2013

Singular Value Inequalities: New Approaches to Conjectures

Peter Chilstrom
University of North Florida, n00446403@unf.edu

SINGULAR VALUE INEQUALITIES: NEW APPROACHES TO CONJECTURES

by

Peter Dylan Chilstrom

A thesis submitted to the Department of Mathematics and Statistics

in partial fulfillment of the requirements of the degree of

Master of Science in Mathematics

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF ARTS AND SCIENCES

August, 2013

Certificate of Approval

The thesis of Peter Dylan Chilstrom is approved: (Date)

Dr. Raluca Dumitru, Thesis Advisor

Dr. Scott Hochwald, Committee Member

Dr. Mei-Qin Zhan, Committee Member

Accepted for the Department of Mathematics and Statistics:

Dr. Scott Hochwald, Chair

Accepted for the College of Arts and Sciences:

Dr. Barbara Hetrick, Dean

Accepted for the University:

Dr. Len Roberson, Dean of the Graduate School

Acknowledgements

Firstly, I'd like to thank Dr. Raluca Dumitru for her tremendous insight and guidance throughout every step of this process. I'd also like to thank Dr. Scott Hochwald and Dr. Mei-Qin Zhan for taking the time to serve on my thesis committee. Finally, I would like to thank my graduate and undergraduate mathematics professors at the University of North Florida for the wonderful education I've received.

Contents

Abstract

Singular values have been found to be useful in the theory of unitarily invariant norms, as well as many modern computational algorithms. In examining singular value inequalities, it can be seen how these can be related to eigenvalues and how several algebraic inequalities can be preserved and written in an analogous singular value form. We examine the fundamental building blocks to the modern theory of singular value inequalities, such as positive matrices, matrix norms, block matrices, and singular value decomposition, then use these to examine new techniques being used to prove singular value inequalities, and also look at existing conjectures.

1 Introduction

Singular values are defined as the square roots of the eigenvalues of the Hermitian, positive semidefinite matrix A^TA (or AA^T) for some square matrix A. They are nonnegative real numbers with useful applications in statistics, functional analysis, and linear algebra. For example, the Ky Fan k-norm is the sum of the first k singular values. There also exists a wide abundance of inequalities involving singular values, which will be examined throughout this paper.

The original motivation which led to singular values was the question of whether two real bilinear forms were equivalent under independent real orthogonal substitutions. I.e. if we consider

$$\phi_A(x,y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j \text{ and } \phi_B(x,y) = \sum_{i,j=1}^{n} b_{ij}x_iy_j$$

where $A = [a_{ij}], B = [b_{ij}] \in M_n(\mathbb{R}), x = [x_i], y = [y_i] \in \mathbb{R}^n$, then we wish to know if there exists real orthogonal $Q_1, Q_2 \in M_n(\mathbb{R})$, such that $\phi_A(x,y) = \phi_B(Q_1x, Q_2y)$ for all $x, y \in \mathbb{R}^n$. In the late nineteenth century, Beltrami and Jordan [17] both independently answered this question with what we now call singular values. Beltrami found that for each A $\in M_n(\mathbb{R})$, there always exist singular values $\sigma_1(A), ..., \sigma_n(A)$ and real orthogonal $Q_1, Q_2 \in M_n(\mathbb{R})$ such that

$${Q_1}^T A Q_2 \equiv \Sigma = diag(\sigma_1(A), \sigma_2(A), ..., \sigma_n(A))$$

is a nonnegative diagonal matrix, where $\sigma_1(A)^2, ..., \sigma_n(A)^2$ are the eigenvalues of AA^T (also A^TA). Beltrami also found that the columns of Q_1 and Q_2 are eigenvectors of AA^T and A^TA, respectively (this is what we now call singular value decomposition). Subsequently, the diagonal bilinear form will provide a convenient canonical form by which any real bilinear form can be reduced by independent orthogonal substitutions, where the eigenvalues of AA^T are a complete set of invariants for this reduction.

1.1 Singular Value Decomposition

The exact development of singular value decomposition (SVD) was first made by L. Autonne. In a 1902 paper, Autonne showed that every nonsingular complex matrix $A \in M_n$ can be written as $A = UP$, where $U \in M_n$ is unitary and $P \in M_n$ is positive definite. Later (in 1915), he used the fact that A^*A and AA^* are similar to show that any square complex matrix $A \in M_n$ (singular or nonsingular) can be written as $A = Q_1 \Sigma Q_2^*$, where $Q_1, Q_2 \in M_n$ are unitary and Σ is a nonnegative diagonal matrix. Autonne also realized that the unitary factors Q_1 and Q_2 could be chosen to be real orthogonal if A is real, thus obtaining the same result as Jordan and Beltrami.

SVD, in particular, has had numerous applications in mathematics and statistics. More precisely, we can define SVD as follows:

Theorem 1.1. (Singular Value Decomposition) *Let $A \in M_{m,n}$ be given, and let $q = min(m,n)$. There exists a matrix $\Sigma = [\sigma_{ij}] \in M_{m,n}$ with $\sigma_{ij} = 0$ for all $i \neq j$, and $\sigma_{11} \geq \sigma_{22} \geq \ldots \geq \sigma_{qq} \geq 0$, and there are two unitary matrices $Q_1 \in M_m$ and $Q_2 \in M_n$ such that $A = Q_1 \Sigma Q_2^*$. If $A \in M_{m,n}(R)$, then Q_1 and Q_2 may be taken to be real orthogonal matrices.*

Some of these applications include finding the rank (when the matrix is not full rank and Gaussian elimination is impractical), range, and null space of a matrix, as well as computing its pseudoinverse. SVD is quite similar to eigendecomposition, so it can be particularly useful when a matrix has no eigendecomposition. Applied use of the SVD includes camera calibration, numerical weather prediction, and quantum information.

1.2 Singular Value Inequalities

The initial research which ultimately led to the modern theory of singular value inequalities began with ideas in integral equations, most notably by E. Schmidt in 1907. Schmidt considered real integral equations with both symmetric and nonsymmetric kernels. In the nonsymmetric case, he asked whether there exist solutions ϕ and ψ to

$$\phi(s) = \lambda \int_a^b K(s,t)\psi(t)\,\mathrm{d}t, \text{ and } \psi(s) = \lambda \int_a^b K(t,s)\phi(t)\,\mathrm{d}t$$

where $\phi(s)$ and $\psi(s)$ are not identically zero ([17], p. 138). Schmidt showed that the scalar λ has to be real, since λ^2 is an eigenvalue of the symmetric and positive semidefinite kernel

$$H(s,t) = \int_a^b K(s,\tau)K(t,\tau)\,\mathrm{d}\tau$$

So if we think of $K(s,t)$ as an analog of a matrix A, then $H(s,t)$ is an analog of AA^T. A natural generalization which Schmidt made was to refer to λ as an eigenvalue and $\phi(s)$ and $\psi(s)$ as eigenfunctions associated with λ (he specifically called them "adjoint eigenfunctions"). Schmidt's research was furthered by Picard (1910), who, at least in the symmetric case, referred to Schmidt's eigenvalues as "singular values." In future research, this became the common way to refer to λ.

The longstanding goal of Schmidt's work was to establish a connection "between the orders of magnitude of the eigenvalues and the singular values when the kernel is not symmetric." ([17], p. 139) His student Chang, in 1949, was able to establish an indirect connection; that "convergence of an infinite series of given powers of the singular values of an integral kernel implies convergence of the infinite series of the same powers of the absolute values of the eigenvalues." ([17], p. 139) This led to Weyl (1949) showing that there exists a direct inequality between partial sums of Chang's two series. Weyl's discovery provided the foundation for singular value inequalities.

Weyl, apparently motivated by Chang, showed that $|\lambda_1 \cdots \lambda_k| \leq \sigma_1 \cdots \sigma_k$ for

$k = 1, ...n$, where $\lambda_1, ..., \lambda_k$ and $\sigma_1, ..., \sigma_k$ denote the eigenvalues and singular values, respectively, of a square matrix, and $|\lambda_1| \geq ... \geq |\lambda_n|$ and $\sigma_1 \geq ... \geq \sigma_k \geq 0$. Further, he deduced $\phi(|\lambda_1|) + ... + \phi(|\lambda_k|) \leq \phi(\sigma_1) + ... + \phi(\sigma_k)$ for any increasing function ϕ on $[0, \infty)$ such that $\phi(e^t)$ is convex on $(-\infty, \infty)$.

Other work which established the modern theory of singular value inequalities includes the work done by A. Horn (1950) and Ky Fan (1951). Horn established the following multiplicative inequalities (and likewise analogous additive inequalities):

$$\sigma_1(AB) \cdots \sigma_k(AB) \leq \sigma_1(A) \cdots \sigma_k(A)\sigma_1(B) \cdots \sigma_k(B)$$

Ky Fan extended his prior work to obtain the fundamental variational characterization of singular value sums. This allowed him to prove several results involving singular value inequalities. He further showed how Von Neumann's characterization of all unitarily invariant norms follow easily. Another feature of Fan's work is that the variational characterizations of singular values are quasilinear functions of A itself, not via A^*A.

2 Definitions and Preliminary Results

2.1 Singular Value Decomposition

Since the singular value decomposition (SVD) is so consistently used when discussing proofs involving singular values, it seems natural to begin here. Above we presented the SVD theorem and now we will prove it.

Theorem 2.1. ([17], pp. 144-145) *Any $m \times n$ real matrix A can be factored into $A = Q_1 \Sigma Q_2^T$, where the columns of Q_1 $(m \times m)$ are the eigenvectors of AA^T, and the columns of Q_2 $(n \times n)$ are the eigenvectors of A^TA. The $r = \min(m, n)$ singular values on the diagonal of Σ $(m \times n)$ are the squareroots of the eigenvalues of both AA^T and A^TA.*

Proof. The matrix A^TA is real symmetric so it has a complete set of orthonormal eigenvectors: $A^TAx_j = \lambda_j x_j$, and

$$x_i^T A^T A x_j = \lambda_j x_i^T x_j = \lambda_j \delta_{ij},$$

where $\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$.

For positive $\lambda_j's$ $(j = 1, ..., r)$, define singular values $\sigma_j = \sqrt{\lambda_j}$ and also vectors q_j as Ax_j/σ_j. Then $q_i^T q_j = \delta_{ij}$. Extend the $q_i's$ to a basis for R^m. Put the $x_i's$ in Q_2 and the $q_i's$ in Q_1, then

$$(Q_1^T A Q_2)_{ij} = q_i^T A x_j = \begin{cases} 0 & j > r \\ \sigma_j q_i^T q_j = \sigma_i \delta_{ij} & j \leq r \end{cases}$$

That is, $Q_1^T A Q_2 = \Sigma$. So $A = Q_1 \Sigma Q_2^T$. ∎

Example 2.2. *Consider* $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. *Then computing A^TA yields*

$\begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$. *This produces eigenvalues* $\lambda_1 = 16$ *and* $\lambda_2 = 4$, *hence the singular values are* $\sigma_1 = 4$ *and* $\sigma_2 = 2$. *Thus* $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. *Then the columns of* Q_2 *are the eigenvectors of* $A^T A$:

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},\ v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix},\ V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

We obtain u_1 *and* u_2 *as the following:*

$$u_1 = \frac{1}{\sigma_1} A v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix},\ u_2 = \frac{1}{\sigma_2} A v_2 = \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix}$$

Then u_3 *and* u_4 *are in the null space of* A^T, *i.e.* $u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ *and* $u_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$.

Thus we have the following singular value decomposition:

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} = Q_1 \Sigma Q_2.$$

Definition 2.3. *A square matrix A is called diagonalizable if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix* P *such that* $P^{-1}AP$ *is a diagonal matrix.*

Definition 2.4. *A square matrix with complex entries is said to* Hermitian *if it is equal to its conjugate transpose. That is, for* $A \in M_n$, A *is Hermitian if* $A = A^*$.

Corollary 2.5. (Spectral Decomposition) ([16], p. 104) *Let A be a square, diagonalizable, Hermitian matrix. Then $A = U\Sigma U^*$, where U is an orthonormal basis of eigenvectors.*

While SVD isn't the focus of this paper, some results involving SVD will be necessary in order to establish desired inequalities. The next result and its corollary provide information about the singular values of A in terms of mins and maxes of the spectral norm of A. To prove this theorem, though, we'll need to use the Courant-Fischer Theorem, whose proof can be found in [16].

Theorem 2.6. (The Courant-Fischer Theorem) ([16], pp. 179-180) *Let $A \in M_n(\mathbb{C})$ be a Hermitian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$, and let k be a given integer with $1 \leq k \leq n$. Then*

$$\min_{w_1,w_2,...,w_{n-k}\in\mathbb{C}^n} \max_{x\in\mathbb{C}^n, x\perp w_1,w_2,...,w_{n-k}} \frac{x^*Ax}{x^*x} = \lambda_k$$

and

$$\max_{w_1,w_2,...,w_{k-1}\in\mathbb{C}^n} \min_{x\in\mathbb{C}^n, x\perp w_1,w_2,...,w_{k-1}} \frac{x^*Ax}{x^*x} = \lambda_k$$

And now we can proceed with the theorem of interest:

Theorem 2.7. ([17], pp. 148-149) *Let $A \in M_{m,n}(\mathbb{C})$ be given, let $\sigma_1(A) \geq \sigma_2(A) \geq ... \geq \sigma_r(A)$ where $r = \min(m,n)$ be the ordered singular values of A, and let k be a given integer with $1 \leq k \leq min(m,n)$. Then*

(a) $\sigma_k(A) = \min_{w_1,...,w_{k-1}\in\mathbb{C}^n} \max_{x\in\mathbb{C}^n, ||x||_2=1, x\perp w_1,...,w_{k-1}} ||Ax||_2$

(b) $\sigma_k(A) = \min_{w_1,...,w_{n-k}\in\mathbb{C}^n} \max_{x\in\mathbb{C}^n, ||x||_2=1, x\perp w_1,...,w_{n-k}} ||Ax||_2$

(c) $\sigma_k(A) = \min_{S\subset\mathbb{C}^n, dim S=n-k+1} \max_{x\in S, ||x||_2=1} ||Ax||_2$

(d) $\sigma_k(A) = \max_{S\subset\mathbb{C}^n, dim S=k} \min_{x\in S, ||x||_2=1} ||Ax||_2$

Proof. Using the "min-max" half of the Courant-Fischer theorem, we can characterize the decreasingly ordered eigenvalues of A^*A:

$$\lambda_k(A^*A) = \min_{w_1,\dots,w_{k-1}\in\mathbb{C}^n} \max_{x\perp w_1,\dots,w_{k-1}} \frac{x^*(A^*A)x}{x^*x}$$
$$= \min_{w_1,\dots,w_{k-1}\in\mathbb{C}^n} \max_{||x||_2=1, x\perp w_1,\dots,w_{k-1}} ||Ax||_2^2$$

which proves (a) since $\lambda_k(A^*A) = \sigma_k(A)^2$. The same argument with the max-min half of the Courant-Fischer theorem proves the characterization in (b). The alternative formulations (c) and (d) are equivalent versions of (a) and (b), in which the specification of x via a stated number of orthogonality constraints is replaced by specification of x via membership in a subspace, the orthogonal complement of the span of the constraints. □

This leads to the following corollaries:

Corollary 2.8. ([17], pp. 149-150) *Let $A \in M_{m,n}$ be given, and let A_r denote a submatrix of A obtained by deleting a total of r rows and/or columns from A. Then*

$$\sigma_k(A) \geq \sigma_k(A_r) \geq \sigma_{k+r}(A), k = 1, ..., min(m, n) \quad ,$$

where for $X \in M_{p,q}$ we set $\sigma_j(X) = 0$ if $j > min(p, q)$.

Proof. It suffices to consider the case $r = 1$, in which any one row or column is deleted, and to show that $\sigma_k(A) \geq \sigma_k(A_1) \geq \sigma_{k+1}(A)$. The general case then follows by repeated application of these inequalities. If A_1 is formed from A by deleting column s, denote by e_s the standard unit basis vector with a 1 in position s. If $x \in \mathbb{C}^n$, denote by $\xi \in \mathbb{C}^{n-1}$ the vector obtained by deleting entry s from x. Now use (a) in the above theorem to write

$$\sigma_k(A) = \min_{w_1,\dots,w_{k-1}\in\mathbb{C}^n} \max_{x\in\mathbb{C}^n, ||x||_2=1, x\perp w_1,\dots,w_{k-1}} ||Ax||_2$$
$$\geq \min_{w_1,\dots,w_{k-1}\in\mathbb{C}^n} \max_{x\in\mathbb{C}^n, ||x||_2=1, x\perp w_1,\dots,w_{k-1}, e_s} ||Ax||_2$$
$$= \min_{w_1,\dots,w_{k-1}\in\mathbb{C}^{n-1}} \max_{\xi\in\mathbb{C}^{n-1}, ||\xi||_2=1, \xi\perp w_1,\dots,w_{k-1}} ||A_1\xi||_2 = \sigma_k(A_1)$$

The second inequality follows similarly by using (b). If a row of A is deleted, apply the same argument to A^*, which has the same singular values as A. $\square$

The next corollary obtains useful inequalities between individual singular values of a matrix and eigenvalues of its Hermitian part.

Corollary 2.9. ([17], p. 151) *Let $A \in M_n(\mathbb{C})$ be given, let $\sigma_1(A) \geq ... \geq \sigma_n(A)$ denote its ordered singular values, let $H(A) = 1/2(A + A^*)$, and let $[\sigma_i(H(A))]$ denote the algebraically decreasingly ordered eigenvalues of $H(A)$, $\lambda_1(H(A)) \geq ... \geq \lambda_n(H(A))$. Then*

$$\sigma_k(A) \geq \lambda_k(H(A)) \text{ for } k = 1, ..., n$$

More generally,

$$\sigma_k(A) \geq \lambda_k(H(UAV)) \text{ for all } k = 1, ..., n \text{ and all unitary } U, V \in M_n(\mathbb{C}).$$

Proof. For any unit vector $x \in \mathbb{C}^n$, we have

$$x^*H(A)x = \frac{1}{2}(x^*Ax + x^*A^*x) = Re(x^*Ax) \leq |x^*Ax| \leq ||x||_2||Ax||_2 = ||Ax||_2,$$

where $|x^*Ax| \leq ||x||_2||Ax||_2$ is the Cauchy-Schwarz Inequality.

