Linear Algebra and Matrix Analysis

Supplementary Material

Pritam Chandra

Monsoon 2025

Table of contents

The material here is supplementary to the lectures, and it is organized with that motive. To use this as a primary source for learning some of the featured topics is not recommended.

Special notations. The notation $S \leq V$ says S is a subspace of V. The spaces $\operatorname{ran} A$ and $\ker A$ are the range and null spaces of A respectively. The set $\operatorname{spec} A$ is the set of eigenvalues of A.

A Hilbert space will generally be denoted by $\mathcal{H}$. The space $\mathbb{H}$ will denote the space of all Hermitian matrices.

Whenever not explicitly mentioned, the coordinates of a vector x will be assumed to be $x_1, \ldots, x_n$. Likewise the entries of the matrix A will be a_{ij} for $1 \leq i, j \leq n$.

1 Subspaces, Linear Independence, Dimension

Let A be a $n \times n$ complex matrix. Then one can look at A as a matrix of column vectors or row vectors respectively as follows

$$A = \begin{bmatrix} A_{01} & A_{02} & \cdots & A_{0n} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} A_{10} — \\ A_{20} — \\ \vdots \\ A_{n0} — \end{bmatrix}.$$

Suppose $x \in \mathbb{C}^n$ is a column vector. Then $Ax = x_1 A_{01} + \cdots + x_n A_{0n}$. Similarly, for the row vector x^T, we have $x^T A = x_1 A_{10} + \cdots + x_n A_{n0}$.

Now consider the product AB. With the above notations for the row and column representations of A and B we get

$$AB = \begin{bmatrix} AB_{01} & \cdots & AB_{0n} \\ | & & | \end{bmatrix} = \begin{bmatrix} A_{10} B — \\ \vdots \\ A_{n0} B — \end{bmatrix} = \sum_{j=1}^{n} A_{0j} B_{j0} = [A_{j0} B_{0k}]_{j,k}. \quad (1)$$

Note that in the above expression $A_{0j} B_{j0}$ is a rank one matrix and $A_{j0} B_{0k}$ is a complex number.

Lemma 1. *Let $V_1, \ldots, V_k$ be subspaces of V. Let $\dim V = n$. Then*

$$\dim (V_1 \cap \cdots \cap V_k) \geq \dim V_1 + \cdots + \dim V_k - (k-1)n.$$

Proof. When $k = 2$, this holds. Then observe

$$\begin{aligned} \dim (V_1 \cap \cdots \cap V_k) &\geq \dim V_1 + \dim (V_2 \cap \cdots \cap V_k) - n \\ &\geq \dim V_1 + [\dim V_2 + \cdots + \dim V_k - (k-2)n] - n \\ &= \dim V_1 + \cdots + \dim V_k - (k-1)n, \end{aligned}$$

where the first inequality follows from the case when $k = 2$, and the second from the inductive hypothesis. ☐

Problem 1. If B is invertible, then B carries linearly independent sets to linearly independent sets. That is, if $y_i = Bx_i$, then $\{x_1, \ldots, x_k\}$ is a linearly independent set if and only if $\{y_1, \ldots, y_k\}$ is a linearly independent set.

Problem 2. Let $\{w_1, \ldots, w_n\}$ be a basis of a vector space V. Let B be the $n \times n$ matrix whose i-th column is w_i. Then B is invertible.

Proposition 1. *For $A \in \mathcal{L}(V, W)$, we know that* $\operatorname{rk} A = \dim \operatorname{ran} A$. *Then* $\operatorname{rk} A$ *is also the number of linearly independent columns in any matrix representation of A.*

Proof. Let $L = [l_{ij}]$ be any matrix representation of A. Suppose this matrix representation is with respect to the basis $\{v_1, \ldots, v_n\}$ of V and $\{w_1, \ldots, w_m\}$ of W. Then by definition of L we have

$$A(v_j) = l_{1j} w_1 + \cdots + l_{mj} w_m.$$

Collect the vectors $\{w_1, \ldots, w_m\}$ into the columns of a matrix B. Then B is invertible (use the problem 2). The above equation then looks like

$$A(v_j) = B L_j,$$

where L_j is the j-th column of L. Then one concludes the proof using problem 1.

Definition 1. Let $\alpha_0, \alpha_1, \ldots, \alpha_k$ be $k+1$ real numbers. Then we call the following $(k+1) \times (k+1)$ matrix

$$V(\alpha_0, \ldots, \alpha_k) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_0 & \alpha_1 & \cdots & \alpha_k \\ \alpha_0^2 & \alpha_1^2 & \cdots & \alpha_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_0^k & \alpha_1^k & \cdots & \alpha_k^k \end{bmatrix}$$

the **Vandermonde matrix** associated with $\alpha_0, \ldots, \alpha_k$.

Lemma 2. $\det V(\alpha_0, \ldots, \alpha_k) = \prod\limits_{i \neq j} (\alpha_i - \alpha_j)$.

Proof. Fix any i. Let $V = V(\alpha_0, \ldots, \alpha_k)$. Consider $\det V$ as a polynomial $p(\alpha_i)$ only in the variable α_i. Then for any $j \neq i$, we have $p(\alpha_j) = 0$. This is because when one puts $\alpha_i = \alpha_j$, the i-th and the j-th columns of V are equal. Thus by factor theorem we get

$$(\alpha_i - \alpha_j) | \, p(\alpha_i) = \det V.$$

This is true for all i, j pairs such that $i \neq j$. Since these are distinct linear factors we get

$$\prod_{1 \leq i < j \leq k} (\alpha_i - \alpha_j) \, | \det V.$$

Since there are $\binom{k+1}{2}$ such pairs, the degree of the polynomial on the left hand side of the above equation is $\binom{k+1}{2}$.

But we also know that

$$\det V = \sum_{\sigma \in S_{k+1}} \operatorname{sgn}(\sigma) \, V_{0,\sigma(0)} V_{1,\sigma(1)} \cdots V_{k,\sigma(k)}.$$

So the determinant is a sum of products of entries of V such that in each summand exactly one factor comes from each row. So the degree of each summand in the determinant expression is

$$0 + 1 + \cdots + k = \frac{k(k+1)}{2} = \binom{k+1}{2}.$$

This is also the degree of $\det V$. Thus $\det V = c \prod_{i \neq j} (\alpha_i - \alpha_j)$. Compare the coefficients of some terms to conclude $c = 1$. $\square$

The usefulness of the Vandermonde matrix is in its ability to characterize a distinct tuple. That is, $\det V(\alpha_0, \ldots, \alpha_k) \neq 0$ if and only if α_js are distinct.

Problem 3. Use the Vandermonde matrix to show

1. Eigenvectors corresponding to distinct eigenvalues are linearly indepedent (not just pairwise).
2. Consider the family of functions $\{e^{\alpha x} | \alpha \in \mathbb{R}\}$ in $C[a, b]$, the space of continuous functions on the interval $[a, b]$. Show that this is a linearly independent family. Conclude that the dimension of $C[a, b]$ is uncountably infinite.

Let V be a vector space of infinite dimension. Then V can be given separate notions of linear independence. For example, let $B = \{b_1, b_2, \ldots\}$ be a countable subset of V. We could say B is a linearly independent set if every finite subset of B is linearly independent in the usual sense. This is called **Hamel** independence. Or we could insist that B is linearly independent if

$$c_1 b_1 + c_2 b_2 + \cdots = 0 \Longrightarrow c_1 = c_2 = \cdots = 0.$$

In this case one needs additional machinery to make sure that the sum makes sense. This is called **Schauder** independence, which is clearly stronger than Hamel. Based on these two different notions of indepedence, V could either have a **Hamel basis** or a **Schauder basis**. We present a curious fact in the following remark.

Remark 1. Every vector space has a Hamel basis. This is equivalent to the axiom of choice and Zorn's lemma. Not every vector space has a Schauder basis.

