Linear Algebra and Geometry

Shan-Hung Wu
shwu@cs.nthu.edu.tw

Department of Computer Science,
National Tsing Hua University, Taiwan

NetDB-ML, Spring 2013

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Notation

- The conditional $A \Rightarrow B$ reads either "if A than B," "A only if B," "A is sufficient for B," or "B is necessary for A"
- The biconditional $A \Leftrightarrow B$ reads "A if and only if (or iff) B"
- $\{x : x \in \mathbb{R}, x > 5\}$ or $\{x \in \mathbb{R} : x > 5\}$ reads "the set of all x such that x is real and x is greater than 5"
- We denote a function as $f : \mathcal{V} \to \mathcal{W}$. The $\mathcal{V}$ and $\mathcal{W}$ are called ***domain*** and ***codomain*** (or ***target***) of f respectively
- Titles marked with ** can be skipped for the first time reading
- Statements marked with [Proof] mean you are encouraged to prove it yourself
- Statements marked with [Homework] are your assignments

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Vector Spaces

Definition (Vector Space)

The set $\mathcal{V} = \{[v_1, v_2, \cdots, v_n]^\top : v_i \in \mathbb{R}\} \subseteq \mathbb{R}^n$ is called a ***vector space*** over $\mathbb{R}$[a] iff there are maps:

1) Vector addition $\mathcal{V} \times \mathcal{V} \to \mathcal{V}$, denoted by $\boldsymbol{v} + \boldsymbol{w}$ for all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$,
2) Scalar multiplication $\mathbb{R} \times \mathcal{V} \to \mathcal{V}$, denoted by $a \cdot \boldsymbol{v}$ or $a\boldsymbol{v}$ for all $a \in \mathbb{R}$ and $\boldsymbol{v} \in \mathcal{V}$;

with the following properties:

a) For all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$, $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$ (commutativity);
b) For all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$, $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{w}) + \boldsymbol{v}$ (associativity);
c) There exists $\mathbf{0} \in \mathcal{V}$ such that $\mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{V}$;
d) For each $\boldsymbol{v} \in \mathcal{V}$, there exists $(-\boldsymbol{v}) \in \mathcal{V}$ with $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$;
e) For all $a \in \mathbb{R}$ and $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$, $a(\boldsymbol{v} + \boldsymbol{w}) = a\boldsymbol{w} + a\boldsymbol{v}$ (distributivity);
f) For all $a, b \in \mathbb{R}$ and $\boldsymbol{v} \in \mathcal{V}$, $(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$ (distributivity);
g) For all $a, b \in \mathbb{R}$ and $\boldsymbol{v} \in \mathcal{V}$, $a \cdot (b \cdot \boldsymbol{v}) = (a \cdot b) \cdot \boldsymbol{v}$ (associativity);
h) For all $\boldsymbol{v} \in \mathcal{V}$, $1 \cdot \boldsymbol{v} = \boldsymbol{v}$.

[a] While any field is applicable, we focus on the real numbers here.

- We call the n-tuple $\boldsymbol{v}$ a ***vector*** and the scalar v_i the ith ***component*** of $\boldsymbol{v}$

Bases and Coordinates (1/2)

Definition (Linear Independence)

A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ in a vector space $\mathcal{V}$ is called ***linear independent*** iff $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_n\boldsymbol{v}_n = 0 \Rightarrow a_1 = a_2 = \cdots = a_n = 0$.

- In other words, there is no vector in this set that can be the ***linear combination*** of others

Definition (Span)

A set of all linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ is called the ***span*** of $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$, i.e., $span(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n) = \{\sum_{i=1}^{n} a_i\boldsymbol{v}_i : a_1, a_2, \cdots, a_n \in \mathbb{R}\}$.

Bases and Coordinates (2/2)

Definition (Basis)

A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ in a vector space $\mathcal{V}$ forms a ***basis*** iff: a) $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ are linear independent; b) $span(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n) = \mathcal{V}$.

- All bases of a space $\mathcal{V}$ must have the same number of vectors [Proof], and this number is called the ***dimension*** of $\mathcal{V}$, denoted as $dim(\mathcal{V})$
- Any $\boldsymbol{v} \in \mathcal{V}$ can be expressed as $\boldsymbol{v} = \sum_{i=1}^{n} a_i \boldsymbol{v}_i$, and the coefficients a_i, $i = 1, 2, \cdots, n$, are called the ***coordinates*** of $\boldsymbol{v}$ with respect to the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$
  - The coordinates of a vector change with the basis
- The ***natural basis*** for $\mathbb{R}^n$ is $\boldsymbol{e}_1 = [1, 0, \cdots, 0]^\top, \boldsymbol{e}_2 = [0, 1, \cdots, 0]^\top, \cdots, \boldsymbol{e}_n = [0, 0, \cdots, 1]^\top$
  - The coordinates of a vector with respect to this basis are identical to the components

Subspaces

Definition (Subspace)

A subset $\mathcal{U}$ of a vector space $\mathcal{V}$ is called a ***subspace*** if $\mathcal{U}$ is closed under the vector addition and scalar multiplication.

- That is, if $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{U}$, then $\boldsymbol{v} + \boldsymbol{w} \in \mathcal{U}$ and $a\boldsymbol{v} \in \mathcal{U}$ for all a
- Every subspace must contain $\mathbf{0}$

Subspaces

Definition (Subspace)

A subset $\mathcal{U}$ of a vector space $\mathcal{V}$ is called a ***subspace*** if $\mathcal{U}$ is closed under the vector addition and scalar multiplication.

- That is, if $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{U}$, then $\boldsymbol{v} + \boldsymbol{w} \in \mathcal{U}$ and $a\boldsymbol{v} \in \mathcal{U}$ for all a
- Every subspace must contain $\mathbf{0}$, as $\forall \boldsymbol{v} \in \mathcal{U}$, $-\boldsymbol{v}$ exists and $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0} \in \mathcal{U}$

Sum Spaces

Definition (Sum Space)

Let $\mathcal{W}$ and $\mathcal{U}$ be two subspaces of $\mathcal{V}$, the set $\{\boldsymbol{w}+\boldsymbol{u} : \boldsymbol{w} \in \mathcal{W}, \boldsymbol{u} \in \mathcal{U}\}$ is called the ***sum space*** of $\mathcal{W}$ and $\mathcal{U}$, denoted by $\mathcal{W}+\mathcal{U}$.