Thus,

$$\begin{aligned} \lambda_k(H(A)) &= \min_{w_1,...,w_{k-1} \in \mathbb{C}^n} \max_{x \in \mathbb{C}^n, ||x||_2 = 1, x \perp w_1,...,w_{k-1}} x^*H(A)x \\ &\leq \min_{w_1,...,w_{k-1} \in \mathbb{C}^n} \max_{x \in \mathbb{C}^n, ||x||_2 = 1, x \perp w_1,...,w_{k-1}} ||Ax||_2 = \sigma_k(A) \end{aligned}$$

The more general assertion follows from the first, since $\sigma_k(A) = \sigma_k(UAV)$ for every $U, V \in M_n$. $\square$

A useful representation which follows from SVD is *polar decomposition*, presented in the following theorem:

Theorem 2.10. (Polar Decomposition) ([17], pp. 152-153) *Let $A \in M_{m,n}(\mathbb{C})$ be given* *(a) If $n \geq m$, then $A = PY$, where $P \in M_m(\mathbb{C})$ is positive semidefinite (see section 2.3), $P^2 = AA^*$, and $Y \in M_{m,n}(\mathbb{C})$ has orthonormal rows.*

(b) If $m \geq n$, then $A = XQ$, where $Q \in M_n(\mathbb{C})$ is positive semidefinite (see section 2.3), $Q^2 = A^*A$, and $X \in M_{m,n}(\mathbb{C})$ has orthonormal columns.

(c) If $m = n$, then $A = PU = UQ$, where $U \in M_n(\mathbb{C})$ is unitary, $P, Q \in M_n(\mathbb{C})$ are positive semidefinite (see section 2.3), $P^2 = AA^*$, and $Q^2 = A^*A$.

In each case, the positive semidefinite factors P and Q are uniquely determined by A and their eigenvalues are the same as the singular values of A.

Proof. If $n \geq m$ and $A = V\Sigma W^*$ is a singular value decomposition, write $\Sigma = [S\ 0]$ and $W = [W_1\ W_2]$, where $S = diag(\sigma_1(A), ..., \sigma_m(A)) \in M_m$ and $W_1 \in M_{n,m}(\mathbb{C})$ is unitary. Then $A = V[S\ 0][W_1\ W_2]^* = VSW_1^* = (VSV^*)(VW_1^*)$. Notice that $P \equiv VSV^*$ is positive semidefinite and $Y \equiv VW_1^*$ satisfies $YY^* = VW_1^*W_1V^* = VIV^* = I$, so Y has orthonormal rows. The assertions in (b) follow from applying (a) to A^*. For (c), notice that $A = V\Sigma W^* = (V\Sigma V^*)(VW^*) = (VW^*)(W\Sigma W^*)$, so we may take $P = V\Sigma V^*$, $Q = W\Sigma W^*$, and $U = VW^*$. □

Example 2.11. *We will find the polar decomposition of $A = \left(\begin{smallmatrix} 2 & 1 \\ 2 & 1 \end{smallmatrix}\right)$ using (c) of the theorem; that is, $A = PU$.*

We begin by computing the SVD of A as

$$A = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} = V\Sigma W^*.$$

The proof of part (c) of the theorem justifies using $P = V\Sigma V^*$ and $U = VW^*$ to arrive at the desired polar decomposition.

Thus we find

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{10}}{2} & \frac{\sqrt{10}}{2} \\ \frac{\sqrt{10}}{2} & \frac{\sqrt{10}}{2} \end{pmatrix}, \text{ and}$$

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{2}{5\sqrt{5}} & \frac{1}{5\sqrt{5}} \\ -\frac{1}{5\sqrt{5}} & 2\sqrt{5} \end{pmatrix} = \begin{pmatrix} \frac{3}{10\sqrt{10}} & -\frac{9\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{11\sqrt{10}}{10} \end{pmatrix}.$$

And we see that $A = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{10}}{2} & \frac{\sqrt{10}}{2} \\ \frac{\sqrt{10}}{2} & \frac{\sqrt{10}}{2} \end{pmatrix} \begin{pmatrix} \frac{3}{10\sqrt{10}} & -\frac{9\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{11\sqrt{10}}{10} \end{pmatrix} = PU.$

There are some interesting results we can now prove involving singular values. The next result is exercise 12 on p. 156 of [17].

Result 2.12. *Let $A \in M_n(\mathbb{C})$. Then $\sigma_1(A) \cdots \sigma_n(A) = |det\ A| = |\lambda_1(A) \cdots \lambda_n(A)|$.*

Proof. We'll start by showing $\sigma_1(A) \cdots \sigma_n(A) = |det\ A|$.

Using the SVD of A, we can say that $A = U\Sigma V$, where U and V are unitary matrices. Since the determinant is multiplicative, this is

$$det(A) = det(U)det(\Sigma)det(V).$$

Then since Σ is defined to be a diagonal of singular values, its determinant will be a product of these. Since U and V are unitary, their determinants have magnitude 1 ([18], p. 103). Hence, we may conclude $\sigma_1(A) \cdots \sigma_n(A) = |det\ A|$.

Now we'll show $|det\ A| = |\lambda_1(A) \cdots \lambda_n(A)|$. In general, the characteristic polynomial of a square matrix M is defined as $f(\lambda) = \det(\lambda I - A)$. Here, the characteristic polynomial of A can be expressed as $f(\lambda) = (-1)^n(\lambda - \lambda_1(A)) \cdots (\lambda - \lambda_n(A))$. Evaluating this at 0, we see

$$detA = f(0) = (-1)^n(0 - \lambda_1) \cdots (0 - \lambda_n) = (-1)^{2n}\lambda_1 \cdots \lambda_n = \lambda_1 \cdots \lambda_n.$$

Thus, $\sigma_1(A) \cdots \sigma_n(A) = |det\ A| = |\lambda_1(A) \cdots \lambda_n(A)|$. □

The next result is exercise 2 on p. 154 of [17].

Result 2.13. *For $A \in M_{m,n}(\mathbb{C})$, the rank of A is exactly the number of its nonzero singular values.*

Proof. Let's say the rank of A is r. Since the rank of a diagonal matrix equals the number of its nonzero entries, using $A = U\Sigma V^T$, where U and V are of full rank, allows us to conclude that $rank(A) = rank(\Sigma) = r$. This follows from certain properties of matrix rank. In general, if $A \in M_{m,n}, B \in M_{n,k}$, and $C \in M_{l,m}$, where B is of rank n and C is of rank m, then ([16], p. 13)

$$rank(AB) = rank(A), \text{ and } rank(CA) = rank(A).$$

□

Corollary 2.14. ([5], pp. 47-48) *For some matrix A,*

$$rank(A) = rank(A^*) = rank(A^*A) = rank(AA^*).$$

Proof. If we apply SVD to the these matrices, we find $A = U\Sigma V^*$, $A^* = V\Sigma^* U^*$, $AA^* = U\Sigma\Sigma^* U^*$, $A^*A = V\Sigma^*\Sigma U$. The singular values of A^* are complex conjugates of the singular values of A, and the singular values of AA^* and A^*A are the absolute values squared of the singular values of A. Hence, zero singular values are preserved everywhere, as are non-zero singular values. Thus we can conclude A, A^*, AA^*, and A^*A all have the same rank. $\square$

2.2 Matrix Norms

Matrix norms come up occasionally throughout matrix analysis and the study of singular values, so it's important to begin with some basic definitions of matrix norms, as well as discussing how they apply directly to the singular values.

A function $||*|| : M_n \mapsto R$ is a *matrix norm* if for all A, $B \in M_n$ it satisfies the following five axioms:

(1) $||A|| \geq 0$ (Nonnegative)

(2) $||A|| = 0$ if and only if $A = 0$ (Positive)

(3) $||cA|| = |c|||A||$ for all complex scalars c (Homogeneous)

(4) $||A + B|| \leq ||A|| + ||B||$(Triangle Inequality)

(5) $||AB|| \leq ||A||||B||$ (Submultiplicative)

Theorem 2.15. ([16], p. 291) *Let the* maximum column sum $||*||_1$ *be defined on* M_n *by*

$$||A||_1 \equiv \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$$

Then $||*||_1$ *is a matrix norm.*

Theorem 2.16. ([16], p. 291) *Let the* maximum row sum $||*||_\infty$ *be defined on* M_n *by*

$$||A||_\infty \equiv \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$$

Then $||*||_\infty$ *is a matrix norm.*

Example 2.17. *Consider* $A=\begin{pmatrix} 3 & -6 & 2 \\ 2 & 5 & 1 \\ -3 & 2 & 2 \end{pmatrix}$ *and* $B=\begin{pmatrix} 3 & 2 & 1 \\ 2 & -3 & 0 \\ 1 & 0 & -1 \end{pmatrix}$.

Then

$$\begin{aligned} ||A||_1 &= \max\left(|3| + |2| + |-3|, (|-6| + |5| + |2|), (|2| + |1| + |2|\right) \\ &= \max\left(8, 13, 5\right) = 13 \end{aligned}$$

$$
\begin{aligned}
||B||_1 &= \max\left(|3|+|2|+|1|\right),\left(|2|+|-3|+|0|\right),\left(|1|+|0|+|-1|\right) \\
&= \max\left(6,5,2\right) = 6
\end{aligned}
$$

$$
\begin{aligned}
||A||_\infty &= \max\left(|3|+|-6|+|2|\right),\left(|2|+|5|+|1|\right),\left(|-3|+|2|+|2|\right) \\
&= \max\left(11,8,7\right) = 11
\end{aligned}
$$

$$
\begin{aligned}
||B||_\infty &= \max\left(|3|+|2|+|1|\right),\left(|2|+|-3|+|0|\right),\left(|1|+|0|+|-1|\right) \\
&= \max\left(6,5,2\right) = 6
\end{aligned}
$$

Another important norm is the *spectral norm.*

Theorem 2.18. ([16], p. 295) *Let the spectral norm $||*||_2$ be defined on M_n by*

$$||A||_2 \equiv \max \sqrt{\lambda}, \text{ where } \lambda \text{ is an eigenvalue of } A^*A. \quad .$$

Then $||*||_2$ is a matrix norm.

The sum of the k largest singular values happens to be a norm, known as the Ky Fan k-norm, which is used commonly in studying singular values. We will show next that they are indeed norms. The last of the Ky Fan norms, the sum of *all* singular values, is the trace norm. In the following theorem and corollaries, we'll present a proof of the Ky Fan k-norm [17].

Definition 2.19. *A matrix $P \in M_{m,n}$ is said to be a rank r partial isometry if $\sigma_1(P) = ... = \sigma_r(P) = 1$ and $\sigma_{r+1}(P) = ... = \sigma_q(P) = 0$, where $q \equiv \min(m,n)$. Two partial isometries $P, Q \in M_{m,n}$ (of unspecified rank) are said to be orthogonal if $P^*Q = 0$ and $PQ^* = 0$.*

Theorem 2.20. ([17], pp. 195-196) *Let $A \in M_{m,n}$ have singular values $\sigma_1(A) \geq ... \geq$*

$\sigma_q(A) \geq 0$, *where* $q = \min(m, n)$. *For each* $k = 1, ..., q$ *we have*

$$\sum_{i=1}^{k} \sigma_i(A) = \max\left[|trX^*AY| : X \in M_{m,k}, Y \in M_{n,k}, X^*X = I = Y^*Y\right]$$

$$= \max\left[|trAC| : C \in M_{n,m} \textit{ is a rank } k \textit{ partial isometry}\right]$$

Proof. If $X \in M_{m,k}$ and $Y \in M_{n,k}$ satisfy $X^*X = I = Y^*Y$, then using a well known property of the trace ($tr(MN) = tr(NM)$, for $M \in M_{m,k}$ and $N \in M_{k,n}$) $trX^*AY = trAYX^*$ and if $C \equiv YX^* \in M_{n,m}$ then $C^*C = XY^*YX^* = XX^*$. Since the k largest singular values of XX^* are the same as those of $X^*X = I \in M_k$, we conclude that $C = YX^*$ is a rank k partial isometry. Conversely, if $C \in M_{n,m}$ is a given rank k partial isometry, then the singular value decomposition of C is

$$C = V\Sigma W^* = [V_k \ *]\left(\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right)[W_k^* \ *]^T = V_k W_k^*$$

where $I_k \in M_k$ is an identity matrix, and $V_k \in M_{n,k}$ and $W_k \in M_{m,k}$ are the first k columns of the unitary matrices $V \in M_n$ and $W \in M_n$, respectively. Hence these two forms are equivalent, so we may show that the second equals the claimed singular value sum.

$$|trAC| = |\sum_{i=1}^{m} \lambda_i(AC)| \leq \sum_{i=1}^{m} |\lambda_i(AC)|$$
$$\leq \sum_{i=1}^{n} \sigma_i(AC) \leq \sum_{i=1}^{q} \sigma_i(A)\sigma_i(C)$$
$$= \sum_{i=1}^{k} \sigma_i(A)$$

in which all indicated eigenvalues λ_i and singular values σ_i are arranged in decreasing absolute value. Note that the second inequality here is from [17], p. 177. If $A = V\Sigma W^*$ is a singular value decomposition of A, let

$$C_{max} \equiv WP_kV^*, \text{ where } P_k \equiv \left(\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right) \in M_{n,m} \text{ and } I_k \in M_k$$

Then $trAC_{max} = trV\Sigma W^*WP_kV^* = trV\Sigma P_k = \sigma_1(A) + ... + \sigma_k(A)$, so our upper bound is achieved. $\square$

Corollary 2.21. ([17], pp. 195-196) *Let A, $B \in M_{m,n}$ have respective ordered singular values $\sigma_1(A) \geq ... \geq \sigma_q(A) \geq 0$ and $\sigma_1(B) \geq ... \geq \sigma_q(B) \geq 0$, $q \equiv \min(m,n)$, and let $\sigma_1(A+B) \geq ... \geq \sigma_q(A+B) \geq 0$ be the ordered singular values of $A+B$. Then*

$$\sum_{i=1}^{k} \sigma_i(A+B) \leq \sum_{i=1}^{k} \sigma_i(A) + \sum_{i=1}^{k} \sigma_i(B),\ k = 1, ..., q$$

Proof. Let $P_{m,n;k}$ denote the rank k partial isometries in $M_{m,n}$. Using the theorem, observe that

$$\begin{aligned}
\sum_{i=1}^{k} \sigma_i(A+B) &= \max\left[|tr(A+B)C| : C \in P_{m,n;k}\right] \\
&= max[|tr(AC+BC)| : C \in P_{m,n;k}] \\
&\leq \max\left[|tr(AC)| + |tr(BC)| : C \in P_{m,n;k}\right] \\
&\leq \max\left[|tr(AC)| : C \in P_{m,n;k}\right] + \max\left[|tr(BC)| : C \in P_{m,n;k}\right] \\
&= \sum_{i=1}^{k} \sigma_i(A) + \sum_{i=1}^{k} \sigma_i(B)
\end{aligned}$$

□

This result says that the sum of the k largest singular values obeys the triangle inequality.

Example 2.22. *Individually, however, the triangle inequality will only hold always for the largest singular value, not necessarily all individual singular values.*
Consider $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $A + B = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and so it's clear that $\sigma_2(A+B) = \sigma_2(I) = 1$, but $\sigma_2(A) + \sigma_2(B) = 0$, so the inequality does not hold.

Now, we'll complete the proof that the Ky Fan k-"norm" is a norm.

Corollary 2.23. ([17], pp. 197-198) *For $A \in M_{m,n}$, let $q \equiv \min(m,n)$, and let $N_k(A) \equiv \sigma_1(A) + ... + \sigma_k(A)$ be the sum of the k largest singular values of A. Then*

(a). $N_k(*)$ *is a norm on* $M_{m,n}$ *for* $k = 1, ..., q$.

(b). *When* $m = n$, $N_k(*)$ *is a matrix norm on* M_n *for* $k = 1, ..., n$.

Proof. To prove (a), we need to show $N_k(*)$ is a positive homogeneous function on $M_{m,n}$ that satisfies the triangle inequality. It's clear that $N_k(A) \geq 0$, and since $N_k(A) \geq \sigma_1(A) = ||A||_2$ is the spectral norm of A it is also clear that $N_k(A) = 0$ if and only if $A = 0$. If c is a given scalar, then $(cA)^*(cA) = |c|^2 A^*A$, so $\sigma_i(cA) = |c|\sigma_i(A)$ for $i = 1, ..., q$. Since the previous corollary says that $N_k(A + B) \leq N_k(A) + N_k(B)$, we conclude that $N_k(*)$ is a norm on $M_{m,n}$. Now let $m = n$. To show that $N_k(*)$ is a matrix norm on M_n, we must show that $N_k(AB) \leq N_k(A)N_k(B)$ for all $A, B \in M_n$. This follows immediately since, using results from our Singular Value Inequalities section,

$$N_k(AB) = \sum_{i=1}^{k} \sigma_i(AB) \leq \sum_{i=1}^{k} \sigma_i(A)\sigma_i(B) \leq \sum_{i=1}^{k} \sigma_i(A) \sum_{i=1}^{j} \sigma_j(B)) = N_k(A)N_k(B)$$

The function $N_k(A) \equiv \sigma_1(A) + ... + \sigma_k(A)$ is often called the *Ky Fan k-norm.* □

Some basic results [16] involving matrix norms follow. The first result is exercise 3 on p. 362 of [16].

Result 2.24. *If* $|| * ||$ *is a matrix norm on* M_n*, then* $c|| * ||$ *is a matrix norm for all* $c \geq 1$.

Proof. Clearly $c|| * ||$ is a vector norm. If we apply the submultiplicative property, $c||AB|| \leq c||A||||B|| \leq c^2||A||||B||$ since $c > 1$, thus $c|| * ||$ is a matrix norm. □

The next result is exercise 5 on p. 311 of [16].

Result 2.25. *The spectral norm* $|| * ||_2$ *is unitarily invariant on* M_n*; that is,* A *and* UAV *have the same norm whenever* U *and* V *are unitary.*

Proof. $||UAV||_2$ is the square root of the largest eigenvalue of $(UAV)^*(UAV) = V^*A^*AV$. And $\sigma(A^*A) = \sigma(V^*A^*AV)$ implies $||UAV||_2 = ||A||_2$, i.e. $|| * ||_2$ is unitarily invariant. □

The following result is exercise 21 on p. 313 of [16].

Result 2.26. *For all $A \in M_n$, we can show that $||A||_2^2 \leq ||A||_1||A||_\infty$.*

Proof. $||A||_2^2$ is the largest eigenvalue of A^*A and so must be less than the trace of A^*A, since eigenvalues of A^*A are nonnegative, hence $||A||_2^2 \leq ||A^*A||_1$. If we apply the submultiplicative property, $||A||_2^2 \leq ||A^*A||_1 \leq ||A^*||_1||A||_1 = ||A||_\infty||A||_1$. $\square$

2.3 Positive Definite Matrices

Positive definite matrices (or simply *positive* matrices) are very commonly used throughout subjects involving applications of linear algebra for their nice properties. The modern theory around singular value inequalities is no exception, and so it feels natural to begin by introducing and examining positive matrices.

A common way to define positive matrices is by inner products defined from an n-dimensional Hilbert space $\mathbb{C}^n$, denoted $\mathcal{H}$. A is *positive semidefinite* (PSD) if

$$\langle x, Ax \rangle \geq 0 \text{ for all } x \in \mathcal{H},$$

and *positive definite* if, in addition,

$$\langle x, Ax \rangle > 0 \text{ for all } x \neq 0$$

A positive semidefinite matrix is positive definite if and only if it is invertible.
We use $A \geq 0$ to mean that A is positive semidefinite, and $A > 0$ to mean it is positive definite (also called *strictly positive*). For Hermitian matrices G and H, we write $G \leq H$ or $H \geq G$ to mean that $H - G$ is positive semidefinite.

