SI231 Matrix Analysis and Computations
Basic Concepts

Ziping Zhao

Spring Term 2022–2023

School of Information Science and Technology
ShanghaiTech University, Shanghai, China

http://si231.sist.shanghaitech.edu.cn

Basic Concepts

- notations and conventions
- subspace, linear independence, basis, dimension
- rank, invertible matrices, determinant
- vector norms, inner product, orthogonality
- projections onto subspaces, orthogonal complements, four fundamental subspaces
- orthogonal and orthonormal bases, orthogonal matrix, Gram-Schmidt
- matrix multiplications and representations, block matrix manipulations
- complexity, floating point operations (flops)

Notations and Conventions

Notations and Conventions

$\mathbb{R}$	the set of real numbers, or real space
$\mathbb{C}$	the set of complex numbers, or complex space
$\mathbb{R}^n$	n-dimensional real space
$\mathbb{C}^n$	n-dimensional complex space
$\mathbb{R}^{m\times n}$	set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m\times n}$	set of all $m \times n$ complex-valued matrices
$\mathbf{x}$	column vector
$x_i, [\mathbf{x}]_i$	ith entry of $\mathbf{x}$
$\mathbf{A}$	matrix
$a_{ij}, [\mathbf{A}]_{ij}$	(i,j)th entry of $\mathbf{A}$
$\mathbb{S}^n$	set of all $n \times n$ real symmetric matrices; i.e, $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $a_{ij} = a_{ji}$ for all i, j
$\mathbb{H}^n$	set of all $n \times n$ complex Hermitian matrices; i.e, $\mathbf{A} \in \mathbb{H}^{n\times n}$ and $a_{ij} = a_{ji}^*$ for all i, j

Notations and Conventions

- vector: $\mathbf{x} \in \mathbb{R}^n$ means that $\mathbf{x}$ is a real-valued n-dimensional column vector; i.e.,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad x_i \in \mathbb{R} \text{ for all } i.$$

  Similarly, $\mathbf{x} \in \mathbb{C}^n$ means that $\mathbf{x}$ is a complex-valued n-dimensional column vector.

- transpose: let $\mathbf{x} \in \mathbb{R}^n$. The notation $\mathbf{x}^T$ ($\mathbf{x}^\top$, $\mathbf{x}^{\mathsf{T}}$) means that

$$\mathbf{x}^T = \begin{bmatrix} x_1, & x_2, & \ldots, & x_n \end{bmatrix}.$$

- conjugate/Hermitian transpose: let $\mathbf{x} \in \mathbb{C}^n$. The notation $\mathbf{x}^H$ ($\mathbf{x}^{\mathsf{H}}$) means that

$$\mathbf{x}^H = \begin{bmatrix} x_1^*, & x_2^*, & \ldots, & x_n^* \end{bmatrix},$$

  where the superscript $*$ denotes the complex conjugate.

Notations and Conventions

- matrix: $\mathbf{A} \in \mathbb{R}^{m \times n}$ means that $\mathbf{A}$ is real-valued $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \qquad a_{ij} \in \mathbb{R} \text{ for all } i, j.$$

  Similarly, $\mathbf{A} \in \mathbb{C}^{m \times n}$ means that $\mathbf{A}$ is a complex-valued $m \times n$ matrix.

- unless specified, we denote the ith column of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ as $\mathbf{a}_i \in \mathbb{R}^m$; i.e.,

$$\mathbf{A} = \left[\mathbf{a}_1, \quad \mathbf{a}_2, \quad \ldots, \quad \mathbf{a}_n\right].$$

  The same notation applies to $\mathbf{A} \in \mathbb{C}^{m \times n}$.

Notations and Conventions

- transpose: let $\mathbf{A} \in \mathbb{R}^{m \times n}$. The notation $\mathbf{A}^T$ means that

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & & & \vdots \\ a_{1n} & a_{m2} & \dots & a_{mn} \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

  - or, we have $\mathbf{B} = \mathbf{A}^T \Longleftrightarrow b_{ij} = a_{ji}$ for all i, j.
  - properties:
    * $(c\mathbf{A})^T = c\mathbf{A}^T$
    * $(\mathbf{A}^T)^T = \mathbf{A}$
    * $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
    * $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

- symmetric and skew-symmetric matrices: a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric if $\mathbf{A}^T = \mathbf{A}$ and skew-symmetric if $\mathbf{A}^T = -\mathbf{A}$.
  - for any matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, it can be decomposed as $\mathbf{A} = \mathbf{T} + \mathbf{S}$ where $\mathbf{T} = \frac{\mathbf{A}+\mathbf{A}^T}{2}$ is symmetric and $\mathbf{S} = \frac{\mathbf{A}-\mathbf{A}^T}{2}$ is skew-symmetric.

Notations and Conventions

- conjugate/Hermitian transpose: let $\mathbf{A} \in \mathbb{C}^{m \times n}$. The notation $\mathbf{A}^H$ means that

$$\mathbf{A}^H = \begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{m1}^* \\ a_{12}^* & a_{22}^* & \cdots & a_{m2}^* \\ \vdots & & & \vdots \\ a_{1n}^* & a_{m2}^* & \cdots & a_{mn}^* \end{bmatrix} \in \mathbb{C}^{n \times m}.$$

  - or, we have $\mathbf{B} = \mathbf{A}^H \Longleftrightarrow b_{ij} = a_{ji}^*$ for all i, j.
  - properties (same as transpose):
    * $(c\mathbf{A})^H = c^* \mathbf{A}^H$
    * $(\mathbf{A}^H)^H = \mathbf{A}$
    * $(\mathbf{A} + \mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H$
    * $(\mathbf{A}\mathbf{B})^H = \mathbf{B}^H \mathbf{A}^H$
- Hermitian and skew-Hermitian matrices: a matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is Hermitian if $\mathbf{A}^H = \mathbf{A}$ (or, equivalently, $\mathbf{A}^T = \mathbf{A}^*$) and skew-Hermitian if $\mathbf{A}^H = -\mathbf{A}$.
  - for any matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, it can be decomposed as $\mathbf{A} = \mathbf{T} + \mathbf{S}$ where $\mathbf{T} = \frac{\mathbf{A} + \mathbf{A}^H}{2}$ is Hermitian and $\mathbf{S} = \frac{\mathbf{A} - \mathbf{A}^H}{2}$ is skew-Hermitian.
  - the Hermitian matrix is also called the self-adjoint matrix

Notations and Conventions

- trace: let $\mathbf{A} \in \mathbb{R}^{n\times n}$. The trace of $\mathbf{A}$ is

