Linear Algebra Review

GIAN course, 18-22 Dec, 2017, IIIT-Delhi

Outline

Chapter 0: Miscellaneous Preliminaries

Unless otherwise noted, all vectors are elements of $\mathbb{R}^n$, although results extend to complex vector spaces.
Let $S = \{\mathbf{v}_i\}_1^k \subseteq \mathbb{C}^n$. We define **span** of S by

$$span\{S\} = \left\{ \mathbf{x} \in \mathbb{C}^n \,|\, \exists \{\alpha_i\}_1^k \subseteq \mathbb{C} \,\text{with}\, \mathbf{x} = \sum_1^k \alpha_i \mathbf{v}_i \right\}$$

- The set S is **linearly independent** if $\sum_1^k \alpha_i \mathbf{v}_i = \mathbf{0}$ if and only if $\alpha_i = 0$ for all i.
- S is a **spanning set** for vector space V if span $\{S\} = V$.
- A linearly independent spanning set for a vector space V is called a **basis**.
- The **dimension** of a vector space V, $\dim(V)$, is the size of the smallest spanning set for V.

Chapter 0: Miscellaneous Preliminaries

- The **rank** of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, rank($\mathbf{A}$), is the size of the largest linearly independent set of columns of $\mathbf{A}$. Rank satisfies rank($\mathbf{A}$) $\leq$ min{m,n}.
- For matrices $\mathbf{A} \in \mathbb{R}^{m \times k}$ and $\mathbf{B} \in \mathbb{R}^{k \times n}$, we have

  $$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - k \leq \text{rank}(\mathbf{AB}) \leq min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}.$$

- For the two matrices of same size, we have

  $$\text{rank}(\mathbf{A} + \mathbf{B}) \leq rank(\mathbf{A}) + rank(\mathbf{B}).$$

- The **trace** of a matrix is the sum of its main diagonal entries, that is, trace($\mathbf{A}$) $= \sum a_{ii}$. For two matrices $\mathbf{A}$ and $\mathbf{B}$, trace($\mathbf{AB}$) $=$ trace($\mathbf{BA}$).

Chapter 0: Miscellaneous Preliminaries

- The **range** of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, range$(\mathbf{A})$, is the set

  $$\text{range}(\mathbf{A}) = \{\mathbf{b} \in \mathbb{R}^m \,|\, \exists\, \mathbf{x} \in \mathbb{R}^n \ \text{ with } \ \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

  Equivalently, the range of $\mathbf{A}$ is the set of all linear combinations of columns of $\mathbf{A}$.
- The **nullspace** of a matrix, null$(\mathbf{A})$ is the set of all vectors $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = 0$.

Chapter 0: Miscellaneous Preliminaries

Suppose we have the matrix-vector equation $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$.

- The equation is **consistent** if there exists a solution $\mathbf{x}$ to this equation; equivalently, we have $\text{rank}([\mathbf{A}, \mathbf{b}]) = \text{rank}(\mathbf{A})$, or $\mathbf{b} \in \text{range}(\mathbf{A})$.
- The equation has a **unique solution** if $\text{rank}([\mathbf{A}, \mathbf{b}]) = \text{rank}(\mathbf{A}) = n$.
- The equation has **infinitely many solutions** if $\text{rank}([\mathbf{A}, \mathbf{b}]) = \text{rank}(\mathbf{A}) < n$.
- The equation has **no solution** if $\text{rank}([\mathbf{A}, \mathbf{b}]) > \text{rank}(\mathbf{A})$.

Chapter 0: Miscellaneous Preliminaries

- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is **nonsingular** if $\mathbf{Ax} = 0$ if and only if $\mathbf{x}= 0$.
- When $m \geq n$ and $\mathbf{A}$ has full rank (rank($\mathbf{A}$)$= n$), $\mathbf{A}$ is nonsingular.
- If $m < n$, then $\mathbf{A}$ must be singular, since rank($\mathbf{A}$) $\leq min\{m, n\} = m < n$.
- If $m = n$ and $\mathbf{A}$ is nonsingular, then there exists a matrix $\mathbf{A}^{-1}$ with $\mathbf{AA}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A}$ and we call $\mathbf{A}$ invertible and the matrix-vector equation $\mathbf{Ax} = \mathbf{b}$ has the unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ in this case.

Chapter 0: Miscellaneous Preliminaries

- The **Euclidean inner product** is a function defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \mathbf{y} = \sum_1^n \bar{x}_i y_i$, where the vectors in use are the same size and could be complex-valued.
- The **Euclidean norm** is a function defined by $\|\mathbf{x}\| = \|\mathbf{x}\|_2 = (\sum_1^n |x_i|^2)^{1/2}$ and satisfies $\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|_2^2$.
- Two vectors x and y are **orthogonal** if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.
- Two vectors are **orthonormal** if they are are orthogonal and $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$.

Chapter 0: Miscellaneous Preliminaries

Given a set $S = \{\mathbf{v}_i\}_1^k \subseteq \mathbb{R}^n$, we define the orthogonal complement of S ("S perp") by

$$S^\perp = \{\mathbf{x} \in \mathbb{R}^n \,|\, \langle \mathbf{x}, \mathbf{v}_i \rangle = 0 \ \text{ for all } \ i\}\,.$$

$$(S^\perp)^\perp = S$$

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, we have the relation

$$\text{range}(\mathbf{A})^\perp = \text{null}(\mathbf{A}^*).$$

Proof: $(\subseteq)$ Let $\mathbf{y} \in \text{range}(\mathbf{A})^\perp$. Then for all $\mathbf{b} = \mathbf{A}\mathbf{x} \in \text{range}(\mathbf{A})$, $0 = \langle \mathbf{b}, \mathbf{y} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}^*\mathbf{y} \rangle$. In particular, if $\mathbf{x} \equiv \mathbf{A}^*\mathbf{y}$, then we have $\|\mathbf{A}^*\mathbf{y}\|_2^2 = 0$, so that $\mathbf{A}^*\mathbf{y} \equiv 0$. Thus $\mathbf{y} \in \text{null}(\mathbf{A}^*)$.
$(\supseteq)$ Let $\mathbf{y} \in \text{null}(\mathbf{A}^*)$. Then for all $\mathbf{x} \in \mathbb{R}^n$, we have $0 = \langle \mathbf{x}, \mathbf{A}^*\mathbf{y} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle$. As this holds for all choices of $\mathbf{x}$, we conclude that $\mathbf{y} \in \text{range}(\mathbf{A})^\perp$. Hence, the set equality follows. $\square$