Some conditions characterizing positive matrices are [10]:

(i) A is positive if and only if it is Hermitian and all its eigenvalues are nonnegative. A is strictly positive if and only if all its eigenvalues are positive.

(ii) A is positive if and only if it is Hermitian and all its principal minors are nonnegative. A is strictly positive if and only if all its principal minors are positive.

(iii) A is positive if and only if $A = B^*B$ for some matrix B. A is strictly positive if and only if B is nonsingular.

(iv) A is positive if and only if $A = T^*T$ for some upper triangular matrix T. Further, T can be chosen to have nonnegative diagonal entries. If A is strictly positive, then T is unique (this is known as the *Cholesky decomposition* of A). A is strictly positive if and only if T is nonsingular.

(v) A is positive if and only if $A = B^2$ for some positive matrix B. Such a B is unique. We write $B = A^{1/2}$ and call it the (positive) square root of A. A is strictly positive if and only if B is strictly positive.

(vi) A is positive if and only if there exists $x_1, ..., x_n \in \mathcal{H}$ such that

$$a_{ij} = \langle x_i, x_j \rangle.$$

A is strictly positive if and only if the vectors x_j, $1 \leq j \leq n$, are linearly independent.

Example 2.27. *To illustrate the ordering on Hermitian matrices notice the PSD matrices* $A = \left(\begin{smallmatrix} 4 & -3 \\ -3 & 6 \end{smallmatrix}\right)$ *and* $B = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ *are Hermitian and* $A - B = \left(\begin{smallmatrix} 3 & -3 \\ -3 & 5 \end{smallmatrix}\right)$. *Since the eigenvalues of* $A - B$ *are positive,* $A - B$ *is positive semidefinite. Hence* $A \geq B$.

A fundamental result on positive definite matrices would be the Loewner-Heinz Theorem:

Theorem 2.28. (Loewner-Heinz) ([23], pp. 2-3) *If* $A \geq B \geq 0$ *and* $0 \leq r \leq 1$ *then*

$$A^r \geq B^r \qquad .$$

Example 2.29. *As a counterexample to this theorem when* $r > 1$*, consider* $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. *Then* A *and* B *are Hermitian and it is easy to see they have nonnegative eigenvalues. Therefore,* A *and* B *are positive semidefinite. Then* $A^2 - B^2 = \left(\begin{smallmatrix} 4 & 3 \\ 3 & 2 \end{smallmatrix}\right)$, *which is not positive. Thus, when* $r > 1$, $A \geq B \geq 0$ *does not imply* $A^2 \geq B^2$.
As another example, consider the same A *with* $B = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. *Subsequently* $A^2 \geq B^2$ *will not hold.*

Another area to look at involving positive matrices would be their maps [23]. A real-valued continuous function $f(t)$ defined on a real interval Ω is said to be *operator monotone* if

$$A \leq B \text{ implies } f(A) \leq f(B)$$

for all such Hermitian matrices A, B of all orders whose eigenvalues are contained in Ω. f is called *operator convex* if for any $0 < \lambda < 1$,

$$f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda)f(B)$$

holds for all Hermitian matrices A, B of all orders with eigenvalues in Ω. f is called operator concave if $-f$ is operator convex.

The Loewner-Heinz inequality says that $f(t) = t^r$ (on $0 < r \leq 1$) is operator monotone on $[0, \infty)$. Another instance of an operator monotone function is $\log t$ on $(0, \infty)$ which an example of an operator convex function is $g(t) = t^r$ on $(0, \infty)$ for $-1 \leq r \leq 0$ or $1 \leq r \leq 2$ ([8], p. 147). The following is a useful integral respresentation for operator monotone and operator convex functions.

Before proceeding with the next theorem, we'll introduce what a *positive measure* is.

Definition 2.30. *A* measure μ *is defined on the power set* $\mathcal{P}(S)$ *of an infinite set* S *such that*

1. $\mu(\emptyset) = 0$ *and* $\mu(S) = 1$

2. If $X \subset Y$*, then* $\mu(X) \leq \mu(Y)$

3. $\mu([a]) = 0$*, for every* $a \in S$

4. If X_n*,* $n = 0, 1, 2, ...$ *are positive disjoint, then* $\mu(\bigcup_{i=0}^{\infty} X_n) = \sum_{n=0}^{\infty} \mu(X_n)$

Subsequently, a positive *measure is a measure which is a function from the measurable sets of a measure space to the nonnegative real numbers.*

Theorem 2.31. ([23], p. 5) *If* f *is an operator monotone function on* $[0, \infty)$*, then there exists a positive measure* μ *on* $[0, \infty)$ *such that*

$$f(t) = \alpha + \beta t + \int_0^\infty \tfrac{st}{s+t}\, \mathrm{d}\mu(s).$$

where α *is a real number and* $\beta \geq 0$ [23]*. If* g *is an operator convex function on* $[0, \infty)$ *then there exists a positive measure* μ *on* $[0, \infty)$ *such that*

$$g(t) = \alpha + \beta t + \gamma t^2 + \int_0^\infty \tfrac{st^2}{s+t}\, \mathrm{d}\mu(s).$$

where α*,* β *are real numbers and* $\gamma \geq 0$*.*

We call a map $\Phi : M_m \to M_n$ *positive* if it maps positive matrices to positive matrices: $A \geq 0 \Rightarrow \Phi(A) \geq 0$. Φ is called *unital* if $\Phi(I_m) = I_n$.

Example 2.32. *Define* $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ *by* $\Phi(A) = A^T$ *for all* $A \in M_n(\mathbb{C})$ *where* A^T *is the transpose of* A*. Then* Φ *is a positive map since*

$$\Phi(A^*A) = (A^*A)^T = A^T(A^*)^T = A^T(A^T)^*$$

for all A, as the transpose and adjoint commute. Note that by condition (iii) above of positive matrices, we knew that A^*A was positive. Further, by condition (iv) all positive matrices are are generated by A^*A.

We now present a few results involving unital positive linear maps.

Lemma 2.33. ([23], p. 6) *Let Φ be a unital positive linear map from $M_m(\mathbb{C})$ to $M_n(\mathbb{C})$. Then*

(a) $$\Phi(A^2) \geq \Phi(A)^2 \ (A \geq 0),$$

(b) $$\Phi(A^{-1}) \geq \Phi(A)^{-1} \ (A > 0)$$

Theorem 2.34. ([23], pp. 6-7) *Let Φ be a unital positive linear map from $M_m(\mathbb{C})$ to $M_n(\mathbb{C})$ and f an operator monotone funtion on $[0, \infty)$. Then for every $A \geq 0$,*

$$f(\Phi(A)) \geq \Phi(f(A)).$$

Theorem 2.35. ([23], p. 7) *Let Φ be a unital positive linear map from $M_m(\mathbb{C})$ to $M_n(\mathbb{C})$ and g an operator convex function on $[0, \infty)$. Then for every $A \geq 0$,*

$$g(\Phi(A)) \leq \Phi(g(A)).$$

We'll now look at results involving inequalities for matrix powers on positive matrices. The following lemma will be helpful.

Lemma 2.36. ([23], p. 11) *Let f be an operator monotone function on $[0, \infty)$, g an operator convex function on $[0, \infty)$ with $g(0) \leq 0$. Then for every contraction C, i.e., $||C||_\infty \leq 1$ and every $A \geq 0$,*

$$f(C^*AC) \geq C^*f(A)C,$$
$$g(C^*AC) \leq C^*g(A)C.$$

Lemma 2.37. ([8], p. 9) *If $A \leq B$, then $PAP \leq PBP$ where P is positive semidefinite.*

Proof. Since P is positive semidefinite, $P = P^*$. Let $M = P(B - A)P$. If we can show $M \geq 0$, we're done. Then $M^* = P^*(B - A)P^* = M$, so M is Hermitian. A simple calculation shows

$$M = P\sqrt{B-A}\sqrt{B-A}P = (P\sqrt{B-A})(P\sqrt{B-A})^* \geq 0.$$

Hence by property (iii) of positive semidefinite matrices, $M \geq 0$. $\square$

Theorem 2.38. ([23], pp. 12-13) *If* $A \geq B \geq 0$ *then*

(a).

$$(B^r A^p B^r)^{1/q} \geq B^{(p+2r)/q}$$

and

(b).

$$A^{(p+2r)/q} \geq (A^r B^p A^r)^{1/q}$$

for $r \geq 0$, $p \geq 0$, $q \geq 1$, *with* $(1+2r)q \geq p+2r$.

Proof. We'll abbreviate the Loewner-Heinz inequality (Theorem 2.28) by LH, and start by proving (a).

If $0 \leq p < 1$, then by LH, $A^p \geq B^p$ and hence $B^r A^p B^r \geq B^{p+2r}$, by Lemma 2.37. Applying LH again with the power $1/q$ gives us (a).

Next we consider the case where $p \geq 1$. It will suffice to prove

$$(B^r A^p B^r)^{(1+2r)/(p+2r)} \geq B^{1+2r}$$

for $r \geq 0$, $p \geq 1$, since by assumption $q \geq (p+2r)/(1+2r)$, and then (a) follows from this inequality via LH. We introduce t to write the above inequality as

$$(B^r A^p B^r)^t \geq B^{1+2r},\ t = \tfrac{1+2r}{p+2r}.$$

Note that $0 < t \leq 1$, as $p \geq 1$. We will prove $(B^r A^p B^r)^t \geq B^{1+2r}$ by induction on $k = 0, 1, 2, ...$ for the intervals $(2^{k-1} - 1/2, 2^k - 1/2]$ containing r. Since $(0, \infty) = \bigcup_{k=0}^{\infty}\{(2^{k-1} - 1/2, 2^k - 1/2]\}$, the inequality will be proved.

By the standard continuity argument (see Definition 3.21), we may and do assume that A, B are positive definite. First consider the case $k = 0$, i.e., $0 < r \leq 1/2$. By LH, $A^{2r} \geq B^{2r}$ and hence $B^r A^{-2r} B^r \leq I$, which means $A^{-r} B^r$ is a contraction.

Applying Lemma 2.36 with $f(x) = x^t$ yields

$$\begin{aligned}(B^r A^p B^r)^t &= [(A^{-r}B^r)^* A^{p+2r}(A^{-r}B^r)]^t \\ &\geq (A^{-r}B^r)^* A^{(p+2r)t}(A^{-r}B^r) \\ &= B^r A B^r \geq B^{1+2r}\end{aligned}$$

proving $(B^r A^p B^r)^t \geq B^{1+2r}$ for the case $k = 0$.

Now suppose the inequality is true for $r \in (2^{k-1} - 1/2, 2^k - 1/2]$. Denote $A_1 = (B^r A^p B^r)^t$, $B_1 = B^{1+2r}$. Then our assumption is

$$A_1 \geq B_1 \text{ with } t = \tfrac{1+2r}{p+2r}.$$

Since $p \equiv 1/t \geq 1$, apply the already proved case $r_1 \equiv 1/2$ to $A_1 \geq B_1$ to get

$$(B_1^{r_1} A_1^{p_1} B_1^{r_1})^{t_1} \geq B_1^{1+2r_1}, t_1 \equiv \tfrac{1+2r_1}{p_1+2r_1}.$$

Note that $t_1 = \frac{2+4r}{p+4r+1}$. Denote $s = 2r + 1/2$. We have $s \in (2^k - 1/2, 2^{k+1} - 1/2]$. Then, explicitly this is

$$(B^s A^p B^s)^{t_1} \geq B^{1+2s},\ t_1 = \tfrac{1+2s}{p+2s},$$

which shows that $(B^r A^p B^r)^t \geq B^{1+2r}$ holds for $r \in (2^k - 1/2, 2^{k+1} - 1/2]$. This completes the inductive argument and proves (a).

$A \geq B > 0$ implies $B^{-1} \geq A^{-1} > 0$. In (a) replacing A, B by B^{-1}, A^{-1} respectively will yield (b). $\square$

The case $q = p \geq 1$ of this theorem is the following corollary:

Corollary 2.39. ([23], p. 13) *If $A \geq B \geq 0$ then*

$$\begin{aligned}(B^r A^p B^r)^{1/p} &\geq B^{(p+2r)/p}, \\ A^{(p+2r)/p} &\geq (A^r B^p A^r)^{1/p}\end{aligned}$$

for all $r \geq 0$ and $p \geq 1$.

Another interesting (and counter-intuitive) case is

Corollary 2.40. ([23], p. 13) *If $A \geq B \geq 0$ then*

$$(BA^2B)^{1/2} \geq B^2 \text{ and } A^2 \geq (AB^2A)^{1/2}.$$

We'll now look at some general results involving PSD (positive semidefinite) matrices, from [16]. The first is exercise 4 on p. 409 of [16].

Result 2.41. *If A and B are positive definite, show that $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ is positive definite.*

Proof. Since A and B are Hermitian, $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ must be as well. The eigenvalues of $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ are the elements of $\sigma(A) \cup \sigma(B)$. Hence they are positive. □

The next result is exercise 3 on p. 421 of [16].

Result 2.42. *$A \in M_n$ has a zero singular value if and only if it has a zero eigenvalue.*

Proof. Notice that A will have a zero singular value iff $|det(A)|^2 = det(AA^*) = 0$, i.e. if $det(A) = 0$, which would mean A has a zero eigenvalue. □

The next result is exercise 5 on p. 421 of [16].

Result 2.43. *If $k \leq \min(m, n)$ and v_k is the kth column of V and w_k is the kth column of W in a singular value decomposition of A, then*

$$A^* v_k = \sigma_k w_k \text{ and } A w_k = \sigma_k v_k$$

where σ_k is the kth singular value of A. In particular, $v_k^* A w_k = \sigma_k$.

Proof. If we apply the SVD of A, $A = V\Sigma W^*$, then $AW = V\Sigma$. So we can say that $Aw_k = \sigma_k v_k$. Then $A^* = W\Sigma^* V^*$, so $A^* v_k = \sigma_k w_k$. Hence $v_k^* \sigma_k v_k = v_k^* \sigma_k v_k = \sigma_k$. □

2.4 Block Matrices

Block matrices play a remarkable role in the study of positive matrices [10]. In particular, block matrices will be useful later on in the examination of Tao's more recent techniques involving singular value inequalities. Besides some basic results on block matrices, we'll look at a few preliminary results to Tao's research.

Definition 2.44. *The direct sum of two matrices A and B, denoted $A \oplus B$, is defined to be a block matrix of the following form:* $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$.

If we denote the norm of A by $||A||$, i.e. $||A|| = \max_{||x||=1} ||Ax|| = \max_{||x||\leq 1} ||Ax||$, we can state the following important properties ([16], pp. 290-293):

(i) $$||AB|| \leq ||A||||B||$$

(ii) $$||A|| = ||A^*||$$

(iii) $$||A|| = ||UAV|| \text{ for all unitary } U,\, V.$$

Recall A is said to be a *contraction* (or A is *contractive*) if $||A|| \leq 1$.

Lemma 2.45. ([10], p. 13) *The operator A is contractive if and only if the operator* $\left(\begin{smallmatrix} I & A \\ A^* & I \end{smallmatrix}\right)$ *is positive.*

Lemma 2.46. ([10], pp. 13-14) *Let A, B be positive. Then the matrix* $\left(\begin{smallmatrix} A & X \\ X^* & B \end{smallmatrix}\right)$ *is positive if and only if $X = A^{1/2}KB^{1/2}$ for some contraction K.*

Theorem 2.47. ([10], p. 14) *Let A, B be strictly positive matrices. Then the block matrix* $\left(\begin{smallmatrix} A & X \\ X^* & B \end{smallmatrix}\right)$ *is positive if and only if $A \geq XB^{-1}X^*$.*

Proof. We have $A \geq XB^{-1}X^*$ if and only if

$$\begin{aligned} I &\geq A^{-1/2}(XB^{-1}X^*)A^{-1/2} \\ &= (A^{-1/2}XB^{-1/2})(B^{-1/2}X^*A^{-1/2}) \\ &= (A^{-1/2}XB^{-1/2})(A^{-1/2}XB^{-1/2})^*. \end{aligned}$$

This is equivalent to saying $||A^{-1/2}XB^{-1/2}|| \leq 1$, or $X = A^{1/2}KB^{1/2}$ where $||K|| \leq 1$. Now we may apply Result 2.46. $\square$

Example 2.48. *Consider* $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 4 \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, $X = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 2 \end{smallmatrix}\right)$, *and* $X^* = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 2 \end{smallmatrix}\right)$. *Then* $\lambda(\left(\begin{smallmatrix} A & X \\ X^* & B \end{smallmatrix}\right)) = (4+\sqrt{5}, 4-\sqrt{5}, 0, 0)$, *meaning this block matrix is positive semidefinite, and as such,* $A \geq XB^{-1}X^*$ *should be true. Indeed, we see that* $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 4 \end{smallmatrix}\right) = XB^{-1}X^*$, *so the inequality holds (in this case, it's an equality).*

Lemma 2.49. ([10], p. 15) *The matrix* A *is positive if and only if* $\left(\begin{smallmatrix} A & A \\ A & A \end{smallmatrix}\right)$ *is positive.*

Proof. We can write

$$\left(\begin{smallmatrix} A & A \\ A & A \end{smallmatrix}\right) = \left(\begin{smallmatrix} A^{1/2} & 0 \\ A^{1/2} & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} A^{1/2} & A^{1/2} \\ 0 & 0 \end{smallmatrix}\right).$$

$\square$

Corollary 2.50. ([10], p. 15) *Let* A *be any matrix. Then the matrix* $\left(\begin{smallmatrix} |A| & A^* \\ A & |A^*| \end{smallmatrix}\right)$ *is positive, where* $|A| = (A^*A)^{1/2}$.

Proof. Use the polar decomposition $A = UP$ to write

$$\left(\begin{smallmatrix} |A| & A^* \\ A & |A^*| \end{smallmatrix}\right) = \left(\begin{smallmatrix} P & PU^* \\ UP & UPU^* \end{smallmatrix}\right) = \left(\begin{smallmatrix} I & 0 \\ 0 & U \end{smallmatrix}\right)\left(\begin{smallmatrix} P & P \\ P & P \end{smallmatrix}\right)\left(\begin{smallmatrix} I & 0 \\ 0 & U^* \end{smallmatrix}\right),$$

and then apply the lemma. $\square$

The following result was proved by Zhan in a paper [22].

Result 2.51. *For positive matrices* A, $B \in M_n$, *with* $j = 1, 2, ..., n$, *we have*

$$\sigma_j(A - B) \leq \sigma_j(A \oplus B).$$

Proof. Note that $\sigma(A \oplus B) = \sigma(A) \bigcup \sigma(B)$. It can be verified using spectral decomposition of A and B that for a fixed j with $1 \leq j \leq n$ there exist H, $F \in M_n$ satisfying $0 \leq H \leq A$, $0 \leq F \leq B$, rank H + rank $F \leq j - 1$ and

$$\sigma_j(A \oplus B) = ||(A - H) \oplus (B - F)||_\infty.$$

Thus $\sigma_j(A \oplus B) = \max(||A - H||_\infty, ||B - F||_\infty) \equiv \gamma$.