- $\mathcal{W}+\mathcal{U}$ is a subspace of $\mathcal{V}$ [Proof]
- $dim(\mathcal{W}+\mathcal{U}) = dim(\mathcal{W}) + dim(\mathcal{U}) - dim(\mathcal{W}\cap\mathcal{U})$
  - Let $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ be a basis for $\mathcal{W}\cap\mathcal{U}$, then we can find $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\}$ and $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ such that $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ and $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ are the bases for $\mathcal{W}$ and $\mathcal{U}$ respectively
  - We can see that $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m, \boldsymbol{u}_1, \cdots, \boldsymbol{u}_n, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ is a basis for $\mathcal{W}+\mathcal{U}$ [Proof]
  - Therefore, $dim(\mathcal{W}+\mathcal{U}) = m+n+k = (m+k)+(n+k)-k = dim(\mathcal{W}) + dim(\mathcal{U}) - dim(\mathcal{W}\cap\mathcal{U})$

Linear Transformation

Definition (Linear Transformation)

A function $\mathcal{L}:\mathcal{V}\to\mathcal{W}$, where $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, is called a ***linear transformation*** iff:
1) $\mathcal{L}(a\boldsymbol{v}) = a\mathcal{L}(\boldsymbol{v})$ for every $\boldsymbol{v}\in\mathcal{V}$ and $a\in\mathbb{R}$;
2) $\mathcal{L}(\boldsymbol{v}+\boldsymbol{w}) = \mathcal{L}(\boldsymbol{v})+\mathcal{L}(\boldsymbol{w})$ for every $\boldsymbol{v},\boldsymbol{w}\in\mathcal{V}$.

Definition (Range)

The ***range*** (or ***image***) of a linear transformation $\mathcal{L}:\mathcal{V}\to\mathcal{W}$ is $\{\mathcal{L}(\boldsymbol{v}):\boldsymbol{v}\in\mathcal{V}\}$, denoted as $\mathcal{R}(\mathcal{L})$ (or $im(\mathcal{L})$).

Definition (Nullspace)

The ***nullspace*** (or ***kernel***) of a linear transformation $\mathcal{L}:\mathcal{V}\to\mathcal{W}$ is $\{\boldsymbol{v}\in\mathcal{V}:\mathcal{L}(\boldsymbol{v})=\boldsymbol{0}\}$, denoted as $\mathcal{N}(\mathcal{L})$ (or $ker(\mathcal{L})$).

- $\mathcal{R}(\mathcal{L})$ and $\mathcal{N}(\mathcal{L})$ are subspaces of $\mathcal{W}$ and $\mathcal{V}$ respectively [Proof]

Dimension Theorem

Theorem

Let $\mathcal{L}:\mathcal{V}\to\mathcal{W}$ be a linear transformation, we have $dim(\mathcal{V}) = dim(\mathcal{R}(\mathcal{L})) + dim(\mathcal{N}(\mathcal{L}))$.

Proof.

Let $\{\boldsymbol{v}_i\}_i$ and $\{\boldsymbol{w}_j\}_j$ be the bases for $\mathcal{N}(\mathcal{L})$ and $\mathcal{R}(\mathcal{L})$ respectively[a]. There exists $\{\boldsymbol{u}_j\}_j \in \mathcal{V}$ such that $\mathcal{L}(\boldsymbol{u}_j) = \boldsymbol{w}_j$. We claim that the set $\{\boldsymbol{v}_i\}_i \cup \{\boldsymbol{u}_j\}_j$ forms a basis of $\mathcal{V}$.

We first prove that $span(\boldsymbol{v}_i, \boldsymbol{u}_j) = \mathcal{V}$. Given any $\boldsymbol{v} \in \mathcal{V}$, there exist scalars $\{y_j\}_j$ such that $\mathcal{L}(\boldsymbol{v}) = \sum_j y_j \boldsymbol{w}_j$. We have

$0 = \mathcal{L}(\boldsymbol{v}) - \sum_j y_j \boldsymbol{w}_j = \mathcal{L}(\boldsymbol{v}) - \sum_j y_j \mathcal{L}(\boldsymbol{u}_j) = \mathcal{L}(\boldsymbol{v} - \sum_j y_j \boldsymbol{u}_j)$. So $\boldsymbol{v} - \sum_j y_j \boldsymbol{u}_j \in \mathcal{N}(\mathcal{L})$, implying that there exists $\{\alpha_i\}_i$ such that $\boldsymbol{v} - \sum_j y_j \boldsymbol{u}_j = \sum_i \alpha_i \boldsymbol{v}_i$. Therefore, $\boldsymbol{v} = \sum_i \alpha_i \boldsymbol{v}_i + \sum_j y_j \boldsymbol{u}_j$.

Next, we prove that $\boldsymbol{v}_i, \boldsymbol{u}_j$ are linear independent. If $\sum_i \alpha_i \boldsymbol{v}_i + \sum_j y_j \boldsymbol{u}_j = \boldsymbol{0}$, we have

$\boldsymbol{0} = \mathcal{L}(\boldsymbol{0}) = \mathcal{L}(\sum_i \alpha_i \boldsymbol{v}_i + \sum_j y_j \boldsymbol{u}_j) = \mathcal{L}(\sum_i \alpha_i \boldsymbol{v}_i) + \mathcal{L}(\sum_j y_j \boldsymbol{u}_j) = \sum_j y_j \boldsymbol{w}_j$, implying that $y_j = 0$ for all j. Substitute y_i back to the equation $\sum_i \alpha_i \boldsymbol{v}_i + \sum_j y_j \boldsymbol{u}_j = \boldsymbol{0}$ we have $\sum_i \alpha_i \boldsymbol{v}_i = \boldsymbol{0}$, meaning $\alpha_i = 0$ for all i. □

[a] Apparently, $\boldsymbol{v}_i$ and $\boldsymbol{w}_j$ are distinct.