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

  - properties:
    * $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$
    * $\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A})$
    * $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
    * $\mathrm{tr}(\mathbf{AB}) = \sum_{i,j} a_{ij}b_{ji} = \mathrm{tr}(\mathbf{BA})$ for $\mathbf{A}, \mathbf{B}$ of appropriate sizes
      · $\mathrm{tr}(\mathbf{b}\mathbf{a}^T) = \mathbf{a}^T\mathbf{b}$
      · $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{BCA}) = \mathrm{tr}(\mathbf{CAB})$ (cyclic property)
    * for three symmetric matrices, $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CBA}) = \mathrm{tr}(\mathbf{ACB})$
    * for symmetric $\mathbf{A}$ and skew-symmetric $\mathbf{B}$, $\mathrm{tr}(\mathbf{AB}) = 0$

- matrix power: let $\mathbf{A} \in \mathbb{R}^{n\times n}$. The notation $\mathbf{A}^2$ means $\mathbf{A}^2 = \mathbf{AA}$, and $\mathbf{A}^k$ means

$$\mathbf{A}^k = \underbrace{\mathbf{AA}\cdots\mathbf{A}}_{k\ \mathbf{A}\text{'s}}.$$

  - $\mathbf{A}^0 = \mathbf{I}$

Notations and Conventions

- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be
  - square if $m = n$;
  - rectangular (or non-square) if $m \neq n$;
    * tall, skinny if $m > n$;
    * short, fat if $m < n$.
- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be
  - upper (or right) triangular if $a_{ij} = 0$ for all $i > j$;
  - lower (or left) triangular if $a_{ij} = 0$ for all $i < j$.

  Examples:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 2 & 0 \\ \frac{1}{8} & 3 & 0 \end{bmatrix}.$$

  The definitions can be extended to non-square matrices, which are also often called the upper (lower) trapezoidal matrix.
- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be
  - a upper Hessenberg matrix if $a_{ij} = 0$ for all $i > j + 1$;
  - a lower Hessenberg matrix if $a_{ij} = 0$ for all $i < j + 1$.

Notations and Conventions

- all-one/unitary vectors: we use the notation

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

  to denote a vector of all 1's.

- zero/null vectors or matrices: we use the notation $\mathbf{0}$ to denote either a vector of all zeros, or a matrix of all zeros.

- unit/canonical vectors: unit vectors are vectors that have only one nonzero element and the nonzero element is 1. We use the notation

$$\mathbf{e}_i = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

  to denote a unit vector with the nonzero element at the ith entry.

Notations and Conventions

- identity matrix:
$$\mathbf{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$
where, as a convention, the empty entries are assumed to be zero. (In some literature, also referred to as the unitary matrix.)
- diagonal matrices: we use the notation
$$\mathrm{Diag}(a_1, \ldots, a_n) = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix}$$
to denote a diagonal matrix with diagonals $a_1, \ldots, a_n$. We also use the shorthand notation $\mathrm{Diag}(\mathbf{a}) = \mathrm{Diag}(a_1, \ldots, a_n)$ with $\mathbf{a} = [a_1, \ldots, a_n]$.
- the (main) diagonal or principal diagonal of a matrix $\mathbf{A}$ is the collection of entries a_{ij} with $i = j$; super-diagonal: a_{ij} with $i \leq j$; sub-diagonal: a_{ij} with $i \geq j$;
note: the notion of diagonal matrices can be extended to rectangular matrices in which case only the main diagonals are nonzero

Notations and Conventions

A vector or matrix is said to be sparse if it contains many zero elements (or, equivalently, few nonzero elements)

- Special data structures can be used to store such matrices.
- A simple strategy is to store every nonzero, a_{ij}, together with its row and column indices, i and j.

A shift matrix is a matrix with ones only on the super-diagonal or sub-diagonal, and zeroes elsewhere.

- A shift matrix with ones on the super-diagonal (sub-diagonal) is an upper (lower) shift matrix.

Notations and Conventions

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be a band/banded (diagonal) matrix (a special type of sparse matrix) if all matrix elements are zero outside a diagonally bordered band

$$a_{ij} = 0 \quad \text{if} \quad i > j + p \;\; \text{or} \;\; j > i + q$$

where $p \geq 0$ and $q \geq 0$ are called the lower bandwidth and upper bandwidth, respectively.

special cases:

- identity matrix, shift matrix
- $p = q = 0 \; (1, \; 2, \; \ldots)$, diagonal (tridiagonal, pentadiagonal, ...) matrix
- $p = 0, q = 1 \; (p = 1, q = 0)$, upper (lower) bidiagonal matrix
- $p = 0, q = n - 1 \; (p = n - 1, q = 0)$, upper (lower) triangular matrix
- $p = 1, q = n - 1 \; (p = n - 1, q = 1)$, upper (lower) Hessenberg matrix
- block diagonal matrices (see the definition for block matrices later)
- ...

Notations and Conventions

- Toeplitz matrices: matrices with constant diagonals (may not be square)

$$\mathbf{A} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

  a special case: circulant matrices (or the transpose of this form)

$$\mathbf{A} = \begin{bmatrix} a_0 & a_{n-1} & a_{n-2} & \cdots & \cdots & a_1 \\ a_1 & a_0 & a_{n-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{n-1} & a_{n-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{n-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

- reference: R. M. Gray, *Toeplitz and Circulant Matrices: A review*, 2006. Available online at `https://ee.stanford.edu/~gray/toeplitz.pdf`.

Notations and Conventions

- Hankel matrices: matrices with constant anti-diagonals (or skew-diagonals), i.e., upside down Toeplitz matrices

$$\mathbf{A} = \begin{bmatrix} a_0 & \cdots & \cdots & a_{n-3} & a_{n-2} & a_{n-1} \\ a_1 & & & a_{n-2} & a_{n-1} & a_n \\ \vdots & & & & a_n & a_{n+1} \\ a_{n-3} & a_{n-2} & & & & \vdots \\ a_{n-2} & a_{n-1} & a_n & & & a_{2n-3} \\ a_{n-1} & a_n & a_{n+1} & \cdots & \cdots & a_{2n-2} \end{bmatrix}$$

- Vandemonde matrices: matrices with the terms of a geometric progression in each row, i.e., $a_{ij} = \alpha_i^{j-1}$

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \cdots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \cdots & \alpha_m^{n-1} \end{bmatrix}$$

(sometimes Vandermonde matrix is referred to as the transpose of the above one)

Notations and Conventions

- involutory matrices: matrix $\mathbf{A}$ is involutory if and only if

$$\mathbf{A}^2 = \mathbf{I}$$

- idempotent matrices: matrix $\mathbf{A}$ is idempotent if and only if

$$\mathbf{A}^2 = \mathbf{A}$$

  hence, $\mathbf{A}^k = \mathbf{A}$ for $k \geq 1$

Notations and Conventions

- nilpotent matrices: matrix $\mathbf{A}$ is nilpotent if

$$\mathbf{A}^k = \mathbf{0}$$

  for some $k > 0$.

- unipotent matrices: matrix $\mathbf{A}$ is unipotent if $\mathbf{A} - \mathbf{I}$ is nilpotent

$$(\mathbf{A} - \mathbf{I})^k = \mathbf{0}$$

  for some $k > 0$.

