Matrix Computations
Chapter 1 Introduction
Section 1.2 Review of Linear Algebra

Jie Lu
ShanghaiTech University

Notation

$\mathbb{R}$	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
$\mathbb{R}^{m\times n}$	the set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m\times n}$	the set of all $m \times n$ complex-valued matrices
a	scalar in $\mathbb{C}$
a^*	conjugate of $a \in \mathbb{C}$
$\mathbf{x}$	vector
x_i, $[\mathbf{x}]_i$	ith entry of $\mathbf{x}$
$\mathbf{A}$	matrix
a_{ij}, $[\mathbf{A}]_{ij}$	(i,j)-entry of $\mathbf{A}$
$\mathbb{S}^n$	the 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 $i,j = 1,\ldots,n$
$\mathbb{H}^n$	the set of all $n \times n$ complex Hermitian matrices, i.e, $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $a_{ij} = a_{ji}^*$ for all i,j

Vector

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

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

- $\mathbf{x} \in \mathbb{C}^n$: $\mathbf{x}$ is a complex-valued n-dimensional column vector

- Transpose: $\mathbf{x}^T = \begin{bmatrix} x_1, & x_2, & \ldots, & x_n \end{bmatrix}$
- Hermitian transpose: $\mathbf{x}^H = \begin{bmatrix} x_1^*, & x_2^*, & \ldots, & x_n^* \end{bmatrix}$

Matrix

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

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

- $\mathbf{A} \in \mathbb{C}^{m \times n}$: $\mathbf{A}$ is a complex-valued $m \times n$ matrix

- We may write

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1, & \mathbf{a}_2, & \dots, & \mathbf{a}_n \end{bmatrix}$$

where $\mathbf{a}_i \in \mathbb{R}^m$ is the ith column of matrix A

Matrix (Cont'd)

- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be
  - square if $m = n$;
  - tall if $m > n$;
  - fat if $m < n$.

- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be
  - upper triangular if $a_{ij} = 0$ for all $i > j$;
  - lower triangular if $a_{ij} = 0$ for all $i < j$.

Examples:

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

Matrix Transpose

- Given a $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & & & \vdots \\ a_{1n} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

  is a $n \times m$ matrix

- The following properties hold:
  - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
  - $(\mathbf{A}^T)^T = \mathbf{A}$
  - $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

Matrix Transpose (Cont'd)

- Hermitian transpose: Given $\mathbf{A} \in \mathbb{C}^{m \times n}$,

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

- The following properties hold:
  - $(\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H$
  - $(\mathbf{A}^H)^H = \mathbf{A}$
  - $(\mathbf{A} + \mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H$

Matrix Trace and Matrix Power

- Given $\mathbf{A} \in \mathbb{R}^{n \times n}$, the trace of $\mathbf{A}$ is

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

  - $\operatorname{tr}(\mathbf{A}^T) = \operatorname{tr}(\mathbf{A})$
  - $\operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B})$
  - $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$ for $\mathbf{A}, \mathbf{B}$ of proper sizes

- Matrix power: Given $\mathbf{A} \in \mathbb{R}^{n \times n}$,

$$\mathbf{A}^k = \underbrace{\mathbf{AA} \cdots \mathbf{A}}_{k \text{ times}}$$

Some Common Vectors and Matrices

- All-one vectors: We use the notation

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

  to denote a vector of all 1's

- Zero vectors or matrices: We use the notation $\mathbf{0}$ to denote either a vector of all zeros or a matrix of all zeros

- Unit vectors: We use the notation

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

  to denote a unit vector whose i-th entry is 1 and other entries are all zero

Some Common Vectors and Matrices (Cont'd)

- Identity matrix:

$$\mathbf{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

  The empty entries are assumed to be zero by default

- 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 whose diagonal entries are $a_1, \ldots, a_n$

  For $\mathbf{a} = \left[a_1, \ldots, a_n\right]^T$, we use the shorthand notation $\mathrm{Diag}(\mathbf{a})$