Matrix Representation (1/2)

- Given two bases $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ in $\mathbb{R}^n$ and $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m\}$ in $\mathbb{R}^m$, $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a matrix $\boldsymbol{A}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, such that for every $\boldsymbol{v} \in \mathcal{V}$ and $\boldsymbol{w} \in \mathcal{W}$, $\mathcal{L}(\boldsymbol{v}) = \boldsymbol{w}$, we have $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$, where $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^\top$ and $\boldsymbol{y} = [y_1, y_2, \cdots, y_m]^\top$ are coordinates of $\boldsymbol{v}$ and $\boldsymbol{w}$ respectively
  - By definition,

$$\begin{aligned}\mathcal{L}(\boldsymbol{v}) &= \mathcal{L}(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n) = x_1\mathcal{L}(\boldsymbol{v}_1) + \cdots + x_n\mathcal{L}(\boldsymbol{v}_n) \\ &= x_1(a_{11}\boldsymbol{w}_1 + \cdots + a_{m1}\boldsymbol{w}_m) + \cdots + x_n(a_{1n}\boldsymbol{w}_1 + \cdots + a_{mn}\boldsymbol{w}_m), \\ \mathcal{L}(\boldsymbol{v}) &= \boldsymbol{w} = y_1\boldsymbol{w}_1 + \cdots + y_m\boldsymbol{w}_m\end{aligned}$$

  - Comparing the coefficients of $\boldsymbol{w}_i$ we have

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Matrix Representation (2/2)

- Rewrite $\boldsymbol{A}$ as $[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$ where $\boldsymbol{a}_i$ denote columns, we have $\boldsymbol{y} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$
  - $\boldsymbol{y}$ is a linear combination of the columns of $\boldsymbol{A}$
- Why a function $\mathcal{L}$ satisfying $\mathcal{L}(a\boldsymbol{v}) = a\mathcal{L}(\boldsymbol{v})$ and $\mathcal{L}(\boldsymbol{v}+\boldsymbol{w}) = \mathcal{L}(\boldsymbol{v}) + \mathcal{L}(\boldsymbol{w})$ for every $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}, a \in \mathbb{R}$ is called "linear?"

Matrix Representation (2/2)

- Rewrite $\boldsymbol{A}$ as $[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$ where $\boldsymbol{a}_i$ denote columns, we have $\boldsymbol{y} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$
  - $\boldsymbol{y}$ is a linear combination of the columns of $\boldsymbol{A}$
- Why a function $\mathcal{L}$ satisfying $\mathcal{L}(a\boldsymbol{v}) = a\mathcal{L}(\boldsymbol{v})$ and $\mathcal{L}(\boldsymbol{v}+\boldsymbol{w}) = \mathcal{L}(\boldsymbol{v}) + \mathcal{L}(\boldsymbol{w})$ for every $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}, a \in \mathbb{R}$ is called "linear?"
  - We can see from the matrix representation that each y_j, $1 \leqslant j \leqslant m$, is mapped from a "linear function" f_j over $x_1, \cdots, x_n$, i.e., $f_j(x_1, \cdots, x_n) = a_{j1}x_1 + \cdots + a_{jn}x_n = y_j$

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Rank of a Matrix

Definition (Rank)

Given an $m \times n$ matrix $\boldsymbol{A}$ and let $\boldsymbol{a}_i$ be the ith column of $\boldsymbol{A}$. The number of linear independent columns of $\boldsymbol{A}$ is called the ***rank*** of $\boldsymbol{A}$, denoted as $rank(\boldsymbol{A})$.

- $rank(\boldsymbol{A}) = dim(span(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)) = dim(\mathcal{R}(\boldsymbol{A}))$
- $rank(\boldsymbol{A}) = rank(\boldsymbol{A}^\top)$ [Proof: Using the Dimension Theorem]
- $rank(\boldsymbol{A}+\boldsymbol{B}) \leqslant rank(\boldsymbol{A}) + rank(\boldsymbol{B})$ [Proof: $\mathcal{R}(\boldsymbol{A}+\boldsymbol{B}) \subseteq \mathcal{R}(\boldsymbol{A}) + \mathcal{R}(\boldsymbol{B})$, and $dim(\mathcal{R}(\boldsymbol{A}) + \mathcal{R}(\boldsymbol{B})) \leqslant dim(\mathcal{R}(\boldsymbol{A}) + dim(\mathcal{R}(\boldsymbol{B}))$]
- $rank(\boldsymbol{AB}) \leqslant \min\{rank(\boldsymbol{A}), rank(\boldsymbol{B})\}$ [Proof: $\mathcal{R}(\boldsymbol{AB}) \subseteq \mathcal{R}(\boldsymbol{A})$]
    - $rank(\boldsymbol{A}^\top \boldsymbol{A}) = rank(\boldsymbol{A})$

Column and Row Operations

- The rank of $\boldsymbol{A}$ is invariant under the column (resp. row) operations [Proof]:
  - Multiplying columns (resp. rows) of $\boldsymbol{A}$ by nonzero scalars
  - Interchanging the columns (resp. rows)
  - Adding to a given column (resp. row) a linear combination of other columns (resp. rows)
- Denote $\boldsymbol{A} \overset{c}{\sim} \boldsymbol{B}$ and $\boldsymbol{A} \overset{r}{\sim} \boldsymbol{B}$ respectively if we can obtain $\boldsymbol{B}$ by performing the column and row operations over $\boldsymbol{A}$
- If $\boldsymbol{A} \overset{c}{\sim} \boldsymbol{B}$ or $\boldsymbol{A} \overset{r}{\sim} \boldsymbol{B}$, then $rank(\boldsymbol{A}) = rank(\boldsymbol{B})$
  - E.g., $[a,b,c]^\top [a,b,c] \overset{r}{\sim} \begin{bmatrix} a & b & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and has rank 1

Trace

Definition (Trace)

Given an $n \times n$ square matrix $\boldsymbol{A}$, the ***trace*** of $\boldsymbol{A}$ is defined as $tr(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,i}$.