Notations and Conventions

- A block/partitioned matrix is a matrix whose entries are themselves matrices. A $q \times r$ block matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1r} \\ \vdots & & \vdots \\ \mathbf{A}_{q1} & \cdots & \mathbf{A}_{qr} \end{bmatrix}$$

  where $\mathbf{A}_{ij} \in \mathbb{R}^{m_i \times n_j}$ with $\sum_{i=1}^{q} m_i = m$ and $\sum_{j=1}^{r} n_j = n$ designates the (i, j) submatrix and is related to $\mathbf{A}$ by $\mathbf{A}_{ij} = \mathbf{A}_{\tau+1:\tau+m_i, \mu+1:\mu+n_j}$ where $\tau = m_1 + \ldots + m_{i-1}$ and $\mu = n_1 + \ldots + n_{j-1}$
  - special cases: partitioning into column vectors or row vectors
  - generally, a submatrix can take any groups of columns (indexed by $\boldsymbol{\alpha}$) and any groups of rows (indexed by $\boldsymbol{\beta}$) from $\mathbf{A}$
    * principal submatrix (a square matrix): $\boldsymbol{\alpha} = \boldsymbol{\beta}$ (length of $\boldsymbol{\alpha}$ determines the order of the principal submatrix)
    * leading (resp. trailing) principal submatrix: $\boldsymbol{\alpha} = \boldsymbol{\beta} = [1, \ldots, k]$ (resp. $\boldsymbol{\alpha} = \boldsymbol{\beta} = [\ell, \ldots, \min\{m, n\}]$)
  - terms used to describe matrices with scalar entries have block analogs
    * block diagonal, block lower/upper triangular, block tridiagonal, block band...

Subspace, Linear Independence, Basis, Dimension

Subspace

A subset $\mathcal{S}$ of $\mathbb{R}^m$ is said to be a linear subspace or subspace if it is nonempty and

$$\begin{matrix} \mathbf{x}, \mathbf{y} \in \mathcal{S}, \\ \alpha, \beta \in \mathbb{R} \end{matrix} \quad \Longrightarrow \quad \alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{S}.$$

- if $\mathcal{S}$ is a subspace and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$, any linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$, i.e., $\sum_{i=1}^n \alpha_i \mathbf{a}_i$ for some $\boldsymbol{\alpha} \in \mathbb{R}^n$, lies in $\mathcal{S}$.
- trivial subspaces: $\{\mathbf{0}\}$ (zero/null subspace) and $\mathbb{R}^m$
- some quick facts: let $\mathcal{S}_1, \mathcal{S}_2$ be subspaces of $\mathbb{R}^m$.
  - the intersection $\mathcal{S}_1 \cap \mathcal{S}_2$ is a subspace
  - the union $\mathcal{S}_1 \cup \mathcal{S}_2$ is only a subspace if $\mathcal{S}_1 \subseteq \mathcal{S}_2$ or $\mathcal{S}_2 \subseteq \mathcal{S}_1$
  - the sum $\mathcal{S}_1 + \mathcal{S}_2$ is a subspace (fact: smallest subspace containing $\mathcal{S}_1 \cup \mathcal{S}_2$) [1]
- if $\mathcal{S}_1, \mathcal{S}_2$ are subspaces of $\mathbb{R}^m$ with $\mathcal{S}_1 \cap \mathcal{S}_2 = \{\mathbf{0}\}$ and $\mathcal{S}_1 + \mathcal{S}_2 = \mathcal{S}_3$, we define the direct sum $\mathcal{S}_3 = \mathcal{S}_1 \oplus \mathcal{S}_2$
  - subspace $\mathcal{S}_1$ ($\mathcal{S}_2$) is the complement (i.e., a complementary subspace) to subspace $\mathcal{S}_2$ ($\mathcal{S}_1$) of space $\mathcal{S}_3$; $\mathcal{S}_1$ and $\mathcal{S}_2$ are mutually complementary
  - complementary subspaces are not unique

[1] note the notation $\mathcal{X} + \mathcal{Y} = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}\}$.

Span

The span of a collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ is defined as

$$\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \left\{ \mathbf{y} \in \mathbb{R}^m \;\middle|\; \mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i,\ \boldsymbol{\alpha} \in \mathbb{R}^n \right\}.$$

- the set of all possible linear combinations of $\mathbf{a}_1, \ldots, \mathbf{a}_n$
- it is a subspace
- Question: any span is a subspace. But can any subspace be written as a span?

  Theorem 1. Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$. There exists a positive integer n and a collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$ such that $\mathcal{S} = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

  – Implication: we can always represent a subspace by a span

  For example, we can represent $\mathbb{R}^m$ by

$$\mathbb{R}^m = \operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}.$$

Range Space and Null Space

The range space (or column space, image space) of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{A}\mathbf{x},\ \mathbf{x} \in \mathbb{R}^n\}.$$

- essentially the same as span, i.e., $\mathcal{R}(\mathbf{A}) = \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$

The null space (nullspace) (or kernel space) of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

- a null space is a subspace (verify as a mini exercise)
- by Theorem 1, we can represent a null space by $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{B})$ for some $\mathbf{B} \in \mathbb{R}^{n\times r}$ with positive integer r.
- Define a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$, $L(\mathbf{x}) = \mathbf{A}\mathbf{x}$. $\mathcal{R}(\mathbf{A})$ is the image/range of L. $\mathcal{N}(\mathbf{A})$ is the kernal/null space of L.
- Also, the row space of $\mathbf{A}$ is $\mathcal{R}(\mathbf{A}^T)$ and the left null space of $\mathbf{A}$ is $\mathcal{N}(\mathbf{A}^T)$.

Linear Independence

A collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ is said to be linearly independent if

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i \neq \mathbf{0}, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{R}^n \text{ with } \boldsymbol{\alpha} \neq \mathbf{0};$$

and linearly dependent otherwise.

- an equivalent way of defining linear dependence: a vector set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly dependent if there exists $\boldsymbol{\alpha} \in \mathbb{R}^m$, $\boldsymbol{\alpha} \neq \mathbf{0}$, such that

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i = \mathbf{0}.$$

- physical meaning: a set of "non-redundant" vectors

Linear Independence

Some known facts (some easy to show, some not):

- if $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly independent, then any $\mathbf{a}_j$ *cannot* be a linear combination of the other $\mathbf{a}_i$'s; i.e., $\mathbf{a}_j \neq \sum_{i \neq j} \alpha_i \mathbf{a}_i$ for any α_i's.
- if $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly dependent, then *there exists* an $\mathbf{a}_j$ such that $\mathbf{a}_j$ is a linear combination of the other $\mathbf{a}_i$'s; i.e., $\mathbf{a}_j = \sum_{i \neq j} \alpha_i \mathbf{a}_i$ for some α_i's.
- $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly dependent if and only if one of them is zero or a linear combination of the others
- if $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly independent, then $n \leq m$ must hold. (an exercise)
- let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ be a linearly independent vector set. Suppose $\mathbf{y} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then the coefficient $\boldsymbol{\alpha}$ for the representation

$$\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$$

  is unique; i.e., there does *not* exist a $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$, such that $\mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i$.