2 Diagonalizability

The **characteristic polynomial** of a matrix A, denoted by $\chi_A(t)$, is defined as $\chi_A(t) = \det(A - t)$. The **Cayley-Hamilton theorem** states that $\chi_A(A) = 0$. That is, every matrix satisfies its characteristic polynomial.

Of all the polynomials which A satisfies, the **minimal polynomial** of A, denoted by $\mu_A(t)$, is the one of smallest degree. A simple consequence of the quotient theorem of polynomials is that a polynomial f enjoys $f(A) = 0$ if and only if $\mu_A | f$. In particular $\mu_A | \chi_A$.

Proposition 2. *The following statements are equivalent, that is, they are valid ways to define diagonalizability of a matrix A.*

1. *A is similar to a diagonal matrix.*
2. *The eigenvectors of A form a basis of V. That is, if $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of V, and S_{λ_j} is the eigenspace corresponding to the eigenvalue λ_j, then*

   $$V = S_{\lambda_1} + \cdots + S_{\lambda_j}.$$

   (Direct sum is implied in the above sum. Why?)
3. *The minimal polynomial of A has distinct roots.*

   (This can be relaxed to say that there exists a polynomial f with distinct roots such that $f(A) = 0$. Why?)

Proof. (of point 3.) Let $\mu(x)$ be the minimal polynomial of x. Suppose

$$\mu(x) = (x - \lambda_1) \ldots (x - \lambda_k)$$

where $\lambda_1, \ldots, \lambda_k$ are distinct. Let

$$f_j(x) = \prod_{\substack{i=1 \\ i \neq j}}^{k} (x - \lambda_i).$$

Then the polynomials $f_j(x)$ are coprime. Therefore, by Bézout's lemma, there exist polynomials $g_j(x)$ such that

$$1 = f_1(x)\, g_1(x) + \cdots + f_k(x)\, g_k(x).$$

Define $h_j(x) = f_j(x) g_j(x)$. Then $m(x) | (x - \lambda_j)\, h_j(x)$. Then for any $v \in V$, we have

$$(A - \lambda_j)\, h_j(A)\, v = 0 \Longrightarrow h_j(A) \in S_{\lambda_j}.$$

Therefore putting A in the equation and applying v gives us

$$v = h_1(A)\, v + \cdots + h_k(A)\, v, \quad \text{where } h_j(A) v \in S_{\lambda_j}.$$

This implies that

$$S_{\lambda_1} + \cdots + S_{\lambda_k} = V.$$

We will give a sketch of the converse statement. Let A be diagonalizable with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then the polynomial $\mu(x) = (x - \lambda_1) \cdots (x - \lambda_k)$ kills A (why?). The minimal polynomial of A must divide $m(x)$, and also have $\lambda_1, \ldots, \lambda_k$ as its roots (why?). So μ must be the minimal polynomial of A. $\square$

A consequence of the above statement is that every k-th root of I is diagonalizable. Suppose A is a k-th root of I, for $k \leq n$. Then A satisfies the polynomial $x^k - 1$ which has distinct roots. So the minimal polynomial of A has distinct roots. Therefore A is diagonalisable. So every k-th root of I has to be similar to a diagonal matrix $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, where each $\lambda_j^k = 1$ for all j.

3 Diagonal Dominance and Disk Theorem

Let V be a vector space equipped with a metric. Let $\sum x_k$ be a series in V. Then if the series converges absolutely, then the sequence of partial sums is Cauchy. Thus, if V is complete, absolute convergence of a series implies its convergence. Further, if the series converges absolutely, then the series formed by any rearrangement also converges, and the convergence is to the same limit.

This can be used to show that when $\|A\| < 1$ then $I - A$ is nonsingular. One could also argue that $I - A$ has no 0 eigenvalue.

Definition 2. Let A be an $n \times n$ complex matrix. Let $r_i = \sum_{j \neq i} |a_{ij}|$. A is said to be **diagonally dominant** if $|a_{ii}| \geq r_i$ for all i. If the inequality is strict then A is **strictly diagonally dominant**.

Theorem 1. *If A is strictly diagonally dominant, then A is nonsingular.*

Problem 4. Show that if $A \neq I$ is idempotent then A cannot be diagonally dominant.

As a consequence of the above theorem we get a glamorous result about the location of eigenvalues of a matrix.

Theorem 2. ***(Gershgorin disk theorem)*** *Let A and r_i be as in definiton 2. Then all eigenvalues of A lie in the set*

$$\bigcup_{i=1}^{n} \{z: |z - a_{ii}| \leq r_i\}.$$

As an application of this theorem, we give a lower bound on the rank of a matrix that can be read off from the matrix entries.

Proposition 3. *Let $A_1, \ldots, A_n$ be the rows of the matrix A. Then*

$$\operatorname{rk} A \geq \sum_{\substack{i=1 \\ A_i \neq 0}}^{n} \frac{|a_{ii}|}{\|A_i\|_1}.$$

Proof. Let S be the diagonal matrix defined as[1]

$$S_{ii} = \begin{cases} \operatorname{sgn}(a_{ii}) \|A_i\|_1^{-1}, & A_i \neq 0 \\ 1, & A_i = 0 \end{cases}.$$

Then S is nonsingular. Let $B = SA$. Then $\operatorname{rk} B = \operatorname{rk} A$. Further, diagonal entries of B are either $b_{ii} = \dfrac{|a_{ii}|}{\|A_i\|_1}$ or $b_{ii} = 0$, and the rows B_i of B are such that $\|B_i\|_1 = 1$.

Then by Gresgorin disk theorem applied to B, every eigenvalue λ of B satisfies

$$|\lambda - b_{ii}| \leq 1 - b_{ii} \Longrightarrow |\lambda| \leq 1.$$

If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of B counted with multiplicity, then

$$\operatorname{rk} B \geq |\lambda_1| + \cdots + |\lambda_n| \geq \operatorname{tr} B = b_{11} + \cdots + b_{nn}.$$

1. If $a = re^{i\theta}$ then $\operatorname{sgn}(a) = e^{-i\theta}$. In particular, $a \operatorname{sgn}(a) = |a|$.

The extreme sides of the above inequality are the extreme sides of the required inequality. □

4 Orthogonality and Projections

Let $\mathcal{H}$ be a Hilbert space. For $x, y \in \mathcal{H}$ we say $x \perp y$, read as x is orthogonal to y, if $\langle x, y \rangle = 0$.

Remark 2. Orthogonality is pairwise concept, unlike linear independence. Linear independence of a set is different from pairwise linear indepedence of the members of the set. It is a much stronger condition. For example consider $\{e_1, e_2, e_1 + e_2\}$. However, orthogonality of a set, by definition, is pairwise orthogonality of its members.

Proposition 4. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Let $S \leq \mathbb{C}^n$. Then if $\langle x, Ax \rangle = 0$ for all $x \in S$, then $A(S) \subset S^{\perp}$.*

Proof. The proposition holds if $\langle x, Ay \rangle = 0$ for all $x, y \in S$. To see this pick any $Ay \in A(S)$. Then $\langle x, Ay \rangle = 0$ for all $x \in S$. So $Ay \in S^{\perp}$.

(Method of polarisation) For all $u, v \in S$ and $\alpha, \beta \in \mathbb{C}$ we have

$$0 = \langle \alpha u + \beta v, A(\alpha u + \beta v) \rangle = \bar{\alpha}\, \beta\, \langle u, Av \rangle + \alpha\, \bar{\beta}\, \langle v, Au \rangle.$$

Put $\alpha = \beta = 1$, then we get

$$0 = \langle u, Av \rangle + \langle v, Au \rangle.$$

Putting $\alpha = 1$ and $\beta = i$ we get

$$0 = i \langle u, Av \rangle - i \langle v, Au \rangle \Longrightarrow 0 = \langle u, Av \rangle - \langle v, Au \rangle.$$

Adding the above equations we get $\langle u, Av \rangle = 0$. □

Let $\mathcal{H} = S_1 \oplus S_2$. Then every $z \in \mathcal{H}$ has a unique representation of the form $z = x + y$, where $x \in S_1$ and $y \in S_2$. Define the operator P_{S_1, S_2}, read as **projection** onto S_1 along S_2, as $P(z) = x$. Further if $S_2 = S_1^{\perp}$, then we call the corresponding operator as the **orthoprojection** onto S_1.