- $tr(\boldsymbol{A}+\boldsymbol{B}) = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$, and $tr(\boldsymbol{A}) = tr(\boldsymbol{A}^\top)$ [Proof]
- $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$ [Proof]
  - $\boldsymbol{A}$ and $\boldsymbol{B}$ need not be square
  - In particular, $tr(\boldsymbol{x}^\top \boldsymbol{x}) = tr(\boldsymbol{x}\boldsymbol{x}^\top)$
- Cyclic property: $tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA})$ [Proof]
  - Generally, $tr(\boldsymbol{CBA}) \neq tr(\boldsymbol{ABC})$, unless both $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are symmetric (i.e., equal to their transpose): $tr(\boldsymbol{ABC}) = tr(\boldsymbol{A}^\top \boldsymbol{B}^\top \boldsymbol{C}^\top) = tr((\boldsymbol{CBA})^\top) = tr(\boldsymbol{CBA})$

Determinant (1/2)

Definition (Determinant)

Given an $n \times n$ square matrix $\boldsymbol{A}$, where $\boldsymbol{A} = [\boldsymbol{a}_i, \cdots, \boldsymbol{a}_n]$, there exists a unique function $det : \mathbb{R}^{n \times n} \to \mathbb{R}$, satisfying the properties:

a) $det(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}, \alpha \boldsymbol{a}_k^{(1)} + \beta \boldsymbol{a}_k^{(2)}, \boldsymbol{a}_{k+1}, \cdots, \boldsymbol{a}_n) =$
$\alpha det(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_k^{(1)}, \boldsymbol{a}_{k+1}, \cdots, \boldsymbol{a}_n) +$
$\beta det(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_k^{(2)}, \boldsymbol{a}_{k+1}, \cdots, \boldsymbol{a}_n)$, $\forall \alpha, \beta \in \mathbb{R}$;

b) $det(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_i, \cdots, \boldsymbol{a}_j, \cdots, \boldsymbol{a}_n) = 0$ if $\boldsymbol{a}_i = \boldsymbol{a}_j$ for some i and j;

c) $det(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n) = 1$.

We call $det(\boldsymbol{A})$ the ***determinant*** of $\boldsymbol{A}$.

- Let $\boldsymbol{I}_n = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n]$ be an ***identity matrix***, we have $det(\boldsymbol{I}_n) = 1$
- $det(\boldsymbol{A})$ changes its sign if we interchanges the columns of $\boldsymbol{A}$ [Proof]

Determinant (2/2)

- The unique function $det : \mathbb{R}^{n\times n} \to \mathbb{R}$ can be written as

$$det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+1} a_{1k} det(\boldsymbol{A}_{1k}),$$

where $\boldsymbol{A}_{ij}$ is the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and jth column [Proof]
- The determinant of $\boldsymbol{A}$ can be also regarded as the ***sign volume of the image of the unit cube***

Theorem

Given any $c \in \mathbb{R}$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$, we have a) $det(c\boldsymbol{A}) = c^n det(\boldsymbol{A})$; b) $det(\boldsymbol{A}^\top) = det(\boldsymbol{A})$; c) $det(\boldsymbol{AB}) = det(\boldsymbol{A})det(\boldsymbol{B})$.

- Can be proved by either tedious calculation or the signed volume interpretation

Linear Equations (1/2)

- Given $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$, and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{Ax} = \boldsymbol{y}$ represents a system of linear equations as follows:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = y_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = y_m \end{cases}$$

Theorem

*Let $[\boldsymbol{A}, \boldsymbol{y}] = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n, \boldsymbol{y}]$ be the **augmented matrix**, the system of linear equations $\boldsymbol{Ax} = \boldsymbol{y}$ has a solution iff $rank(\boldsymbol{A}) = rank([\boldsymbol{A}, \boldsymbol{y}])$.*

Linear Equations (2/2)

Proof.

$\Rightarrow$: $\boldsymbol{y}$ is a linear combination of the columns of $\boldsymbol{A}$, so $rank([\boldsymbol{A}, \boldsymbol{y}]) = dim(span(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n, \boldsymbol{y})) = dim(span(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)) = rank(\boldsymbol{A})$.
$\Leftarrow$: Let $rank(\boldsymbol{A}) = rank([\boldsymbol{A}, \boldsymbol{y}]) = r$ and $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$ be the linear independent columns of both $\boldsymbol{A}$ and $[\boldsymbol{A}, \boldsymbol{y}]$. Since $\boldsymbol{y}$ is not one of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$, it is their linear combination; that is, there exists $x_1, \cdots, x_r$ such that $\boldsymbol{y} = x_1\boldsymbol{a}_1 + \cdots + x_r\boldsymbol{a}_r$. So $\boldsymbol{x} = [x_1, \cdots, x_r]^\top$ is the solution. □

Definition (Linear Variety)

The set $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}$ is called the ***linear variety*** for $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{y} \in \mathbb{R}^m$.

- If $\boldsymbol{x}_0$ is a solution, then for all $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$, $\boldsymbol{x}_0 + \boldsymbol{x}$ is also a solution
- Is linear variety a subspace of $\mathbb{R}^n$?

Linear Equations (2/2)

Proof.

$\Rightarrow$: $\boldsymbol{y}$ is a linear combination of the columns of $\boldsymbol{A}$, so $rank([\boldsymbol{A}, \boldsymbol{y}]) = dim(span(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n, \boldsymbol{y})) = dim(span(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)) = rank(\boldsymbol{A})$.
$\Leftarrow$: Let $rank(\boldsymbol{A}) = rank([\boldsymbol{A}, \boldsymbol{y}]) = r$ and $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$ be the linear independent columns of both $\boldsymbol{A}$ and $[\boldsymbol{A}, \boldsymbol{y}]$. Since $\boldsymbol{y}$ is not one of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$, it is their linear combination; that is, there exists $x_1, \cdots, x_r$ such that $\boldsymbol{y} = x_1\boldsymbol{a}_1 + \cdots + x_r\boldsymbol{a}_r$. So $\boldsymbol{x} = [x_1, \cdots, x_r]^\top$ is the solution. $\square$

Definition (Linear Variety)

The set $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}$ is called the ***linear variety*** for $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{y} \in \mathbb{R}^m$.