Linear Independence

Let $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$, and denote $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ as an index subset with $k \leq n$ and $i_j \neq i_l$ for all $j \neq l$.

A vector subset $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is called a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ if

1. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is linearly independent;
2. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is not contained by any other linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$.

- physical meaning: find an irreducibly non-redundant set of vectors for representing the whole vector set $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$

Linear Independence

- example:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The linearly independent subets of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ are

$$\begin{gathered} \{\mathbf{a}_1\}, \{\mathbf{a}_2\}, \{\mathbf{a}_3\}, \{\mathbf{a}_4\}, \\ \{\mathbf{a}_1, \mathbf{a}_2\}, \{\mathbf{a}_1, \mathbf{a}_3\}, \{\mathbf{a}_1, \mathbf{a}_4\}, \{\mathbf{a}_2, \mathbf{a}_3\}, \{\mathbf{a}_2, \mathbf{a}_4\}, \{\mathbf{a}_3, \mathbf{a}_4\}, \\ \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}. \end{gathered}$$

But the maximal linearly independent subsets are

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}.$$

Linear Independence

Facts:

- $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ if and only if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}, \mathbf{a}_j\}$ is linearly dependent for any $j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$

- if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$, then

$$\mathrm{span}\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\} = \mathrm{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}.$$

Basis

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace with $\mathcal{S} \neq \{\mathbf{0}\}$.

A vector set $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\} \subset \mathbb{R}^m$ is called a basis for $\mathcal{S}$ if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ is linearly independent and

$$\mathcal{S} = \mathrm{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}.$$

- examples: let $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ be a maximal linearly independent vector subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then, $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a basis for $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

Some facts:

- we may have more than one basis for $\mathcal{S}$
- all bases for $\mathcal{S}$ have the same number of elements; i.e., if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_l\}$ are bases for $\mathcal{S}$, then $k = l$

Dimension

The dimension of a subspace $\mathcal{S}$, with $\mathcal{S} \neq \{\mathbf{0}\}$, is defined as the number of elements of a basis for $\mathcal{S}$.

- The dimension of $\{\mathbf{0}\}$ is defined as 0.
- $\dim \mathcal{S}$ will be used as the notation for denoting the dimension of $\mathcal{S}$
- physical meaning: effective degrees of freedom of the subspace
- examples:
  - $\dim \mathbb{R}^m = m$
  - if k is the number of maximal linearly independent vectors of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then $\dim \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = k$.

Dimension

Properties:

- let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. If $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $\dim \mathcal{S}_1 \leq \dim \mathcal{S}_2$.
- let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. If $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $\dim \mathcal{S}_1 = \dim \mathcal{S}_2$, then $\mathcal{S}_1 = \mathcal{S}_2$.
- let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace. Then $\dim \mathcal{S} = r \Longleftrightarrow \mathcal{S} = \mathbb{R}^r$.
- let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. We have $\dim(\mathcal{S}_1 + \mathcal{S}_2) \leq \dim \mathcal{S}_1 + \dim \mathcal{S}_2$.
  - as a more advanced result, we also have
$$\dim(\mathcal{S}_1 + \mathcal{S}_2) = \dim \mathcal{S}_1 + \dim \mathcal{S}_2 - \dim(\mathcal{S}_1 \cap \mathcal{S}_2).$$

  (proof as an excerise)

Rank, Invertible Matrices, Determinant

Rank

The rank of a set of vectors $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$, denoted by $\operatorname{rank}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$, is defined as the number of elements of a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$.

- if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$, then

$$\operatorname{rank}\{\mathbf{a}_1, \ldots \mathbf{a}_n\} = \operatorname{rank}\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\} = k.$$

- equal to the dimension of the subspace $\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$

$$\operatorname{rank}\{\mathbf{a}_1, \ldots \mathbf{a}_n\} = \dim \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

- if $\mathbf{a}_i = \mathbf{0}$ for all i, $\operatorname{rank}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ is defined as 0

Rank

The rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted by $\mathrm{rank}(\mathbf{A})$, is defined as the number of elements of a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

- or, $\mathrm{rank}(\mathbf{A})$ is the maximum number of linearly independent columns of $\mathbf{A}$
- $\dim \mathcal{R}(\mathbf{A}) = \mathrm{rank}(\mathbf{A})$ by definition, i.e., rank is the dimension of the image of $\mathbf{A}$
- the rank of $\mathbf{0}$ is defined as 0

Rank

Facts:

- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$, i.e., the rank of $\mathbf{A}$ is also the maximum number of linearly independent rows of $\mathbf{A}$ or the dimension of the column space of $\mathbf{A}$ is equal to the dimension of the row space of $\mathbf{A}$ (proof as an exercise)
- $\text{rank}(\mathbf{A}) \leq \min\{m, n\}$
- $\text{rank}(k\mathbf{A}) = \text{rank}(\mathbf{A})$ for $k \neq 0$
- $\text{rank}(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$
- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^T) = \text{rank}(\mathbf{A}^T\mathbf{A})$
- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$ (proof)
- $\text{rank}(\mathbf{A}\mathbf{B}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{A}\mathbf{B}) \leq \text{rank}(\mathbf{B})$
- $\text{rank}(\mathbf{A}\mathbf{B}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$. Also, the equality above holds if the columns of $\mathbf{A}$ are linearly independent *or* the rows of $\mathbf{B}$ are linearly independent. (proof as an exercise)
- $\text{rank}(\mathbf{A}\mathbf{B}) \geq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n$ for $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$ (Sylvester's rank inequality)
  - if $\mathbf{A}\mathbf{B} = \mathbf{0}$, $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \leq n$

Rank

- $\mathbf{A}$ is said to have
  - full column rank if the columns of $\mathbf{A}$ are linearly independent (more precisely, the collection of *all* columns of $\mathbf{A}$ is linearly independent)
    * $\mathbf{A} \in \mathbb{R}^{m \times n}$ being of full-column rank $\Longleftrightarrow m \geq n, \operatorname{rank}(\mathbf{A}) = n$
  - full row rank if the rows of $\mathbf{A}$ are linearly independent
    * $\mathbf{A} \in \mathbb{R}^{m \times n}$ being of full-row rank $\Longleftrightarrow m \leq n, \operatorname{rank}(\mathbf{A}) = m$
  - full rank if $\operatorname{rank}(\mathbf{A}) = \min\{m, n\}$; i.e., it has either full column rank or full row rank
  - rank deficient if $\operatorname{rank}(\mathbf{A}) < \min\{m, n\}$

- $\mathbf{A}$ is said to have low rank or reduced rank when its rank is significantly less than the maximum rank possible for the matrix

Invertible Matrices

A square matrix $\mathbf{A}$ is said to be invertible (a.k.a. nonsingular or nondegenerate) if the columns of $\mathbf{A}$ are linearly independent (i.e., $\mathbf{A}$ has full rank), and noninvertible (a.k.a. singular or degenerate) otherwise.