Subspace

A subset $\mathcal{S}$ of $\mathbb{R}^m$ is said to be a subspace if for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ and any $\alpha, \beta \in \mathbb{R}$,

$$\alpha \mathbf{x} + \beta \mathbf{y} \in \mathcal{S}$$

- If $\mathcal{S}$ is a subspace and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$, then any linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$, i.e., $\sum_{i=1}^n \alpha_i \mathbf{a}_i$ for some $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, lies in $\mathcal{S}$
- Let $\mathcal{S}_1, \mathcal{S}_2$ be subspaces of $\mathbb{R}^m$
  - $\mathcal{S}_1 + \mathcal{S}_2 := \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{S}_1, \mathbf{y} \in \mathcal{S}_2\}$ is a subspace
  - $\mathcal{S}_1 \cap \mathcal{S}_2$ is a subspace

Span

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

$$\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \left\{ \mathbf{y} \in \mathbb{R}^m \;\middle|\; \mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i,\ \alpha_1, \ldots, \alpha_n \in \mathbb{R} \right\}$$

- $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is the set of all linear combinations of $\mathbf{a}_1, \ldots, \mathbf{a}_n$
- $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a subspace

Theorem

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

- We can always represent a subspace by a span

Range and Nullspace

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

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

- $\mathcal{R}(\mathbf{A})$ is the span of the columns of $\mathbf{A}$

The nullspace of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as

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

- $\mathcal{N}(\mathbf{A})$ is a subspace
- $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{B})$ for some integer $r > 0$ and $\mathbf{B} \in \mathbb{R}^{n \times r}$

Linear Independence

$\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are said to be linearly independent if

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i \neq \mathbf{0} \quad \text{for all } \alpha = \left[\alpha_1, \ldots, \alpha_n\right]^T \in \mathbb{R}^n \text{ with } \alpha \neq \mathbf{0}$$

and linearly dependent otherwise

- Equivalent definition of linear dependence: $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are linearly dependent if there exists $\alpha \in \mathbb{R}^n$, $\alpha \neq \mathbf{0}$ such that

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

Linear Independence (Cont'd)

- If $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$ are linearly independent, then any $\mathbf{a}_j$ *cannot* be a linear combination of the other $\mathbf{a}_i$'s
- If $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$ are 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
- If $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$ are linearly independent, then $n \leq m$

Linear Independence (Cont'd)

- If $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are linearly independent and $\mathbf{y} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then the coefficient $\alpha = [\alpha_1, \ldots, \alpha_n]^T$ for the representation

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

  is unique, i.e., there does *not* exist $\beta = [\beta_1, \ldots, \beta_n]^T \in \mathbb{R}^n$, $\beta \neq \alpha$ such that $\mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i$

Linear Independence (Cont'd)

Let $\{\mathbf{a}_1, \dots \mathbf{a}_n\} \subset \mathbb{R}^m$, and denote $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ as an index subset with $k \leq n$ and $i_j \neq i_\ell$ for all $j \neq \ell$.
A vector subset $\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\}$ is called a maximal linearly independent subset of $\{\mathbf{a}_1, \dots \mathbf{a}_n\}$ if both of the following conditions hold:

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

- A set of non-redundant vectors from $\{\mathbf{a}_1, \dots \mathbf{a}_n\}$

Linear Independence (Cont'd)

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 subsets of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ are

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

The maximal linearly independent subsets are

Linear Independence (Cont'd)

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 subsets of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ are

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

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 (Cont'd)

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 both of the following hold:

1. $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are linearly independent
2. $\mathcal{S} = \mathrm{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$

- If $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is 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 $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$
- Given $\mathcal{S}$, there can be multiple bases
- All bases for $\mathcal{S}$ have the same number of elements, i.e., if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_\ell\}$ are bases for $\mathcal{S}$, then $k = \ell$

Dimension of a Subspace

The dimension of a subspace $\mathcal{S}$ with $\mathcal{S} \neq \{\mathbf{0}\}$, denoted by $\dim \mathcal{S}$, is the number of elements of any basis for $\mathcal{S}$