- If $\boldsymbol{x}_0$ is a solution, then for all $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$, $\boldsymbol{x}_0 + \boldsymbol{x}$ is also a solution
- Is linear variety a subspace of $\mathbb{R}^n$? No, as $\boldsymbol{0}$ is not included
- However, we still say that the linear variety has dimension r if $dim(\mathcal{N}(\boldsymbol{A})) = r$

Cramer's Rule

Theorem (Cramer's Rule)

Given a square, invertible matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the solution to a system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ can be obtained by $x_i = det(\boldsymbol{A}_i)/det(\boldsymbol{A})$ for $i = 1, \cdots, n$, where $\boldsymbol{A}_i$ is the matrix formed by replacing the ith column of $\boldsymbol{A}$ by the column vector $\boldsymbol{y}$.

- The proof is easy [Proof]

Invertibility

Definition (Nonsingular Matrix)

A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is ***nonsingular*** (or ***invertible***) if there exists another matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}_n$. We call $\boldsymbol{B}$ the ***inverse*** of $\boldsymbol{A}$ and denote it as $\boldsymbol{A}^{-1}$.

- $(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$ and $det(\boldsymbol{A}^{-1}) = det(\boldsymbol{A})^{-1}$ [Proof]

Theorem

Given $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the following conditions are equivalent:
a) $\boldsymbol{A}$ is invertible;
b) There exists a unique solution $\boldsymbol{x}$ satisfying $\boldsymbol{Ax} = \boldsymbol{y}$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$;
c) $\mathcal{N}(\boldsymbol{A}) = \boldsymbol{0}$ (trivial kernel);
d) The columns are linearly independent (i.e., $rank(\boldsymbol{A}) = n$);
e) $det(\boldsymbol{A}) \neq 0$;
f) $\boldsymbol{A}^\top$ is invertible;
g) The rows of $\boldsymbol{A}$ are linearly independent;
h) All of the eigenvalues of $\boldsymbol{A}$ are nonzero (explained later).

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Change of Basis

- Recall that given the bases of domain and range, a linear transformation can be represented by a matrix
  - What's the relation between matrices obtained from different bases?

Definition (Change of Basis Matrix)

Consider two bases $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ and $\{\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n\}$ for $\mathbb{R}^n$ and a vector $\boldsymbol{v} \in \mathbb{R}^n$. There are two sets of coordinates $\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$, $1 \leqslant i \leqslant n$, such that $[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n][x_1, \cdots, x_n]^\top = \boldsymbol{v} = [\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n][x'_1, \cdots, x'_n]^\top$. We call $[\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n]^{-1}[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n]$ the ***change of basis matrix*** (or ***transition matrix***) from $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ to $\{\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n\}$.

Similar Matrices

Definition (Similar Matrices)

Two square matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ are ***similar*** if there exists nonsingular matrices $\boldsymbol{C} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{A} = \boldsymbol{C}^{-1}\boldsymbol{B}\boldsymbol{C}$.

- If $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar, then $tr(\boldsymbol{A}) = tr(\boldsymbol{B})$ and $det(\boldsymbol{A}) = det(\boldsymbol{B})$ [Proof]
- Let $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ and $\{\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n\}$ be two bases of domain, $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\}$ and $\{\boldsymbol{w}'_1, \cdots, \boldsymbol{w}'_m\}$ be two bases of range, and $\boldsymbol{S}$ and $\boldsymbol{T}$ be the change of basis matrices from $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ to $\{\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n\}$ and $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\}$ to $\{\boldsymbol{w}'_1, \cdots, \boldsymbol{w}'_m\}$ respectively. We have the following relations:

$$
\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{\boldsymbol{A}} & \mathbb{R}^m \\
\boldsymbol{S} \downarrow & & \downarrow \boldsymbol{T} \\
\mathbb{R}^n & \xrightarrow[\boldsymbol{B}]{} & \mathbb{R}^m
\end{array}
$$

- Similar matrices correspond to the same linear transform with respect to different bases

Eigen Decomposition

- Why do we need eigenvalues and eigenvectors?
  - Given a linear transformation, we want to find a basis (if existing) such that the corresponding matrix representation $\boldsymbol{D}$ is diagonal
  - So, given coordinates $\boldsymbol{x} \in \mathbb{R}^n$ with respect to this basis, the effect of the transformation is just a scaling to each coordinate, as $\boldsymbol{Dx} = [d_{11}x_1, \cdots, d_{nn}x_n]^\top$
  - An example application to compression: We can drop small d_{ii}s without changing the original transformation too much

Eigenvalues and Eigenvectors (1/3)

Definition (Eigenvalues and Eigenvectors)

Given $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, a nonzero vector $\boldsymbol{x}$ satisfying $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$, where λ is a scalar (possibly complex), is called the ***eigenvector*** of $\boldsymbol{A}$, and λ is called the ***eigenvalue***.

- $\boldsymbol{x}$ is an eigenvector iff the matrix $\lambda \boldsymbol{I} - \boldsymbol{A}$ is singular, as $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \Rightarrow \lambda \boldsymbol{x} - \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow (\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ and $\lambda \boldsymbol{I} - \boldsymbol{A}$ has nontrivial kernel (note $\boldsymbol{x}$ is nonzero by definition)
- We have $0 = det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$; that is, the ***characteristic polynomial*** of $\boldsymbol{A}$ equals 0
- The eigenvalues are the roots (possibly with multiplicity) of the above equation
- For each eigenvalue λ_i, we can obtain its corresponding eigenvectors by solving $(\lambda_i \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$

Multiplicities

- The eigenvector (i.e., solution to $(\lambda_i \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$) of an eigenvalue λ_i is not unique
  - If $\boldsymbol{A}\boldsymbol{x} = \lambda_i \boldsymbol{x}$, so does $\boldsymbol{A}(c\boldsymbol{x}) = \lambda_i(c\boldsymbol{x})$ for any $c \in \mathbb{R}$
  - $\mathcal{N}(\lambda_i \boldsymbol{I} - \boldsymbol{A})$, called the ***eigenspace*** of λ_i, has dimension at least 1
- ***Algebraic multiplicity*** of an eigenvalue λ_i is the multiplicity of the corresponding root of the characteristic polynomial
- ***Geometric multiplicity*** of λ_i is the dimension of $\mathcal{N}(\lambda_i \boldsymbol{I} - \boldsymbol{A})$, the number of linear independent eigenvectors we solve from $(\lambda_i \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$
  - Geometric multiplicity must be less than or equal to the algebraic multiplicity
  - We may not be able to find n linear independent eigenvectors for a matrix

Eigenvalues and Eigenvectors (2/3)

Theorem

If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ has n linear independent eigenvectors $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$, then $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ form a basis of $\mathbb{R}^n$.