- alternatively, we say $\mathbf{A}$ is singular if $\mathbf{A}\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$.

The inverse of an invertible $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$, is a square matrix that satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Invertible Matrices

Facts (for a nonsingular $\mathbf{A}$):

- $\mathbf{A}^{-1}$ always exists and is unique (or there are no two inverses of $\mathbf{A}$)
- $\mathbf{A}^{-1}$ is nonsingular
- $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1}$ for $k > 0$
- $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ (can be proved by "if $\mathbf{AB} = \mathbf{I}$ for square $\mathbf{A}$ and $\mathbf{B}$, then $\mathbf{BA} = \mathbf{I}$")
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, where $\mathbf{A}, \mathbf{B}$ are (square and) nonsingular
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
  - as a shorthand notation, we sometimes denote $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T = \mathbf{A}^{-T}$
  - similar result holds for complex matrices, i.e., $(\mathbf{A}^H)^{-1} = (\mathbf{A}^{-1})^H = \mathbf{A}^{-H}$
  - and $(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^* = \mathbf{A}^{-*}$
- $(\mathbf{A}^k)^{-1} = (\mathbf{A}^{-1})^k = \mathbf{A}^{-k}$
- for nonsingular $\mathbf{P}$ and $\mathbf{Q}$, $\mathrm{rank}(\mathbf{PM}) = \mathrm{rank}(\mathbf{M}) = \mathrm{rank}(\mathbf{MQ}) = \mathrm{rank}(\mathbf{PMQ})$

Invertible Matrices

Matrix inversion lemma/Sherman-Morrison-Woodbury formula (Woodbury formula, Woodbury matrix identity): for nonsingular matrices $\mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{C} \in \mathbb{R}^{k\times k}$ and $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times k}$

$$\left(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V}^T\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^T\mathbf{A}^{-1}$$

- intuition: the inverse of a rank-k correction to $\mathbf{A}$ can be computed by doing a rank-k correction to the inverse of $\mathbf{A}$
- (Sherman-Morrison formula) when $k = 1$ and $\mathbf{C} = 1$ we have

$$\left(\mathbf{A} + \mathbf{u}\mathbf{v}^T\right)^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}$$

  - when $n = k = 1$ and $\mathbf{C} = 1$, we have

$$\frac{1}{a + uv} = \frac{1}{a} - \frac{uv}{a(a + vu)}$$

Determinant

Let $\mathbf{A} \in \mathbb{R}^{m \times m}$. The determinant of $\mathbf{A}$, denoted by $\det(\mathbf{A})$, is defined inductively.

- if $m = 1$, $\det(\mathbf{A}) = a_{11}$.
- if $m \geq 2$, we have the following:
  - let $\mathbf{A}_{ij} \in \mathbb{R}^{(m-1)\times(m-1)}$ be a submatrix of $\mathbf{A}$ obtained by deleting the ith row and jth column of $\mathbf{A}$.
  - let
$$c_{ij} = (-1)^{i+j} \det(\mathbf{A}_{ij}).$$
  - cofactor expansion:
$$\det(\mathbf{A}) = \sum_{j=1}^{m} a_{ij} c_{ij}, \quad \text{for any } i = 1, \ldots, m$$
$$\det(\mathbf{A}) = \sum_{i=1}^{m} a_{ij} c_{ij}, \quad \text{for any } j = 1, \ldots, m$$
  - remark: c_{ij}'s are called the cofactors, $\det(\mathbf{A}_{ij})$'s are called the minors

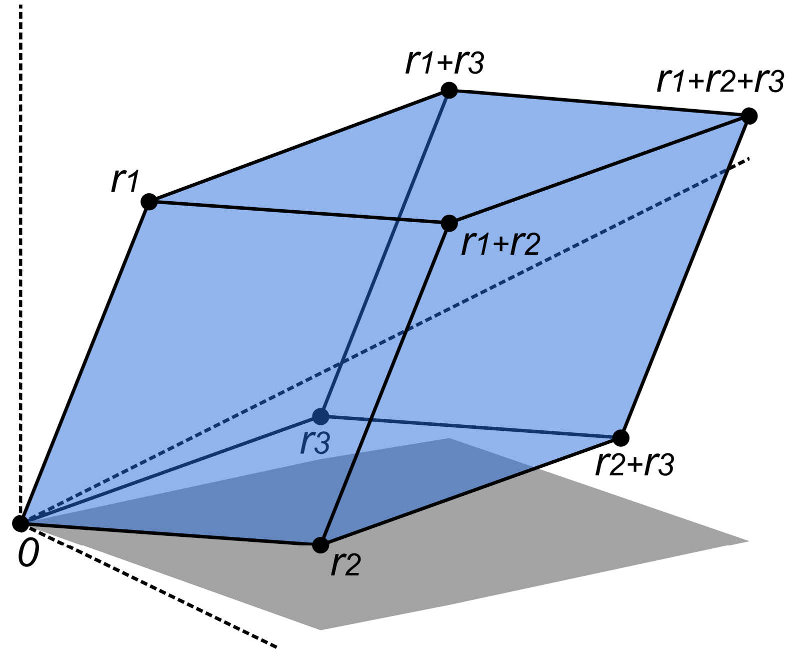
Determinant

- principal minor: the determinant of a principal submatrix
- leading principal minor: the determinant of a leading principal submatrix
- Given an $m \times n$ matrix and rank r, then there exists at least one non-zero $r \times r$ minor, while all larger minors are zero.
- leading principal submatrix can be used to determine the definiteness of a matrix (cf. PSD Topic)

Determinant

Some interpretations of determinant:

- a matrix $\mathbf{A}$ is nonsingular if and only if $\det(\mathbf{A}) \neq 0$
- (important) $\mathbf{A}\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$ if and only if $\det(\mathbf{A}) = 0$, i.e., $\mathbf{A}$ is singular
- $|\det(\mathbf{A})|$ is the volume of the parallelepiped $\mathcal{P} = \{\mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{a}_i \mid \alpha_i \in [0,1]\ \forall i\}$



Source: Wiki. r_1, r_2, r_3 are $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ on $\mathbb{R}^3$.