- $\dim\{\mathbf{0}\} = 0$
- represent 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 of a Subspace (Cont'd)

Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces

- If $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $\dim \mathcal{S}_1 \leq \dim \mathcal{S}_2$
- If $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $\dim \mathcal{S}_1 = \dim \mathcal{S}_2$, then $\mathcal{S}_1 = \mathcal{S}_2$
- $\dim \mathcal{S}_1 = m$ if and only if $\mathcal{S}_1 = \mathbb{R}^m$
- $\dim(\mathcal{S}_1 + \mathcal{S}_2) \leq \dim \mathcal{S}_1 + \dim \mathcal{S}_2$
    - $\dim(\mathcal{S}_1 + \mathcal{S}_2) = \dim \mathcal{S}_1 + \dim \mathcal{S}_2 - \dim(\mathcal{S}_1 \cap \mathcal{S}_2)$

Rank

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

- Equivalently, $\mathrm{rank}(\mathbf{A})$ is the maximum number of linearly independent columns of $\mathbf{A}$
- $\dim \mathcal{R}(\mathbf{A}) = \mathrm{rank}(\mathbf{A})$

Facts:

- $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}^T)$, i.e., the rank of $\mathbf{A}$ is also the maximum number of linearly independent rows of $\mathbf{A}$
- $\mathrm{rank}(\mathbf{A} + \mathbf{B}) \le \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$
- $\mathrm{rank}(\mathbf{AB}) \le \min\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\}$
  - The equality holds when the columns of $\mathbf{A}$ are linearly independent and the rows of $\mathbf{B}$ are linearly independent

Rank (Cont'd)

- Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to
  - have full column rank if all the columns of $\mathbf{A}$ are linearly independent
    - $\mathbf{A} \in \mathbb{R}^{m \times n}$ full column rank $\iff m \geq n, \mathrm{rank}(\mathbf{A}) = n$
  - have full row rank if all the rows of $\mathbf{A}$ are linearly independent
    - $\mathbf{A} \in \mathbb{R}^{m \times n}$ full row rank $\iff m \leq n, \mathrm{rank}(\mathbf{A}) = m$
  - have full rank if $\mathrm{rank}(\mathbf{A}) = \min\{m, n\}$, i.e., it has either full column rank or full row rank
  - be rank deficient if $\mathrm{rank}(\mathbf{A}) < \min\{m, n\}$

Invertible Matrices

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be nonsingular or invertible if the columns of $\mathbf{A}$ are linearly independent, and singular or non-invertible otherwise

- Alternatively, $\mathbf{A}$ is singular if $\mathbf{Ax} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$

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

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

Invertible Matrices (Cont'd)

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a nonsingular matrix

- $\mathbf{A}^{-1}$ always exists and is unique
- $\mathbf{A}^{-1}$ is nonsingular
- $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ are nonsingular
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
    - As a shorthand notation, we denote $\mathbf{A}^{-T} = (\mathbf{A}^T)^{-1}$

Determinant

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

- For $m = 1$: $\det(\mathbf{A}) = a_{11}$
- For $m \geq 2$:
  - 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:

$$\det(\mathbf{A}) = \sum_{j=1}^{m} a_{ij} c_{ij}, \quad \text{for any } i = 1, \ldots, m$$

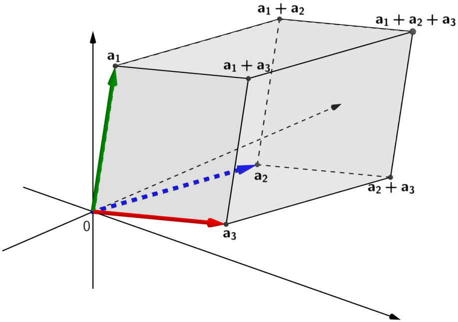
$$\det(\mathbf{A}) = \sum_{i=1}^{m} a_{ij} c_{ij}, \quad \text{for any } j = 1, \ldots, m$$

where c_{ij}'s are the cofactors and $\det(\mathbf{A}_{ij})$'s are the minors

Determinant (Cont'd)