Since

$$A - H \geq 0,\ B - F \geq 0,\ rank(H - F) \leq rankH + rankF \leq j - 1 \qquad .$$

Using this approximation characterization that

$\sigma_j(G) = min(||G - X||_\infty : rankX \leq j - 1. X \in M_n)$, we have

$$\begin{aligned}
\sigma_j(A - B) &\leq ||A - B - (H - F)||_\infty \\
&= ||(A - H - \frac{\gamma}{2}I) - (B - F - \frac{\gamma}{2}I)||_\infty \\
&\leq ||(A - H) - \frac{\gamma}{2}I||_\infty + ||(B - F) - \frac{\gamma}{2}I||_\infty \\
&\leq \frac{\gamma}{2} + \frac{\gamma}{2} = \gamma = \sigma_j(A \oplus B)
\end{aligned}$$

$\square$

Another preliminary result we'll need is a well known property of eigenvalues of block matrices with zeros on the main diagonal. It is exercise 2 on p. 141 of [17].

Result 2.52. *Let $A \in M_{m,n}$ be given, let $q = \min(m, n)$, and let $\mathcal{A} = \left(\begin{smallmatrix} 0 & A \\ A^* & 0 \end{smallmatrix}\right) \in M_{m+n}$.*
We can say the following

$$\begin{aligned}
det(tI_{m+n} - \mathcal{A}) &= det\left(\begin{smallmatrix} I_m & t^{-1}A \\ 0 & I_n \end{smallmatrix}\right) det\left(\begin{smallmatrix} tI_m & -A \\ -A^* & tI_n \end{smallmatrix}\right) \\
&= det\left(\begin{smallmatrix} tI_m - t^{-1}AA^* & 0 \\ -A^* & tI_n \end{smallmatrix}\right) \\
&= t^{n-m} det(t^2 I_m - AA^*)
\end{aligned}$$

I.e. we may conclude the eigenvalues of $\mathcal{A}$ are $-\sigma_1(A), ..., -\sigma_q(A), 0, ..., 0, \sigma_1(A), ..., \sigma_q(A)$ with $|m - n|$ zeros.

Proof. $det(tI_{m+n} - \mathcal{A}) = det\left(\begin{smallmatrix} I_m & t^{-1}A \\ 0 & I_n \end{smallmatrix}\right) det\left(\begin{smallmatrix} tI_m & -A \\ -A^* & tI_n \end{smallmatrix}\right)$ is true since $det\left(\begin{smallmatrix} I_m & t^{-1}A \\ 0 & I_n \end{smallmatrix}\right)$ is 1.
The proof follows from recognizing that, since the determinant is multiplicative, $\left(\begin{smallmatrix} I_m & t^{-1}A \\ 0 & I_n \end{smallmatrix}\right)\left(\begin{smallmatrix} tI_m & -A \\ -A^* & tI_n \end{smallmatrix}\right) = \left(\begin{smallmatrix} tI_m - t^{-1}AA^* & 0 \\ -A^* & tI_n \end{smallmatrix}\right)$.

And further,

$$\begin{aligned}
det\begin{pmatrix} tI_m - t^{-1}AA^* & 0 \\ -A^* & tI_n \end{pmatrix} &= det\begin{pmatrix} t^{-1}(I_m t^2 - AA^*) & 0 \\ -A^* & tI_n \end{pmatrix} \\
&= t^n det\begin{pmatrix} t^{-1}(I_m t^2 - AA^*) & 0 \\ -A^* & I_n \end{pmatrix} \\
&= t^{n-m} det\begin{pmatrix} I_m t^2 - AA^* & 0 \\ -A^* & I_n \end{pmatrix} \\
&= t^{n-m} det(I_m t^2 - AA^*)
\end{aligned}$$

where the last step follows from the known property of determinants: $\det\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} = \det A \det B$. □

A natural question involving a block matrix might be, what is its determinant? We'll answer this with the below result.

Result 2.53. ([1], p. 114) *If A is invertible, then the determinant of the block matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is given by*

$$det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = det(A)det(D - CA^{-1}B).$$

Proof. Since A is invertible we can write

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & O \\ CA^{-1} & I \end{pmatrix}\begin{pmatrix} A & B \\ O & D - CA^{-1}B \end{pmatrix}$$

Hence

$$det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = det\begin{pmatrix} I & O \\ CA^{-1} & I \end{pmatrix} det\begin{pmatrix} A & B \\ O & D - CA^{-1}B \end{pmatrix}.$$

Since the determinant of the first matrix will be I, the determinant of the original matrix will equal the determinant of the second matrix, thus

$$det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = det(A)det(D - CA^{-1}B).$$

□

2.5 Some Essential Singular Value Inequalities

Some results in this section use the concept relating to vectors known as *majorization*, the state where all the elements of one vector are either larger or smaller than the respective elements of another vector. Before proceeding with the bulk of this section, we will introduce majorization and show some basic results.

Definition 2.54. (Majorization) *Let $x = [x_i]$, $y = [y_i] \in R^n$ be given vectors, and denote their algebraically decreasingly ordered entries by $x_{[1]} \geq ... \geq x_{[n]}$ and $y_{[1]} \geq ... \geq y_{[n]}$. We say that y* weakly majorizes *x if*

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}.$$

There's a relationship between weak majorization and doubly stochastic matrices.

Definition 2.55. *A* doubly stochastic matrix *is a square matrix of nonnegative real numbers, each of whose rows and columns sum to 1.*

Theorem 2.56. ([17], pp. 166-167) *Let $x = [x_i]$, $y = [y_i] \in R^n$ be given vectors with nonnegative entries. Then y weakly majorizes x if and only if there is a doubly stochastic $Q \in M_n(R)$ such that $x = Qy$.*

Corollary 2.57. ([17], pp. 167-168) *Let $x = [x_i], y = [y_i] \in R^n$ be given. Then y weakly majorizes x if and only if there is a doubly stochastic $S \in M_n(R)$ such that the entrywise inequalities $x \leq Sy$ hold.*

Proof. If there is a doubly stochastic S such that $x \leq Sy$, then since there is a (strong) majorization relationship between Sy and y, x must be weakly majorized by y. Conversely, suppose x is weakly majorized by y, let $e = [1, ..., 1]^T \in R^n$, and let $k \geq 0$ be large enough that $x + ke$ and $y + ke$ are both positive vectors. Since $x + ke$ is weakly majorized by $y + ke$, Theorem 2.56 guarantees there is a doubly stochastic $Q \in M_n(R)$ such that $x + ke = Q(y + ke) = Qy + kQe \leq Qy + ke$. The conclusion is that $x \leq Qy$. □

The most fundamental result involving singular value inequalities would be a theorem first proved by Weyl in 1949, establishing that the product of the eigenvalues is less than the product of the singular values. Before looking at the theorem, we present an important lemma which also has use in proving a variety of other results:

Lemma 2.58. ([17], pp. 170-171) *Let $C \in M_{m,n}(\mathbb{C}), V_k \in M_{m,k}(\mathbb{C})$, and $W_k \in M_{n,k}(\mathbb{C})$ be given, where $k \leq min(m,n)$ and V_k, W_k have orthonormal columns. Then*

(a) $\quad \sigma_i(V_k^*CW_k) \leq \sigma_i(C), i = 1, ..., k,$ *and*

(b) $\quad |detV_k^*CW_k| \leq \sigma_1(C)...\sigma_k(C)$

Proof. Since V_k and W_k are defined to have orthonormal columns, they may be extended to form orthonormal bases of C^m and C^n, respectively. As such, there exist respective unitary matrices $V \in M_m$ and $W \in M_n$. We'll denote these $V = [V_k *]$ and $W = [W_k *]$. Since $V_k^*CW_k$ is the upper left k-by-k submatrix of V^*CW, by Corollary 2.8 and unitary invariance of singular values, we know $\sigma_i(V_k^*CW_k) \leq \sigma_i(V^*CW) = \sigma_i(C), i = 1, ..., k$, and hence $|detV_k^*CW_K| = \sigma_1(V_k^*CW_k)...\sigma_k(V_k^*CW_k) \leq \sigma_1(C)...\sigma_k(C)$. $\square$

Another useful result in proving Weyl's theorem is the Schur triangularization theorem (also known as Schur decomposition), which states for $A \in M_n(\mathbb{C})$ with complex entries, we may say $A = QUQ^*$, where Q is a unitary matrix and U is an upper triangular matrix.

Theorem 2.59. ([20], p. 288) (Schur Decomposition) *If A is an $n \times n$ square matrix with complex entries, then A can be expressed as*

$$A = QTQ^{-1}$$

where Q is a unitary matrix (so that its inverse Q^{-1} is also the conjugate transpose Q^* of Q), and T is an upper triangular matrix, called the Schur form *of A.*

Now we can prove the following result by Weyl:

Theorem 2.60. (Weyl's Theorem) ([17], p. 171) *Let $A \in M_n(\mathbb{C})$ have singular values $\sigma_1(A) \geq ... \geq \sigma_n(A) \geq 0$ and eigenvalues $(\lambda_1(A), ..., \lambda_n(A)) \in \mathbb{C}$ ordered so that $|\lambda_1(A)| \geq ... \geq |\lambda_n(A)|$. Then*

$$|\lambda_1(A) \cdots \lambda_k(A)| \leq \sigma_1(A) \cdots \sigma_k(A) \textit{ for } k = 1, ..., n, \textit{ with equality for } k = n.$$

Proof. From the Schur triangularization theorem, there exists a unitary $U \in M_n(\mathbb{C})$ such that $U^*AU = \Delta$ is upper triangular and diag $\Delta = (\lambda_1, ..., \lambda_n)$. If we let $U_k \in M_{n,k}(\mathbb{C})$ be the first k columns of U, we compute

$$U^*AU = [U_k*]^*A[U_k*] = \left(\begin{smallmatrix} U_k^*AU_k & * \\ * & * \end{smallmatrix}\right) = \Delta$$

Hence, $U_k^*AU_k = \Delta_k$ is upper triangular. Now if we apply Lemma 2.58 where $C = A$ and $V_k = W_k = U_k$ we may conclude

$$|\lambda_1(A) \cdots \lambda_k(A)| = |\det \Delta_k| = |\det U_k^*AU_k| \leq \sigma_1(A) \cdots \sigma_k(A)$$

Also, if $k = n$ it follows from SVD that $|\det A| = \sigma_1(A) \cdots \sigma_n(A)$, and we know $\det A = \lambda_1(A) \cdots \lambda_n(A)$. □

Example 2.61. *Consider* $A=\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$. *This has (ordered) eigenvalues $\lambda_1 \approx 3.41421$, $\lambda_2 = 2$, and $\lambda_3 \approx .585786$.*

To find the singular values of A, we begin by computing

$$AA^T = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}^T = \begin{pmatrix} 5 & -4 & 1 \\ -4 & 6 & -4 \\ 1 & -4 & 5 \end{pmatrix}.$$

Subsequently, the singular values are the square roots of the eigenvalues of AA^T, i.e. $\sigma_1 \approx \sqrt{11.6569}$, $\sigma_2 = \sqrt{4}$, and $\sigma_3 \approx \sqrt{0.343146}$. And we check that Weyl's theorem holds, since

$$|\lambda_1(A)\lambda_2(A)\lambda_3(A)| \approx 3.9999 \leq 4.0000009 \approx \sigma_1(A)\sigma_2(A)\sigma_3(A).$$

These calculations were performed in Matlab using the "eig" function with a roundoff error of 10^{-16}.

Another natural question would be what can be said about the singular values of two matrices, if we consider their product? Once such result was proved by A. Horn in 1950 with the following theorem, which also follows from the same lemma as in Weyl's Theorem:

Theorem 2.62. ([17], pp. 171-172) *Let $A \in M_{m,p}(\mathbb{C})$ and $B \in M_{p,n}(\mathbb{C})$ be given, let $q = \min(n, p, m)$, and if we order the singular values of A, B, and AB, respectively, as $\sigma_1(A) \geq ... \geq \sigma_{\min(m,p)}(A) \geq 0$, $\sigma_1(B) \geq ... \geq \sigma_{\min(p,n)}(B) \geq 0$, and $\sigma_1(AB) \geq ... \geq \sigma_{\min(m,n)}(AB) \geq 0$. Then*

$$\prod_{i=1}^{k} \sigma_i(AB) \leq \prod_{i=1}^{k} \sigma_i(A)\sigma_i(B), k = 1, ..., q$$

If $n = p = m$ the above equality holds for $k = n$.

Proof. Using SVD, we may write $AB = V\Sigma W*$, with $V_k \in M_{m,k}$ and $W_k \in M_{n,k}$ being the first k columns of V and W, respectively. Then $V_k^*(AB)W_k = diag(\sigma_1(AB), ..., \sigma_k(AB))$ since it's the upper left $k - by - k$ submatrix of $V^*(AB) = \Sigma$. Since $p \geq k$, we can use polar decomposition to write $BW_k \in M_{p,k}$ as $BW_k = X_kQ$, where $X_k \in M_{p,k}$ has orthonormal columns, $Q \in M_k$ is positive semidefinite, $Q^2 = (BW_k)^*(BW_k) = W_k^*B^*BW_k$, and so $\det Q^2 = \det W_k^*(B^*B)W_k \leq \sigma_1(B^*B) \cdots \sigma_k(B^*B) = \sigma_1(B)^2 \cdots \sigma_k(B)^2$ using the Lemma 2.58. Applying the Lemma 2.58 again,

$$\begin{aligned}
\sigma_1(AB) \cdots \sigma_k(AB) &= |\det V_k^*(AB)W_k| \\
&= |\det V_k^*AX_kQ| \\
&= |\det V_k^*AX_k \det Q| \\
&\leq (\sigma_1(A) \cdots \sigma_k(A))(\sigma_1(B) \cdots \sigma_k(B))
\end{aligned}$$

If $n = p = m$, then $\sigma_1(AB) \cdots \sigma_n(AB) = |\det AB| = |\det A||\det B|$
$= \sigma_1(A) \cdots \sigma_k(A)\sigma_1(B) \cdots \sigma_k(B)$. ☐

Example 2.63. *To illustrate this theorem, consider $A = \begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $AB = \begin{pmatrix} 3 & 0 \\ -3 & 2 \end{pmatrix}$. We need to compute the singular values of these,* *so compute*

$$AA^T = \begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix}^T = \begin{pmatrix} 18 & -24 \\ -24 & 34 \end{pmatrix}$$

$$BB^T = B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^T = B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$(AB)(AB)^T = \begin{pmatrix} 3 & 0 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ -3 & 2 \end{pmatrix}^T = \begin{pmatrix} 9 & -9 \\ -9 & 13 \end{pmatrix}$$

Subsequently, these produce singular values $\sigma_1(A) = \sqrt{26 + 8\sqrt{10}}, \sigma_2(A) = \sqrt{26 - 8\sqrt{10}}, \sigma_1(B) = \sqrt{.5(3 + \sqrt{5})}, \sigma_2(B) = \sqrt{.5(3 - \sqrt{5}}, \sigma_1(AB) = \sqrt{11 + \sqrt{85}}$, *and* $\sigma_2(AB) = \sqrt{11 - \sqrt{85}}$.

Now if we consider the inequality given in the theorem and square both sides, this is

$$\begin{aligned} \sigma_1(AB)^2\sigma_2(AB)^2 &= [11 + \sqrt{85}][11 - \sqrt{85}] = 36 \\ &= [26 + 8\sqrt{10}][26 - 8\sqrt{10}][.5(3 + \sqrt{5})][.5(3 - \sqrt{5}] \\ &= \sigma_1(A)^2\sigma_2(A)^2\sigma_1(B)^2\sigma_2(B)^2 \end{aligned}$$

The following is exercise 9 on page 182 in [17].

Result 2.64. *Theorem 2.62 implies Weyl's Theorem.*

Proof. If we apply the result of 2.62 to $U_k^* A U_k = \Delta_k$, as defined in the proof of Weyl's Theorem (in particular, note that Δ was defined as a diagonal matrix of the eigenvalues), we find that

$$\lambda_1 \cdots \lambda_k \leq [\sigma_1(U_k^*) \cdots \sigma_k(U_k^*)][\sigma_1(A) \cdots \sigma_k(A)][\sigma_1(U_k) \cdots \sigma_k(U_k)] = \sigma_1(A) \cdots \sigma_k(A),$$

since U is unitary (i.e. U has the property that $\lambda^2 = 1$). $\square$

The inequalities we derived in the above two theorems are a form of multiplicative majorization (see section 2.2). If A is nonsingular, we can take logarithms for the result in Theorem 2.60 to obtain equivalent ordinary (strong) majorization inequalities

$\sum_{i=1}^k log|\lambda_i(A)| \leq \sum_{i=1}^k log\sigma_i(A)$, $k = 1, ..., n$, with equality for $k = n$

Our next goal is to show these can be exponentiated to form weak majorization inqualities such as

$$\sum_{i=1}^{k} |\lambda_i(A)| \leq \sum_{i=1}^{k} \sigma_i(A),\ k = 1, ..., n.$$

To do this, it's useful to show that there are a wide class of functions (including $f(t) = e^t$) that preserve systems of inequalities such as the above. These are called *Schur-convex* or *isotone*. The following lemma shows that any increasing convex function with appropriate domain preserves weak majorization.

Lemma 2.65. ([17], pp. 173-174) *Let $x_1, ..., x_n$, $y_1, ..., y_n$ be $2n$ given real numbers such that $x_1 \geq x_2 \geq ... \geq x_n$, $y_1 \geq y_2 \geq ... \geq y_n$, and*

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i,\ k = 1, ..., n$$

If $f(t)$ is a given real-valued increasing convex function on the interval $[\min(x_n, y_n), y_1]$, *then $f(x_1) \geq ... \geq f(x_n)$, $f(y_1) \geq ... \geq f(y_n)$, and*

$$\sum_{i=1}^{k} f(x_i) \leq \sum_{i=1}^{k} f(y_i),\ k = 1, ..., n$$

If equality holds when $k = n$ for $\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$ (that is, if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$) and if f is convex (but not necessarily increasing) on $[y_n, y_1]$, then

$$\sum_{i=1}^{n} f(x_i) \leq \sum_{i=1}^{n} f(y_i)$$

Proof. Clearly it is trivial when $k = 1$, so assume k is an integer such that $2 \leq k \leq n$. Since $x \equiv [x_i]_{i=1}^{k} \in R^k$ is weakly majorized by $y \equiv [y_i]_{i=1}^{k} \in R^k$, by Corollary 2.57 we know there is a doubly stochastic $S = [S_{ij}]_{i=1}^{k} \in M_n(\mathbb{R})$ such that $x \leq Sy$, that is, all $s_{ij} \geq 0$,

$$x_i \leq \sum_{i=1}^{k} s_{ij} y_j, \text{ and } \sum_{i=1}^{n} s_{ij} = \sum_{i=1}^{n} s_{ji} = 1 \text{ for } i = 1, ..., n$$

The entries of Sy are in the interval $[y_n, y_1]$ and hence in the domain of f. Using the monotonicity and convexity of f,

$$\begin{aligned}\sum_{i=1}^{k} f(x_i) &\leq \sum_{i=1}^{k} f[\sum_{i=1}^{k} s_{ij}y_j] \\ &\leq \sum_{i=1}^{k}\sum_{i=1}^{k} s_{ij}f(y_j) \\ &= \sum_{i=1}^{k}[\sum_{i=1}^{k} s_{ij}]f(y_j) \\ &= \sum_{i=1}^{k} f(y_j)\end{aligned}$$

If equality holds for $k = n$ in $\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$, there's a doubly stochastic $S \in M_n(R)$ such that $[x_i]_{i=1}^{n} = S[y_i]_{i=1}^{n}$ and hence convexity of f gives (without needing to assume f is increasing):

$$\begin{aligned}\sum_{i=1}^{n} f(x_i) &= \sum_{i=1}^{n} f[\sum_{i=1}^{n} s_{ij}y_j] \\ &\leq \sum_{i=1}^{n}\sum_{i=1}^{n} s_{ij}f(y_j) \\ &= \sum_{i=1}^{n} f(y_j)\end{aligned}$$

□

We'll now present a corollary with some general results for real numbers that we'll apply to singular values and functions of singular values.