The operator P is a projection if and only if P is idempotent. Here are some facts about orthoprojections.

1. P is an orthoprojection if and only if P is idempotent and Hermitian.
2. $P \geq 0$ and $\|P\| = 1$.
3. If P is the orthoprojection onto S, then $\bar{P} = I - P$ is the orthoprojection onto $S^{\perp}$.

4. Suppose P_1, P_2 are orthoprojections onto S_1, S_2 respectively. Then
   a. $P_1 P_2 = 0$ if and only if $S_1 \perp S_2$.
   b. $P_1 + P_2$ is an orthoprojection if and only if $P_1 P_2 = 0$.

Proposition 5. *Let P_1, P_2 be the orthoprojections onto S_1 and S_2 respectively. Then, the following are equivalent.*

1. $S_1 \subseteq S_2$
2. $P_2 P_1 = P_1$
3. $P_1 P_2 = P_1$
4. $\|P_1 x\| \leq \|P_2 x\|$ *for all* x
5. $P_1 \leq P_2$

Proof. Statement 1 implies statement 2. Let $x \in S_1^{\perp}$. Then $P_2 P_1 x = 0 = P_1 x$. Let $x \in S_1$, then $x \in S_2$. Then $P_2 P_1 x = P_2 x = x = P_1 x$.

Statement 2 implies statement 3. Take adjoint of statement 2.

Statement 3 implies statement 4. Observe

$$\|P_1 x\| = \|P_1 P_2 x\| \leq \|P_1\| \, \|P_2 x\| = \|P_2 x\|.$$

Statement 4 implies statement 5. Note that

$$\langle x, P_1 x \rangle = \langle P_1 x, P_1 x \rangle = \|P_1 x\|^2 \leq \|P_2 x\|^2 = \langle P_2 x, P_2 x \rangle = \langle x, P_2 x \rangle.$$

Statement 5 implies statement 1. Let $z \in S_1$. Then $z = x + y$ where $x \in S_2$ and $y \in S_2^{\perp}$. Then note that

$$0 \leq \langle z, (P_2 - P_1) z \rangle = \langle x + y, x \rangle - \|z\|^2 = -\|y\|^2 \Longrightarrow \|y\|^2 = 0 \Longrightarrow z \in S_2. \qquad \square$$

Proposition 6. *Suppose P is a projection ($P^2 = P$) and $\|P\| \leq 1$. Then P is an orthoprojection.*

Proof. As P is already a projection, to show that it is an orthoprojection, it is sufficient to show that $\operatorname{ran} P \perp \ker P$. Suppose this was not the case. Assume $P \neq 0$, as otherwise the statement is trivially true. Then there exist unit vectors $x \in \ker P$ and $y \in \operatorname{ran} P$ such that $\langle x, y \rangle = \alpha \neq 0$. Let $z = y - \alpha x$. Then observe that

$$\langle \alpha x, z \rangle = \langle \alpha x, y - \alpha x \rangle = |\alpha|^2 - |\alpha|^2 = 0.$$

Therefore

$$\|Pz\|^2 = \|y\|^2 = \|z + \alpha x\|^2 = \|z\|^2 + \|\alpha x\|^2 > \|z\|^2.$$

Thus $\|P\| > 1$. This is a contradiction. $\square$

Proposition 7. *Let $u_1, \ldots, u_k$ be the orthonormal basis of a subspace S. Let U be the $n \times k$ column matrix $U = [\, u_1 \ \cdots \ u_k \,]$. Then the orthoprojection onto S is given by UU^*. Moreover $UU^* = \sum_{j=1}^{k} u_j u_j^*$.*

Proof. (sketch) Let $P = UU^*$. Show that $P^* = P$ and $P^2 = P$. Then use $\operatorname{ran} P \subset \operatorname{ran} U$ and $Pu_j = u_j$. For the second part look at equation 1. $\square$

5 Normality and Spectral Theorem

A matrix A is called **normal** if $A^*A = AA^*$. Before proceeding we prove the following facts.

Problem 5. S is invariant under A if and only if $S^\perp$ is invariant under A^*.

Problem 6. If λ, v are an eigenpair of A if and only if $\bar{\lambda}, v$ an are eigenpair of A^*.

Problem 7. Eigenvectors corresponding to distinct eigenvalues of a normal matrix are mutually orthogonal.

Proposition 8. *The following statements are equivalent. That is, they are all valid ways to define normality.*

1. $A^*A = AA^*$
2. $\|A^*x\| = \|Ax\|$ *for all* x.
3. $A = UDU^*$ *where U unitary and D diagonal (**spectral theorem**).*
4. *One can choose eigenvectors of A to form an orthonormal basis of $\mathcal{H}$. (Can one choose eigenvectors that do not form an orthonormal basis?)*

   Equivalently, the eigenspaces $S_1, \ldots, S_k$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ are mutually orthogonal and satisfy

   $$\mathcal{H} = S_1 + \cdots + S_k.$$

5. *There exist $\lambda_j \in \mathbb{C}$ and $\{u_j\}$ orthonormal such that* $A = \sum_{j=1}^{n} \lambda_j u_j u_j^*$.
6. *There exist distinct $\lambda_1, \ldots, \lambda_k$ and disjoint orthoprojections $P_1, \ldots, P_k$ such that*

   $$A = \lambda_1 P_1 \oplus \cdots \oplus \lambda_k P_k, \quad \text{with} \quad P_1 \oplus \cdots \oplus P_k = I.$$

Proof. (sketch of the proof of spectral theorem: normal matrices are diagonalizable in an orthonormal basis) For this we use the statements of the problems in the beginning of the section. Now, let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of A. Let $S_1, \ldots, S_k$ be the corresponding eigenspaces. Let

$$S = S_1 \oplus \cdots \oplus S_k.$$

To show A is "diagonalizable" it is sufficient to show that $S = \mathcal{H}$. Orthogonality will follow from the mutual orthogonality of eigenvectors.

Now if $S \neq \mathcal{H}$, then $S^{\perp} \neq \{0\}$. Since $S_1, \ldots, S_k$ are also eigenspaces of A^*, the space S is also invariant under A^*. This means that $S^{\perp}$ is invariant under A.

Then look at the operator A restricted to $S^{\perp}$. Call it $\bar{A}$. Then $\bar{A}$ will have an eigenvalue λ. And λ is also an eigenvalue of A. So $\lambda = \lambda_j$ for some j. Therefore the corresponding eigenvector x will lie both in S and $S^{\perp}$. This is a contradiction. □

Note that the equivalence of statement 5 and 6 follows from proposition 7.

Note 1. Why is invariance of $S^{\perp}$ necessary in the proof of spectral theorem? A similar argument was also used in proving that if A and B commute they share an eigenvector. In general, why do we demand invariance to construct operator restrictions? Suppose a subspace S is not invariant under the operator A. Then A: $S \to S'$, where S' may not be S. But this operator may still have eigenvalues and eigenvectors. Potentially, there could be $x \in S \cap S'$ for which $Ax = \lambda x$.

So far there is no problem. But in this case, the eigenvalue λ of the restriction $A|_S$ may no longer be an eigenvalue of A. Now suppose S is invariant under A. Choose a basis of the entire space whose first few vectors span S and the rest span $S^{\perp}$. In this basis A has the form

$$A \sim \begin{bmatrix} A|_S & * \\ 0 & B \end{bmatrix}.$$

Then $\det(A - t) = \det(A|_S - t)\det(B - t)$. This ensures that $\operatorname{spec} A|_S \subset \operatorname{spec} A$. If S was not invariant under A, the bottom left submatrix would not necessarily be 0, and hence the determinant relation would not follow.