Chapter 0: Miscellaneous Preliminaries

- **Projection Matrix:** Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then for all $\mathbf{y} \in \mathbb{R}^m$, there exists unique vectors $\mathbf{y}_A$ and $\mathbf{y}_\perp$ in $\mathbb{R}^m$ such that $\mathbf{y} = \mathbf{y}_A + \mathbf{y}_\perp$, where $\mathbf{y}_A \in \text{range}(\mathbf{A})$ and $\mathbf{y}_\perp \in \text{range}(\mathbf{A})^\perp \equiv \text{null}(\mathbf{A}^*)$
- A **normal matrix** is a matrix $\mathbf{N}$ such that $\mathbf{N}\mathbf{N}^* = \mathbf{N}^*\mathbf{N}$.
- A **Hermitian matrix** is one such that $\mathbf{A}^* = \mathbf{A}$. A real valued Hermitian matrix is called a **symmetric matrix**.
- A **skew Hermitian matrix** is one such that $\mathbf{A}^* = -\mathbf{A}$.
- A **unitary matrix** is a square matrix with $\mathbf{U}\mathbf{U}^* = \mathbf{I} = \mathbf{U}^*\mathbf{U}$.
- A real-valued unitary matrix is called an **orthogonal matrix**.
- An **idempotent matrix** satisfies $\mathbf{A}^2 = \mathbf{A}$.

Chapter 0: Miscellaneous Preliminaries

- A **Projection matrix** $\mathbf{P}$ satisfies $\mathbf{P}^2 = \mathbf{P}$ ($\mathbf{P}$ is idempotent). If $\mathbf{P} = \mathbf{P}^*$, then $\mathbf{P}$ is called an **orthogonal projection**.
- For any projection $\mathbf{P}$ which projects onto a subspace S, the projector onto subspace $S^\perp$ is given by $(\mathbf{I} - \mathbf{P})$. Given a matrix $\mathbf{U}$ with orthonormal columns, the (orthogonal) projector onto column space of $\mathbf{U}$ is given by $\mathbf{P} = \mathbf{U}\mathbf{U}^*$.
- **Classical Gram Schmidt Algorithm** is a theoretical tool which takes a set of vectors $\{\mathbf{v}_i\}_1^k$ and creates a set of orthonormal vectors $\{\mathbf{q}_i\}_1^k$ which span the same space as the original set.
  Let $\{\mathbf{v}_i\}_1^k$ be a linearly independent set of vectors.
  Initialize $\mathbf{z}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|_2}$. For $l = 2...k$, compute
  $\mathbf{y}_l = \left(\mathbf{v}_l - \sum_{i=1}^{l-1} \langle \mathbf{z}_i, \mathbf{v}_l \rangle \mathbf{z}_i\right)$ and then let $\mathbf{z}_l = \frac{\mathbf{y}_l}{\|\mathbf{y}_l\|_2}$.

Chapter 0: Miscellaneous Preliminaries

- A **permutation matrix** is a matrix obtained by permuting rows and(or) columns of an identity matrix. Permutation matrices satisfy $\mathbf{P}^2 = \mathbf{I}$, so that a permutation matrix is its own inverse.
- Permutation matrices are symmetric and orthogonal.
- A **circulant matrix** has the general form

$$\begin{pmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \ddots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \dots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_2 & a_3 & \dots & a_n & a_1 \end{pmatrix}$$

Each row of a circulant matrix is a cyclic permutation of the first row.

Chapter 0: Miscellaneous Preliminaries

- A **Toeplitz matrix** has the general form

$$\begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_{-1} & a_0 & a_1 & \ddots & a_{n-1} & a_{n-1} \\ a_{-2} & a_{-2} & a_0 & \ddots & \ddots & a_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{-n+1} & a_{-n+2} & \ddots & \ddots & a_0 & a_1 \\ a_{-n} & a_{-n+1} & a_{-n+2} & \dots & a_{-1} & a_0 \end{pmatrix}$$

  where $\{a_i\}_{-n}^{n}$ is any collection of scalars.
- An **upper triangular matrix** is a matrix whose entries below (above) the main diagonal are all zero.
- A diagonal matrix is one whose only nonzero entries lie on the main diagonal. The eigen values of a triangular matrix or a diagonal matrix are the diagonal entries.

Chapter 0: Miscellaneous Preliminaries

- Let $\mathbf{T}$ be an upper-triangular matrix. $\mathbf{T}$ is invertible if and only if its diagonal entries are nonzero (since these are its eigenvalues).
- The matrix $\mathbf{T}^{-1}$ is also upper-triangular.
- Given any two upper-triangular matrices $\mathbf{T}_1$ and $\mathbf{T}_2$, their sum $\mathbf{T}_1 + \mathbf{T}_2$ and their product $\mathbf{T}_1\mathbf{T}_2$ are also upper-triangular.
- A Hermitian upper-triangular matrix is necessarily diagonal (and real-valued).
- More generally, any normal upper-triangular matrix is diagonal.

These results also hold for lower-triangular matrices.

Chapter 1: Eigen Values and Similarity

Suppose $\mathbf{A} \in \mathbb{C}^{n \times n}$ and there exist $\lambda \in \mathbb{C}$ and $\mathbf{x} \in \mathbb{C}^n$ (with $\mathbf{x} \neq 0$) such that $\mathbf{Ax} = \lambda\mathbf{x}$. Then we call λ an eigenvalue of $\mathbf{A}$ with the corresponding **eigenvector** $\mathbf{x}$.

- The **spectrum** of $\mathbf{A}$, $\sigma(\mathbf{A})$, is the set of all eigenvalues of $\mathbf{A}$.
- The **spectral radius** of $\mathbf{A}$, $\rho(\mathbf{A})$, is defined as $\rho(\mathbf{A}) = max_{\lambda \in \sigma(\mathbf{A})}\{|\lambda|\}$.
- If $\mathbf{Ax} = \lambda\mathbf{x}$, then $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ for $k \in \mathbb{N} \cup \{0\}$.
  Proof: By induction
  For the base case $k = 0$, the result is obvious. (Note $\mathbf{A}^0 = \mathbf{I}$). Suppose the result is true for $k = m \geq 0$. Consider $\mathbf{A}^{m+1}\mathbf{x} = \mathbf{A}^m(\mathbf{Ax}) = \mathbf{A}^m(\lambda\mathbf{x}) = \lambda(\mathbf{A}^m\mathbf{x}) = \lambda.\lambda^m\mathbf{x} = \lambda^{m+1}\mathbf{x}$. By the Principle of Mathematical Induction, the desired result follows. $\square$

