r(i-j)} \\
&= \begin{cases} 0 & \text{if } i \neq j \quad \text{(follows from properties of } \omega) \\ a_{ii} & \text{if } i = j \end{cases} .
\end{aligned}$$

This shows that the matrix D is a diagonal matrix whose diagonal entries are the diagonal entries of A. So,

$$\begin{aligned}
\sum_{j=1}^{n} |a_{jj}| = \|D\|_1 &= \left\| \frac{1}{n}\sum_{r=0}^{n-1} U^r A (U^*)^r \right\|_1 \\
&\leq \frac{1}{n}\sum_{r=0}^{n-1} \|U^r A (U^*)^r\|_1 \\
&= \frac{1}{n}\sum_{r=0}^{n-1} \| A \|_1 \qquad \because \text{Unitary invariance} \\
&= \|A\|_1 .
\end{aligned}$$

This completes the proof of the result.

Proof of result 6.

Let us first observe the following. Let $x = (x_1, \ldots, x_n)$. Say $x_1, \ldots, x_n \geq 0$ and $r \geq 1$. Then we have,

$$\begin{aligned}
(x_1 + \cdots + x_n)^r &= \sum_{\substack{1 \leq i_j \leq n \\ 1 \leq j \leq r}} x_{i_1} x_{i_2} \cdots x_{i_r} \\
&\leq \sum_{\substack{1 \leq i_j \leq n \\ 1 \leq j \leq r}} \frac{x_{i_1}^r + \cdots + x_{i_r}^r}{r} \qquad \because \text{AM-GM inequality} \\
&= \sum_{j=1}^{r} \sum_{\substack{1 \leq i_k \leq n \\ k \neq j}} \sum_{i_j=1}^{n} \frac{x_{i_j}^r}{r} \qquad \because \text{Taking the inner sum outside} \\
&= \sum_{j=1}^{r} n^{r-1} \frac{x_1^r + \cdots + x_n^r}{r} \\
&= n^{r-1}(x_1^r + \cdots + x_n^r) .
\end{aligned}$$

We can assume the coordinates of x to be positive without loss of generality. So the above implies

$$\|x\|_1 \;\leq\; n^{1-\frac{1}{r}}\|x\|_r \text{ for all } x. \tag{35}$$

Now let $q \geq p$. Then $\frac{q}{p} \geq 1$. Consider $y = (x_1^p, \ldots, x_n^p)$. Then from N we have

$$\begin{aligned}
\|y\|_1 &\;\leq\; n^{1-\frac{p}{q}}\|y\|_{\frac{q}{p}}\\
\Longrightarrow\; (x_1^p + \cdots + x_n^p) &\;\leq\; n^{1-\frac{p}{q}}\left[(x_1^p)^{\frac{q}{p}} + \cdots + (x_n^p)^{\frac{q}{p}}\right]^{\frac{p}{q}}\\
&\text{take } p\text{-th root} \ldots\\
\Longrightarrow\; \|x\|_p &\;\leq\; n^{\frac{1}{p}-\frac{1}{q}}\|x\|_q.
\end{aligned}$$

This proves the required result. □

Proof of result 7.

Let A, B be doubly stochastic matrices. For $t \in [0, 1]$ let $C = (1-t)A + tB$ and $D = AB$. Then

$$\sum_j c_{ij} = \sum_j [(1-t)a_{ij} + t b_{ij}] = (1-t)\sum_j a_{ij} + t\sum_j b_{ij} = (1-t) + t = 1.$$

And again

$$\sum_j d_{ij} = \sum_j \sum_k a_{ik} b_{kj} = \sum_k a_{ik} \sum_j b_{kj} = 1.$$

So the i-th row of both C and D have sum 1, and their entries are clearly nonnegative. Now, with the similar argument conclude that the i-th row of C^T and D^T also have sum 1 (as A^T and B^T are also doubly stochastic).

This shows that all rows and columns of C and D have nonnegative entries that sum to 1. So both C and D are doubly stochastic. □

Proof of result 8.

One can quickly verify that for $\mathcal{I}, \mathcal{J} \subset \{1, \ldots, n\}$ of same size, we have $(T^*)_{\mathcal{I},\mathcal{J}} = (T_{\mathcal{J},\mathcal{I}})^*$. So, given $T = A + iB$ we have

$$\frac{T_{\mathcal{I},\mathcal{I}} + (T_{\mathcal{I},\mathcal{I}})^*}{2} \;=\; \frac{(T + T^*)_{\mathcal{I},\mathcal{I}}}{2} \;=\; A_{\mathcal{I},\mathcal{I}}\,.$$

Similarly, we have

$$\frac{T_{\mathcal{I},\mathcal{I}} - (T_{\mathcal{I},\mathcal{I}})^*}{2i} \;=\; \frac{(T - T^*)_{\mathcal{I},\mathcal{I}}}{2i} \;=\; B_{\mathcal{I},\mathcal{I}}\,.$$

This proves the required result. □