Corollary 2.66. ([17], pp. 174-175) *Let $\alpha_1, ..., \alpha_n$, $\beta_1, ..., \beta_n$ be $2n$ given nonnegative real numbers such that $\alpha_1 \geq ... \geq \alpha_n \geq 0$ and $\beta_1 \geq ... \geq \beta_n \geq 0$. If*

$$\prod_{i=1}^{k} \alpha_i \leq \prod_{i=1}^{k} \beta_i,\ k = 1, ..., n$$

then

$$\sum_{i=1}^{k} \alpha_i \leq \sum_{i=1}^{k} \beta_i,\ k = 1, ..., n$$

Generally, suppose f is a given real-valued function such that $f(e^t)$ is increasing and *convex on the interval* $[\min(\alpha_n, \beta_n), \beta_1]$. *Then*

$$\sum_{i=1}^k f(\alpha_i) \leq \sum_{i=1}^k f(\beta_i), \text{ for } k = 1, ..., n$$

If $\prod_{i=1}^n \alpha_i = \prod_{i=1}^n \beta_i$, *and if* $f(e^t)$ *is convex (but not necessarily increasing) on the interval* $[\beta_n, \beta_1]$, *then*

$$\sum_{i=1}^n f(\alpha_i) \leq \sum_{i=1}^n f(\beta_i)$$

Proof. Notice that $\sum_{i=1}^k \alpha_i \leq \sum_{i=1}^k \beta_i$ follows from $\sum_{i=1}^k f(\alpha_i) \leq \sum_{i=1}^k f(\beta_i)$ when $f(t) = t$, so it just remains to prove $\sum_{i=1}^k f(\alpha_i) \leq \sum_{i=1}^k f(\beta_i)$. Firstly, consider when all α_i and β_i are positive. Then

$$\prod_{i=1}^k \alpha_i \leq \prod_{i=1}^k \beta_i \text{ is equivalent to } \sum_{i=1}^k log\alpha_i \leq \sum_{i=1}^k log\beta_i$$

The inequalities we want to prove now follow from Lemma 2.65 using $f(e^t)$.

Now suppose not all α_i, β_i are positive. If $\beta_1, ..., \beta_p > 0$ and $\beta_{p+1} = ... = \beta_n = 0$, then $\alpha_{p+1} = ... = \alpha_n = 0$ and the inequalities $\sum_{i=1}^k \alpha_i \leq \sum_{i=1}^k \beta_i$ are valid for all k if they are valid for $k = 1, ..., p$. So it is sufficient to consider $\beta_i > 0$, $\alpha_1 \geq ... \geq \alpha_p > 0$, and $\alpha_{p+1} = ... = \alpha_n = 0$. The validity of the inequalities for $k = 1, ..., p$ has been established, so consider $k = p + 1, ..., n$. For $\delta > 0$, the validity of the product inequalities is preserved if $\alpha_{p+1}, ..., \alpha_n$ are replaced by ϵ for any $\epsilon \in (0, \delta]$. Since monotonicity of f implies that $f(\epsilon) \geq f(0)$ for any $\epsilon \in [0, \delta]$, we conclude that

$$\begin{aligned}\sum_{i=1}^{p+r} f(\beta_i) &\geq \sum_{i=1}^{p} f(\alpha_i) + \sum_{i=p+1}^{p+r} f(\epsilon) \\ &\geq \sum_{i=1}^{p} f(\alpha_i) + \sum_{i=p+1}^{p+r} f(0) \\ &= \sum_{i=1}^{p+r} f(\alpha_i)\end{aligned}$$

for $r = 1, ..., n - p$. □

Using this corollary, along with Weyl's Theorem (Theorem 2.60) we can state a variety of singular value inequalities:

Theorem 2.67. ([17], pp. 175-176) *Let* $A \in M_n(\mathbb{C})$ *have ordered singular values*

$\sigma_1(A) \geq ... \geq \sigma_n \geq 0$ and eigenvalues $[\lambda_1(A), ..., \lambda_n(A)]$ ordered so that $|\lambda_1(A)| \geq ... \geq |\lambda_n(A)|$. Then

(a) $$\sum_{i=1}^{k} |\lambda_i(A)| \leq \sum_{i=1}^{k} \sigma_i(A) \text{ for } k = 1, ..., n$$

In particular,

(a') $$|trA| \leq \sum_{i=1}^{n} \sigma_i(A)$$

(b) $$\sum_{i=1}^{k} |\lambda_i(A)|^p \leq \sum_{i=1}^{k} \sigma_i(A)^p \text{ for } k = 1, ..., n \text{ and any } p > 0$$

More generally, for any real-valued function f such that $\phi(t) \equiv f(e^t)$ is increasing and convex on $[\sigma_n(A), \sigma_1(A)]$, then

(c) $$\sum_{i=1}^{k} f(|\lambda_i(A)|) \leq \sum_{i=1}^{k} f(\sigma_i(A)) \text{ for } k = 1, ..., n$$

If if is a real-valued function such that $\phi(t) \equiv f(e^t)$ is convex, but not necessarily increasing on $[\sigma_n(A), \sigma_1(A)]$, then

(d) $$\sum_{i=1}^{n} f(|\lambda_i(A)|) \leq \sum_{i=1}^{n} f(\sigma_i(A))$$

In particular, if A is nonsingular, then

(e) $$\sum_{i=1}^{n} |\lambda_i(A)|^p \leq \sum_{i=1}^{n} \sigma_i(A)^p \text{ for all } p \in R$$

Example 2.68. *We'll illustrate the first three parts of this theorem with the following matrix: $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, which has ordered eigenvalues $\lambda_1 = 3$, $\lambda_2 = 2$, and $\lambda_3 = 1$. Also note, the singular values of A are found from*

$$AA^T = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}^T = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & -4 \\ 0 & -4 & 5 \end{pmatrix}$$

which gives us $\sigma_1 \approx \sqrt{9.92434}$, $\sigma_2 \approx \sqrt{4.21507}$, and $\sigma_3 \approx \sqrt{0.86056}$.

Then for (a), we see that $|\lambda_1(A) + \lambda_2(A) + \lambda_3(A)| = |3 + 2 + 1| = 6 \leq 6.131 \approx \sqrt{9.92434} + \sqrt{4.21507} + \sqrt{0.86056}$.

Likewise, (a)' is true since $|tr\ A| = 6$, as well.

(b) is true for any $p > 0$, so in this case we'll select $p = 2$. Then

$$\begin{aligned}|\lambda_1(A)^2 + \lambda_2(A)^2 + \lambda_3(A)^2| &= |9 + 4 + 1| = 14 \\ &\leq 14.99997 \approx 9.92434 + 4.21507 + 0.86056 \\ &\approx \sigma_1(A)^2 + \sigma_2(A)^2 + \sigma_3(A)^2\end{aligned}$$

To illustrate (c) we'll choose the convex function $f(x) = e^x$. Then

$$\begin{aligned}f(|\lambda_1(A)|) + f(|\lambda_2(A)|) + f(|\lambda_3(A)|) &\approx 20.0855 + 7.38905 + 2.71828 = 30.19283 \\ &\leq 32.475531 \\ &= 23.34288 + 7.791735 + 1.340916 \\ &\approx f(\sigma_1(A)) + f(\sigma_2(A)) + f(\sigma_3(A))\end{aligned}$$

By the same reasoning, and by the use of Theorem 2.62, we can deduce inequalities relating to singular values of a product.

Theorem 2.69. ([17], pp. 176-177) *Let $A \in M_{n,p}(\mathbb{C})$ and $B \in M_{p,m}(\mathbb{C})$ be given, let $q \equiv \min(n, p, m)$, and denote the ordered singular values of A, B, and AB by $\sigma_1(A) \geq ... \geq \sigma_{\min(n,p)} \geq 0$, $\sigma_1(B) \geq ... \geq \sigma_{\min(p,m)}(B) \geq 0$, and $\sigma_1(AB) \geq ... \geq \sigma_{\min(n,m)}(AB) \geq 0$. Then*

(a) $\sum_{i=1}^{k} \sigma_i(AB) \leq \sum_{i=1}^{k} \sigma_i(A)\sigma_i(B)$ *for* $k = 1, ..., q$

(b) $\sum_{i=1}^{k} [\sigma_i(AB)]^p \leq \sum_{i=1}^{k} [\sigma_i(A)\sigma_i(B)]^p$ *for* $k = 1, .., q$ *and any* $p > 0$

More generally, for any real-valued function f such that $\phi(t) \equiv f(e^t)$ is increasing and convex on the interval $[\min[\sigma_q(AB), \sigma_q(A)\sigma_q(B)], \sigma_1(A)\sigma_1(B)]$,

(c) $\sum_{i=1}^{k} f(\sigma_i(AB)) \leq \sum_{i=1}^{k} f(\sigma_i(A)\sigma_i(B))$, *for* $k = 1, ..., q$

If $m = n = p$ and if f is a real-valued function such that $\phi(t) \equiv f(e^t)$ is convex, but not necessarily increasing, on the interval $[\sigma_n(A)\sigma_n(B), \sigma_1(A)\sigma_1(B)]$, then

(d) $$\sum_{i=1}^{n} f(\sigma_i(AB)) \leq \sum_{i=1}^{n} f(\sigma_i(A)\sigma_i(B))$$

In particular, if $m = n = p$ and if A and B are nonsingular, then

(e) $$\sum_{i=1}^{n} \sigma_i(AB)^p \leq \sum_{i=1}^{n} \sigma_i(A)^p \sigma_i(B)^p \text{ for all } p \in R$$

A few interesting results involving singular value inequalities follow. The first is exercise 13 on p. 183 of [17].

Result 2.70. *Let $A \in M_{m,n}(\mathbb{C})$ and $X \in M_{n,k}(\mathbb{C})$ be given with $k \leq \min(m,n)$. Then $det(X^*A^*AX) \leq [\sigma_1(A)\cdots\sigma_k(A)]^2 det X^*X$.*

Proof. Using singular value decomposition, we can start by setting the determinant equal to a product of its singular values:

$$det(X^*A^*AX) = \sigma_1(X^*A^*AX)\cdots\sigma_k(X^*A^*AX)$$

We can now apply Theorem 2.62 to say

$$\begin{aligned}\sigma_1(X^*A^*AX)\cdots\sigma_k(X^*A^*AX) &\leq \sigma_1(X^*)\sigma_1(A^*A)\sigma_1(X)\cdots\sigma_k(X^*)\sigma_k(A^*A)\sigma_k(X)\\ &= [\sigma_1(A)\cdots\sigma_k(A)]^2\sigma_1(X^*)\sigma_1(x)\cdots\sigma_k(X^*)\sigma_k(X)\\ &\leq [\sigma_1(A)\cdots\sigma_k(A)]^2 det|X^*X|\end{aligned}$$

□

The next result is exercise 10 on pp. 182-183 of [17].

Result 2.71. *Let $A, B \in M_{m,n}(\mathbb{C})$ be given and let $q = \min(m,n)$. Using the inequalities of Theorem 2.67 and 2.69 we can show that $|tr(A^*B)| \leq \sum_{i=1}^{q} \sigma_i(A)\sigma_i(B)$.*

Proof. In general, by the above theorems, we have that

$$|trX| = |\sum_{i=1}^{n} \lambda_i(x)| \leq \sum_{i=1}^{n} |\lambda_i(x)|$$

So

$$|tr(A^*B)| = |\sum_{i=1}^{n} \lambda_i(A^*B)| \leq \sum_{i=1}^{n} |\lambda_i(A^*B)|.$$

Since

$$\sum_{i=1}^{n} |\lambda_i(A^*B)| \leq \sum_{i=1}^{n} \sigma_i(A^*B), \text{ by 2.67(a)},$$
$$\text{and } \sum_{i=1}^{n} \sigma_i(A^*B) \leq \sum_{i=1}^{n} \sigma_i(A^*)\sigma_i(B), \text{ by 2.69(a)}.$$

then we get $|tr(A^*B)| \leq \sum_{i=1}^{q} \sigma_i(A)\sigma_i(B)$. □

The next result is exercise 4 on p. 182 of [17].

Result 2.72. *Let $A_1, A_2, ..., A_m \in M_n(\mathbb{C})$ for some integer $m \geq 2$. Using the notation of Theorem 2.69, we may show that*

$$\sum_{i=1}^{k} \sigma_i(A_1 \cdots A_m) \leq \sum_{i=1}^{k} \sigma_i(A_1) \cdots \sigma_i(A_m) \text{ for } k = 1, ..., n$$

Proof. By Theorem 2.62, we know for $A \in M_{m,p}(\mathbb{C})$, $B \in M_{p,n}(\mathbb{C})$,

$$\prod_{i=1}^{k} \sigma_i(AB) \leq \prod_{i=1}^{k} \sigma_i(A)\sigma_i(B),\ n = p = m$$

If we can generalize this result, our result will follow. Consider $A_1, ..., A_m \in M_n$. By singular value decomposition, $A_1 \cdots A_m = V\Sigma W^*$. Then $V^*(A_1 \cdots A_m)W = \Sigma = diag(\sigma_1(A_1 \cdots A_m) \cdots \sigma_n(A_1 \cdots A_m))$.

Now applying Lemma 2.58, we can say the following:

$$\begin{aligned} \sigma_1(A_1 \cdots A_m) \cdots \sigma_n(A_1 \cdots A_m) &= |detV^*(A_1 \cdots A_m)W| \\ &\leq (\sigma_1(A_1) \cdots \sigma_n(A_1)) \cdots (\sigma_1(A_m) \cdots \sigma_n(A_m)) \end{aligned}$$

From here the result follows easily. □

3 Conjectures and New Ways of Looking at Singular Values

Now we look at some recent results involving singular values. These include inequalities of singular values derived by positive semidefinite block matrices (Tao) [21] and matrix monotone functions (Audenaert) [3]. These papers are presented in a logical order, since Tao's paper includes a proof of a special case of the result later proved by Audenaert.

Definition 3.1. *Heinz means are means that interpolate in a certain way between the arithmetic and geometric mean. They are defined over all positive reals as*

$$H_v(a,b) = (a^v b^{1-v} + a^{1-v} b^v)/2,$$

for $0 \leq v \leq 1$.

Note that it can be easily shown that the Heinz means are "in between" the geometric and arithmetic mean:

$$\sqrt{ab} \leq H_v(a,b) \leq (a+b)/2.$$

To see this, notice

$$H_v(a,b) = (a^v b^{1-v} + a^{1-v} b^v)/2 \geq \sqrt{a^v b^{1-v} a^{1-v} b^v} = \sqrt{ab}$$

and for the other side,

$$H_v(a,b) \leq \tfrac{1}{2}(\tfrac{a+a+...+a+b+...+b}{v+(1-v)})^{v+1-v} + (\tfrac{a+a+...+a+b+...+b}{(1-v)+v})^{(1-v)+v} = \tfrac{1}{2}(a+b).$$

This was then extended to the matrix case by Bhatia and Davis [7], who showed these inequalities remain true for PSD (positive semidefinite) matrices:

$$||A^{1/2}B^{1/2}|| \leq ||H_v(A,B)|| \leq ||(A+B)/2||,$$

where $||*||$ is any unitarily invariant norm and the Heinz means for matrices is defined identically as it was over the positive real numbers, namely $H_v(a,b) = \frac{A^v B^{1-v} + A^{1-v} B^v}{2}$. Zhan [22] conjectured that this inequality would also hold for singular values, i.e. for A, $B \geq 0$,

$$\sigma_j(H_v(A,B)) \leq \sigma_j((A+B)/2),$$

was conjectured to hold for all j. Tao proved this inequality in the special case where $v = 1/4$ (and $v = 3/4$) in the paper we'll examine below (section 3.1). Finally, Audenaert proved the inequality holds for all $0 \leq v \leq 1$. To do so, he first proved a general matrix inequality for matrix monotone functions, from which the proof of the conjecture is relatively straightforward.

Note that counterexamples for the first inequality here, $\sigma_j(A^{1/2}B^{1/2}) \leq \sigma_j(H_v(A,B))$, can be readily found. That is, the analog of $||A^{1/2}B^{1/2}|| \leq ||H_v(A,B)||$ does not hold for all PSD matrices.

Remark 3.2. *We'll present computer-generated counterexamples to the inequality $\sigma_j(A^{1/2}B^{1/2}) \leq \sigma_j(H_v(A,B))$ for both 2×2 and 3×3 PSD matrices A and B, both in the case where $v = .10$.*

(a) Consider $A = \begin{pmatrix} 49 & 14 \\ 14 & 53 \end{pmatrix}$ *and* $B = \begin{pmatrix} 100 & 70 \\ 70 & 50 \end{pmatrix}$. *Then*

$A^{1/2}B^{1/2} \approx \begin{pmatrix} 63.77 & 41.75 \\ 47.07 & 38.50 \end{pmatrix}$ *and* $H_v(A,B) \approx \begin{pmatrix} 182.06 & 126.91 \\ 130.43 & 93.94 \end{pmatrix}$. *Then when j=2 the inequality does not hold:*

$$\sigma_2(A^{1/2}B^{1/2}) \approx 5.039 > 2 = \sigma_2(H_v(A,B)).$$

(b) Consider $A = \begin{pmatrix} 144 & 56 & 80 \\ 56 & 69 & 40 \\ 80 & 40 & 80 \end{pmatrix}$ *and* $B = \begin{pmatrix} 21 & 44 & 10 \\ 44 & 153 & 53 \\ 10 & 53 & 37 \end{pmatrix}$. *Then*

$A^{1/2}B^{1/2} \approx \begin{pmatrix} 47.19 & 74.22 & 30.18 \\ 31.45 & 102.38 & 34.28 \\ 20.66 & 58.13 & 46.98 \end{pmatrix}$ *and* $H_v(A,B) \approx \begin{pmatrix} 50.18 & 105.16 & 27.84 \\ 79.98 & 278.75 & 95.35 \\ 22.22 & 109.12 & 76.99 \end{pmatrix}$. *When j=3 the inequality does not hold:*

$$\sigma_3(A^{1/2}B^{1/2}) \approx 19.05 > 13.78 \approx \sigma_3(H_v(A,B)).$$

3.1 Tao's Block Matrices

In [21] Tao looked at existing conjectures and results involving singular values and proved that these may in fact be converted into equivalent forms that involved block matrices, which he subsequently used to prove an existing conjecture by Zhan [22] in 2000.