Chapter 1: Eigen Values and Similarity

- If $\mathbf{A}$ is a Hermitian Matrix, then $\sigma(\mathbf{A}) \subset \mathbb{R}$.
  Proof: Let $\mathbf{Ax} = \lambda\mathbf{x}$. Then $\langle \mathbf{Ax}, \mathbf{x} \rangle = \langle \lambda\mathbf{x}, \mathbf{x} \rangle = \bar{\lambda}\langle \mathbf{x}, \mathbf{x} \rangle$.
  However,
  $\langle \mathbf{Ax}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{A}^*\mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{Ax} \rangle = \langle \mathbf{x}, \lambda\mathbf{x} \rangle = \lambda\langle \mathbf{x}, \mathbf{x} \rangle$. Since, $\mathbf{x}$ is an eigenvector, $\mathbf{x} \neq \mathbf{0}$, so $\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|_2^2 \neq 0$. Thus $\lambda = \bar{\lambda}$, so $\lambda \in \mathbb{R}$. $\square$
- A square matrix $\mathbf{A}$ is invertible if and only if 0 is not an eigenvalue of $\mathbf{A}$.
- $\mathbf{A} \sim \mathbf{B}$ if there exists a nonsingular matrix $\mathbf{S}$ such that $\mathbf{A} = \mathbf{S}^{-1}\mathbf{BS}$.
- $\mathbf{A} \sim \mathbf{B}$ implies $\sigma(\mathbf{A}) = \sigma(\mathbf{B})$.
  Proof: Let $\lambda \in \sigma(\mathbf{A})$. Then there exists $\mathbf{x} \neq 0$ such that $\mathbf{Ax} = \lambda\mathbf{x}$. Applying similarity, we have $\mathbf{Ax} = \mathbf{S}^{-1}\mathbf{BSx} = \lambda\mathbf{x}$, which implies that $\mathbf{B}(\mathbf{Sx}) = \lambda(\mathbf{Sx})$. Since $\mathbf{S}$ is invertible and $\mathbf{x} \neq 0, \mathbf{Sx} \neq 0$, so $\lambda \in \sigma(\mathbf{B})$. $\square$

Chapter 1: Eigen Values and Similarity

- Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. $\mathbf{A}$ is **diagonalizable** if $\mathbf{A}$ is similar to a diagonal matrix Λ whose (diagonal) entries are the eigenvalues of $\mathbf{A}$, that is, there exists $\mathbf{S}$ such that $\mathbf{A} = \mathbf{S}^{-1}\Lambda\mathbf{S}$.
- A matrix $\mathbf{A}$ is **unitarily diagonalizable** if $\mathbf{S}$ is a unitary matrix: $\mathbf{A} = \mathbf{U}^{*}\Lambda\mathbf{U}$ for $\mathbf{U}$ unitary.
- A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors.
  Proof: Suppose $\mathbf{A} = \mathbf{S}^{-1}\Lambda\mathbf{S}$. Then $\mathbf{A}\mathbf{S}^{-1} = \mathbf{S}^{-1}\Lambda$. The matrix $\mathbf{A}\mathbf{S}^{-1}$ has columns $(\mathbf{A}\mathbf{S}^{-1})_j = \mathbf{A}(\mathbf{S}^{-1})_j$ and the matrix $\mathbf{S}^{-1}\Lambda$ has columns $(\mathbf{S}^{-1}\Lambda)_j = \lambda_j(\mathbf{S}^{-1})_j$. Therefore, the columns of $\mathbf{S}^{-1}$ are the eigenvectors of $\mathbf{A}$. Since $\mathbf{S}^{-1}$ is invertible, it has n linearly independent columns, which proves the result. $\square$

Chapter 2: Triangularization and Factorizations

- Two matrices $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{C}^{n\times n}$ are **unitarily equivalent** if there exists unitary matrices $\mathbf{U}$ and $\mathbf{V}$ such that $\mathbf{A}$=$\mathbf{UBV}$.
- The matrices are **unitarily similar** if $\mathbf{A} = \mathbf{U}^*\mathbf{BU}$ for some unitary $\mathbf{U}$.
- **Schur's Unitary Triagularization Theorem**: Every matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is unitarily similar to an upper-triangular matrix $\mathbf{T}$ whose diagonal entries are the eigenvalues of $\mathbf{A}$; that is, there exist $\mathbf{U}$ unitary and $\mathbf{T}$ upper-triangular with $t_{ii} = \lambda_i(\mathbf{T})$ such that $\mathbf{A} = \mathbf{U}^*\mathbf{TU}$. If $\mathbf{A}$ is real and has only real eigenvalues, then $\mathbf{U}$ can be chosen real (orthogonal).

- A consequence of Schurs theorem is that if $\mathbf{A}$ is normal, then $\mathbf{T}$ is also normal.
  Proof: Let $\mathbf{A} = \mathbf{U}^*\mathbf{T}\mathbf{U}$. Then
  $\mathbf{A}\mathbf{A} = \mathbf{U}^*\mathbf{T}^*\mathbf{U}\mathbf{U}^*\mathbf{T}\mathbf{U} = \mathbf{U}^*\mathbf{T}\mathbf{T}^*\mathbf{U}$ and
  $\mathbf{A}\mathbf{A}^* = \mathbf{U}^*\mathbf{T}\mathbf{U}\mathbf{U}^*\mathbf{T}^*\mathbf{U} = \mathbf{U}^*\mathbf{T}\mathbf{T}^*\mathbf{U}$. Therefore,
  $\mathbf{U}^*\mathbf{T}\mathbf{U} = \mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^* = \mathbf{U}^*\mathbf{T}\mathbf{T}^*\mathbf{U}$, so $\mathbf{T}^*\mathbf{T} = \mathbf{T}\mathbf{T}^*$, as desired.
- Another consequence of Schur's theorem is that
  $\text{trace}(\mathbf{A}) = \sum_{\sigma(\mathbf{A})} \lambda_i$.
  Proof: There exist $\mathbf{U}$ unitary and $\mathbf{T}$ upper-triangular such that $\mathbf{A} = \mathbf{U}^*\mathbf{T}\mathbf{U}$ with $t_{ii} = \lambda_i$. So,
  $\text{trace}(\mathbf{A}) = \text{trace}(\mathbf{U}^*\mathbf{T}\mathbf{U}) = \text{trace}(\mathbf{T}\mathbf{U}\mathbf{U}^*) = \text{trace}(\mathbf{T}) = \sum t_{ii} = \sum_{\sigma(\mathbf{A})} \lambda_i$.

Chapter 2: Triangularization and Factorizations

The following are equivalent:

1. $\mathbf{A}$ is normal.
2. $\mathbf{A}$ is unitarily diagonalizable.
3. $\mathbf{A}$has n orthonormal eigenvectors.
4. $\sum |a_{ij}|^2 = \sum |\lambda_i|^2$.