Determinant

Properties:

- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ for any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times m}$
- $\det(\mathbf{A}^T) = \det(\mathbf{A})$, but $\det(\mathbf{A}^H) = (\det(\mathbf{A}^T))^* = (\det(\mathbf{A}))^*$,
- $\det(\alpha\mathbf{A}) = \alpha^m \det(\mathbf{A})$ for any $\alpha \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{m\times m}$
- $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$ for any nonsingular $\mathbf{A}$
- $\det(\mathbf{B}^{-1}\mathbf{AB}) = \det(\mathbf{A})$ for any nonsingular $\mathbf{B}$
- $\mathbf{A}^{-1} = \dfrac{1}{\det(\mathbf{A})}\tilde{\mathbf{A}}$, where $\tilde{a}_{ij} = c_{ji}$ (the cofactor) for all i, j ($\mathbf{A}$ is nonsingular)
  - remark: $\tilde{\mathbf{A}}$ is called the adjoint of $\mathbf{A}$

Determinant

More properties:

- if $\mathbf{A} \in \mathbb{R}^{m \times m}$ is triangular, either upper or lower,

$$\det(\mathbf{A}) = \prod_{i=1}^{m} a_{ii}$$

  - proof: apply cofactor expansion inductively

- if $\mathbf{A} \in \mathbb{R}^{m \times m}$ takes a block (upper or lower) triangular form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \quad \text{or} \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

  where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square (and can be of different sizes), then

$$\det(\mathbf{A}) = \det(\mathbf{A}_{11})\det(\mathbf{A}_{22}).$$

Determinant

More properties:

- Matrix determinant lemma: Suppose $\mathbf{A}$ is invertible

$$\det(\mathbf{A} + \mathbf{u}\mathbf{v}^T) = \det(\mathbf{A})\,(1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}).$$

  more generally, we have

$$\det(\mathbf{A} + \mathbf{U}\mathbf{V}^T) = \det(\mathbf{A})\det(\mathbf{I} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}).$$

  and

$$\det(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V}^T) = \det(\mathbf{A})\det(\mathbf{C}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})\det(\mathbf{C}).$$

  – Special case: Weinstein-Aronszajn identity (Sylvester's determinant theorem) Given $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$

$$\det(\mathbf{I}_m + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_n + \mathbf{B}\mathbf{A})$$

Vector Norms, Inner Product, Orthogonality

Vector Norms

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a vector norm if

1. $f(\mathbf{x}) \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$ (positive definiteness)
2. $f(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$ (positive definiteness)
3. $f(\mathbf{x}+\mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (triangle inequality)
4. $f(\alpha\mathbf{x}) = |\alpha| f(\mathbf{x})$ for any $\alpha \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$ (homogeneity)

- used to measure the length of a vector
- we usually use the notation $\|\cdot\|$ to denote a norm
- also used to measure the distance of two vectors, specifically, via $\|\mathbf{x}-\mathbf{y}\|$ where $\mathbf{x}, \mathbf{y}$ are the two vectors

Vector Norm

Examples of norm:

- 2-norm or Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = (\mathbf{x}^T\mathbf{x})^{1/2}$

- 1-norm or Taxicab norm or Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$

- ∞-norm or maximum norm or sup-norm: $\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$

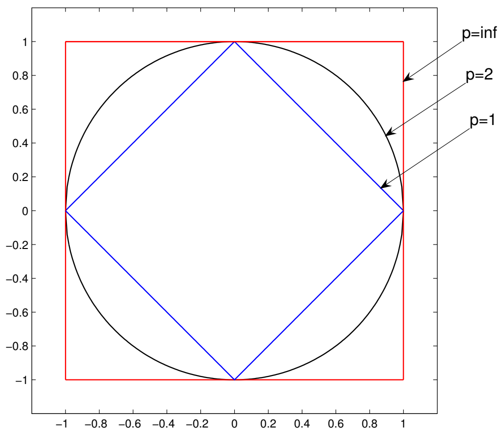
- p-norm $(p \geq 1)$ or Hölder norm: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

They are also called ℓ_1-norm, ℓ_2-norm, ℓ_∞-norm, and ℓ_p-norm, respectively.

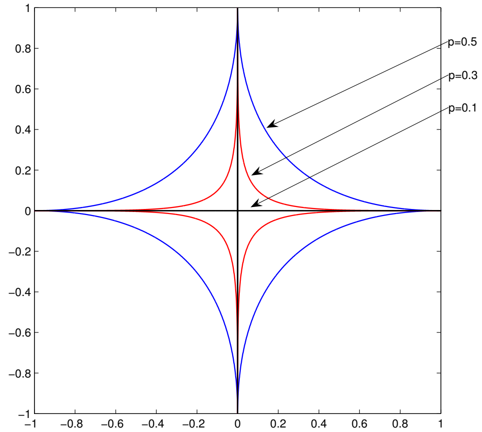
ℓ_p Function

Let

$$f_p(\mathbf{x}) = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \qquad p > 0.$$



(a) Region of $f_p(\mathbf{x}) = 1$, $p \geq 1$.



(b) Region of $f_p(\mathbf{x}) = 1$, $0 < p < 1$.

- f_p is *not* a norm for $0 < p < 1$
- when $p \to 0$, f_p is like the cardinality function $\mathrm{card}(\mathbf{x}) = \|\mathbf{x}\|_0 = \sum_{i=1}^{n} \mathbb{1}\{x_i \neq 0\}$,

  where $\mathbb{1}\{x \neq 0\} = 1$ if $x \neq 0$ and $\mathbb{1}\{x \neq 0\} = 0$ if $x = 0$.
- the 1-norm is the convex envelope of cardinality function

Inner Product and Angle

The inner product or dot product of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} y_i x_i = \mathbf{y}^T \mathbf{x} = \mathbf{y}^T \cdot \mathbf{x}.$$

- $\mathbf{x}, \mathbf{y}$ are said to be orthogonal or perpendicular to each other if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, denoted by $\mathbf{x} \perp \mathbf{y}$
- $\mathbf{x}, \mathbf{y}$ are said to be parallel if $\mathbf{x} = t\mathbf{y}$ for some t
  - for parallel $\mathbf{x}, \mathbf{y}$ we have $\langle \mathbf{x}, \mathbf{y} \rangle = \pm \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$

The angle between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\theta = \angle(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \right).$$

- $\mathbf{x}, \mathbf{y}$ are orthogonal if $\theta = \pm\pi/2$
- $\mathbf{x}, \mathbf{y}$ are parallel if $\theta = 0$ or $\theta = \pm\pi$

Important Inequalities for Inner Product

Cauchy-Schwartz inequality:

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2.$$

Also, the above equality holds if and only if $\mathbf{x} = \alpha\mathbf{y}$ for some $\alpha \in \mathbb{R}$.

- proof: suppose $\mathbf{y} \neq \mathbf{0}$; the case of $\mathbf{y} = \mathbf{0}$ is trivial. For any $\alpha \in \mathbb{R}$,

  $$0 \leq \|\mathbf{x} - \alpha\mathbf{y}\|_2^2 = (\mathbf{x} - \alpha\mathbf{y})^T(\mathbf{x} - \alpha\mathbf{y}) = \|\mathbf{x}\|_2^2 - 2\alpha\mathbf{x}^T\mathbf{y} + \alpha^2\|\mathbf{y}\|_2^2. \qquad (*)$$

  Also, the equality above holds if and only if $\mathbf{x} = t\mathbf{y}$ for some t.