- Given coordinates $\boldsymbol{x} \in \mathbb{R}^n$ with respect to this basis, the effect of the transformation is just a scaling to each coordinate, as $\boldsymbol{A}(x_1\boldsymbol{u}_1 + \cdots + x_n\boldsymbol{u}_n) = x_1\boldsymbol{A}(\boldsymbol{u}_1) + \cdots + x_n\boldsymbol{A}(\boldsymbol{u}_n) = x_1\lambda_1\boldsymbol{u}_1 + \cdots + x_n\lambda_n\boldsymbol{u}_n$
- Under this basis, the transformation can be represented by a diagonal matrix $\boldsymbol{D}$, where $d_{ii} = \lambda_i$ (counting the multiplicity)
- We say $\boldsymbol{A}$ is ***diagonalizable*** if there exists a basis such that $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{D}\boldsymbol{T} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{-1}$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n]$ and $\boldsymbol{T} = \boldsymbol{U}^{-1}[\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n]$
- $\boldsymbol{T}$ is the change of basis matrix from the natural basis to $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$:

$$
\begin{array}{rcccl}
 & \mathbb{R}^n & \xrightarrow{\boldsymbol{A}} & \mathbb{R}^n & \\
\boldsymbol{T} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n]^{-1} & \downarrow & & \downarrow & \boldsymbol{T} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n]^{-1} \\
 & \mathbb{R}^n & \xrightarrow[\boldsymbol{D}]{} & \mathbb{R}^n &
\end{array}
$$

Eigenvalues and Eigenvectors (3/3)

- $tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$ and $det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$ [Proof]
- If two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ are similar, then their characteristic polynomials (and eigenvalues) are equal, as $det(\lambda \boldsymbol{I} - \boldsymbol{A}) = det(\lambda \boldsymbol{I} - \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}) = det(\lambda \boldsymbol{T}^{-1}\boldsymbol{T} - \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}) = det(\boldsymbol{T}^{-1}) det(\lambda \boldsymbol{I} - \boldsymbol{B}) det(\boldsymbol{T}) = det(\lambda \boldsymbol{I} - \boldsymbol{B})$

Theorem

A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is invertible iff all eigenvalues of $\boldsymbol{A}$ are nonzero.

- The above theorem dose ***not*** imply any consequence between the diagonalizability and invertibility of a matrix
  - E.g., $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is diagonalizable but not invertible, yet $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ is invertible but not diagonalizable

Outline

1. **Linear Algebra**
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. **Geometry**
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. **Point Set Topology****
   - Topological Spaces
   - Manifolds

Inner Products

Definition (Inner Product)

A function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ is called the ***inner product*** if it satisfies:
a) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0, \forall \boldsymbol{x} \in \mathcal{V}$ and the equality holds iff $\boldsymbol{x} = \boldsymbol{0}$ (positivity);
b) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ (conjugate symmetry);
c) $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle, \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V}$ (additivity);
d) $\langle r\boldsymbol{x}, \boldsymbol{y} \rangle = r \langle \boldsymbol{x}, \boldsymbol{y} \rangle, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}, r \in \mathbb{C}$ (homogeneity).

- Note we have $\langle \boldsymbol{x}, r\boldsymbol{y} \rangle = \bar{r} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ based on properties b) and d)
- A common example is the ***Euclidean inner product***: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^\top \boldsymbol{y}$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$
- Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be ***orthogonal*** if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$
- The ***Euclidean norm*** of $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$
- A vector space with an inner product/norm defined is called the ***inner product/normed space*** respectively

Cauchy-Schwarz Inequality

Theorem (Cauchy-Schwarz Inequality)

For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*, we have* $|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\| \|\mathbf{y}\|$ *and the equality holds iff* $\mathbf{x} = \alpha \mathbf{y}$ *for some* $\alpha \in \mathbb{R}$*.*

Proof.

The proof is obvious when $\mathbf{x} = 0$ or $\mathbf{y} = 0$. Otherwise, consider the case where $\mathbf{x}$ and $\mathbf{y}$ are unit vectors; that is, $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. Then
$0 \leqslant \|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = \|\mathbf{x}\|^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 = 2 - 2\langle \mathbf{x}, \mathbf{y} \rangle$,
implying $\langle \mathbf{x}, \mathbf{y} \rangle \leqslant 1$. The equality holds iff $\mathbf{x} = \mathbf{y}$. Similarly, by
$0 \leqslant \|\mathbf{x} + \mathbf{y}\|^2$ we have $\langle \mathbf{x}, \mathbf{y} \rangle \geqslant -1$ and the equality holds iff $\mathbf{x} = -\mathbf{y}$. For any nonzero vectors $\mathbf{x}$ and $\mathbf{y}$, we have
$-1 \leqslant \langle \mathbf{x}/\|\mathbf{x}\|, \mathbf{y}/\|\mathbf{y}\| \rangle \leqslant 1 \Rightarrow |\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\| \|\mathbf{y}\|$ and the equality holds iff
$\mathbf{x}/\|\mathbf{x}\| = \pm \mathbf{y}/\|\mathbf{y}\|$; that is, $\mathbf{x} = \alpha \mathbf{y}$ for some $\alpha \in \mathbb{R}$. □

- Since $-1 \leqslant \langle \mathbf{x}, \mathbf{y} \rangle / \|\mathbf{x}\| \|\mathbf{y}\| \leqslant 1$, we can define the ***included angle*** θ of $\mathbf{x}$ and $\mathbf{y}$ by $\cos\theta = \langle \mathbf{x}, \mathbf{y} \rangle / \|\mathbf{x}\| \|\mathbf{y}\|$

Norms

Definition (Vector Norm)

A function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ is called the ***vector norm*** if it satisfies:
a) $\|\boldsymbol{x}\| \geqslant 0, \forall \boldsymbol{x} \in \mathcal{V}$ and the equality holds iff $\boldsymbol{x} = \boldsymbol{0}$ (positivity);
b) $\|r\boldsymbol{x}\| = |r| \|\boldsymbol{x}\|, \forall \boldsymbol{x} \in \mathcal{V}, r \in \mathbb{R}$ (homogeneity);
c) $\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ (triangle inequality).