- **Fact**: $\mathbf{Ax} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$ if and only if $\det(\mathbf{A}) = 0$

- Interpretation: $|\det(\mathbf{A})|$ is the volume of the parallelepiped $\mathcal{P} = \{\mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{a}_i \mid \alpha_i \in [0,1] \; \forall i = 1, \ldots, m\}$



Determinant (Cont'd)

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times m}$

- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- $\det(\mathbf{A}) = \det(\mathbf{A}^T)$
- $\det(\alpha\mathbf{A}) = \alpha^m \det(\mathbf{A})$ for any $\alpha \in \mathbb{R}$
- $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$ 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)
  - $\tilde{\mathbf{A}}$ is the adjoint or adjugate matrix of $\mathbf{A}$

Determinant (Cont'd)

- 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 $\mathbf{A} \in \mathbb{R}^{m \times m}$ is *block* upper or lower triangular

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

where $\mathbf{B}$ and $\mathbf{D}$ are square (and can be of different sizes), then

$$\det(\mathbf{A}) = \det(\mathbf{B})\det(\mathbf{D})$$

Vector Norms

A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a vector norm if all of the following hold:

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

- Usually $\|\cdot\|$ denotes a norm
- $\|\mathbf{x}\|$ represents the "length" of vector $\mathbf{x}$
- $\|\mathbf{x} - \mathbf{y}\|$ represents the "distance" of vectors $\mathbf{x}, \mathbf{y}$

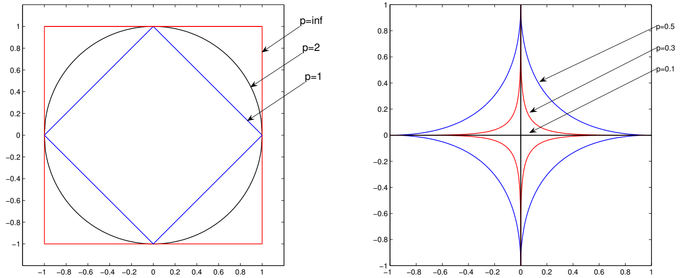
Vector Norms (Cont'd)

Examples:

- 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 Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- ∞-norm: $\|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$
- p-norm, $p \geq 1$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$

ℓ_p Function

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



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

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

Inner Product

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

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

- $\mathbf{x}, \mathbf{y}$ are said to be orthogonal to each other if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$
- $\mathbf{x}, \mathbf{y}$ are said to be parallel if $\mathbf{x} = \alpha \mathbf{y}$ for some α
  - $\langle \mathbf{x}, \mathbf{y} \rangle = \pm \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$ for parallel $\mathbf{x}, \mathbf{y}$

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

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

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

Hölder Inequality

Hölder Inequality: For any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

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

Proof. **Young's Inequality**: For any $a, b \geq 0$ and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Hölder Inequality (Cont'd)

Hölder Inequality: For any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

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

- Cauchy-Schwartz Inequality: Let $p = q = 2$ in Hölder Inequality

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

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

- Hölder Inequality holds for $p = 1$ and $q = \infty$

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

Equivalence of Norms

All norms on $\mathbb{R}^n$ are equivalent in the sense that if $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are norms on $\mathbb{R}^n$, then there exist $c1, c_2 > 0$ such that

$$c_1\|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq c_2\|\mathbf{x}\|_\alpha, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

- $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$
- $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$
- $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$

Projections on Subspaces

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty closed set (not necessarily a subspace)
Given $\mathbf{y} \in \mathbb{R}^m$, 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 point in $\mathcal{S}$ that is closest to $\mathbf{y}$
  - Projection of $\mathbf{y} \in \mathcal{S}$ onto $\mathcal{S}$ is $\mathbf{y}$ itself
- If for *any* $\mathbf{y} \in \mathbb{R}^m$, there always exists a *unique* 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 $\mathcal{S}$