  Let
  $$f(\alpha) = \|\mathbf{x}\|_2^2 - 2\alpha\mathbf{x}^T\mathbf{y} + \alpha^2\|\mathbf{y}\|_2^2.$$
  The function f is minimized when $\alpha = (\mathbf{x}^T\mathbf{y})/\|\mathbf{y}\|_2^2$. Plugging this α back to $(*)$ leads to the desired result.

Important Inequalities for Inner Product

Hölder inequality:

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_p\|\mathbf{y}\|_q,$$

for any p, q such that $1/p + 1/q = 1$, $p \geq 1$.

- examples:
  - $(p, q) = (2, 2)$: Cauchy-Schwartz inequality
  - $(p, q) = (1, \infty)$: $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty$. This can be easily verified to be true:

$$|\mathbf{x}^T\mathbf{y}| \leq \sum_{i=1}^n |x_i y_i| \leq \max_j |y_j| \left(\sum_{i=1}^n |x_i|\right) = \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty.$$

Dual Norm

For a given norm $\|\cdot\|$ on $\mathbb{R}^n$, the dual norm, denoted $\|\cdot\|_*$, is the function from $\mathbb{R}^n$ to $\mathbb{R}$ with values

$$\|\mathbf{y}\|_* = \max_{\|\mathbf{x}\|\leq 1} \langle \mathbf{x}, \mathbf{y} \rangle$$

- The above definition indeed corresponds to a norm: it is convex, as it is the pointwise maximum of convex (in fact, linear) functions $\mathbf{x} \to \langle \mathbf{x}, \mathbf{y} \rangle$; it is homogeneous of degree 1, that is, $\|\alpha\mathbf{x}\|_* = \alpha\|\mathbf{x}\|_*$ for every $\mathbf{x}$ in $\mathbb{R}^n$ and $\alpha \geq 0$.
- By definition of the dual norm,

$$\mathbf{y}^T\mathbf{x} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|_*.$$

- Examples:
  - The norm dual to the 2-norm is itself. (Cauchy-Schwartz inequality)
  - The norm dual to the ∞-norm is the 1-norm. $(\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\|_\infty \cdot \|\mathbf{y}\|_1)$
  - The norm dual to the p-norm is the q-norm where $1/p + 1/q = 1$.
- The dual of the dual norm is the original norm.

Orthogonality

- a vector $\mathbf{x} \in \mathbb{R}^n$ is said to be orthogonal to a nonempty set $\mathcal{S} \subseteq \mathbb{R}^n$ if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for any $\mathbf{y} \in \mathcal{S}$, denoted by $\mathbf{x} \perp \mathcal{S}$

- nonempty sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$ are said to be orthogonal to each other if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for any $\mathbf{x} \in \mathcal{S}_1$ and $\mathbf{y} \in \mathcal{S}_2$, denoted by $\mathcal{S}_1 \perp \mathcal{S}_2$

- properties:
  - given a nonempty set $\mathcal{S} \subseteq \mathbb{R}^n$, for any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \perp \mathcal{S} \Rightarrow \mathbf{x} \perp \operatorname{span} \mathcal{S}$.
  - if $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^n$ are orthogonal sets, $\mathcal{S}_1 \cap \mathcal{S}_2 = \{\mathbf{0}\}$ or $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$ (disjoint).

- note: the above results can be generalized to subspaces $\mathcal{S} \subseteq \mathbb{R}^n$

Projections Onto Subspaces, Orthogonal Complements, Four Fundamental Subspaces

Projection

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty closed set (not necessarily a subspace) and let $\mathbf{y} \in \mathbb{R}^m$ be given. A projection of $\mathbf{y}$ onto $\mathcal{S}$ is any solution to

$$\min_{\mathbf{z} \in \mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2^2$$

- a projection of $\mathbf{y}$ onto $\mathcal{S}$ is any point that is closest to $\mathbf{y}$ and lies in $\mathcal{S}$
- intepratation: to find a point in $\mathcal{S}$ that is closest to $\mathbf{y}$ in the Euclidean sense.

- in general, there may be more than one such closest point.
- notation: if, for every $\mathbf{y} \in \mathbb{R}^m$, there is always *only one* projection of $\mathbf{y}$ onto $\mathcal{S}$, then we denote
$$\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z} \in \mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2^2$$
and $\Pi_{\mathcal{S}}$ is called *the* projection (or projection operator) of $\mathbf{y}$ onto $\mathcal{S}$.

- we are interested in projections onto subspaces, which play a crucial role in linear algebra and matrix analysis.

Orthogonal Projections onto Subspaces

Theorem 2 (Projection Theorem)**.** Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$.

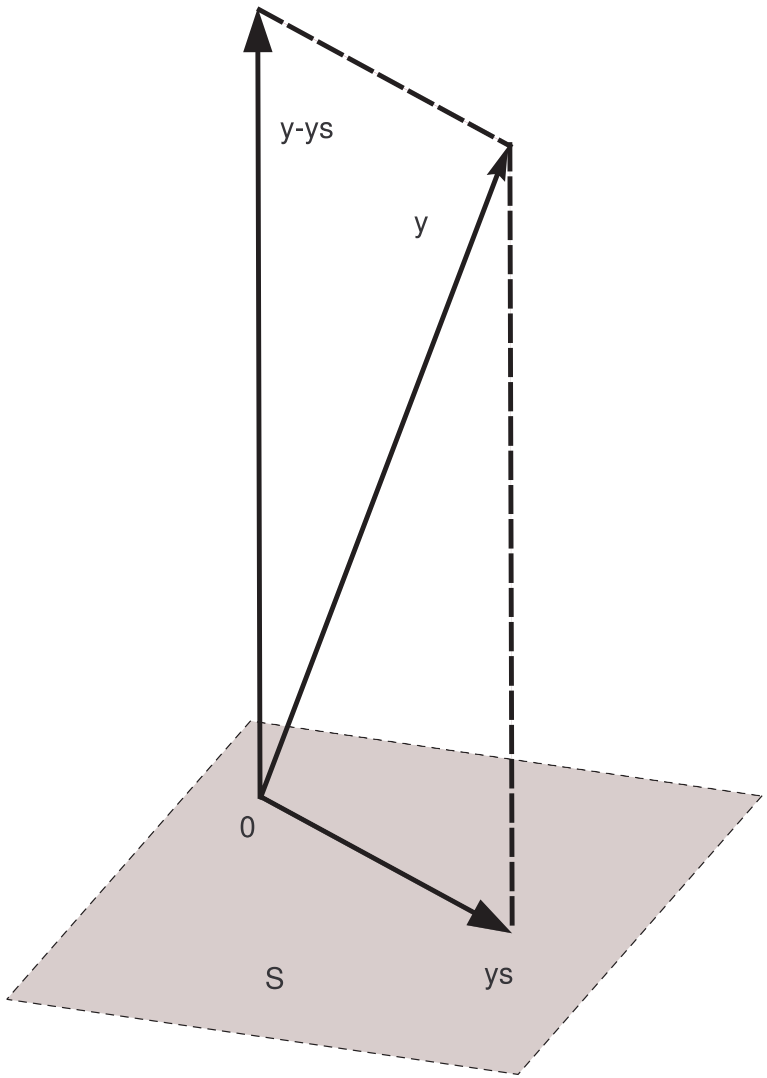
1. for every $\mathbf{y} \in \mathbb{R}^m$, there exists a unique vector $\mathbf{y}_s \in \mathcal{S}$ that minimizes $\|\mathbf{z} - \mathbf{y}\|_2^2$ over $\mathbf{z} \in \mathcal{S}$. Thus, we can use the notation $\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z}\in\mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2^2$.