Proposition 9. *An upper (or lower) triangular matrix is normal if and only if it is diagonal.*

Proof. We will proceed by induction on the dimension n. For any n use statement 2 in the definition of normality to observe that

$$|t_{11}|^2 = \|Te_1\|^2 = \|T^* e_1\|^2 = |t_{11}|^2 + \cdots + |t_{1n}|^2 \Longrightarrow t_{12} = \cdots = t_{1n} = 0.$$

This proves the result when $n = 2$. For an arbitrary n, the result above says

$$T = \begin{bmatrix} t_{11} & \\ & H \end{bmatrix},$$

where H is an $(n-1)\times(n-1)$ upper triangular matrix. But T is normal if and only if H is normal (show this). By induction hypothesis H is diagonal. Hence T is diagonal. □

Theorem 3. ***(Schur's triangularization theorem)*** *Let A be any matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that $A = UTU^*$. That is every matrix is triangularizable in an orthonormal basis.*

Equivalently, for every matrix A there exists an orthonormal basis $\{v_1,\ldots,v_n\}$ such that

$$Av_j \in \{v_1,\ldots,v_j\} \quad \text{for } j = 1,\ldots,n. \tag{2}$$

Note that normality is invariant under unitary similarity. Then Schur's theorem along with proposition 9 gives another proof of the spectral theorem.

Proposition 10. *Suppose $\operatorname{rk} A = r$. Then A can be triangularized to a matrix whose first r rows are linearly independent and the last $n-r$ rows are 0.*

Proof. Let P be the projection onto $\operatorname{ran} A$ and $\bar{P}$ onto $\ker A$. Then by Schur's theorem applied to PA, there exists $v_1,\ldots,v_r \in \operatorname{ran} A$ such that

$$PAv_j \in \operatorname{span}\{v_1,\ldots,v_j\}.$$

Club these vectors with a basis $\{v_{r+1},\ldots,v_n\}$ of $\ker A$. Then for $j \leq r$

$$Av_j = PAv_j + \bar{P}Av_j = PAv_j,$$

where the last equality follows from the fact that $Av_j \in \operatorname{ran} A = \ker \bar{P}$. For $j > r$, $Av_j = 0$ so the condition 2 is immediately satisfied. The linear independence of the first r rows and the rest of the rows being 0 follows from the choice of v_js. □

In the spirit of proposition 3, we state another lower bound for the rank of a matrix.

Proposition 11. *For any matrix A we have $\operatorname{rk} A \geq \dfrac{|\operatorname{tr} A|^2}{\|A\|_F^2}$.*

Proof. Let k be the number of nonzero eigenvalues of A. Then, the rank-nullity theorem gives us that $\operatorname{rk} A \geq k$. Also note that either side of the required inequality is unchanged by unitary conjugation. Then using Schur's theorem we can assume A to be upper triangular with diagonal entries $\lambda_1,\ldots,\lambda_k,0,\ldots,0$. Then note that

$$|\operatorname{tr} A|^2 = (\lambda_1 + \cdots + \lambda_k)^2 \leq k\,(\lambda_1^2 + \cdots + \lambda_n^2) \leq k\,\|A\|_F^2,$$

where the first inequality is the Cauchy-Schwarz inequality. □

Proposition 12. *Suppose A is upper triangular and invertible. Then A^{-1} is also upper triangular.*

Proof. (*Vishnu*) Let A^{-1}. We will show that $A^{-1}e_j \in \operatorname{span}\{e_1, \ldots, e_j\}$. This is equivalent to showing A^{-1} is upper triangular. We will proceed by induction on j.

As A is upper triangular and invertible, we have $A e_1 = \lambda_1 e_1$ for some $\lambda_1 \neq 0$. Then applying A^{-1} to this equation we get $e_1 = \lambda_1 A^{-1} e_1 \Longrightarrow A^{-1} e_1 = \frac{e_1}{\lambda_1} \in \operatorname{span}\{e_1\}$. This proves the base case. By the same argument we see that

$$A e_j = c_1 e_1 + \cdots + c_{j-1} e_{j-1} + \lambda_j e_j \quad \text{for some } \lambda_j \neq 0.$$

Applying A^{-1} to the above equation we get

$$A^{-1} e_j = \frac{1}{\lambda_j}[e_j - c_1 e_1 - \cdots - c_{j-1} e_{j-1}] \in \operatorname{span}\{e_1, \ldots, e_j\}.$$

This shows that A^{-1} is upper triangular.

Alternate proof. (*Anwesha*) Let us proceed by induction on the dimension of A. The 1 dimensional case is trivial. Let A and A^{-1} be the $n \times n$ block matrices

$$A = \begin{bmatrix} \alpha & x^* \\ & A_1 \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} \beta & y^* \\ z & B_1 \end{bmatrix}$$

where A_1, B_1 are $(n-1) \times (n-1)$ blocks. Then

$$I = A A^{-1} = \begin{bmatrix} * & * \\ A_1 z & A_1 B_1 \end{bmatrix} \Longrightarrow A_1 z = 0 \text{ and } B_1 = A_1^{-1}.$$

Note that A_1 is a nonsingular matrix (as it is upper triangular with nonzero diagonal entries). So $A_1 z = 0 \Rightarrow z = 0$. And B_1 is upper triangular by inductive hypothesis. Therefore A^{-1} is upper triangular. □

6 AB, BA and Commuting Matrices

The following are true for operators A and B.

1. If $AB = BA$ then they share an eigenvector.
2. AB and BA have the same eigenvalues.
3. If $A, B \geq 0$ then all eigenvalues of AB are nonnegative.
4. A, B normal. Then AB normal $\Longleftrightarrow$ BA normal.

5. A normal, B any matrix, then $\|AB - BA\|_F = \|A^*B - BA^*\|_F$. In particular A and B commute if and only if A^* and B commute.

Theorem 4. *Commuting matrices are simultaneously triangularizable. That is, if $AB = BA$ then there exists a unitary U for which both U^*AU and U^*BU are upper triangular.*

Proof. Let A, B commute. Then they share an eigenvector, say v_1. Then

$$A = PA + \bar{P}A \quad \text{ and } \quad B = PB + \bar{P}B.$$

where $P + \bar{P} = I$ and $\operatorname{ran} P = \operatorname{span}\{v_1\} = \mathcal{M}$ and $\operatorname{ran} \bar{P} = \mathcal{N}$. We will show that $\bar{P}A$ and $\bar{P}B$ commute on $\mathcal{N}$. To show this it is sufficient to show that

$$\bar{P}A\bar{P}B = \bar{P}AB.$$

Now note that $\mathcal{M}$ is invariant under B, therefore $AP = PAP$. That is, P commutes with BP. Then

$$\bar{P}A\bar{P}B = AB - APB - PAB + PAPB = \bar{P}AB.$$

Therefore, by inductive hypothesis, there exists an orthonormal basis $v_2, \ldots, v_n$ of $v^\perp$ such that both Av_j and Bv_j lie in $\operatorname{span}\{v_2, \ldots, v_j\}$. As $PAx \in \operatorname{span}\{v_1\}$ for all x, the vectors $\{v_1, \ldots, v_n\}$ satisfy condition 2. □

Remark 3. The converse of theorem 4 need not hold. Consider the following example

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \text{and} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = 0.$$

Commuting is a strictly stronger condition than simultaneous triangularizability.

We quickly also recall that two commuting matrices need not be simultaneously diagonalizable. Commuting ensures one common eigenvector. Diagonalizability requires a basis of eigenvectors.

Schur's triangular form gives us access to the eigenvalues of a matrix, as the eigenvalues can be read off from the diagonals. Here is a consequence of the theorem. Suppose A and B are simultaneously triangularizable. Then every eigenvalue λ of $A + B$ is of the form $\alpha + \beta$ where α and β are some eigenvalues of A and B respectively. Similarly, any eigenvalue λ of AB is of the form $\alpha\beta$.