Projection Theorem

Theorem (Projection Theorem)

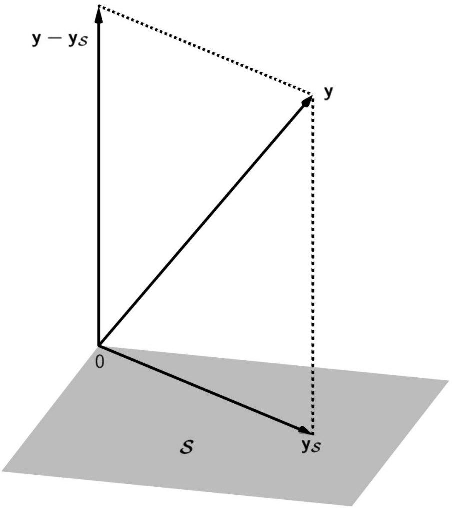
Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$.

1. *For any $\mathbf{y} \in \mathbb{R}^m$, there exists a unique vector $\mathbf{y}_s \in \mathcal{S}$ that minimizes $\|\mathbf{z} - \mathbf{y}\|_2^2$ over $\mathbf{z} \in \mathcal{S}$ (so that 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 $\mathbf{y} \in \mathbb{R}^m$,*

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

- Statement 1 of Projection Theorem also holds for closed convex set (more general)
  - Very important to convex optimization

Projection Theorem (Cont'd)



Orthogonal Complement

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty closed set

The orthogonal complement of $\mathcal{S}$ is defined as

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

- $\mathcal{S}^\perp$ is a subspace (Why?)
- Any $\mathbf{z} \in \mathcal{S}$ and any $\mathbf{y} \in \mathcal{S}^\perp$ are orthogonal
- Either $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$ or $\mathcal{S} \cap \mathcal{S}^\perp = \emptyset$
- **Facts**:
  - $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$
  - $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^\perp$
  - Recall that range and nullspace of a matrix are subspaces

Orthogonal Complement of Subspace

Theorem

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace. For any $\mathbf{y} \in \mathbb{R}^m$, there uniquely exists $(\mathbf{y}_s, \mathbf{y}_c) \in \mathcal{S} \times \mathcal{S}^\perp$ such that

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

In particular, $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}), \mathbf{y}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y}) = \Pi_{\mathcal{S}^\perp}(\mathbf{y})$.

- Proof sketch: From Statement 2 of the Projection Theorem,

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

Orthogonal Complement of Subspace (Cont'd)

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace. It follows from the above theorem that

- $\mathcal{S} + \mathcal{S}^\perp = \mathbb{R}^m$
- $\dim \mathcal{S} + \dim \mathcal{S}^\perp = m$
  - Proof: $\dim \mathcal{S} + \dim \mathcal{S}^\perp = \dim(\mathcal{S} + \mathcal{S}^\perp) + \dim(\mathcal{S} \cap \mathcal{S}^\perp) = \dim(\mathcal{S} + \mathcal{S}^\perp) + 0 = \dim \mathbb{R}^m$
- $(\mathcal{S}^\perp)^\perp = \mathcal{S}$

Example: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$

- $\dim \mathcal{R}(\mathbf{A}) + \dim \mathcal{R}(\mathbf{A})^\perp = \dim \mathcal{R}(\mathbf{A}) + \dim \mathcal{N}(\mathbf{A}^T) = m$
- **Rank-Nullity Theorem**: $\dim \mathcal{N}(\mathbf{A}) = n - \dim \mathcal{R}(\mathbf{A}^T) = n - \mathrm{rank}(\mathbf{A})$

Orthogonal and Orthonormal Vectors

A collection of *nonzero* vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are said to be

- orthogonal if $\mathbf{a}_i^T \mathbf{a}_j = 0$ for all i, j with $i \neq j$
- orthonormal if they are orthogonal and $\|\mathbf{a}_i\|_2 = 1$ for all i