Tao notes that the well known arithmetic-geometric mean inequalities proved by Bhatia and Kittaneh, namely,

$$2\sigma_j(AB^*) \leq \sigma_j(A^*A + B^*B), \text{ for } j = 1, 2, ..., n$$

for any $A, B \in M_n$ are equivalent to the following result by Zhan:

$$\sigma_j(A - B) \leq \sigma_j(A \oplus B), \text{ for } j = 1, 2, ..., n$$

where $A, B \in M_n$ are positive semidefinite and the direct sum $A \oplus B$ denotes the block diagonal matrix $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ (we'll prove this in Theorem 3.5). He goes on to show that these inequalities are equivalent to:

$$2\sigma_j(K) \leq \sigma_j\left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right), \text{ where } M \in M_m, N \in M_n, \text{ and } r = \min(m, n) \qquad .$$

First, we'll prove this inequality. Note the following definition:

Theorem 3.3. ([8], p. 63) (*Weyl's Monotonicity Principle*) *$A \leq B$ implies $\lambda_k(A) \leq \lambda_k(B)$.*

Theorem 3.4. *Given any positive semidefinite block matrix $\left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right)$, where $M \in M_m, N \in M_n$, and $r = \min(m, n)$. We have*

$$2\sigma_j(K) \leq \sigma_j\left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right) \text{ for } j = 1, ..., r$$

Proof. If $Q = \left(\begin{smallmatrix} 0 & K \\ K^* & 0 \end{smallmatrix}\right)$, observe that

$$0 \leq \left(\begin{smallmatrix} I_m & 0 \\ 0 & -I_n \end{smallmatrix}\right)\left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right)\left(\begin{smallmatrix} I_m & 0 \\ 0 & -I_n \end{smallmatrix}\right) = \left(\begin{smallmatrix} M & -K \\ -K^* & N \end{smallmatrix}\right) = \left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right) - 2Q,$$

thus $2Q \leq \left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right)$. Weyl's monotonicity principle says that for $G \leq H$, we have $\lambda_j(G) \leq \lambda_j(H)$ for $j = 1, 2, ..., n$. Hence, we may apply it here to conclude

$$2\lambda_j\left(\begin{smallmatrix} 0 & K \\ K^* & 0 \end{smallmatrix}\right) = 2\lambda_j(Q) \leq \lambda_j\left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right), \text{ for } j = 1, ..., m + n.$$

Since $\lambda(Q) = (\sigma_1(K), ..., \sigma_r(K), 0, ..., 0, -\sigma_r(K), ... - \sigma_1(K))^T$, we can obtain the desired result

$$2\sigma_j(K) \leq \sigma_j\begin{pmatrix} M & K \\ K^* & N \end{pmatrix} \text{ for } j = 1, ..., r.$$

$\square$

For an alternate proof, see [21].

We now establish that the earlier inequalities we introduced are equivalent to this one:

Theorem 3.5. *The following statements are equivalent:*

(i) Let $A, B \in M_n$ be positive semidefinite. Then

$$\sigma_j(A - B) \leq \sigma_j(A \oplus B),\ j = 1, 2, ..., n.$$

(ii) For any $X, Y \in M_n$,

$$2\sigma_j(XY^*) \leq \sigma_j(X^*X + Y^*Y),\ j = 1, ..., n.$$

(iii) Given any positive semidefinite block matrix $\begin{pmatrix} M & K \\ K^* & N \end{pmatrix}$, where $M, N \in M_n$, we have

$$2\sigma_j(K) \leq \sigma_j\begin{pmatrix} M & K \\ K^* & N \end{pmatrix},\ j = 1, ..., n.$$

Proof. $(i) \implies (ii)$. For any $X, Y \in M_n$, let $C = \begin{pmatrix} X \\ Y \end{pmatrix}$, $D = \begin{pmatrix} X \\ -Y \end{pmatrix}$. By (i),

$$\begin{aligned} 2\sigma_j\begin{pmatrix} YX^* & 0 \\ 0 & XY^* \end{pmatrix} &= 2\sigma_j\begin{pmatrix} 0 & XY^* \\ YX^* & 0 \end{pmatrix} \\ &= \sigma_j(CC^* - DD^*) \leq \sigma_j\begin{pmatrix} CC^* & 0 \\ 0 & DD^* \end{pmatrix} \\ &= \sigma_j\begin{pmatrix} C^*C & 0 \\ 0 & D^*D \end{pmatrix} \\ &= \sigma_j\begin{pmatrix} X^*X+Y^*Y & 0 \\ 0 & X^*X+Y^*Y \end{pmatrix}. \end{aligned}$$

To see these steps in more detail, notice $CC^* = \begin{pmatrix} XX^* & XY^* \\ YX^* & YY^* \end{pmatrix}$ and $DD^* = \begin{pmatrix} XX^* & -XY^* \\ -YX^* & YY^* \end{pmatrix}$, so $CC^* - DD^* = \begin{pmatrix} 0 & 2XY^* \\ 2YX^* & 0 \end{pmatrix}$.

So $\sigma_j(CC^* - DD^*) = \sigma_j\begin{pmatrix} 0 & 2XY^* \\ 2YX^* & 0 \end{pmatrix} = 2\sigma_j\begin{pmatrix} 0 & XY^* \\ YX^* & 0 \end{pmatrix} = 2\sigma_j(XY^*)$. On the other hand, $\sigma_j\begin{pmatrix} CC^* & 0 \\ 0 & DD^* \end{pmatrix} = \sigma_j\begin{pmatrix} C^*C & 0 \\ 0 & D^*D \end{pmatrix}$.

Observe $C^*C = X^*X + Y^*Y$ and $D^*D = X^*X + Y^*Y$.

Thus, we have $2\sigma_j(XY^*) \leq \sigma_j(X^*X + Y^*Y)$, for $j = 1, ..., n$.

$(ii) \implies (iii)$. Since $\begin{pmatrix} M & K \\ K^* & N \end{pmatrix} \geq 0$, there must exist $S, T \in M_{2n,m}$ such that

$$(S,T)^*(S,T) = \left(\begin{smallmatrix} S^*S & S^*T \\ T^*S & T^*T \end{smallmatrix}\right) = \left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right) \geq 0,\ j = 1,...,n.$$

(ii) also holds for rectangular matrices X, Y. By this version of (ii) we have

$$2\sigma_j(K) = 2\sigma_j(S^*T) \leq \sigma_j(SS^* + TT^*) = \sigma_j\left(\begin{smallmatrix} S^*S & S^*T \\ T^*S & T^*T \end{smallmatrix}\right) = \sigma_j\left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right),$$

$j = 1,...,n$.

$(iii) \implies (i)$. For any positive semidefinite matrices $A, B \in M_n$, note the following unitary similarity transforms:

$$\tfrac{1}{\sqrt{2}}\left(\begin{smallmatrix} I & I \\ -I & I \end{smallmatrix}\right)\left(\begin{smallmatrix} \frac{A+B}{2} & \frac{A-B}{2} \\ \frac{A-B}{2} & \frac{A+B}{2} \end{smallmatrix}\right)\tfrac{1}{\sqrt{2}}\left(\begin{smallmatrix} I & -I \\ I & I \end{smallmatrix}\right) = \left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right) \geq 0.$$

Now the desired result follows from (iii)

$$\sigma_j(A-B) \leq \sigma_j\left(\begin{smallmatrix} \frac{A+B}{2} & \frac{A-B}{2} \\ \frac{A-B}{2} & \frac{A+B}{2} \end{smallmatrix}\right) = \sigma_j\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right),$$

$j = 1,...,n$. □

Tao now proves the following conjecture by Zhan [22]

$$\sigma_j(A^r B^{1-r} + A^{1-r} B^r) \leq \sigma_j(A+B), \text{ for } j = 1,...,n$$

in the case where $r = \frac{1}{4}$.

Theorem 3.6. *Let $A, B \in M_n$ be positive semidefinite and m be a positive integer. Then,*

$$2\sigma_j(A^{1/2}(A+B)^{m-1}B^{1/2}) \leq \sigma_j((A+B)^m), \textit{ for } j = 1,...,n,$$

$$\sigma_j(A^{1/4}B^{3/4} + A^{3/4}B^{1/4}) \leq \sigma_j(A+B), \textit{ for } j = 1,...,n.$$

Proof. Let X $= \left(\begin{smallmatrix} A^{1/2} & 0 \\ B^{1/2} & 0 \end{smallmatrix}\right)$. Then,

$$(X^*X)^m = \left(\begin{smallmatrix} (A+B)^m & 0 \\ 0 & 0 \end{smallmatrix}\right).$$

$$(XX^*)^m = X(X^*X)^{m-1}X^* = \left(\begin{smallmatrix} A^{1/2}(A+B)^{m-1}A^{1/2} & A^{1/2}(A+B)^{m-1}B^{1/2} \\ B^{1/2}(A+B)^{m-1}A^{1/2} & B^{1/2}(A+B)^{m-1}B^{1/2} \end{smallmatrix}\right).$$

Then from Theorem 3.5(iii),

$$2\sigma_j(A^{1/2}(A+B)^{m-1}B^{1/2}) \leq \sigma_j((XX^*)^m) = \sigma_j((X^*X)^m) = \sigma_j((A+B)^m),$$

for $j = 1,...,n$.

When $m = 2$, then

$$2\sigma_j(A^{1/2}B^{1/2} + A^{3/2}B^{1/2}) \leq \sigma_j((A+B)^2)),\ j = 1,...,n.$$

Since $0 \leq (A-B)^2 = A^2 + B^2 - AB - BA$ we derive $(A+B)^2 \leq 2(A^2+B^2)$. Hence,

$$2\sigma_j(A^{1/2}B^{3/2} + A^{3/2}B^{1/2}) \leq \sigma_j((A+B)^2) \leq 2\sigma_j(A^2+B^2).$$

Now let $A^{1/2}, B^{1/2}$ substitute for A and B, respectively. Then,

$$\sigma_j(A^{1/4}B^{3/4} + A^{3/4}B^{1/4}) \leq \sigma_j(A+B),\ j = 1, ..., n.$$

$\square$

Finally, Tao proves an interesting theorem involving singular values for block matrices.

Lemma 3.7. *Let $A, B \in M_n$ be positive semidefinite matrices and let f, g be nonnegative functions on $[0,\infty)$ which are continuous and satisfy the relation $f(t)g(t) = t$ for all $t \in [0,\infty)$. Then for $j = 1, ..., n$*

(i) $2\sigma_j(A+B) \leq \left(\begin{smallmatrix} f^2(A)+g^2(A)+|g(B)f(A)+f(B)g(A)| & 0 \\ 0 & f^2(B)+g^2(B)+|g(A)f(B)+f(A)g(B)| \end{smallmatrix}\right)$

(ii) $2\sigma_j\left(\begin{smallmatrix} A & f(A)f(B) \\ g(B)g(A) & B \end{smallmatrix}\right) \leq \left(\begin{smallmatrix} |f(A)|^2+|g(A)|^2+|g(B)|^2+|f(B)|^2 & 0 \\ 0 & 0 \end{smallmatrix}\right)$

For a proof of this lemma, see [15].

Theorem 3.8. *Let $A, B \in M_n$ be positive semidefinite. Then,*

$$\begin{aligned}\sigma_j\left(\begin{smallmatrix} A^{1/2}B^{1/2} & A^{1/2}B^{1/2} \\ A^{1/2}B^{1/2} & A^{1/2}B^{1/2} \end{smallmatrix}\right) &= 2\sigma_j(A^{1/2}B^{1/2}) \\ &\leq \sigma_j\left(\begin{smallmatrix} A & A^{1/2}B^{1/2} \\ B^{1/2}A^{1/2} & B \end{smallmatrix}\right) \\ &\leq \sigma_j(A+B) \\ &\leq \sigma_j\left(\begin{smallmatrix} A+|B^{1/2}A^{1/2}| & 0 \\ 0 & B+|A^{1/2}B^{1/2}| \end{smallmatrix}\right)\end{aligned}$$

for $j = 1, ..., n$.

Proof. For the equality identity, observe that

$$\tfrac{1}{\sqrt{2}}\left(\begin{smallmatrix} I & I \\ -I & I \end{smallmatrix}\right)\left(\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right)\tfrac{1}{\sqrt{2}}\left(\begin{smallmatrix} I & -I \\ I & I \end{smallmatrix}\right) = \left(\begin{smallmatrix} A+B & 0 \\ 0 & A-B \end{smallmatrix}\right).$$

That is, $\left(\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} A+B & 0 \\ 0 & A-B \end{smallmatrix}\right)$ are unitarily equivalent. Likewise, $\left(\begin{smallmatrix} A^{1/2}B^{1/2} & A^{1/2}B^{1/2} \\ A^{1/2}B^{1/2} & A^{1/2}B^{1/2} \end{smallmatrix}\right)$

and $\left(\begin{smallmatrix} 2A^{1/2}B^{1/2} & 0 \\ 0 & 0 \end{smallmatrix}\right)$ are unitarily equivalent. If we apply Theorem 3.5(iii) and note that

$$\left(\begin{smallmatrix} A & A^{1/2}B^{1/2} \\ B^{1/2}A^{1/2} & B \end{smallmatrix}\right) = [A^{1/2}, B^{1/2}]^*[A^{1/2}, B^{1/2}] \geq 0,$$

we may derive the first inequalities $(2\sigma_j(A^{1/2}B^{1/2}) \leq \sigma_j\left(\begin{smallmatrix} A & A^{1/2}B^{1/2} \\ B^{1/2}A^{1/2} & B \end{smallmatrix}\right))$. The second inequalities may be derived by using Lemma 3.7(ii) with the nonnegative functions $f(t) = g(t) = t^{1/2}$. Likewise, the last inequalities follow from Lemma 3.7(i) with $f(t) = g(t) = t^{1/2}$. $\square$

3.2 Audenaert's Matrix Monotone Functions

In [3], Audenaert proves a matrix inequality for matrix monotone functions, and uses it to prove a singular value inequality for Heinz means conjectured by Zhan [22].

Here, Audenaert proves the inequality will hold for all $0 \leq v \leq 1$. To do so, he first proves a general matrix inequality for matrix monotone functions, from which the proof of the conjecture is relatively straightforward.

We begin by noting that $A \geq B$ implies $XAX^* \geq XBX^*$, for all A, $B \geq 0$ and where f is matrix convex,

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B).$$

Matrix monotone functions, as discussed previously (see Theorem 2.31) are characterized by the integral representation

$$f(t) = \alpha + \beta t + \textstyle\int_0^\infty \frac{\lambda t}{t+\lambda} d\mu(\lambda),$$

where $d\mu(\lambda)$ is any positive measure on the interval $\lambda \in [0, \infty)$, α is a real scalar and β is a non-negative scalar. When applied to matrices, this gives, for $A \geq 0$,

$$f(A) = \alpha I + \beta A + \textstyle\int_0^\infty \lambda A (A + \lambda I)^{-1} d\mu(\lambda)$$

The primary function $x \mapsto x^p$ is matrix convex for $1 \leq p \leq 2$, matrix monotone and matrix concave for $0 \leq p \leq 1$, and inversely matrix monotone and matrix convex for $-1 \leq p \leq 0$.

With this in mind, we can present the proof of the matrix inequality which precedes the proof of Zhan's conjecture.

Theorem 3.9. *For A, $B \geq 0$, and any matrix monotone function f:*

$$Af(A) + Bf(B) \geq (\tfrac{A+B}{2})^{1/2}(f(A) + f(B))(\tfrac{A+B}{2})^{1/2}.$$

Proof. Let A and B be PSD. $f : t \mapsto t^{-1}$ is matrix convex. It follows that

$$\tfrac{A^{-1}+B^{-1}}{2} \geq (\tfrac{A+B}{2})^{-1}.$$

If we replace A by $A + I$ and B by $B + I$, we get

$$(A + I)^{-1} + (B + I)^{-1} \geq 2(I + (A + B)/2)^{-1}.$$

Let us now define

$$C_k \equiv A^k(A+I)^{-1} + B^k(B+I)^{-1},$$

and

$$M = (A+B)/2.$$

If we apply these notations to $(A+I)^{-1} + (B+I)^{-1} \geq 2(I+(A+B)/2)^{-1}$, we get

$$C_0 \geq 2(1+M)^{-1}.$$

Thus

$$\begin{aligned} C_0 + \sqrt{M}C_0\sqrt{M} &\geq 2(I+M)^{-1} + 2\sqrt{M}(I+M)^{-1}\sqrt{M} \\ &= 2(I+M)^{-1} + 2M(I+M)^{-1} \\ &= 2 + 2M(I+M)^{-1} = 2(I+M)(I+M)^{-1} = 2I. \end{aligned}$$

Since $C_k + C_{k+1} = A^k + B^k$ (in particular, $C_0 + C_1 = 2I$) this reduces to

$$\sqrt{M}(2I - C_1)\sqrt{M} \geq C_1.$$

Further, since $C_1 + C_2 = 2M$, this is equivalent to

$$C_2 \geq \sqrt{M}C_1\sqrt{M},$$

or, written out in full:

$$A^2(A+I)^{-1} + B^2(B+I)^{-1} \geq (\tfrac{A+B}{2})^{1/2}(A(A+I)^{-1} + B(B+I)^{-1})(\tfrac{A+B}{2})^{1/2}$$

holds for all $\lambda \geq 0$.

If we replace A by $\lambda^{-1}A$ and B by $\lambda^{-1}B$, for a positive scalar λ, and if we multiply both sides by λ^2, we obtain

$$\begin{gathered} \lambda A^2(A+\lambda I)^{-1} + \lambda B^2(B+\lambda I)^{-1} \\ \geq (\tfrac{A+B}{2})^{1/2}(\lambda A(A+\lambda I)^{-1} + \lambda B(B+\lambda I)^{-1})(\tfrac{A+B}{2})^{1/2} \end{gathered}$$

holds for all $\lambda \geq 0$. Hence we can integrate this inequality over $\lambda \in [0,\infty)$ using any positive measure $d\mu(\lambda)$.

By matrix convexity of the square function, $((A+B)/2)^2 \leq (A^2+B^2)/2$, we have, for $\beta \geq 0$,

$$A(\alpha I + \beta A) + B(\alpha I + \beta B) \geq (\tfrac{A+B}{2})^{1/2}(2\alpha I + \beta(A+B))(\tfrac{A+B}{2})^{1/2}.$$

If we sum this with the integral expression just obtained, and recognize the representation $f(A) = \alpha I + \beta A + \int_0^\infty \lambda A(A+\lambda I)^{-1} d\mu(\lambda)$ in both sides, we obtain the desired result. $\square$

Using Weyl monotonicity, we get the following corollary.