Question 1. When is a pair of matrices A, B simultaneously triangularizable — is there another "clean" characterization of simultaneously triangularizable matrices?

Definition 3. A pair of matrices A and B are said to have the **additive eigenvalue property** if their eigenvalues can be ordered as $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ such that the eigenvalues of $A+B$ are exactly $\alpha_1+\beta_1, \ldots, \alpha_n+\beta_n$.

A pair of simultaneously triangularizable matrices have the additive eigenvalue property. However the converse is not true. Consider

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The only eigenvalue of both A and B is 0. And

$$\delta A + \gamma B = \begin{bmatrix} 0 & \delta & 0 \\ \gamma & 0 & -\delta \\ 0 & \gamma & 0 \end{bmatrix}.$$

Characteristic polynomial of this matrix (in the variable t) is t^3, so the only eigenvalue of this matrix is also 0. So the pair $\delta A, \gamma B$ has the additive eigenvalue property for any $\delta, \gamma \in \mathbb{C}$. But observe that

$$AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

If A and B were simultaneously triangularizable, then the only eigenvalue of AB would also be 0. But that is not the case. We summarize these findings in the following remark.

Remark 4. Let A, B be two matrices. Then

$$A, B \,\text{commute} \;\begin{array}{c}\Rightarrow \\ \not\Leftarrow\end{array}\; \begin{array}{c} A, B\,\text{simultaneously} \\ \text{triangularizable}\end{array} \;\begin{array}{c}\Rightarrow \\ \not\Leftarrow\end{array}\; \begin{array}{c} A, B\,\text{has additive} \\ \text{eigenvalue property}\end{array}.$$

We state a result that gives *an* answer to question 1.

Theorem 5. *(Radjavi and Rosenthal) A pair of matrices A and B are simultaneously triangularizable if $w(A, B)\,[A, B]$ is nilpotent for every word w.*

A word $w(A, B)$ is simply a product of a sequence containing As and Bs. This naturally includes commuting matrices.

7 A Variational Characterization for Singular Values

Quasilinearization or, more generally, variational characterization, is a very important idea. The principle is to write a non-linear function as an extremum of linear (or "nicer") functions, and exploit the linearity. Many elegant matrix quantities, like eigenvalues, singular values, some norms etc., admit variational characterizations, making their manipulations, at the very least, easier, and sometimes even feasible.

Definition 4. Every matrix A can be decomposed as $A = USV^*$ where U, V are unitary, and $S = \operatorname{diag}(s_1, \ldots, s_n)$ with $s_1 \geq \ldots \geq s_n \geq 0$. This is called the **singular value decomposition (SVD)** of A. Some facts are as follows:

1. The columns of U, called the **left singular vectors** of A are eigenvectors of the positive matrix AA^*.
2. The columns of V, called the **right singular vectors** of A are the eigenvectors of the positive matrix A^*A.
3. The numbers $s_1, \ldots, s_n$, called the **singular values** of A, are the positive square roots of the eigenvalues of A^*A or AA^*.
4. This decomposition is equivalent to the **polar decomposition**, where one writes an arbitrary matrix A as $A = QP$ where Q is unitary and P positive.

In the rest of the discussion $s_j(T)$ will denote the j-th singular value of the matrix T. When obvious from the context, $s_j(T)$ may simply be written as s_j. Now we proceed to state a variational characterisation of the singular values.

Theorem 6. *Let A be any matrix with singular value decomposition $A = USV^*$. Then*

$$s_{r+1}(A) = \min_{T:\operatorname{rk}T \leq r} \|A - T\|.$$

The minimum of the above equation is obtained at $T = A_r$ where

$$A_r = US_rV^* \text{ and } S_r = \operatorname{diag}(s_1, \ldots, s_r, 0, \ldots, 0).$$

Further, the operator norm in the above minimization can be replaced by any unitarily invariant norm.

Using this we can immediately go on to prove some important and useful results.

Proposition 13. *Suppose A and B be any two matrices. Then for i, j, k satisfying $j + k = i + 1$, we get*

$$s_i(A + B) \leq s_j(A) + s_k(B) \text{ and } s_i(AB) \leq s_j(A)\, s_k(B).$$

Proof. First note that $\operatorname{rk}(A_{j-1}+B_{k-1}) \leq \operatorname{rk} A_{j-1} + \operatorname{rk} B_{k-1} = i-1$. So we get

$$\begin{aligned} s_i(A+B) &\leq \|A+B-(A_{j-1}+B_{k-1})\| \\ &\leq \|A-A_{j-1}\| + \|B-B_{k-1}\| \\ &= s_j(A)+s_k(B). \end{aligned}$$

In a similar vein even the other result is obtained. Note that

$$\operatorname{rk}(AB_{k-1}+A_{j-1}[B-B_{k-1}]) \leq \operatorname{rk} B_{k-1} + \operatorname{rk} A_{j-1} = i-1\,.$$

So we have

$$\begin{aligned} s_i(AB) &\leq \|AB-AB_{k-1}-A_{j-1}(B-B_{k-1})\| \\ &= \|(A-A_{j-1})(B-B_{k-1})\| \\ &\leq \|A-A_{j-1}\|\,\|B-B_{k-1}\| \\ &= s_j(A)\,s_k(B)\,. \end{aligned}$$

□

Proposition 13 gives us the following substantial theorem as a corollary.

Theorem 7. *For every i, the function $s_i(X)$ is (Lipschitz) continuous in X.*

Proof. Apply the first inquality of proposition 13 to $i=k$ and $j=1$, and replace A by $A-B$. This gives

$$s_i(A) \leq s_1(A-B)+s_i(B) \Longrightarrow |s_i(A)-s_i(B)| \leq \|A-B\|\,.$$

□

Along these lines we also have the following reverse inequalities.

Proposition 14. *Let A, B be any two matrices. Suppose i, j, k satisfy $j+k=i+n$. Then we have*

$$s_i(AB) \geq s_j(A)\,s_k(B)\,.$$

Proof. Let $B=USV^*$ be the singular value decomposition of B. Then for any $0<\varepsilon$, we define $B_\varepsilon = U\,(S+\varepsilon)\,V^*$. Note that B_ε is invertible. Then we get

$$\begin{aligned} s_j(A) &= s_j\,(AB_\varepsilon\,B_\varepsilon^{-1}) \\ &\leq s_{j+k-n}(AB_\varepsilon)\,s_{n-k+1}\,(B_\varepsilon^{-1}) \qquad \text{(proposition 13)} \\ &= s_i(AB_\varepsilon)\,\frac{1}{s_k(B_\varepsilon)} \\ \Longrightarrow \quad s_j(A)\,s_k\,(B_\varepsilon) &\leq s_i(AB_\varepsilon). \end{aligned}$$

We take limit $\varepsilon \longrightarrow 0$. Theorem 7 tells us that the limit can be taken inside s_k and s_i. This concludes the proof of the required inequality. □

Remark 5. When $j+k=i+1$, the inequality $s_i(A+B)\leq s_j(A)+s_k(B)$ is similar to Weyl's inequality for eigenvalues of a Hermitian matrix, and indeed they can be derived from each other. There is a reverse Weyl's inequality that says, when $j+k=i+n$ we have

$$\lambda_i(A+B)\geq\lambda_j(A)+\lambda_k(B),$$

where the eigenvalues of each matrix are counted in a decreasing order. We have seen something like a multiplicative analogue of this inequality for singular values in proposition 14. However, the immediate analogue of the Weyl's reverse inequality does not hold for singular values. For example, consider

$$A=\begin{bmatrix} 1 & \\ & 0 \end{bmatrix} \text{ and } B=\begin{bmatrix} -1 & \\ & 0 \end{bmatrix}.$$

Then $A+B=0$. Here we get $s_1(A+B)<s_1(A)+s_2(B)$.

Question 2. Do either of the multiplicative inequalities for singular values have an analogue for the eigenvalues of a Hermitian matrix?

8 A Glance at Ky Fan Norms of Matrices

For any matrix A, we can define its **absolute value**, denoted by $|A|$, in any of the following ways.