- The Euclidean norm is a vector norm [Proof]
- We can define the ***p-norm*** directly without going through the inner product first: $\|\boldsymbol{x}\|_p = \begin{cases} \left(\sum_i |x_i|^p\right)^{1/p} & 1 \leqslant p < \infty \\ \max\{|x_i|\}_i & p = \infty \end{cases}$
  - Euclidean norm is also known as the 2-norm

Symmetric and Hermitian Matrices (1/2)

- A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is ***symmetric*** if $\boldsymbol{A}^\top = \boldsymbol{A}$; and ***antisymmetric*** if $\boldsymbol{A}^\top = -\boldsymbol{A}$
- A matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is ***Hermitian*** if $\boldsymbol{A} = \boldsymbol{A}^*$ (conjugate transpose); and ***antihermitian*** if $\boldsymbol{A}^* = -\boldsymbol{A}$

Theorem

All eigenvalues of a real symmetric matrix are real.

Proof.

Let $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$, where $\boldsymbol{x} \neq \boldsymbol{0}$. We have $\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \lambda \boldsymbol{x}, \boldsymbol{x} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{x} \rangle$. On the other hand, $\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{A}^T \boldsymbol{x} \rangle = \overline{\lambda} \langle \boldsymbol{x}, \boldsymbol{x} \rangle$. This implies $\lambda \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \overline{\lambda} \langle \boldsymbol{x}, \boldsymbol{x} \rangle \Rightarrow (\lambda - \overline{\lambda}) \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{0}$. Since $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$ for any $\boldsymbol{x} \neq \boldsymbol{0}$, $\lambda - \overline{\lambda}$ must be 0; that is, λ is real. □

Symmetric and Hermitian Matrices (2/2)

Theorem

Any real symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ has n eigenvectors that are mutually orthogonal.

Proof.

Here we only prove a special case where the n eigenvalues are distinct. Suppose $\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $\boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$, where $\lambda_1 \neq \lambda_2$. Then $\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \lambda_1 \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \lambda_1 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$. However, $\left\langle \boldsymbol{x}_1, \boldsymbol{A}^T \boldsymbol{x}_2 \right\rangle = \langle \boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \lambda_2 \boldsymbol{x}_2 \rangle = \lambda_2 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$. Therefore we have $\lambda_1 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \lambda_2 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$. Since $\lambda_1 \neq \lambda_2$, $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = 0$. □

- Real symmetric matrices are always diagonalizable
- $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n]$ and $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{A}$
- Since the columns of $\boldsymbol{U}$ are orthogonal with each other, $\boldsymbol{U}^\top \boldsymbol{U}$ is diagonal
- By picking the eigenvectors of unit norm, we have $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$, and therefore $\boldsymbol{U}^{-1} = \boldsymbol{U}^T$

Orthogonal and Unitary Matrices

- A matrix $\boldsymbol{U}$ having inverse as $\boldsymbol{U}^\top$ is called the ***orthogonal matrix***
- If $\boldsymbol{U} \in \mathbb{C}^{n \times n}$ and $\boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I}$, then $\boldsymbol{U}$ is called the ***unitary matrix***
- Unitary (and orthogonal) matrices are always invertible and diagonalizable [Proof]
- Given any orthogonal (or unitary) matrix $\boldsymbol{U}$, we have $\|\boldsymbol{U}\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{x}} = \|\boldsymbol{x}\|_2$
  - As a linear transformation, $\boldsymbol{U}$ preserves distance so the "shape" of a set of vectors in the domain can be preserved in the range
  - Examples?

Orthogonal and Unitary Matrices

- A matrix $\boldsymbol{U}$ having inverse as $\boldsymbol{U}^\top$ is called the ***orthogonal matrix***
- If $\boldsymbol{U} \in \mathbb{C}^{n\times n}$ and $\boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I}$, then $\boldsymbol{U}$ is called the ***unitary matrix***
- Unitary (and orthogonal) matrices are always invertible and diagonalizable [Proof]
- Given any orthogonal (or unitary) matrix $\boldsymbol{U}$, we have $\|\boldsymbol{U}\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{x}} = \|\boldsymbol{x}\|_2$
  - As a linear transformation, $\boldsymbol{U}$ preserves distance so the "shape" of a set of vectors in the domain can be preserved in the range
  - Examples? Rotation, reflection etc.
  - On the other hand, the Euclidean norm is ***unitarily invariant***

Orthogonal Projection (1/3)

Definition (Orthogonal Complement)

Given a subspace $\mathcal{V}$ of $\mathbb{R}^n$. The ***orthogonal complement*** of $\mathcal{V}$ is defined by $\mathcal{V}^\perp = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{v}, \boldsymbol{x} \rangle = 0, \forall \boldsymbol{v} \in \mathcal{V}\}$.

Definition (Orthogonal Projector)

A matrix $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is called a ***orthogonal projector*** onto $\mathcal{V}$ if $\boldsymbol{P}\boldsymbol{x} \in \mathcal{V}$ and $\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x} \in \mathcal{V}^\perp$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Orthogonal Projection (2/3)

Theorem

Given a matrix $\boldsymbol{A}$*, we have* $\mathcal{R}(\boldsymbol{A})^{\perp} = \mathcal{N}(\boldsymbol{A}^{\top})$ *and* $\mathcal{N}(\boldsymbol{A})^{\perp} = \mathcal{R}(\boldsymbol{A}^{\top})$*.*

Proof.