2. given $\mathbf{y} \in \mathbb{R}^m$, we have the equivalence

$$\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}) \quad \Longleftrightarrow \quad \mathbf{y}_s \in \mathcal{S}, \quad \mathbf{z}^T(\mathbf{y} - \mathbf{y}_s) = 0 \text{ for all } \mathbf{z} \in \mathcal{S}.$$

- a special case of the projection theorem for convex sets
  - the latter plays a key role in convex analysis and optimization
- the subspace projection theorem above is very useful, as we will see

Orthogonal Projections onto Subspaces



Orthogonal Complements

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty set. The orthogonal complement of $\mathcal{S}$ is defined as

$$\mathcal{S}^\perp = \{\mathbf{z} \in \mathbb{R}^m \mid \mathbf{y}^T\mathbf{z} = 0 \text{ for all } \mathbf{y} \in \mathcal{S}\},$$

i.e., $\mathcal{S}^\perp$ is the largest subset of $\mathbb{R}^m$ orthogonal to $\mathcal{S}$.

- $\mathcal{S}^\perp$ is a subspace in $\mathbb{R}^m$ (easy to verify) and is unique
- properties:
  - any $\mathbf{y} \in \mathcal{S}, \mathbf{z} \in \mathcal{S}^\perp$ are orthogonal
  - either $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$ or $\mathcal{S} \cap \mathcal{S}^\perp = \emptyset$, i.e., if we exclude $\mathbf{0}$, the sets $\mathcal{S}$ and $\mathcal{S}^\perp$ are non-intersecting.
  - $(\mathcal{S}^\perp)^\perp = \text{span } \mathcal{S}$
- (Fundamental Subspace Theorem) for any $\mathbf{A} \in \mathbb{R}^{m\times n}$,
  - $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$
  - $\mathcal{N}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}^T)$

Orthogonal Complements

What happens to the orthogonal complement if $\mathcal{S}$ is a subspace?

Theorem 3 (Orthogonal Decomposition Theorem)**.** Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace.

1. for every $\mathbf{y} \in \mathbb{R}^m$, there exists a unique $(\mathbf{y}_s, \mathbf{y}_c) \in \mathcal{S} \times \mathcal{S}^\perp$ such that

$$\mathbf{y} = \mathbf{y}_s + \mathbf{y}_c.$$

   Also, such a $(\mathbf{y}_s, \mathbf{y}_c)$ is $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}), \mathbf{y}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})$.

2. the projection of $\mathbf{y}$ onto $\mathcal{S}^\perp$ can be determined by $\Pi_{\mathcal{S}^\perp}(\mathbf{y}) = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})$.

- proof sketch: by the Theorem 2. We can rephrase the projection theorem as

$$\mathbf{y}_s \in \Pi_{\mathcal{S}}(\mathbf{y}) \quad \Longleftrightarrow \quad \mathbf{y}_s \in \mathcal{S},\ \mathbf{y} - \mathbf{y}_s \in \mathcal{S}^\perp.$$

  This leads us to Statement 1.

Orthogonal Complements

Consequences of Theorem 3:

Property 1. Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace.

1. $\mathcal{S} + \mathcal{S}^\perp = \mathbb{R}^m$ or $\mathcal{S} \oplus \mathcal{S}^\perp = \mathbb{R}^m$;
2. $\dim \mathcal{S} + \dim \mathcal{S}^\perp = m$;
3. $(\mathcal{S}^\perp)^\perp = \mathcal{S}$.

- examples: let $\mathbf{A} \in \mathbb{R}^{m\times n}$.
  - (Orthogonal Decomposition Theorem)
    * $\mathcal{R}(\mathbf{A}) \oplus \mathcal{R}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^T) = \mathbb{R}^m$
    * $\mathcal{N}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}) \oplus \mathcal{R}(\mathbf{A}^T) = \mathbb{R}^n$
  - $\dim \mathcal{R}(\mathbf{A}) + \dim \mathcal{R}(\mathbf{A})^\perp = \dim \mathcal{R}(\mathbf{A}) + \dim \mathcal{N}(\mathbf{A}^T) = m$
    * $\operatorname{rank}(\mathbf{A}) = m - \dim \mathcal{N}(\mathbf{A})^\perp$
  - $\dim \mathcal{N}(\mathbf{A}) + \dim \mathcal{N}(\mathbf{A})^\perp = \dim \mathcal{N}(\mathbf{A}) + \dim \mathcal{R}(\mathbf{A}^T) = n$
    * $\operatorname{rank}(\mathbf{A}) = n - \dim \mathcal{N}(\mathbf{A})$
    * (Rank-Nullity Theorem) $\operatorname{rank}(\mathbf{A}) + \operatorname{nullity}(\mathbf{A}) = n$, $\operatorname{nullity}(\mathbf{A}) = \dim \mathcal{N}(\mathbf{A})$, i.e., nullity is the dimension of the kernal of $\mathbf{A}$

Orthogonal Complements

Property 2. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be a subspace.

1. $\mathcal{S}_1 \subset \mathcal{S}_2 \Longrightarrow \mathcal{S}_2^\perp \subset \mathcal{S}_1^\perp$;

2. $(\mathcal{S}_1 \cap \mathcal{S}_2)^\perp = \mathcal{S}_1^\perp + \mathcal{S}_2^\perp$;

3. $(\mathcal{S}_1 + \mathcal{S}_2)^\perp = \mathcal{S}_1^\perp \cap \mathcal{S}_2^\perp$.

Four Fundamental Subspaces

The subspaces $\mathcal{N}(\mathbf{A}), \mathcal{R}(\mathbf{A}^T) \subseteq \mathbb{R}^n$ and $\mathcal{R}(\mathbf{A}), \mathcal{N}(\mathbf{A}^T) \subseteq \mathbb{R}^m$ are the four fundamental subspaces associated with matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$.