1. $|A|$ is the unique positive square root of A^*A.
2. $|A|$ is the positive matrix appearing in the polar decomposition of $A=U\,|A|$.
3. If $A=USV^*$ is the singular value decomposition of A, then $|A|=VSV^*$.

Proposition 15. *Let A be any matrix. Then*

$$s_1(A)+\cdots+s_k(A)=\max_{U,V}|\operatorname{tr}U^*AV|$$

where U,V vary over $n\times k$ matrices with orthonormal columns.

Proof. For this, first note that

$$|\operatorname{tr}U^*AV|=|\operatorname{tr}VU^*A|\leq\operatorname{tr}|VU^*A|=s_1(VU^*A)+\cdots+s_k(VU^*A).$$

Here we can truncate at k, as $\operatorname{rk}U^*AV\leq k$. Then proposition 13 tells us

$$s_j(VU^*A)\leq s_1(VU^*)\,s_k(A)\,.$$

Now, $(VU^*)^*VU^*=UV^*VU^*=UU^*$. This is an orthoprojection, so the top eigenvalue is 1. Hence $s_1(VU^*)=1$. This completes the proof. ☐

Corollary 1. $s_1(A+B)+\cdots+s_k(A+B)\leq s_1(A)+\cdots+s_k(A)+s_1(B)+\cdots+s_k(B)$.

Proof. The statement of the corollary is like a triangule inequality of for function $s_1(X)+\cdots+s_k(X)$. Observe that

$$\begin{aligned}
& s_1(A+B)+\cdots+s_k(A+B) \\
= & \max_{U,V} |\operatorname{tr} U^*(A+B)V| \\
\leq & \max_{U,V} |\operatorname{tr} U^* AV| + \max_{U,V} |\operatorname{tr} U^* BV| \\
= & \|A\|_{(k)} + \|B\|_{(k)}. \\
= & s_1(A)+\cdots+s_k(A)+s_1(B)+\cdots+s_k(B).
\end{aligned}$$

□

This justifies the following definition with some elementary checks.

Definition 5. For any matrix A, we define $\|A\|_{(k)}=s_1(A)+\cdots+s_k(A)$. This is a norm on matrices, and is called the **Ky-Fan** k-norm of A.

Theorem 8. *(Ky Fan, Von Neumann) An inequality holds for every unitarily invariant norm if and only if it holds for the Ky-Fan norms for every k.*

9 Some Results about Positive Matrices

We use the term "positive" for positive semidefinite matrices. They are denoted by $A\geq 0$. A positive definite matrix will be called "strictly positive" and be denoted by $A>0$.

The following are equivalent definitions for a matrix being positive.

1. A is Hermitian and all eigenvalues of A are nonnegative.
2. A is Hermitian and all principal minors are nonnegative (proof uses Descarte's rule of signs).

   The matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ shows that the Hermitian property has to be explicitly added for the two criteria above.
3. $\langle x, Ax \rangle \geq 0$ for all x.
4. A is a correlation matrix. That is there exist vectors $x_1,\ldots,x_n$ such that $a_{ij}=\langle x_i, x_j\rangle$.
5. $A=B^*B$ for some matrix B.
6. $A=R^*R$ for upper triangular R. The diagonal entries of R can be forced to be nonnegative.
7. $A=e^H$ for some Hermitian H.

Proposition 16. *Let $A\geq 0$. Suppose $a_{ii}=0$ for some i. Then the entire i-th row and i-th column is 0.*

Proof. Consider the the principal submatrix corresponding to i, j for any $j \neq i$. Then this submatrix (up to a permutation, depending on where $j > i$) is

$$\begin{bmatrix} 0 & a_{ij} \\ \bar{a}_{ij} & a_{jj} \end{bmatrix}.$$

The corresponding principal minor is $-|a_{ij}|^2$, which is nonnegative only if $a_{ij} = 0$. This concludes the proof. $\square$

Proposition 17. *Suppose* $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \geq 0$. *Then* $\operatorname{rk} A \leq \operatorname{rk} A_{11} + \operatorname{rk} A_{22}$.

Proof. (*Rushil*) Since $A \geq 0$, there exists a matrix X such that $A = X^* X$. Write X as the block matrix

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

such that the block sizes are identical to the corresponding blocks of A. Then

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = X^* X = \begin{bmatrix} X_{11}^* X_{11} + X_{21}^* X_{21} & * \\ * & X_{12}^* X_{12} + X_{22}^* X_{22} \end{bmatrix}.$$

Let B be defined as

$$B = \begin{bmatrix} A_{11} & \\ & A_{22} \end{bmatrix} = \begin{bmatrix} X_{11}^* X_{11} + X_{21}^* X_{21} & \\ & X_{12}^* X_{12} + X_{22}^* X_{22} \end{bmatrix}.$$

Suppose $x \in \ker B$. Then the vector x can be decomposed as $x = [\, x_1 \;\; x_2 \,]$ such that x_1 and x_2 have the same sizes as A_{11} and A_{22} respectively. Then

$$\begin{aligned} 0 = \langle x, Bx \rangle &= \langle x_1, A_{11} x_1 \rangle + \langle x_2, A_{22} x_2 \rangle \\ &= \|X_{11} x_1\|^2 + \|X_{21} x_1\|^2 + \|X_{12} x_2\|^2 + \|X_{22} x_2\|^2 \end{aligned}$$

implying that $X_{11} x_1 = X_{21} x_1 = X_{12} x_2 = X_{22} x_2 = 0$. So

$$Xx = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} X_{11} x_1 + X_{12} x_2 \\ X_{21} x_1 + X_{22} x_2 \end{bmatrix} = 0.$$

Or $Ax = X^* X x = 0 \Longrightarrow x \in \ker A$. Thus $\ker B \subset \ker A$. So $\operatorname{rk} A \leq \operatorname{rk} B$. But B is a block diagonal matrix. Hence $\operatorname{rk} B = \operatorname{rk} A_{11} + \operatorname{rk} A_{22}$.

Alternate proof. Suppose A_{11} and A_{22} are m and n dimensional respectively. Since A_{11}, A_{22} are principal submatrices of a positive matrix, they themselves are positive. So there exist unitary matrices U, V such that $U^* A_{11} U = D_1$ and $V^* A_{22} V = D_2$, where D_1 and D_2 are positive diagonals. Further, exactly $m - \operatorname{rk} A_{11}$ diagonal entries of D_1 are 0, and so are $n - \operatorname{rk} A_{22}$ many of D_2.

Consider the matrix $\bar{A}$ which is a unitary conjugate of A as follows

$$\bar{A} = \begin{bmatrix} U & \\ & V \end{bmatrix}^* A \begin{bmatrix} U & \\ & V \end{bmatrix} = \begin{bmatrix} D_1 & * \\ * & D_2 \end{bmatrix}.$$

Note that $\bar{A}$ has the same rank as A. And $\bar{A}$ is positive. So whenever a diagonal entry of $\bar{A}$ is 0, that entire column of $\bar{A}$ is 0 (proposition 16). But the diagonal entries of $\bar{A}$ are entries of D_1 and D_2. So, at least $m + n - (\operatorname{rk} A_{11} + \operatorname{rk} A_{22})$ many columns of $\bar{A}$ are 0. So $\operatorname{rk} \bar{A} = \operatorname{rk} A$ is at most $\operatorname{rk} A_{11} + \operatorname{rk} A_{22}$. □

Theorem 9. *(Hadamard Inequality) Let $A \geq 0$. Then*

$$\det A \leq \prod_j a_{jj}.$$

Proof. Let $A = R^* R$ where R is upper triangular. Then $a_{jj} = \|r_j\|^2$ where r_j is the j-th row of R. Then

$$\det A = |\det R|^2 = \prod_j |r_{jj}|^2 \leq \prod_j \|r_j\|^2 = \prod_j a_{jj}.$$ □

Corollary 2. *Let A be any matrix. Let a_j be the j-th column of A*

$$|\det A| \leq \prod_j \|a_j\|.$$

Theorem 10. *(Schur Product Theorem) Let $A, B \geq 0$. Then $A \circ B \geq 0$.*

In the lectures we have seen a proof of this by first proving the result when A, B are rank 1 matrices, and then extending this to the general case via the singular value decomposition. Here we shall see two more proofs of this.