Corollary 3.10. *For A, $B \geq 0$, and any matrix monotone function f:*

$$\lambda_j(Af(A) + Bf(B)) \geq \lambda_j(\tfrac{A+B}{2}(f(A) + f(B))).$$

Now we may now present Audenaert's proof of Zhan's conjecture.

Theorem 3.11. *For A, $B \in M_n(\mathbb{C})$, A, $B \geq 0$, $j = 1, ..., n$, and $0 \leq s \leq 1$,*

$$\sigma_j(A^s B^{1-s} + A^{1-s} B^s) \leq \sigma_j(A + B).$$

Proof. Applying Corollary 3.10 to $f(A) = A^r$, for $0 \leq r \leq 1$, yields

$$\begin{aligned}
\lambda_j(A^{r+1} + B^{r+1}) &\geq \frac{1}{2}\lambda_j((A+B)(A^r + B^r)) \\
&= \frac{1}{2}\lambda_j((A^{r/2}B^{r/2})^T(A+B)(A^{r/2}B^{r/2})) \\
&= \frac{1}{2}\lambda_j((A^{1/2}B^{1/2})^T(A^r + B^r)(A^{1/2}B^{1/2}))
\end{aligned}$$

Note: the above steps followed using $(A^r + B^r)$=$[A^{r/2}\ B^{r/2}][A^{r/2}\ B^{r/2}]^T$ and likewise $(A+B)$=$[A^{1/2}\ B^{1/2}][A^{1/2}\ B^{1/2}]^T$.

Using Tao's Theorem (Theorem 3.5), it can be said that, for a 2×2 PSD block matrix $Z = \left(\begin{smallmatrix} M & K \\ K^* & N \end{smallmatrix}\right) \geq 0$ (with $M \in M_m$ and $N \in M_n$) the following relation holds between the singular values of the off-diagonal block K and the eigenvalues of Z, for $j \leq m$,n:

$$\sigma_j(K) \leq \tfrac{1}{2}\lambda_j(Z).$$

The inequality therefore yields

$$\lambda_j(A^{r+1} + B^{r+1}) \geq \sigma_j(A^{r/2}(A+B)B^{r/2}) = \sigma_j(A^{(1+r/2}B^{r/2} + A^{r/2}B^{1+r/2}).$$

Replacing A by $A^{1/(r+1)}$ and B by $B^{1/(r+1)}$ then yields the desired result for $s = (1+r/2)/(1+r)$, hence for $0 \leq s \leq 1/4$ and $3/4 \leq s \leq 1$.

If, instead, we start from the derived inequality and proceed in an identical way as above, then we obtain the result for $s = (r+1/2)/(1+r)$, which covers the remaining case of $1/4 \leq s \leq 3/4$. $\square$

3.3 A Recent Singular Value Inequality by Drury

A long-standing operator arithmetic-geometric mean inequality (conjectured in 1990 by Bhatia and Kittaneh [6]) that has received enormous attention was recently proved by Drury [13] in 2012. A proof of this conjecture had eluded mathematicians for several years, so Drury's proof is quite significant. In a brief article in support of his paper [14], Drury remarks that he "nearly fell out of his chair" discovering an important element in his proof one day, seemingly by chance.

The inequality of interest is the generalization of the two variable arithmetic geometric mean inequality

$$\sqrt{ab} \leq \tfrac{1}{2}(a+b) \text{ for a, b} \geq 0$$

to the singular value setting. The generalization Drury proves is, for positive semidefinite $n \times n$ matrices P and Q, the inequality

$$\sqrt{\sigma_r(PQ)} \leq \tfrac{1}{2}\lambda_r(P+Q)$$

holds for $r = 1, 2, ..., n$.

Example 3.12. *To visualize how this inequality functions before we proceed further, consider the PSD matrices $P = \left(\begin{smallmatrix} 49 & 14 \\ 14 & 53 \end{smallmatrix}\right)$ and $Q = \left(\begin{smallmatrix} 100 & 70 \\ 70 & 50 \end{smallmatrix}\right)$. Since these are 2 dimensional matrices, they'll each yield 2 respective pairs of singular values and eigenvalues. Since these are PSD matrices, we can write them with the respective diagonalizations*

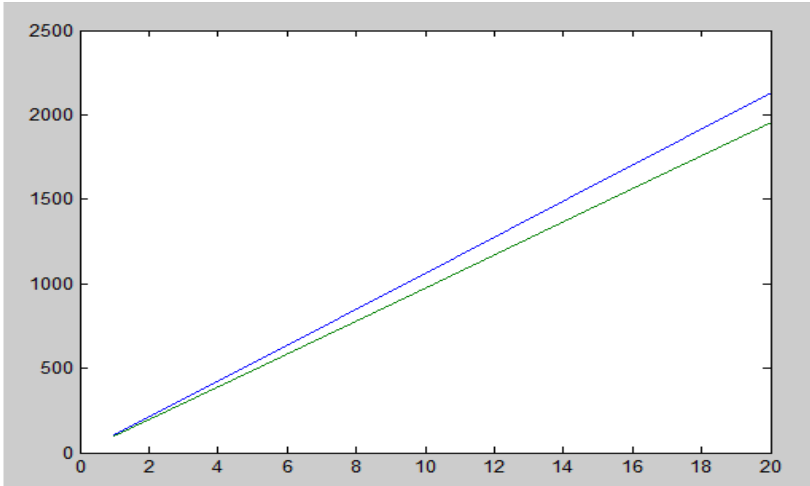
$$\left(\begin{smallmatrix} 49 & 14 \\ 14 & 53 \end{smallmatrix}\right) = \left(\begin{smallmatrix} \frac{1}{7}(-1-5\sqrt{2}) & \frac{1}{7}(-1+5\sqrt{2}) \\ 1 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 51-10\sqrt{2} & 0 \\ 0 & 51+10\sqrt{2} \end{smallmatrix}\right)\left(\begin{smallmatrix} \frac{1}{7}(-1-5\sqrt{2}) & \frac{1}{7}(-1+5\sqrt{2}) \\ 1 & 1 \end{smallmatrix}\right)^{-1}$$

and

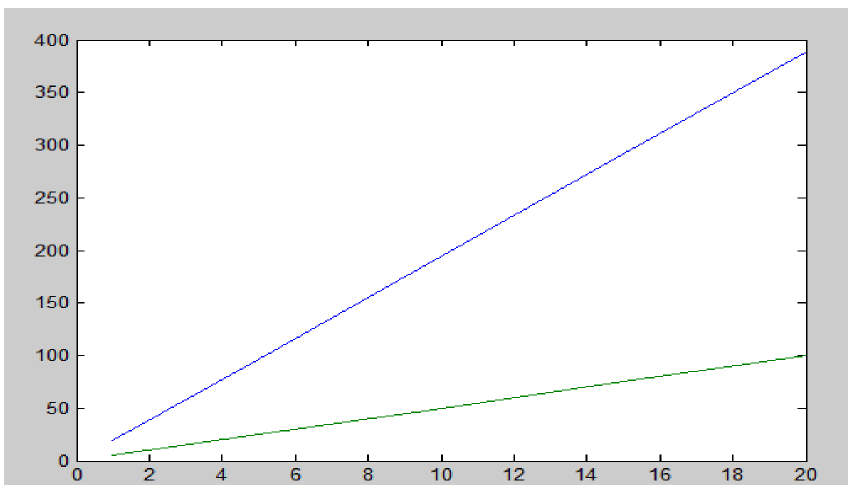
$$\left(\begin{smallmatrix} 100 & 70 \\ 70 & 50 \end{smallmatrix}\right)$$

$$= \left(\begin{smallmatrix} \frac{1}{14}(5-\sqrt{221}) & \frac{1}{14}(5+\sqrt{221}) \\ 1 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} -5(-15+\sqrt{221}) & 0 \\ 0 & 5(15+\sqrt{221}) \end{smallmatrix}\right)\left(\begin{smallmatrix} \frac{1}{14}(5-\sqrt{221}) & \frac{1}{14}(5+\sqrt{221}) \\ 1 & 1 \end{smallmatrix}\right)^{-1}.$$

Then, we may scale the diagonal matrices in these decompositions by 1, 5, 10, 15, and 20, respectively, and plot the resulting eigenvalues and singular values that they produce for the left and right hand sides of this inequality. For the larger set of singular values and eigenvalues, this yielded



That is, we can notice that the rate of change is such that the right hand side of the inequality is increasing at a faster rate, as we scale the diagonals of both matrices. Similarly, looking at the smaller set of points, we see that this is even more pronounced:



Now we'll proceed to prove the inequality. To do so, Drury first proves several preliminary results.

To begin, we'll introduce some notation. For two positive definite matrices B and X of the same size, denote the geometric mean of B and X as $B\#X$. This is the unique positive definite matrix such that $B = (B\#X)X^{-1}(B\#X)$ or equivalently $X = (B\#X)B^{-1}(B\#X)$. Then the geometric mean is given by $B\#X = B^{1/2}(B^{-1/2}XB^{-1/2})^{1/2}B^{1/2}$ or equivalently $B\#X = X^{1/2}(X^{-1/2}BX^{-1/2})^{1/2}X^{1/2}$. This is symmetric, i.e. $B\#X = X\#B$.

Example 3.13. *Consider the PSD matrices* $B = \left(\begin{smallmatrix} 49 & 14 \\ 14 & 53 \end{smallmatrix}\right)$ *and* $X = \left(\begin{smallmatrix} 100 & 70 \\ 70 & 50 \end{smallmatrix}\right)$. *Then the geometric mean is given by*

$$B\#X = B^{1/2}(B^{-1/2}XB^{-1/2})^{1/2}B^{1/2} \approx \left(\begin{smallmatrix} 65.508 & 43.3884 \\ 43.3884 & 36.2128 \end{smallmatrix}\right)$$

Lemma 3.14. *For all PSD matrices* A, $\lambda_j(A + A^{-1}) \geq 2$ *for all* j.

Proof. To demonstrate this, we'll consider a result in [9]; namely

$$\lambda_j(AB) \leq \lambda_j(\tfrac{A^2+B^2}{2}), \text{ for all } j.$$

Consider if we take A to be $A^{1/2}$ and B to be $A^{-1/2}$. Then this inequality becomes

$$2\lambda_j(A^{1/2}A^{-1/2}) \leq \lambda_j(A + A^{-1})$$

which simplifies to the desired result. ☐

Result 3.15. *Let* B *and* X *be positive definite* $r \times r$ *matrices. Let*

$$R = \left(\begin{smallmatrix} B & (B\#X)^{-1} \\ (B\#X)^{-1} & X \end{smallmatrix}\right)$$

be a $2r \times 2r$ *matrix. Then* $\lambda_r(R) \geq 2$.

Proof. Let $S = B\#X$. Using $B = (B\#X)X^{-1}(B\#X)$ with Theorem 2.47, we see that

$$R_1 = \left(\begin{smallmatrix} B & -S \\ -S & X \end{smallmatrix}\right)$$

is positive semidefinite, and in fact has rank r. Then

$$R - R_1 = \left(\begin{smallmatrix} 0 & S+S^{-1} \\ S+S^{-1} & 0 \end{smallmatrix}\right).$$

The eigenvalues of $S+S^{-1}$ are all ≥ 2 since S is positive semidefinite. The eigenvalues of $R - R_1$ are the eigenvalues of $S + S^{-1}$ and their negatives. Therefore $R - R_1$ has exactly r eigenvalues ≥ 2, hence R has at least r eigenvalues ≥ 2. ☐

A well known result involving the determinant of a block matrix is presented below [13].

Result 3.16. *Let* M_{11}, M_{12}, M_{21}, *and* M_{22} *be* $r \times r$ *matrices and assume that* M_{12} *is invertible. If*

$$M = \left(\begin{smallmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{smallmatrix}\right),$$

then $det(M) = det(M_{12}M_{22}M_{12}^{-1}M_{11} - M_{12}M_{21})$.

Corollary 3.17. *If* M_{11}, M_{12}, M_{21}, M_{22}, *and* M *are as above, then*

$$\begin{aligned} &det(\lambda I - M) \\ &= det(\lambda^2 I - \lambda(M_{11} + M_{12}M_{22}M_{12}^{-1}) + (M_{12}M_{22}M_{12}^{-1}M_{11} - M_{12}M_{21})). \end{aligned}$$

With this, Drury proves the following result.

Result 3.18. *Let* A *and* B *be* $r \times r$ *positive definite matrices, and let* Z *be an* $r \times r$ *matrix such that* $BA(I + ZZ^*)AB = I$. *Let*

$$T = \left(\begin{smallmatrix} A+B & AZ \\ Z^*A & Z^*AZ \end{smallmatrix}\right).$$

Then $det(\lambda I - T) = det(\lambda^2 I - \lambda(B + B^{-1}A^{-1}B^{-1}) + (B^{-1}A^{-1} - BA))$.

Proof. The hypotheses imply that $BA^2B \leq I$ or equivalently $A^{-1}B^{-2}A^{-1} \geq I$. First, assume Z is nonsingular. Then Corollary 3.17 will apply with $M = T$. Note that $M_{12}M_{22}M_{12}^{-1} = AZZ^*AZ(AZ)^{-1} = AZZ^* = B^{-2}A^{-1} - A$. Then

$$\begin{aligned} det(\lambda I - T) &= det(\lambda^2 I - \lambda(A + B + AZZ^*) + (AZZ^*(A + B) - AZZ^*A)) \\ &= det(\lambda^2 I - \lambda(A + B + AZZ^*) + AZZ^*B) \\ &= det(\lambda^2 I - \lambda(B + B^{-2}A^{-1}) + (B^{-2}A^{-1}B - AB)) \\ &= det(\lambda^2 I - \lambda(B + B^{-1}A^{-1}B^{-1}) + (B^{-1}A^{-1} - BA)) \end{aligned}$$

using a similarity for the last step (multiplying the left of a term by B and the right by B^{-1}). This completes the proof in the case that Z is nonsingular.

In the general case, observe that, without loss of generality, Z may be replaced by $W = (A^{-1}B^{-2}A^{-1} - I)^{1/2}$, which we found by solving for Z in $BA(I + ZZ^*)AB = I$. In fact, the polar decomposition (explained in Theorem 2.10) of Z is $Z = WU$ where U is unitary. Then

$$T = \left(\begin{smallmatrix} A+B & AWU \\ U^*W^*A & U^*W^*AWU \end{smallmatrix}\right) \text{ and } \left(\begin{smallmatrix} A+B & AW \\ W^*A & W^*AW \end{smallmatrix}\right)$$

are unitarily similar via the unitary $\left(\begin{smallmatrix} I & 0 \\ 0 & U^* \end{smallmatrix}\right)$ and therefore have the same characteristic polynomial. Next, we approximate B by $B_k = \mu_k B$ where $0 < \mu_k < 1$ and where μ_k increases to 1. Note that the corresponding $W_k = (A^{-1}B_k^{-2}A^{-1} - I)^{1/2}$ is now invertible (as it is positive definite). This follows using the following arguments: Since $BA^2B \leq I$, then $B_kA^2B_k = \mu_k^2 BA^2B \leq \mu_k^2 I < I$ where the last inequality is easily obtained from $0 < \mu_k < 1$. Hence $B_kA^2B_k < I$ or equivalently $A^{-1}B_k^{-2}A^{-1} > I$. Then $A^{-1}B_k^{-2}A^{-1} - I > 0$ which implies that $W_k = (A^{-1}B_k^{-2}A^{-1} - I)^{1/2} > 0$, so we conclude that W_k is invertible.

If we apply the previous argument to the approximating sequence, we obtain

$$det(\lambda I - T_k) = det(\lambda^2 I - \lambda(B_k + B_k^{-1}A^{-1}B_k^{-1}) + (B_k^{-1}A^{-1} - B_kA))$$

where

$$T_k = \left(\begin{smallmatrix} A+B & AW_k \\ W_k^*A & W_k^*AW_k \end{smallmatrix}\right)$$

Then taking the limit as k approaches ∞ on both sides will obtain the desired result.

□

Again using Corollary 3.17, we can establish another result.

Result 3.19. *Let B and S be $r \times r$ positive definite matrices. Let $X = SB^{-1}S$ (so that $S = B\#X$) and*

$$R = \left(\begin{smallmatrix} B & S^{-1} \\ S^{-1} & X \end{smallmatrix}\right).$$

Then $det(\lambda I - R) = det(\lambda^2 I - \lambda(B + B^{-1}S^2) + (B^{-1}S^2B - S^{-2}))$.

Proof. Note that since S^{-1} is invertible, the proof becomes much simpler than the above result.

$$\begin{aligned} det(\lambda I - R) &= det(\lambda^2 I - \lambda(B + S^{-1}XS) + (S^{-1}XSB - S^{-1}S^{-1})) \\ &= det(\lambda^2 I - \lambda(B + S^{-1}(SB^{-1}S)S) + (S^{-1}(SB^{-1}S)SB - S^{-2})) \\ &= det(\lambda^2 I - \lambda(B + B^{-1}S^2) + (B^{-1}S^2B - S^{-2})) \end{aligned}$$

□

Theorem 3.20. *Let A and B be $r\times r$ positive definite matrices, and let Z be an $r\times r$ matrix such that $BA(I+ZZ^*)AB = I$. Then*

$$\lambda_r\left(\begin{smallmatrix} A+B & AZ \\ Z^*A & Z^*AZ \end{smallmatrix}\right) \geq 2.$$

Proof. Let $S = (B^{1/2}A^{-1}B^{1/2})^{1/2}$ and let $X = SB^{-1}S$ as in the above result, and let R also be defined as there. If we again have $T = \left(\begin{smallmatrix} A+B & AZ \\ Z^*A & Z^*AZ \end{smallmatrix}\right)$, then

$$\begin{aligned}
&det(\lambda I - R)\\
&= det(\lambda^2 I - \lambda(B + B^{3/2}A^{-1}B^{-1/2}) + (B^{-3/2}A^{-1}B^{-1/2}B - B^{1/2}AB^{1/2}))\\
&= det(\lambda^2 I - \lambda(B + B^{-1}A^{-1}B^{-1}) + (B^{-1}A^{-1} - BA))\\
&= det(\lambda I - T)
\end{aligned}$$

Thus R and T have the same eigenvalues, and using Result 3.15 it follows that $\lambda_r(T) \geq 2$. □

Definition 3.21. *Before proceeding, we should introduce the* standard continuity argument*; that is, in many cases to prove some conclusion on positive semidefinite matrices, it suffices to show it for positive definite matrices by considering $B = A + \epsilon I$ as ϵ approaches zero (since $(A+\epsilon I) - A = \epsilon I > 0$, we have that $A + \epsilon I > A \geq 0$), where A is positive semi-definite and B is positive definite.*

Now Drury proceeds to resolve the question. Without loss of generality, we can assume P is positive definite (and hence invertible) since the general case is obtainable by approximating with such matrices (the standard continuity argument). Fix r in the range $1 \leq r \leq n$ and we normalize so that $\sigma_r(PQ) = 1$; that is, divide P and Q by $\sqrt{\sigma_k(PQ)}$, then $P \equiv P/\sqrt{\sigma_k(PQ)}$ and $Q \equiv Q/\sqrt{\sigma_k(PQ)}$. Thus the objective is to show that $\lambda_r(P+Q) \geq 2$. Using the definition of singular values, restate $\sigma_r(PQ)$ as $\lambda_r(PQ^2P) = 1$.