Proof:

- $(1 \Rightarrow 2)$ If $\mathbf{A}$ is normal, then $\mathbf{A} = \mathbf{U}^*\mathbf{T}\mathbf{U}$ implies $\mathbf{T}$ is also normal. A normal triangular matrix is diagonal, so $\mathbf{A}$ is unitarily diagonalizable.
- $(2 \Rightarrow 1)$ Let $\mathbf{A} = \mathbf{U}^*\mathbf{\Lambda}\mathbf{U}$. Since diagonal matrices commute, $\mathbf{\Lambda}^*\mathbf{\Lambda} = \mathbf{\Lambda}\mathbf{\Lambda}^*$, so $\mathbf{A}^*\mathbf{A} = \mathbf{U}^*\mathbf{\Lambda}^*\mathbf{\Lambda}\mathbf{U} = \mathbf{U}^*\mathbf{\Lambda}\mathbf{\Lambda}^*\mathbf{U} = \mathbf{A}\mathbf{A}^*$, and thus $\mathbf{A}$ is normal.

- $(2 \Rightarrow 3)$ $\mathbf{A} = \mathbf{U}^*\mathbf{\Lambda}\mathbf{U}$ if and only if $\mathbf{A}\mathbf{U}^* = \mathbf{U}^*\mathbf{\Lambda}$. As we saw in the section on eigenvalues, this is true if and only if the columns of $\mathbf{U}^*$ are the eigenvectors of $\mathbf{A}$. These eigenvectors are orthonormal since $\mathbf{U}^*$ is unitary.
- $(2 \Rightarrow 4)$ Let $\mathbf{A} = \mathbf{U}^*\mathbf{\Lambda}\mathbf{U}$. Consider $\sum |a_{ij}|^2 = \text{trace}(\mathbf{A}^*\mathbf{A}) = \text{trace}(\mathbf{U}^*\mathbf{\Lambda}^*\mathbf{\Lambda}\mathbf{U}) = \text{trace}(\mathbf{\Lambda}^*\mathbf{\Lambda}\mathbf{U}\mathbf{U}^*) = \text{trace}(\mathbf{\Lambda}^*\mathbf{\Lambda}) = \sum |\lambda_i|^2$.
- $(4 \Rightarrow 2)$) Suppose that $\sum |a_{ij}|^2 = \text{trace}(\mathbf{A}^*\mathbf{A}) = \sum |\lambda_i|^2$. By Schurs thereom, $\mathbf{A} = \mathbf{U}^*\mathbf{T}\mathbf{U}$ for some upper-triangular $\mathbf{T}$. We have $\text{trace}(\mathbf{A}^*\mathbf{A}) = \text{trace}(\mathbf{T}^*\mathbf{T}) = \sum_{i,j} |t_{ij}|^2 = \sum |t_{ii}|^2 + \sum_{i \neq j} |t_{ij}|^2$. Since $t_{ii} \equiv \lambda_i$, we see that $\sum_{i \neq j} |t_{ij}|^2 = 0$, which implies that $t_{ij} \equiv 0$ for all $i \neq j$. Thus, $\mathbf{T}$ is diagonal and $\mathbf{A}$ is, therefore, unitarily diagonalizable. $\square$

Chapter 2: Triangularization and Factorizations

- A matrix is Hermitian if and only if $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$ with Λ diagonal and real. Further, a normal matrix whose eigenvalues are real is neccesarily Hermitian.
  Proof: $\mathbf{A} = \mathbf{A}^* \Leftrightarrow \mathbf{U}^*\mathbf{T}\mathbf{U} = \mathbf{U}^*\mathbf{T}^*\mathbf{U} \Leftrightarrow \mathbf{T} = \mathbf{T}^* \Leftrightarrow \mathbf{T} = \boldsymbol{\Lambda}$ is diagonal and real valued, which proves the first result. Since normal matrices are unitarily diagonizable, the second result follows.
- **QR factorization:** Let $\mathbf{A} \in \mathbb{C}^{m\times n} with\, m \geq n$. There exist matrices $\mathbf{Q} \in \mathbb{C}^{m\times m}$ unitary and $\mathbf{R} \in \mathbb{C}^{m\times n}$ upper-triangular such that $\mathbf{A} = \mathbf{Q}\mathbf{R}$.
  If $\mathbf{A}$ is nonsingular, then the diagonal entries of $\mathbf{R}$ can be chosen positive and the resulting QR factorization is unique. $\mathbf{R}$ is invertible in this case.

- **QR factorisation:** If $m > n$, then we can form the reduced QR factorization $\mathbf{A} = \hat{\mathbf{Q}}\hat{\mathbf{R}}$,where $\hat{\mathbf{Q}} \in \mathbb{C}^{m\times n}$ has orthonormal columns and $\hat{\mathbf{R}} \in \mathbb{C}^{n\times n}$ is upper-triangular. Lastly, if $\mathbf{A}$ is nonsingular, then the columns of $\mathbf{Q}$ span the same space as the columns of $\mathbf{A}$.
- **Cholesky factorisation:** Suppose $\mathbf{B} = \mathbf{A}^*\mathbf{A}$ for some matrix $\mathbf{A}$. Then $\mathbf{B}$ has a Cholesky factorization $\mathbf{B} = \mathbf{L}\mathbf{L}^*$, where $\mathbf{L}$ is lower-triangular.
  Proof: Since $\mathbf{A}$ has a full $\mathbf{QR}$ factorization,
  $\mathbf{B} = \mathbf{A}^*\mathbf{A} = \mathbf{R}^*\mathbf{Q}^*\mathbf{Q}\mathbf{R} = \mathbf{R}^*\mathbf{R} = \mathbf{L}\mathbf{L}^*$, where $\mathbf{L} = \mathbf{R}^*$.

Chapter 3: Variational Characteristics of Hermitian Matrices

In this section, all matrices are $(n \times n)$ Hermitian matrices unless otherwise noted. Since the eigenvalues of Hermitian matrices are real-valued, we can order the eigenvalues, $\lambda_{min} = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n = \lambda_{max}$.