Proof. Choose any $x \in \mathcal{H}$. Let $X = \operatorname{Diag}(x)$. Note that

$$\begin{aligned} \langle x, (A \circ B)\, x \rangle &= \sum_{j,k=1}^{n} \bar{x}_j\, (A \circ B)_{jk}\, x_k \\ &= \sum_{j,k=1}^{n} (\bar{x}_j\, A_{jk}\, x_k)\, B_{kj} \\ &= \sum_{j,k=1}^{n} (X^* A X)_{jk}\, B_{kj} \\ &= \operatorname{tr}\, (X^* A X)\, B. \end{aligned}$$

Then $X^* A X$ and B are both positive, and a product of positive matrices has positive eigenvalues. This completes the proof.

Alternate proof. Since B is positive we know that there exists vectors $v_1, \ldots, v_n$ such that $b_{ij} = \langle v_i, v_j \rangle$. Suppose $v_i = [\, v_{i1} \ \cdots \ v_{in} \,]^T$. Observe that

$$\begin{aligned}
\langle x, (A \circ B)\, x \rangle &= \sum_{i,j=1}^{n} \bar{x}_i \, a_{ij} \, \langle v_i, v_j \rangle \, x_j \\
&= \sum_{i,j=1}^{n} \bar{x}_i \, a_{ij} \, x_j \sum_{k=1}^{n} \bar{v}_{ik} v_{jk} \\
&= \sum_{k=1}^{n} \sum_{i,j=1}^{n} (\bar{x}_i \, \bar{v}_{ik}) \, a_{ij} \, (x_j \, v_{jk}) \\
&= \sum_{k=1}^{n} \langle w_k, A w_k \rangle \\
&\geq 0,
\end{aligned}$$

where $w_k = [\, x_1 v_{1k} \ \cdots \ x_n v_{nk} \,]^T$. The last inequality follows from the positivity of A. □

Definition 6. A matrix A is said to be **congruent** to another matrix B, if there exists an X nonsingular, for which $B = X^* A X$.

Congruence is an equivalence relation on the space of matrices. Some other examples of equivalence relations for matrices are similarity, unitary equivalence etc. For relations it is meaningful to talk about invariants and complete invariants.

Definition 7. Let A be a Hermitian matrix. Then let $\pi(A), \zeta(A), \nu(A)$ denote respectively the number of positive, zero and negative eigenvalues of A. The the tuple $(\pi(A), \zeta(A), \nu(A))$ is denoted by $\mathrm{in}(A)$, and is called the **inertia** of A.

Theorem 11. *(Sylvester's law of inertia) The inertia is a complete invariant for congruence in the space of Hermitian matrices. That is, if A is Hermitian, then A and X^*AX have the same inertia, for any invertible matrix X.*

We will see a proof of this using minmax principles

10 Minmax Principles and Conseqeuences

In this discussion, the eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$ of a Hermitian matrix A are always arranged in decreasing order.

Remark 6. *How to go back and forth from eigenvalue inequalities of Hermitian matrices and singular value inequalities of general matrices?*

Any inequality that holds for singular values of arbitrary matrices holds for the eigenvalues of positive matrices. Additive inequalties of such nature can therefore be extended to eigenvalues of Hermitian matrices by noting the following. For any Hermitian matrix A, there exists α such that $A+\alpha$ is positive. And $\lambda(A+\alpha)=\lambda(A)+\alpha$.

In the reverse direction one notes that for any matrix A, the matrix

$$\tilde{A}=\begin{bmatrix} & A \\ A^* & \end{bmatrix}$$

is Hermitian, and

$$\frac{1}{2}\begin{bmatrix} U & -U \\ V & V \end{bmatrix}\begin{bmatrix} S & \\ & -S \end{bmatrix}\begin{bmatrix} U^* & V^* \\ -U^* & V^* \end{bmatrix}$$

is its eigendecomposition. Here U,S,V come from the singular value decomposition $A=USV^*$ of A. Thus the eigenvalues of $\tilde{A}$ are the singular values of A.

Theorem 12. *(Minmax principles, Courant-Fisher-Weyl) Let A be a Hermitian matrix with eigenvalues $\lambda_1(A)\geq\cdots\geq\lambda_n(A)$. Then for $1\leq k\leq n$*

$$\lambda_k=\max_{\dim\mathcal{M}=k}\ \min_{\substack{x\in\mathcal{M}\\ \|x\|=1}}\langle x,Ax\rangle=\min_{\dim\mathcal{M}=n-k+1}\ \max_{\substack{x\in\mathcal{M}\\ \|x\|=1}}\langle x,Ax\rangle\,.$$

In the above equation the optimization over unit vectors can be replaced by nonzero vectors by normalizing the objective $\langle x,Ax\rangle$ to $\left\langle \frac{x}{\|x\|},A\frac{x}{\|x\|}\right\rangle$.

The minmax principles also extend to the singular values where the kth singular value s_k of an arbitary matrix A is obtained by replacing the objective $\langle x,Ax\rangle$ by $\|Ax\|$ in the equation of theorem 12.

Some corollaries of the minmax principles are follows.

1. Weyl's monotonicity principle. This says that when $A\leq B$ then

   $$\lambda_j(A)\leq\lambda_j(B)\quad\text{for all } j=1,\ldots,n.$$

2. Weyl's inequalities. These are the following set of inequalities for Hermitian matrices A,B with eigenvalues arranged in decreasing order.

   $$\begin{aligned}\lambda_i(A+B)&\leq\lambda_j(A)+\lambda_k(B)\quad\text{when } i+1=j+k,\\ \lambda_i(A+B)&\geq\lambda_j(A)+\lambda_k(B)\quad\text{when } i+n=j+k.\end{aligned}$$

   These inequalities further imply Weyl's perturbation theorem, which states that

   $$|\lambda_j(A)-\lambda_j(B)|\leq\|A-B\|\,.$$

3. Sylvester's law of inertia.

4. Cauchy's interlacing principles.

We will see the proof of last two corollaries here.

Proof. (Theorem 11) Let X be any nonsingular matrix. Let $B = X^*AX$. Firstly, as X is nonsingular, we know that $\operatorname{rk} A = \operatorname{rk} B$. This immediately gives $\zeta(A) = \zeta(B)$. Since $\pi(Z) + \zeta(Z) + \nu(Z) = n$ for any Hermitian Z, it sufficies to show that $\pi(B) = \pi(A)$.

Suppose the eigenvalues of A and B are arranged in the decreasing order. Let $k = \pi(A)$. We will first show that $\pi(A) \geq \pi(B)$. To show this it is sufficient to show that $\lambda_k(A) > 0$, where $k = \pi(B)$.