Same definition applies to complex $\mathbf{a}_i$'s by replacing transpose (T) with Hermitian transpose (H)

Example: Any vectors from $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ are orthonormal and $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subset \mathbb{R}^m$ is an orthonormal basis for $\mathbb{R}^m$

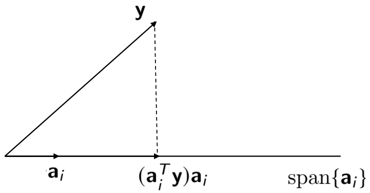
Orthonormal vectors are linearly independent

Orthogonal and Orthonormal Vectors (Cont'd)

Fact: Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ be an orthonormal set of vectors and $\mathbf{y} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then, the coefficient α for the representation

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

is uniquely given by $\alpha_i = \mathbf{a}_i^T \mathbf{y}$, $i = 1, \ldots, n$



Fact: Every subspace $\mathcal{S}$ with $\mathcal{S} \neq \{\mathbf{0}\}$ has an orthonormal basis

- It can be shown using Gram-Schmidt

Orthogonal Matrix

A real matrix $\mathbf{Q}$ is said to be

- orthogonal if it is square and its columns are orthonormal
- semi-orthogonal if its columns are orthonormal
    - a semi-orthogonal $\mathbf{Q}$ must be tall or square

A complex matrix $\mathbf{Q}$ is said to be unitary if it is square and its columns are orthonormal, and semi-unitary if its columns are orthonormal

Example: Consider the transformation $\mathbf{y} = \mathbf{Q}\mathbf{x}$ with

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad \text{rotation counterclock-wise by } \theta \in [0, 2\pi)$$

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \quad \text{reflection about the } \theta/2 \text{ line, } \theta \in [0, 2\pi)$$

The rotation and reflection matrices are orthogonal

Orthogonal Matrix (Cont'd)

Facts:

- $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ and $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ for orthogonal $\mathbf{Q}$
- $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ (but *not* necessarily $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) for semi-orthogonal $\mathbf{Q}$
- $\|\mathbf{Q}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ for orthogonal $\mathbf{Q}$
  - For example, rotation and reflection do not change the vector length
- For any tall and semi-orthogonal matrix $\mathbf{Q}_1 \in \mathbb{R}^{n\times k}$, there exists a matrix $\mathbf{Q}_2 \in \mathbb{R}^{n\times(n-k)}$ such that $[\ \mathbf{Q}_1\ \mathbf{Q}_2\ ]$ is orthogonal

Matrix Product Representations

Let $\mathbf{A} \in \mathbb{R}^{m \times k}$ and $\mathbf{B} \in \mathbb{R}^{k \times n}$. Consider

$$\mathbf{C} = \mathbf{A}\mathbf{B}$$

- Column representation:

$$\mathbf{c}_i = \mathbf{A}\mathbf{b}_i, \quad i = 1, \ldots, n$$

  where $\mathbf{c}_i$ and $\mathbf{b}_i$ are the ith columns of $\mathbf{C}$ and $\mathbf{B}$

- Inner-product representation: Let $\tilde{\mathbf{a}}_i^T \in \mathbb{R}^{1 \times k}$ be the ith row of $\mathbf{A}$

$$\mathbf{A}\mathbf{B} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \vdots \\ \tilde{\mathbf{a}}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{b}_1 & \cdots & \tilde{\mathbf{a}}_1^T \mathbf{b}_n \\ \vdots & & \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{b}_1 & \cdots & \tilde{\mathbf{a}}_m^T \mathbf{b}_n \end{bmatrix}$$

  Thus,

$$c_{ij} = \tilde{\mathbf{a}}_i^T \mathbf{b}_j, \quad \text{for all } i, j$$

Matrix Product Representations (Cont'd)

- Outer-product representation: Let $\tilde{\mathbf{b}}_i^T \in \mathbb{R}^{1\times n}$ be the ith row of $\mathbf{B}$