- For $\mathbf{x} \neq \mathbf{0}$, the value $\frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \frac{\langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}$ is called a **Rayleigh quotient**.
- **Rayleigh-Ritz Theorem:** we have the following relations:
  $\lambda_{max} = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \max_{\|\mathbf{x}\|_2 = 1} \mathbf{x}^*\mathbf{A}\mathbf{x}$.
  $\lambda_{min} = \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \min_{\|\mathbf{x}\|_2 = 1} \mathbf{x}^*\mathbf{A}\mathbf{x}$.
- **Courant-Fisher Theorem:** Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Let $\{\mathbf{w}_i\}$ be arbitrary sets of linearly independent vectors in $\mathbb{C}^n$. Then the following characterizations of λ_k hold:
  $\lambda_k(\mathbf{A}) = \min_{\{\mathbf{w}_1, \ldots, \mathbf{w}_{n-k}\}} \max_{\mathbf{x} \neq 0; \mathbf{x} \perp \{\mathbf{w}_1, \ldots, \mathbf{w}_{n-k}\}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$.
  $\lambda_k(\mathbf{A}) = \max_{\{\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}\}} \min_{\mathbf{x} \neq 0; \mathbf{x} \perp \{\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}\}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$

Chapter 3: Variational Characteristics of Hermitian Matrices

- **Courant-Fisher Theorem:** To simplify previous notations, we can express them in terms of an arbitary subspace S:

  $$\lambda_k(\mathbf{A}) = \min_{dim(S)=n-k} \max_{\mathbf{x}\neq 0;\mathbf{x}\in S^{\perp}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$$

  $$\lambda_k(\mathbf{A}) = \max_{dim(S)=k-1} \min_{\mathbf{x}\neq 0;\mathbf{x}\in S^{\perp}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$$

  One final equivalent version of the theorem (Horn and Johnson 2e) is given by:

  $$\lambda_k(\mathbf{A}) = \min_{dim(S)=k} \max_{\mathbf{x}\neq 0;\mathbf{x}\in S} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$$

  $$\lambda_k(\mathbf{A}) = \max_{dim(S)=n-k+1} \min_{\mathbf{x}\neq 0;\mathbf{x}\in S} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$$

Chapter 3: Variational Characteristics of Hermitian Matrices

- **Weyl's Inequality:** Let $\mathbf{A}$, $\mathbf{B}$ be Hermitian matrices. Then $\lambda_k(\mathbf{A}) + \lambda_{min}(\mathbf{B}) \le \lambda_k(\mathbf{A}+\mathbf{B}) \le \lambda_k(\mathbf{A}) + \lambda_{max}(\mathbf{B})$. Using the fact that for a Hermitian matrix, $\|\mathbf{B}\|_2 = \max(|\lambda_{min}(\mathbf{B})|, |\lambda_{max}(\mathbf{B})|)$, we have that $-\|\mathbf{B}\|_2 \le \lambda_{min}(\mathbf{B}) \le \lambda_{max}(\mathbf{B}) \le \|\mathbf{B}\|_2$. Using this, Weyl implies that $\lambda_k(\mathbf{A}) - \|\mathbf{B}\|_2 \le \lambda_k(\mathbf{A}+\mathbf{B}) \le \lambda_k(\mathbf{A}) + \|\mathbf{B}\|_2$. In general, we have
  $\lambda_{j+k-n}(\mathbf{A}+\mathbf{B}) \le \lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B})$
  $\lambda_{j+k-1}(\mathbf{A}+\mathbf{B}) \ge \lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B})$.
- **Ostrowski's Theorem:** Let $\mathbf{A}$ be Hermitian and $\mathbf{S}$ be nonsingular. Then there exists $\theta_k \in [\lambda_{min}(\mathbf{S}\mathbf{S}^*), \lambda_{max}(\mathbf{S}\mathbf{S}^*)]$ such that $\lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) = \theta_k \lambda_k(\mathbf{A})$.
  Corollary:
  $\lambda_{min}(\mathbf{S}\mathbf{S}^*)\lambda_k(\mathbf{A}) \le \lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) \le \lambda_{max}(\mathbf{S}\mathbf{S}^*)\lambda_k(\mathbf{A})$.

Chapter 3: Variational Characteristics of Hermitian Matrices

- **Interlacing of eigenvalues:** Let $\mathbf{A}$ be Hermitian and $\mathbf{z}$ be a vector. Then $\lambda_k(\mathbf{A}+\mathbf{z}\mathbf{z}^*) \leq \lambda_{k+1}(\mathbf{A}) \leq \lambda_{k+2}(\mathbf{A}+\mathbf{z}\mathbf{z}^*)$.
- **Bordered matrices:** Let $\mathbf{B} \in \mathbb{C}^{n\times n}, \mathbf{a} \in \mathbb{R}, \mathbf{y} \in \mathbb{C}^n$ and define

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{y} \\ \mathbf{y}^* & \mathbf{a} \end{pmatrix}$$

Then $\lambda_1(\mathbf{A}) \leq \lambda_1(\mathbf{B}) \leq \lambda_2(\mathbf{A}) \leq \lambda_2(\mathbf{B}) \leq \ldots \leq \lambda_n(\mathbf{B}) \leq \lambda_{n+1}(\mathbf{A})$. If no eigenvector of $\mathbf{B}$ is orthogonal to $\mathbf{y}$, then every inequality is a strict inequality.

Theorem: If $\mu_1 \leq \lambda_1 \leq \mu_2 \leq \ldots \leq \mu_n \leq \lambda_n \leq \mu_{n+1}$ then there exist $\mathbf{y} \in \mathbb{R}^n \, and \, \mathbf{a} \in \mathbb{R}$ such that

$$\mathbf{M} = \begin{pmatrix} \Lambda & \mathbf{y} \\ \mathbf{y}^* & \mathbf{a} \end{pmatrix}$$

has the eigenvalues $\{\mu_i\}_1^{n+1}$, where $\lambda = \text{diag}\,(\{\lambda_i\}_1^n)$.

Chapter 4: Norms and Inner Products

A function $||.|| : \mathbb{C}^n \to \mathbb{R}$ is a **vector norm** if it satisfies:

1. $||\mathbf{x}|| \geq 0$ for all $\mathbf{x} \in \mathbb{C}^n$ and $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} \equiv \mathbf{0}$;
2. $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$ for all $\alpha \in \mathbb{C}$, $\mathbf{x} \in \mathbb{C}^n$;
3. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ (Triangle Inequality)

Chapter 4: Norms and Inner Products

Another useful form of the Triangle Inequality is

$$||\mathbf{x} - \mathbf{y}|| \geq \big|\, ||\mathbf{x}|| - ||\mathbf{y}|| \,\big|$$

Proof:
Let $\mathbf{z} = \mathbf{x} - \mathbf{y}$. Then

$$||\mathbf{x}|| = ||\mathbf{z} + \mathbf{y}|| \leq ||\mathbf{z}|| + ||\mathbf{y}|| = ||\mathbf{x} - \mathbf{y}|| + ||\mathbf{y}||,$$

so that $||\mathbf{x}|| - ||\mathbf{y}|| \leq ||\mathbf{x} - \mathbf{y}||$.
Swapping the roles of $\mathbf{x}$ and $\mathbf{y}$, we see

$$||\mathbf{y}|| - ||\mathbf{x}|| \leq ||\mathbf{y} - \mathbf{x}|| = ||\mathbf{x} - \mathbf{y}||;$$