Let $u, v \in \mathbb{R}^n$, so that $u = [u_1 ... u_n]^T$ and $v = [v_1 ... v_n]^T$. We'll use the notation

$u \otimes v^* = [u_1 ... u_n]^T \otimes [v_1 ... v_n] = \left(\begin{smallmatrix} u_1v_1 & u_1v_2 & ... & u_1v_n \\ ... & ... & ... & ... \\ u_nv_1 & u_nv_2 & ... & u_nv_n \end{smallmatrix}\right) \in M_n$. Using Corollary 2.5, we get $PQ^2P = U\Sigma U^*$, where $U = [v_1 ... v_n] \in M_n$ with $(e_1, ..., e_n)$ an orthonormal basis of eigenvectors, and Σ is a diagonal matrix with entries $\lambda_1(PQ^2P), ..., \lambda_n(PQ^2P)$. Note that, using the tensor notation above, we can now write

$$PQ^2P = \textstyle\sum_{k=1}^n \lambda_k(PQ^2P)(e_k \otimes e_k^*).$$

Thus $\lambda_k(PQ^2P) \geq 1$ for $k = 1, 2, ..., r$ and $\lambda_k(PQ^2P) \leq 1$ for $k = r, ..., n$.

We now define a positive semidefinite matrix Q_1 by

$$Q_1 = (P^{-1}(\textstyle\sum_{k=1}^r e_k \otimes e_k^*)P^{-1})^{1/2}$$

Then

$$\begin{aligned} Q_1^2 &= P^{-1}(\sum_{k=1}^r e_k \otimes e_k^*)P^{-1} \\ &\leq P^{-1}(\sum_{k=1}^r \lambda_k(PQ^2P)e_k \otimes e_k^*)P^{-1} \\ &\leq P^{-1}(\sum_{k=1}^n \lambda_k(PQ^2P)e_k \otimes e_k^*)P^{-1} = Q^2 \end{aligned}$$

i.e. $Q_1^2 \leq Q^2$. Now, we can use the fact that the square root is a matrix monotone function to assert that $Q_1 \leq Q$. This is a special case of the Lowner-Heinz inequality. Therefore, $P + Q_1 \leq P + Q$ and if the statement $\lambda_r(P + Q_1) \geq 2$ is true then $\lambda_r(P+Q) \geq 2$ follows. Therefore, without loss of generality, we will use the expression $Q_1 = [P^{-1}(\sum_{k=1}^r e_k \otimes e_k^*)P^{-1}]^{-1}$ for Q.

Since P is invertible, it can be concluded that Q has rank r, since Q can be written as the sum from 1 to r of the product of orthonormal vectors. If $Q : S \mapsto S$, say, then split this space of Q as $S = \ker Q \oplus \text{range}(Q)$ (note that the dimension of the range of Q equals r, the rank of Q). Then after applying a unitary similarity, we can assume that

$$Q = \left(\begin{smallmatrix} Q_{11} & 0 \\ 0 & 0 \end{smallmatrix}\right) \text{ and } P = \left(\begin{smallmatrix} P_{11} & P_{12} \\ P_{12}^* & P_{22} \end{smallmatrix}\right)$$

where the top blocks are size $r \times r$ and the bottom are $(n - r) \times (n - r)$. Note that P_{11} is necessarily invertible. Now by Corollary 2.14 we know that PQ^2P and QP^2Q

are both of rank r, and thus we see that

$$Q_{11}(P_{11}^2 + P_{12}P_{12}^*)Q_{11} = I$$

since this is the only non-zero block of QP^2Q. It possesses full rank and is thus similar to I.

Let

$$P_1 = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^* & P_{12}^*P_{11}^{-1}P_{12} \end{pmatrix}$$

Then P_1 has rank r, satisfies $P_1 \leq P$ (by Schur complements) and $QP^2Q = QP_1^2Q$, so that $\lambda_r(P+Q) \geq \lambda_r(P_1+Q)$. Hence, it may be assumed that $P_{22} = P_{12}^*P_{11}^{-1}P_{12}$ at the expense of no longer being able to assert that P is necessarily invertible (it's simply PSD now, i.e. assume P looks like P_1).

Now, Drury obtains matrices A, B, and Z for which Theorem 3.20 can be applied. This depends on the relative sizes of n and r.

If $n = 2r$, we set $A = P_{11}$, $B = Q_{11}$, and $Z = P_{11}^{-1}P_{12}$. So if we check that this satisfies the theorem, we need to verify if

$$Q_{11}P_{11}(I + (P_{11}^{-1}P_{12})(P_{11}^{-1}P_{12})^*) = I$$

and also if the rth eigenvalue of the matrix is ≥ 2. The latter part follows since, if we examine the matrix of the theorem in this case, we see that it simplifies to

$$\begin{pmatrix} P_{11}+Q_{11} & P_{12} \\ P_{12}^* & P_{12}^*P_{11}^{-1}P_{12} \end{pmatrix}$$

and this is precisely what we defined the matrix P_1 above to be, which we know satisfies this condition (its rth eigenvalue is ≥ 2). So all that remains to be shown is that the equality $Q_{11}P_{11}(I + (P_{11}^{-1}P_{12})(P_{11}^{-1}P_{12})^*) = I$ is true. Notice that

$$\begin{aligned} Q_{11}P_{11}(I + (P_{11}^{-1}P_{12})(P_{11}^{-1}P_{12})^*) &= Q_{11}P_{11}(I + P_{11}^{-1}P_{12}P_{12}^*P_{11}^{-1}) \\ &= Q_{11}(P_{11}^2 + P_{12}P_{12}^*)Q_{11} = I \end{aligned}$$

So the theorem holds for this case.

The next two cases, which we'll present below, could be shown similarly.

If $n < 2r$, then set $A = P_{11}$, $B = Q_{11}$, and take Z to be the matrix obtained by appending $2r - n$ zero columns to $P_{11}^{-1}P_{12}$.

If $n > 2r$, set $A = P_{11}$ and $B = Q_{11}$. The matrix Z is then taken to be the matrix

$P_{11}^{-1}P_{12}U$ but with the last $n-2r$ columns deleted.

Thus, if all 3 cases are shown to be true, the hypothesis of Theorem 3.20 is satisfied, and applying the Theorem 3.20 will yield $\lambda_r(P+Q) = \lambda_r(T) \geq 2$, as required.

3.4 Existing Conjectures

In a recent paper, Audenaert and Kittaneh [4] examined several conjectures and open problems in the theory of matrix and operator inequalities. Naturally, we're particularly interested in those which involve singular values.

The area of matrix subadditivity inequalities, in particular, has several open problems that are relevant to this paper. We'll begin by looking at a few known results.

If $f : [0, \infty) \mapsto [0, \infty)$ is concave, then the subadditivity relation $f(a + b) \leq f(a) + f(b)$ holds for all $a, b \geq 0$. A non-commutative version of this inequality is true for all PSD matrices.

The following theorem was proved in [11].

Theorem 3.22. *Let $f : [0, \infty) \mapsto [0, \infty)$ be concave. If A and B are PSD matrices, then for any unitarily invariant norm $||.||$*

$$||f(A + B)|| \leq ||f(A) + f(B)||.$$

To prove this theorem, we'll use the technique of Ando and Zhan [2].

Lemma 3.23. *Let $A, B \geq 0$ and let u_j be the orthonormal eigenvectors of $A + B$ corresponding to $\lambda_j(A + B)$, $j = 1, 2, ..., n$. Then the following inequalities hold:*

$$\sum_{j=1}^{k} \langle [A(A + I)^{-1} + B(B + I)^{-1}]u_j, u_j\rangle \geq \sum_{j=1}^{k} \langle (A + B)(A + B + I)^{-1}u_j, u_j\rangle$$

where $k = 1, 2, ..., n$.

The proof of this lemma is contained in [2]. Now we proceed to prove the theorem:

Proof. From the lemma, we have that

$$\sum_{j=1}^{k} \langle [sA(A + sI)^{-1} + sB(B + sI)^{-1}]u_j, u_j\rangle$$
$$\geq \sum_{j=1}^{k} \langle s(A + B)(A + B + sI)^{-1}u_j, u_j\rangle.$$

Now, if we use the specific integral representation $\langle f(A)u, u\rangle = \alpha\langle u, u\rangle + \beta\langle Au, u\rangle + \int_0^\infty s\langle A(A + sI)^{-1}u, u\rangle \,\mathrm{d}\mu(s)$ (see Theorem 2.31) we have

$$\sum_{j=1}^{k} \langle [f(A) + f(B)]u_j, u_j\rangle \geq \sum_{j=1}^{k} \langle f(A + B)u_j, u_j\rangle.$$

Since $f(t)$ is non-decreasing, any unit eigenvector u_j of A (for for $j = 1, 2, ..., n$), which corresponds to $\lambda_j(A)$, will become a unit eigenvector of $f(A)$, which corresponds to $\lambda_j(f(A)) = f(\lambda_j(A))$. Hence, applying Ky Fan norms:

$$||f(A)||_k = \sum_{j=1}^{k} \langle f(A)u_j, u_j \rangle,$$

for $k = 1, 2, ..., n$. Taking this and applying it to A+B in place of A, we see

$$||f(A + B)||_k = \sum_{j=1}^{k} \langle f(A + B)u_j, u_j \rangle.$$

The Ky Fan maximum principle shows that

$$||f(A) + f(B)||_k \geq \sum_{j=1}^{k} \langle [f(A) + f(B)]u_j, u_j \rangle,$$

for $k = 1, 2, ..., n$. Hence

$$||f(A) + f(B)||_k \geq ||f(A + B)||_k.$$

Then if we apply the Ky Fan dominance principle (for matrices A, B the inequalities $||A||_k \geq ||B||_k$ imply the same relation for unitarily invariant norms) the result follows.

□

Further, it has been shown that this result may be generalized to normal matrices [12].

Theorem 3.24. *Let A and B be normal matrices and let $f : [0, \infty) \mapsto [0, \infty)$ be concave. Then*

$$||f(|A + B|)|| \leq ||f(|A|) + f(|B|)||$$

for every unitarily invariant norm.

If A and/or B are non-normal the inequality no longer holds, but the following problem could be asked:

Conjecture 3.25. *For a given unitarily invariant norm is there a constant c such that*

$$||f(|A + B|)|| \leq c||f(|A|) + f(|B|)||$$

holds for all $A, B \in M_n(C)$ for all concave functions $f : [0, \infty) \mapsto [0, \infty)$.

If the above theorem is specialized to fractional powers, we have

$$|||A + B|^p|| \leq |||A|^p + |B|^p|| \text{ for } 0 < p \leq 1.$$

This suggests several related questions.

Conjecture 3.26. *Given $A, B \geq 0$ and $p, q > 0$, is it true that*

$$||A^{p+q} + B^{p+q}|| \leq ||(A^p + B^p)(A^q + B^q)||?$$

Conjecture 3.27. *Given $A, B \geq 0$ and $p, q > 0$, is it true that*

$$||A^p B^q + B^p A^q|| \leq ||A^{p+q} + B^{p+q}||?$$

For Heinz means, the related inequality

$$||A^p B^{1-q} + A^{1-p} B^p|| \leq ||A + B||$$

is known to be true. In fact, even stronger inequalities exist. If X is any matrix and $0 \leq p \leq 1$, for any unitarily invariant norm

$$||A^p X B^{1-p} + A^{1-p} X B^p|| \leq ||AX + XB||.$$

Further, it was proven by Audeneart in a prior paper we examined [3] that the inequality holds for each singular value:

$$\sigma_i(A^p B^{p-1} + A^{1-p} B^p) \leq \sigma_i(A + B).$$

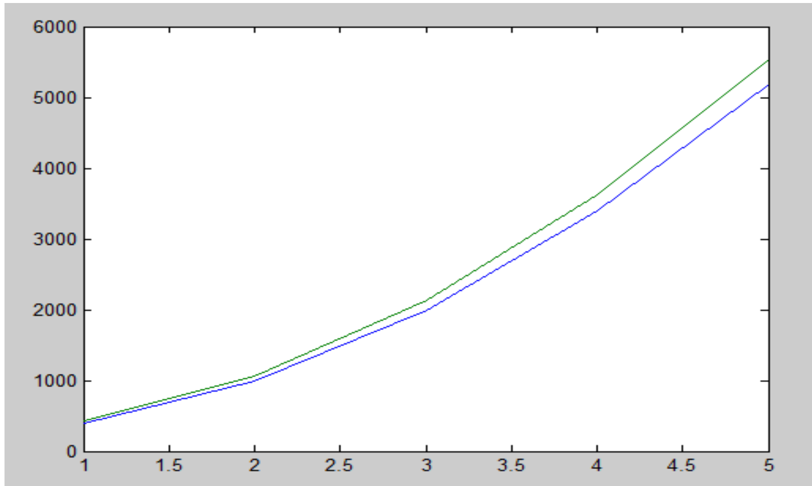
This suggests that the inequality of Conjecture 3.26 may similarily have a strong counterpart for singular values:

Conjecture 3.28. *Given $A, B \geq 0$ and $p > 0$, is it true that*

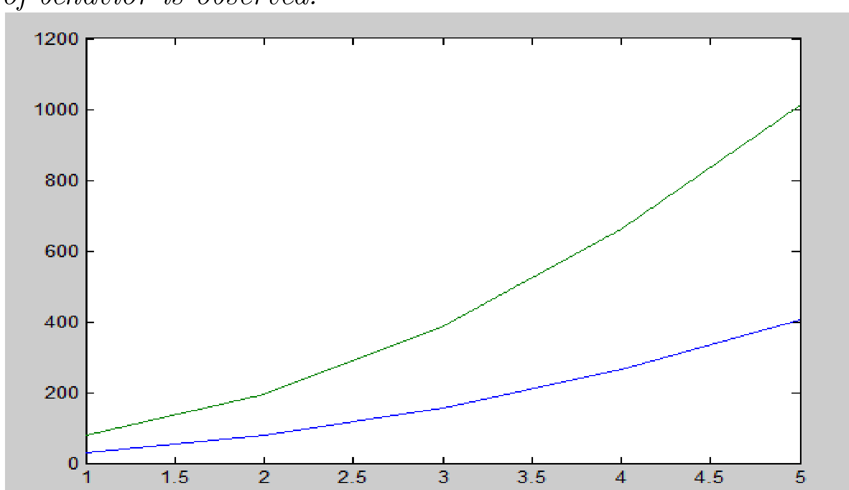
$$\sigma_i(A^p B^{1-p} + B^p A^{1-p}) \leq \sigma_i(A + B)?$$

Example 3.29. *We can illustrate this conjecture using a similar technique as we did in Example 3.12; that is, we'll consider the diagonalizations of the PSD matrices $A = \left(\begin{smallmatrix} 49 & 14 \\ 14 & 53 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 100 & 70 \\ 70 & 50 \end{smallmatrix}\right)$. Unlike in that example, however, here we'll scale by $x = t^2 + 1$ for t=1,2,...,5 to create a non-linear graph. Here we select $p = .75$.*

For the larger set of singular values and eigenvalues, we notice that there's a small distance between both sides of the inequality, which increases as we scale the diagonals:



For the smaller set there's a larger initial distance between the lines, however the same type of behavior is observed:



References

[1] Karim Abadir and Jan Magnus *Matrix Algebra* York, Tilburg 2005.

[2] Tsuyoshi Ando and Xingzhi Zhan *Norm inequalities related to operator monotone functions* Math. Ann. 315, 771-780 (1999)

[3] Koenraad M.R. Audenaert *A Singular Value Inequality for Heinz Means* Linear Algebra and Its Applications 422 (2007) 279-283

[4] Koenraad M.R. Audenaert and Fuad Kittaneh *Problems and Conjectures in Matrix and Operator Inequalities* arXiv:1201.5232v3 1 May 2012

[5] Adi Ben-Israel and Thomas Greville *Generalized Inverses: Theory and Applications* Springer-Verlag 2003.

[6] R. Bhatia and F. Kittaneh *On the singular values of a product of operators* SIAM J. Matrix Anal. Appl. 11 (1990) 272-277.

[7] R. Bhatia and C. Davis *More matrix forms of the arithmetic-geometric mean inequality"* SIAM J. Matrix Anal. Appl. 14, 132-136 (1993).

[8] Rajendra Bhatia *Matrix Analysis* Springer-Verlag 1997.

[9] R. Bhatia and F. Kittaneh *Notes on matrix arithmetic-geometric mean inequalities* Lin. Alg. Appl. 308, 203-211 (2000).

[10] Rajendra Bhatia *Positive Definite Matrices* Princeton University Press, 2007.

[11] J.-C. Bourin and M. Uchiyama *A matrix subadditivity inequality for f(A+B) and f(A)+f(B)* Linear Alegbra Appl. 423 (2007), 512-518.

[12] J.-C. Bourin *A matrix subadditivity inequality for symmetric norms* Proc. AMS 138 (2010), 495-504.

[13] S.W. Drury *On a question of Bhatia and Kittaneh* Linear Algebra and its Applications 437 (2012) 1955-1960.

[14] S.W. Drury *An Operator Arithmetic-Geometric Mean Inequality* http://www.math.mcgill.ca/drury/research/bhatiakittaneh/BKBlurb.html

[15] O. Hirzallah *Inequalities for sums and products of operators* Linear Algebra Appl. 407 (2005) 32-42

[16] R.A. Horn and C.R. Johnson *Matrix Analysis* Cambridge Univeristy Press, 1985.

[17] Roger A. Horn, and Charles R. Johnson *Topics in Matrix Analysis*. Cambridge, 1st Edition, 1991.

[18] Anthony W. Knapp *Basic Algebra* Birkhauser 2006.

[19] E.-Y. Lee *How to compare the absolute values of operator sums and the sums of absolute values?* In press, Operators and Matrices 2012

[20] Thomas Shores *Applied Linear Algebra and Matrix Analysis* Springer 2007.

[21] Yunxing Tao *More results on singular value inequalities of matrices* Linear Algebra and Its Applications 416 (2006) 724-729

[22] Xingzhi Zhan *Some research problems on the Hadamard product and singular values of matrices* Linear and Multilinear Algebra 47 (2000) 191-194.

[23] Xingzhi Zhan *Matrix Inequalities* Springer, 2002.

VITA

Peter Dylan Chilstrom was born After completing his work at University Christian High School, Jacksonville, Florida, in , he entered The University of North Florida in Jacksonville, Florida. In August of 2011 he completed a Bachelor of Arts in Mathematical Sciences. In August, 2011, he entered The Graduate School at The University of North Florida.