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{I} \cdot \mathbf{B} = \mathbf{A}\left(\sum_{i=1}^{k} \mathbf{e}_i \mathbf{e}_i^T\right)\mathbf{B} = \sum_{i=1}^{k} (\mathbf{A}\mathbf{e}_i)(\mathbf{e}_i^T \mathbf{B})$$

Thus,

$$\mathbf{C} = \sum_{i=1}^{k} \mathbf{a}_i \tilde{\mathbf{b}}_i^T$$

Matrix Product Representations (Cont'd)

- A matrix of the form $\mathbf{X} = \mathbf{a}\mathbf{b}^T$ for some $\mathbf{a}, \mathbf{b}$ is called a rank-one outer product
- $\mathrm{rank}(\mathbf{X}) \leq 1$, and $\mathrm{rank}(\mathbf{X}) = 1$ if $\mathbf{a} \neq \mathbf{0}, \mathbf{b} \neq \mathbf{0}$
- The outer-product representation $\mathbf{C} = \sum_{i=1}^{k} \mathbf{a}_i \tilde{\mathbf{b}}_i^T$ is a sum of k rank-one outer product
- $\mathrm{rank}(\mathbf{C}) =?$
  - $\mathrm{rank}(\mathbf{C}) \leq \sum_{i=1}^{k} \mathrm{rank}(\mathbf{a}_i \mathbf{b}_i^T) \leq k$ is true [1]
  - $\mathrm{rank}(\mathbf{C}) = \sum_{i=1}^{k} \mathrm{rank}(\mathbf{a}_i \mathbf{b}_i^T)$ may not be true
    - Counterexample: $k = 2$, $\mathbf{a}_1 = \mathbf{a}_2$, $\mathbf{b}_1 = -\mathbf{b}_2$ leads to $\mathbf{C} = \mathbf{0}$
  - $\mathrm{rank}(\mathbf{C}) = k$ only when $\mathbf{A}$ has full-column rank and $\mathbf{B}$ has full-row rank

[1] $\mathrm{rank}(\mathbf{A} + \mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$

Block Matrix Manipulations

It is more convenient to manipulate matrices in block forms

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$. By partitioning

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$$

  where $\mathbf{A}_1 \in \mathbb{R}^{m \times n_1}, \mathbf{A}_2 \in \mathbb{R}^{m \times n_2}, \mathbf{x}_1 \in \mathbb{R}^{n_1}, \mathbf{x}_2 \in \mathbb{R}^{n_2}$ with $n_1 + n_2 = n$, we can write

$$\mathbf{A}\mathbf{x} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2$$

- Similarly, by partitioning properly,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix},$$

  we can write

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 \\ \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2 \end{bmatrix}$$

Block Matrix Manipulations (Cont'd)

- Consider **AB**. By partitioning properly,

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_2$$

- Similarly,

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1\mathbf{B}_1 & \mathbf{A}_1\mathbf{B}_2 \\ \mathbf{A}_2\mathbf{B}_1 & \mathbf{A}_2\mathbf{B}_2 \end{bmatrix}$$

- Easily extended to multi-block partitioning

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1q} \\ \vdots & & \vdots \\ \mathbf{A}_{p1} & \cdots & \mathbf{A}_{pq} \end{bmatrix}$$

Extension from $\mathbb{R}^n$ to $\mathbb{C}^n$

- The previous concepts for vectors apply to the complex case
- Only need to replace every "$\mathbb{R}$" with "$\mathbb{C}$", and every "T" with "H"
- **Examples**:
  - $\mathrm{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \{\mathbf{y} \in \mathbb{C}^m \mid \mathbf{y} = \sum_{i=1}^n \alpha_i \mathbf{a}_i,\ \alpha \in \mathbb{C}^n\}$
  - $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^H \mathbf{x}$
  - $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^H \mathbf{x}}$