Thus,

$$-||\mathbf{x} - \mathbf{y}|| \leq ||\mathbf{x}|| - ||\mathbf{y}|| \leq ||\mathbf{x} - \mathbf{y}||$$

Chapter 4: Norms and Inner Products

Common vector norms:

- $||\mathbf{x}||_1 = \sum_i |\mathbf{x}_i|$
- $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^*\mathbf{x}} = \sqrt{\sum_i |\mathbf{x}_i|^2}$
- $||\mathbf{x}||_\infty = \max_i\{|\mathbf{x}_i|\}$
- $||\mathbf{x}||_p = (\sum |\mathbf{x}_i|^p)^{\frac{1}{p}}$ (the l_p-norm,$p \in \mathbb{N}$; these norms are convex)
- ◇ The three norms above are the l_1,l_2 and l_∞ norms.
- $\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ for any inner product $\langle ., . \rangle$.
- $\|\mathbf{x}\|_\mathbf{A} = \|\mathbf{A}\mathbf{x}\|$ for $\mathbf{A}$ nonsingular, $\|.\|$ any vector norm.
- The **support** of a vector, $\text{supp}(\mathbf{x})$, is the set of indices i such that $\mathbf{x}_i \neq 0$. The size of the support of $\mathbf{x}$ (that is, the number of nonzero entries of $\mathbf{x}$), $|\text{supp}(\mathbf{x})|$, is often denoted $\|\mathbf{x}\|_0$, although $\|.\|_0$ is not a vector norm.

Equivalence of norms:
Let $||.||_\alpha$ and $||.||_\beta$ be any two norms on $\mathbb{C}^n$.
There exist constants m and M such that

$$m||\mathbf{x}||_\alpha \leq ||\mathbf{x}||_\beta \leq M||\mathbf{x}||_\alpha \text{ for all } \mathbf{x} \in \mathbb{C}^n$$

The best attainable bounds for $||\mathbf{x}||_\alpha \leq C_{\alpha,\beta}||\mathbf{x}||_\beta$ are given for $\alpha, \beta \in \{1, 2, \infty\}$;

		β		
	$C_{\alpha,\beta}$	1	2	∞
	1	1	$\sqrt{n}$	n
α	2	1	1	$\sqrt{n}$
	∞	1	1	1

Chapter 4: Norms and Inner Products

- **Pseudonorm:** A function $f(.)$ on $\mathbb{C}^n$ that satisfies all the norm condition except that $f(\mathbf{x})$ may equal 0 for a nonzero $\mathbf{x}$ (i.e. (1) is not totally satisfied).
- **Pre-norm:** A continuous function $f(.)$ which satisfies $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, $f(\mathbf{x}) = 0$ if and only if $\mathbf{x} \equiv \mathbf{0}$ and $f(\alpha\mathbf{x}) = |\alpha| f(\mathbf{x})$(i.e. f satisfies (1) and (2) but not necessarily (3)).
  Note: All norms are also pre-norms but pre-norms are not norms.
- **Dual norm:** $f^D(\mathbf{y}) = \max_{f(\mathbf{x})=1} |\mathbf{y}^*\mathbf{x}|$, given any pre-norm $f(\mathbf{x})$
  Note: f could be a vector norm,as all norms are pre-norms. The dual norm of a pre-norm is always a norm,regardless of whether $f(\mathbf{x})$ is a (full) norm or not. If $f(\mathbf{x})$ is a norm, then $(f^D)^D = f$.

Chapter 4: Norms and Inner Products

An **Inner product** is a function

$$\langle ., . \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$$

such that:

1. $\langle \mathbf{x}, \mathbf{x} \rangle \in \mathbb{R}$ with $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ for all $\mathbf{x} \in \mathbb{C}^n$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} \equiv 0$
2. $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^n$
3. $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$ for all $\alpha \in \mathbb{C}, \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$
4. $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$

Note: Condition (4) and (3) together imply

$$\langle \mathbf{x}, \alpha \mathbf{y} \rangle = \overline{\alpha} \langle \mathbf{x}, \mathbf{y} \rangle$$

It should be noted that the engineering convention of writing $\mathbf{x}^* \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle$ (as opposed to the mathematically accepted notation $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$) results in property (3) being re-defined as $\langle \mathbf{x}, \alpha \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$.

Cauchy-Schwartz Inequality:

$$|\langle \mathbf{x}, \mathbf{y} \rangle|^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle$$

Proof:
Let $\mathbf{v} = a\mathbf{x} - b\mathbf{y}$, where $a = \langle \mathbf{y}, \mathbf{y} \rangle$ and $b = \langle \mathbf{x}, \mathbf{y} \rangle$. WLOG assume $\mathbf{y} \neq \mathbf{0}$. Consider
$0 \leq \langle \mathbf{v}, \mathbf{v} \rangle$
$= \langle a\mathbf{x}, a\mathbf{x} \rangle + \langle a\mathbf{x}, -b\mathbf{y} \rangle + \langle -b\mathbf{y}, a\mathbf{x} \rangle + \langle -b\mathbf{y}, -b\mathbf{y} \rangle$
$= |a|^2 \langle \mathbf{x}, \mathbf{x} \rangle - a\overline{b} \langle \mathbf{x}, \mathbf{y} \rangle - \overline{a} b \overline{\langle \mathbf{x}, \mathbf{y} \rangle} + |b|^2 \langle \mathbf{y}, \mathbf{y} \rangle$
$=$
$\langle \mathbf{y}, \mathbf{y} \rangle^2 \langle \mathbf{x}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle \overline{\langle \mathbf{x}, \mathbf{y} \rangle} \langle \mathbf{x}, \mathbf{y} \rangle - \overline{\langle \mathbf{y}, \mathbf{y} \rangle} \langle \mathbf{x}, \mathbf{y} \rangle \overline{\langle \mathbf{x}, \mathbf{y} \rangle} + |\langle \mathbf{x}, \mathbf{y} \rangle|^{\mathbf{2} \langle \mathbf{y}, \mathbf{y} \rangle}$
$= \langle \mathbf{y}, \mathbf{y} \rangle^2 \langle \mathbf{x}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle |\langle \mathbf{x}, \mathbf{y} \rangle|^2$
Add $\langle \mathbf{y}, \mathbf{y} \rangle |\langle \mathbf{x}, \mathbf{y} \rangle|^2$ to both sides and divide by $\langle \mathbf{y}, \mathbf{y} \rangle$.