Let $\mathcal{M}$ be the span of top k eigenvectors of B. As X is nonsingular, we get $\dim X\mathcal{M} = k$. Then by the minmax principles

$$\lambda_k(A) \geq \min_{\substack{y \in X\mathcal{M} \\ y \neq 0}} \frac{\langle y, Ay \rangle}{\|y\|^2} = \min_{\substack{x \in \mathcal{M} \\ x \neq 0}} \frac{\langle Xx, AXx \rangle}{\|Xx\|^2} \geq \frac{1}{\|X\|^2} \min_{\substack{x \in \mathcal{M} \\ x \neq 0}} \frac{\langle x, Bx \rangle}{\|x\|^2} = \frac{\lambda_k(B)}{\|X\|^2} > 0.$$

But if $B = X^* AX$, then $A = (X^{-1})^* B\, (X^{-1})$. Thus applying the same argument to this congruence we get $\pi(A) \geq \pi(B)$. This shows that two congruent matrices have the same inertia. $\square$

Proposition 18. *(Cauchy's interlacing principles) Let $A \in \mathbb{H}$ be of the form*

$$A = \begin{bmatrix} B & * \\ * & * \end{bmatrix},$$

where $B \in \mathbb{H}$ is $k \times k$. Then for all $1 \leq j \leq k$ we get

$$\lambda_{j+(n-k)}(A) \leq \lambda_j(B) \leq \lambda_j(A)$$

Proof. Let $\mathcal{M}$ be the span of the top j eigenvectors of B. Identify any $x \in \mathcal{M}$ with $\tilde{x} \in \mathcal{H}$ where $\tilde{x} = (x, 0, \ldots, 0)$, where the 0s repeat $n - k$ times. Then the collection of $\tilde{x}$s form a j-dimensional subspace, say $\tilde{\mathcal{M}}$ of $\mathcal{H}$. Further, we get $\|x\| = \|\tilde{x}\|$ and $\langle x, Bx \rangle = \langle \tilde{x}, A\tilde{x} \rangle$.

Then, by the minmax principles,

$$\lambda_j(B) = \min_{\substack{x \in \mathcal{M} \\ \|x\| = 1}} \langle x, Bx \rangle = \min_{\substack{\tilde{x} \in \tilde{\mathcal{M}} \\ \|\tilde{x}\| = 1}} \langle \tilde{x}, A\tilde{x} \rangle \leq \lambda_j(A).$$

This shows one side of the inequality. Now, apply this inequality to the matrices $-A$ and $-B$ which also satisfy the hypothesis of the proposition. This gives

$$\lambda_j(-B) \le \lambda_j(-A) \Longrightarrow \lambda_{k-j+1}(B) \ge \lambda_{n-j+1}(A).$$

By replacing $k-j+1$ by j, we get the required inequality. □

When $k=n-1$, the above inequalities say

$$\lambda_1(A) \ge \lambda_1(B) \ge \lambda_2(A) \ge \lambda_2(B) \ge \cdots \ge \lambda_{n-1}(B) \ge \lambda_n(A)\,.$$

The theorem gets its name from this interlacing phenomenon.

Proposition 19. *Let A be any matrix. Then $\lambda_j(\operatorname{Re} A) \le s_j(A)$.*

Proof. (*Pratyush*) The inequality is obtained by taking minmax on the following inequality for any unit vector x:

$$\langle x, (\operatorname{Re} A)x\rangle = \operatorname{Re}\langle x, Ax\rangle \le |\langle x, Ax\rangle| \le \|Ax\|.$$

Here the last inequality is Cauchy-Schwarz inequality. □

As we mentioned earlier, eigenvalue inequalities for Hermitian matrices usually correspond to singular value inequalities of arbitrary matrices. The following is a variational characterization for the sum of top eigenvalues of a Hermitian matrix, which is an analogue of the variational characterization for the Ky Fan norms.

Proposition 20. *Let A be any matrix. Then for any $k \le n$*

$$\lambda_1(A) + \cdots + \lambda_k(A) = \max \operatorname{tr} U^* A U$$

where U varies over $n \times k$ matrices with orthonormal columns.

Proof. Complete the orthonormal basis containing the columns of U to form the unitary matrix $W = [\, U \;\; V\,]$. Then

$$W^* A W = \begin{bmatrix} U^* A U & * \\ * & * \end{bmatrix}.$$

Then, by Cauchy's interlacing principles applied to $U^* A U$ and $W^* A W$

$$\operatorname{tr} U^* A U = \lambda_1(U^* A U) + \cdots + \lambda_k(U^* A U) \le \lambda_1(A) + \cdots + \lambda_k(A).$$

Further, choose the columns of U as the top k eigenvectors of A fetches equality in the above inequality. $\square$

Corollary 3. *The function* $X \longmapsto \lambda_1(X) + \cdots + \lambda_k(X)$ *on Hermitian matrices is convex.*

Remark 7. Let f be a function on the vector space V over $\mathbb{R}$. Then f is **convex** if

$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y)$$

for all $x, y \in V$ and $\alpha, \beta \in \mathbb{R}$ such that $\alpha + \beta = 1$. If the above condition holds for all x, y but only for $\alpha = \beta = \frac{1}{2}$, then f is called **mid-point convex**. If f is continuous, convexity and mid-point convexity are equivalent. If f is affine, that is $f(\alpha x) = \alpha f(x)$ for all $\alpha \in \mathbb{R}$ and $x \in V$, the convexity is equivalent to subadditivity, that is,

$$f(x + y) \leq f(x) + f(y)\,.$$

11 A Brief Introduction to Majorisation

Let A be Hermitian. By some unitary conjugation, assume that the diagonal entries of A are arranged in decreasing order. Further, let λ_js be its eigenvalues in decreasing order. Then hiding in proposition 20 are the following relations.

$$\begin{array}{rcll} a_{11} + \cdots + a_{kk} & \leq & \lambda_1 + \cdots + \lambda_k & \text{for all } 1 \leq k \leq n \\ a_{11} + \cdots + a_{nn} & = & \lambda_1 + \cdots + \lambda_n & . \end{array}$$

Such relationship has a special name in the theory of convex functions. Let x, $y \in \mathbb{R}^n$. Then let $x^{\downarrow} = (x_1^{\downarrow}, \ldots, x_n^{\downarrow})$ be the vector obtained from x such that $x^{\downarrow}$ and x have the same set of coordinates except that $x^{\downarrow}$ arranges them in decreasing order. If x is a sample, then $x^{\downarrow}$ is the ordered statistic.

Definition 8. For $x, y \in \mathbb{R}^n$ we say x is **majorised** by y, and denote it by $x \prec y$ if the following relations hold.

$$\begin{array}{rcll} x_1^{\downarrow} + \cdots + x_k^{\downarrow} & \leq & y_1^{\downarrow} + \cdots + y_k^{\downarrow} & \text{for all } 1 \leq k \leq n \\ x_1^{\downarrow} + \cdots + x_n^{\downarrow} & = & y_1^{\downarrow} + \cdots + y_k^{\downarrow} & . \end{array}$$

If the last equality is replaced by $\leq$, then we say x is **weakly majorised** by y, and denote this by $x \prec_w y$.

For example, let x be any vector such that sum of its coordinates is 1. Then we have $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \prec x \prec (1, 0, \ldots, 0)$.

The above discussion about Hermitian matrices gives the following result.

Theorem 13. *(Schur's Theorem) Let A be a Hermitian matrix. Let $d(A)$ be the vector of the diagonal entries of A, and let $\lambda(A)$ be the vector of eigenvalues of A. Then $d(A) \prec \lambda(A)$.*

Majorisation of vectors interacts well with convex functions. A concrete example of this is the following theorem.

Lemma 3. *Let $\varphi: \mathbb{R} \to \mathbb{R}$ be convex. If $x \prec y$, then $\varphi(x) \prec_w \varphi(y)$.*

The triangle inequality for the Ky-Fan norms can be alternatively read as the relation $s(A+B) \prec s(A) + s(B)$. Note that $t \to t^p$ is a convex function for $p \geq 1$. As a corollary we get the triangle inequality for the Schatten-p norms.

Proposition 21. *Define $\|A\|_p = \|s(A)\|_p$ where $s(A)$ is the vector of singular values of A. Then $\|A\|_p$ is a norm on $\mathbb{M}(n)$.*

Proof. Sufficient to show the triangle inequality for this function, as the other requirements are easy. Note that

$$\begin{aligned}
\|A+B\|_p^p &= \sum_{j=1}^{n} s_j^p(A+B) \\
&\leq \sum_{j=1}^{n} [s(A)+s(B)]^p \\
&\leq (\|s(A)\|_p + \|s(B)\|_p)^p \\
&= (\|A\|_p + \|B\|_p)^p,
\end{aligned}$$

where the second last inequality follows from triangle inequality for the p-norm on vectors. □