Extension from $\mathbb{R}^n$ to $\mathbb{R}^{m \times n}$

- The previous concepts for vectors apply to the matrix case
  - For example,

$$\operatorname{span}\{\mathbf{A}_1, \ldots, \mathbf{A}_k\} = \{\mathbf{Y} \in \mathbb{R}^{m \times n} \mid \mathbf{Y} = \textstyle\sum_{i=1}^{k} \alpha_i \mathbf{A}_i,\ \alpha \in \mathbb{R}^k\}.$$

  - Sometimes it is more convenient to *vectorize* $\mathbf{X}$ as a vector $\mathbf{x} \in \mathbb{R}^{mn}$, and use the same treatment as in the vector case
  - Inner product for $\mathbb{R}^{m \times n}$:

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \textstyle\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} y_{ij} = \operatorname{tr}(\mathbf{Y}^T \mathbf{X})$$

  - The matrix version of the Euclidean norm is called the Frobenius norm:

$$\|\mathbf{X}\|_F = \sqrt{\textstyle\sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij}|^2} = \sqrt{\operatorname{tr}(\mathbf{X}^T \mathbf{X})}$$

- Likewise, we can extend the above to $\mathbb{C}^{m \times n}$

Matrix Norm

A function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is a matrix norm if all of the following hold:

1. $f(\mathbf{A}) \geq 0$ for any $\mathbf{A} \in \mathbb{R}^{m\times n}$
2. $f(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$
3. $f(\mathbf{A}+\mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B})$ for any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$
4. $f(\alpha\mathbf{A}) = |\alpha| f(\mathbf{A})$ for any $\alpha \in \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{m\times n}$

Most commonly used matrix norms:

- Frobenius norm: $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\mathrm{tr}(\mathbf{A}^T\mathbf{A})}$
- p-norm ($p \geq 1$):

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$
$$= \sup_{\mathbf{x}\neq\mathbf{0}} \left\|\mathbf{A}\frac{\mathbf{x}}{\|\mathbf{x}\|_p}\right\|_p = \sup_{\mathbf{x}:\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_p$$

  - $\|\mathbf{A}\|_p$ is the p-norm of the largest vector obtained by applying $\mathbf{A}$ to a unit p-norm vector
  - Induced by the vector p-norm

Matrix p-Norm

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$

$$\|\mathbf{A}\|_1 = \sup_{\mathbf{x}:\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}| \quad \text{the largest absolute column sum}$$

$$\|\mathbf{A}\|_\infty = \sup_{\mathbf{x}:\|\mathbf{x}\|_\infty=1} \|\mathbf{A}\mathbf{x}\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}| \quad \text{the largest absolute row sum}$$

$$\|\mathbf{A}\|_2 = \sup_{\mathbf{x}:\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 \quad \text{spectral norm}$$

- $\|\mathbf{A}\|_2^2$ is equal to the largest eigenvalue of A^TA (will be discussed later)

Matrix p-Norm (Cont'd)

Facts: $\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \cdot \|\mathbf{B}\|_p$, $\forall \mathbf{A} \in \mathbb{R}^{m\times n}$, $\forall \mathbf{B} \in \mathbb{R}^{n\times q}$

Properties: For any $\mathbf{A} \in \mathbb{R}^{m\times n}$,

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{\min\{m,n\}}\|\mathbf{A}\|_2$$

$$\max_{i,j} |a_{ij}| \leq \|\mathbf{A}\|_2 \leq \sqrt{mn}\max_{i,j} |a_{ij}|$$

$$\frac{1}{\sqrt{n}}\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\|_2 \leq \sqrt{m}\|\mathbf{A}\|_\infty$$

$$\frac{1}{\sqrt{m}}\|\mathbf{A}\|_1 \leq \|\mathbf{A}\|_2 \leq \sqrt{n}\|\mathbf{A}\|_1$$

$$\|\mathbf{A}(i_1 : i_2, j_1 : j_2)\|_p \leq \|\mathbf{A}\|_p, \quad 1 \leq i_1 \leq i_2 \leq m,\ 1 \leq j_1 \leq j_2 \leq n$$

$$\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_1\|\mathbf{A}\|_\infty}$$