Chapter 4: Norms and Inner Products

- The l_2 inner-product is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \mathbf{y} = \sum \overline{\mathbf{x}_i} \mathbf{y}_i$$

  which induces the l_2-norm: $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^* \mathbf{x}} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$
- If $\mathbf{A}$ is a Hermitian positive definite matrix, then $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle = \mathbf{x}^* \mathbf{A} \mathbf{y}$ is also an inner product.
- Any vector norm induced by an inner product must satisfy the **Parallelogram Law:**

$$||\mathbf{u} + \mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2 = 2(||\mathbf{u}||^2 + ||\mathbf{v}||^2)$$

Chapter 4: Norms and Inner Products

A **matrix norm** is a function $||.|| : \mathbb{C}^{n\times n} \to \mathbb{R}$ which satisfies:

1. $||\mathbf{A}|| \geq 0$ for all $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $||\mathbf{A}|| = 0$ if and only if $\mathbf{A} \equiv 0$
2. $||\alpha\mathbf{A}|| = |\alpha|\,||\mathbf{A}||$ for all $\alpha \in \mathbb{C}, \mathbf{A} \in \mathbb{C}^{n\times n}$
3. $||\mathbf{A}+\mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}||$ for all $\mathbf{A}, B \in \mathbb{C}^{n\times n}$(Triangle Inequality)
4. $||\mathbf{AB}|| \leq ||\mathbf{A}||\,||\mathbf{B}||$ for all $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$

Chapter 4: Norms and Inner Products

Common matrix norms:

- $||\mathbf{A}||_1 = \max_j \sum_i |a_{ij}|$ = maximum absolute column sum
- $||\mathbf{A}||_2 = \sigma_1(\mathbf{A}) = \sqrt{\lambda_{max}(\mathbf{A}^*\mathbf{A})}$
- $||\mathbf{A}||_\infty = \max_i \sum_j |a_{ij}|$ = maximum absolute row sum
- Matrix norms induced by vector norms:
  $||\mathbf{A}||_\beta = \max_{||\mathbf{x}||_\beta=1} ||\mathbf{Ax}||_\beta = \max_{\mathbf{x}\neq 0} \frac{||\mathbf{Ax}||_\beta}{||\mathbf{x}||_\beta}$
- ◇ The three norms above are alternate characterizations of the matrix norms induced by the vector norms $||.||_1$, $||.||_2$, and $||.||_\infty$, respectively.
- $||\mathbf{A}||_* = \sum_i \sigma_i(\mathbf{A}) = \sum_i \sqrt{\lambda_i(\mathbf{A}^*\mathbf{A}}$
- $||\mathbf{A}||_F = \sqrt{\sum |a_{ij}|^2} = \sqrt{\text{trace}(\mathbf{A}^*\mathbf{A})} = \sqrt{\sum_i \lambda_i(\mathbf{A}^*\mathbf{A})}$, sometimes denoted by $||\mathbf{A}||_{2,vec}$

Chapter 4: Norms and Inner Products

Statement: For any invertible matrix $\mathbf{A}$, we have

$$||\mathbf{A}^{-1}|| \geq ||\mathbf{A}||^{-1}$$

Proof:
$\mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$
$\Rightarrow 1 \leq ||\mathbf{I}|| = ||\mathbf{A}\mathbf{A}^{-1}|| \leq ||\mathbf{A}||\,||\mathbf{A}^{-1}||$
$\Rightarrow ||\mathbf{A}||^{-1} = \frac{1}{||\mathbf{A}||} \leq ||\mathbf{A}^{-1}||$
Statement: For any matrix $\mathbf{A}$, we have

$$||\mathbf{A}||_2 \leq ||\mathbf{A}||_F$$

Proof:
$||\mathbf{A}||_2^2 = \max_i \sigma_i^2(\mathbf{A}) \leq \sum_i \sigma_i^2(\mathbf{A}) = \sum_i \lambda_i(\mathbf{A}^*\mathbf{A}) =$
$\text{trace}(\mathbf{A}^*\mathbf{A}) = ||\mathbf{A}||_F^2.$ $\square$

- A matrix $\mathbf{B}$ is an **isometry** for the norm $||.||$ if $||\mathbf{Bx}|| = ||\mathbf{x}||$ for all $\mathbf{x}$
- If $\mathbf{U}$ is a unitary matrix, then $||\mathbf{Ux}||_2 = ||\mathbf{x}||_2$ for all vectors $\mathbf{x}$
  Proof: $||\mathbf{Ux}||_2^2 = \langle \mathbf{Ux}, \mathbf{Ux} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle = ||\mathbf{x}||_2^2$. $\square$
- If $\mathbf{A} = \mathbf{U}^*\mathbf{BU}$, then $||\mathbf{A}||_F = ||\mathbf{B}||_F$
  Proof: $||\mathbf{A}||_F^2 = trace(\mathbf{A}^*\mathbf{A}) = trace(\mathbf{U}^*\mathbf{B}^*\mathbf{BU}) = trace(\mathbf{B}^*\mathbf{BUU}^*) = trace(\mathbf{B}^*\mathbf{B}) = ||\mathbf{B}||_F^2$. $\square$

- $\rho(\mathbf{A}) = \max_i |\lambda_i(\mathbf{A})|$
- For any matrix $\mathbf{A}$, matrix norm $||.||$, and eigenvalue $\lambda = \lambda(\mathbf{A})$, we have

$$\lambda \leq \rho(\mathbf{A}) \leq ||\mathbf{A}||$$

Proof: Let $\lambda \in \sigma(\mathbf{A})$ with corresponding eigenvector $\mathbf{x}$ and let $\mathbf{X} = [\mathbf{x}, \mathbf{x}, \cdots, \mathbf{x}]$ (n copies of $\mathbf{x}$).
Consider
$\mathbf{AX} = [\mathbf{Ax}, \mathbf{Ax}, \cdots, \mathbf{Ax}] = [\lambda\mathbf{x}, \lambda\mathbf{x}, \cdots, \lambda\mathbf{x}] = \lambda X$.
$\Rightarrow |\lambda| \, ||\mathbf{X}|| = ||\lambda\mathbf{X}|| = ||\mathbf{AX}|| \leq ||\mathbf{A}|| \, ||\mathbf{X}||$.
Since, $\mathbf{x}$ is an eigenvector, $\mathbf{x} \neq 0$, so $||\mathbf{X}|| \neq 0$.
On dividing by $||\mathbf{X}||$, we obtain $|\lambda| \leq \|\mathbf{A}\|$. Since λ is arbitrary, we conclude that $|\lambda| \leq \rho(\mathbf{A}) \leq \|\mathbf{A}\|$, as desired. $\square$

Chapter 4: Norms and Inner Products

- Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ and let $\varepsilon > 0$ be given. Then there exists a matrix norm $||.||$ such that

  $$\rho(\mathbf{A}) \leq ||\mathbf{A}|| \leq \rho(\mathbf{A}) + \varepsilon$$

  As a consequence, if $\rho(\mathbf{A}) < 1$, then $\exists$ some matrix norm such that $||\mathbf{A}|| < 1$
- Let $\mathbf{A} \in \mathbb{C}^{n \times n}$. If $||\mathbf{A}|| < 1$ for some matrix norm, then $\lim_{k \to \infty} \mathbf{A}^k = 0$.
- $\lim_{k \to \infty} \mathbf{A}^k = 0$ if and only if $\rho(\mathbf{A}) < 1$.

Chapter 5: SVD and Pseudoinverse

Let $\mathbf{A} \in \mathbb{C}^{m \times n}$. Then, a **singular value decomposition (SVD)** of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$, where

- $\mathbf{U} \in \mathbb{C}^{m \times n}$ is unitary
- $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_p) \in \mathbb{R}^{m \times n}$ is diagonal with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0 \, [p = \min(m, n)]$ and
- $\mathbf{V} \in \mathbb{C}^{n \times n}$ is unitary

The values $\sigma_i = \sigma_i(\mathbf{A})$ are called the **singular values** of $\mathbf{A}$ and are uniquely determined as the positive square roots of the eigenvalues of $\mathbf{A}^*\mathbf{A}$.

- If $rank(\mathbf{A}) = r$, then $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_p = 0$
- A **reduced SVD** of $\mathbf{A}$ is given by

$$\mathbf{A} = \hat{\mathbf{U}}\hat{\Sigma}\hat{\mathbf{V}}^*, where$$

  - $\hat{\mathbf{U}} \in \mathbb{C}^{m \times r}$ has orthonormal columns
  - $\hat{\Sigma} = diag(\sigma_1, \cdots, \sigma_r) \in \mathbb{R}^{r \times r}$ is diagonal
  - $\hat{\mathbf{V}} \in \mathbb{C}^{n \times r}$ has orthonormal columns
- In particular, given an SVD $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$, $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ in the reduced SVD are the first r columns of $\mathbf{U}$ and $\mathbf{V}$.

Chapter 5: SVD and Pseudoinverse

- One useful identity is that if $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ with $rank(\mathbf{A}) = r$, then $A = \sum_1^r \sigma_i \mathbf{u}_i \mathbf{v}_i^*$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of $\mathbf{U}$ and $\mathbf{V}$ (or , $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$), respectively

Statement: The first r columns of $\mathbf{U}$ in the SVD span the same space as the columns of $\mathbf{A}$.
Proof:
Let $\mathbf{x} \in \mathbb{C}^n$. Then $\mathbf{Ax} = \sum_1^n a_i x_i$ is in the columns space of $\mathbf{A}$.
However,

$$\mathbf{Ax} = (\sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*)x = \sum_{i=1}^{r} \beta_i \mathbf{u}_i, where\, \beta_i = \sigma_i \mathbf{v}_i^* \mathbf{x}$$

. $\Rightarrow \mathbf{Ax} = \sum_1^r \beta_i \mathbf{u}_i$ lies in the span of the first r columns of $\mathbf{U}$.

Statement: A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ has 0 as an eigenvalue if and only if 0 is also a singular value of $\mathbf{A}$.
Proof: ($\Rightarrow$) Suppose that 0 is an eigenvalue of $\mathbf{A}$, that is, $\mathbf{Ax} = \mathbf{0}$ for some nonzero $\mathbf{x}$.
Then $\mathbf{A}^*\mathbf{Ax} = 0$, so 0 is also an eigenvalue of $\mathbf{A}^*\mathbf{A}$.
The $0 = \sqrt{0}$ is a singular value of $\mathbf{A}$.($\Leftarrow$) Suppose 0 is a singular value of $\mathbf{A}$. Then there exists some nonzero $\mathbf{x}$ such that $\mathbf{A}^*\mathbf{Ax} = \mathbf{0}$. This implies that
$\mathbf{x}^*\mathbf{A}^*\mathbf{Ax} = 0 = (\mathbf{Ax})^*(\mathbf{Ax}) = ||\mathbf{Ax}||_2^2$.
By the properties of norms, we must have $\mathbf{Ax} = \mathbf{0}$

Chapter 5: SVD and Pseudoinverse

- The **Moore-Penrose pseudoinverse** of $\mathbf{A}$ is the matrix

$$\mathbf{A}^\dagger = \mathbf{V}\Sigma^\dagger\mathbf{U}^* (\text{"}A\,dagger\text{"})$$

  where $\Sigma^\dagger$ is obtained by replacing the nonzero singular values of $\mathbf{A}$ (in Σ) with their inverses and then transposing the resulting matrix.
- In terms of a reduced SVD, $\mathbf{A}^\dagger = \hat{\mathbf{V}}\hat{\Sigma}^{-1}\hat{\mathbf{U}}^*$.
- The pseudoinverse is uniquely determined by the following three properties:
  1. $\mathbf{A}\mathbf{A}^\dagger$ and $\mathbf{A}^\dagger\mathbf{A}$ are Hermitian;
  2. $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$;
  3. $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$.

- Additionally,
  - $(\mathbf{A}^{\dagger})^{\dagger} = \mathbf{A}$
  - $\mathbf{A}^{\dagger} = \mathbf{A}^{-1}$,if $\mathbf{A}$ is square and nonsingular
  - $\mathbf{A}^{\dagger} = (\mathbf{A}^{*}\mathbf{A})^{-1}\mathbf{A}^{*}$, if $\mathbf{A}$ has full column rank
- One use of the pseudoinverse is to compute least-squares solutions of $\mathbf{A}\mathbf{x} = \mathbf{b}$
- A least-squares solution $\mathbf{x}$ satisfies,
  $||\mathbf{x}||_2$ is minimal among all $\mathbf{z}$, such that $||\mathbf{A}\mathbf{z} - \mathbf{b}||_2$ is also minimal.
- In this setup, the unique minimizer is computed as
  $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$.

For Further Reading I

Horn and Johnson.
Matrix Analysis.
2nd Edition, Cambridge University Press, 2012.

F.R. Gantmacher.
The Theory of Matrices.
Chelsea Pub. Co., 1960.

K.Hoffman and R.Kunze.
Linear Algebra.
Pearson Education, 1971.