$\subseteq$: Suppose that $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{A})^{\perp}$, we have $(\boldsymbol{A}\boldsymbol{y})^{\top}\boldsymbol{x} = \boldsymbol{y}^{\top}(\boldsymbol{A}^{\top}\boldsymbol{x}) = 0$ for all $\boldsymbol{y} \in \mathbb{R}^{n}$, implying that $\boldsymbol{A}^{\top}\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}^{\top})$. So $\mathcal{R}(\boldsymbol{A})^{\perp} \subseteq \mathcal{N}(\boldsymbol{A}^{\top})$.
$\supseteq$: If now $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}^{\top})$, then $\boldsymbol{y}^{\top}(\boldsymbol{A}^{\top}\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{y})^{\top}\boldsymbol{x} = 0$ for all $\boldsymbol{y} \in \mathbb{R}^{n}$, implying $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{A})^{\perp}$ and $\mathcal{R}(\boldsymbol{A})^{\perp} \supseteq \mathcal{N}(\boldsymbol{A}^{\top})$.
Thus $\mathcal{R}(\boldsymbol{A})^{\perp} = \mathcal{N}(\boldsymbol{A}^{\top})$. □

- The proof of $\mathcal{N}(\boldsymbol{A})^{\perp} = \mathcal{R}(\boldsymbol{A}^{\top})$ follows from the above and the fact that $(\mathcal{V}^{\perp})^{\perp} = \mathcal{V}$ [Proof].

Orthogonal Projection (3/3)

Theorem

A matrix $\boldsymbol{P}$ *is an orthogonal projector (on to* $\mathcal{R}(\boldsymbol{P})$*) iff* $\boldsymbol{P}^2 = \boldsymbol{P} = \boldsymbol{P}^\top$*.*

Proof.

$\Rightarrow$: Since $\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x} \in \mathcal{R}(\boldsymbol{P})^\perp$ for all $\boldsymbol{x} \in \mathbb{R}^n$, we have $\mathcal{R}(\boldsymbol{I} - \boldsymbol{P}) \subseteq \mathcal{R}(\boldsymbol{P})^\perp$. But from the previous theorem $\mathcal{R}(\boldsymbol{P})^\perp = \mathcal{N}(\boldsymbol{P}^\top)$. This implies that $\mathcal{R}(\boldsymbol{I} - \boldsymbol{P}) \subseteq \mathcal{N}(\boldsymbol{P}^\top)$ and therefore $\boldsymbol{P}^\top(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y} = \boldsymbol{0}$ for all $\boldsymbol{y} \in \mathbb{R}^n$. We have $\boldsymbol{P}^\top(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{O} \Rightarrow \boldsymbol{P}^\top = \boldsymbol{P}^\top\boldsymbol{P}$. It is easy to verify that $\boldsymbol{P} = \boldsymbol{P}^\top = \boldsymbol{P}^2$.

$\Leftarrow$: For any $\boldsymbol{x} \in \mathbb{R}^n$ we have $(\boldsymbol{P}\boldsymbol{y})^\top(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} = \boldsymbol{y}^\top\boldsymbol{P}^\top(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} = \boldsymbol{y}^\top\boldsymbol{O}\boldsymbol{x} = 0$ for all $\boldsymbol{y} \in \mathbb{R}^n$. Thus, $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} \in \mathcal{R}(\boldsymbol{P})^\perp$ and $\boldsymbol{P}$ is an orthogonal projector. □

Normal Equations (1/2)

- Linear varity includes all solutions of $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$
  - What if $\boldsymbol{Ax} = \boldsymbol{b}$ has no solution (that is, $\boldsymbol{b}$ is not a linear combination of the columns of $\boldsymbol{A}$, or $\boldsymbol{b} \notin \mathcal{R}(\boldsymbol{A})$)?

Normal Equations (1/2)

- Linear varity includes all solutions of $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$
  - What if $\boldsymbol{Ax} = \boldsymbol{b}$ has no solution (that is, $\boldsymbol{b}$ is not a linear combination of the columns of $\boldsymbol{A}$, or $\boldsymbol{b} \notin \mathcal{R}(\boldsymbol{A})$)?
- We can instead find $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{A})$ which is closest to $\boldsymbol{b}$

Theorem

Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, finding $\boldsymbol{x} \in \mathbb{R}^n$ minimizing $\|\boldsymbol{Ax} - \boldsymbol{b}\|$ is equivalent to solving $\boldsymbol{A}^\top \boldsymbol{Ax} = \boldsymbol{A}^\top \boldsymbol{b}$.

Proof.

We can see that $\|\boldsymbol{Ax} - \boldsymbol{b}\|$ is minimized when the $\boldsymbol{Ax} - \boldsymbol{b}$ is normal to $\mathcal{R}(\boldsymbol{A})$. That is, $\langle \boldsymbol{Ax} - \boldsymbol{b}, \boldsymbol{w} \rangle = 0, \forall \boldsymbol{w} \in \mathcal{R}(\boldsymbol{A}) \Leftrightarrow \langle \boldsymbol{Ax} - \boldsymbol{b}, \boldsymbol{Ay} \rangle = 0, \forall \boldsymbol{y} \in \mathbb{R}^n \Leftrightarrow (\boldsymbol{Ay})^\top (\boldsymbol{Ax} - \boldsymbol{b}) = 0, \forall \boldsymbol{y} \in \mathbb{R}^n \Leftrightarrow \boldsymbol{y}^\top \boldsymbol{A}^\top \boldsymbol{Ax} - \boldsymbol{y}^\top \boldsymbol{A}^\top \boldsymbol{b} = 0, \forall \boldsymbol{y} \in \mathbb{R}^n \Leftrightarrow \boldsymbol{y}^\top (\boldsymbol{A}^\top \boldsymbol{Ax} - \boldsymbol{A}^\top \boldsymbol{b}) = 0, \forall \boldsymbol{y} \in \mathbb{R}^n \Leftrightarrow \boldsymbol{A}^\top \boldsymbol{Ax} - \boldsymbol{A}^\top \boldsymbol{b} = 0 \Leftrightarrow \boldsymbol{A}^\top \boldsymbol{Ax} = \boldsymbol{A}^\top \boldsymbol{b}$. □

Normal Equations (2/2)

- $\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{b}$ is called the ***normal equation*** (as $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$ is normal to $\mathcal{R}(\boldsymbol{A})$) and must have at least one solution
  - $\boldsymbol{A}^\top \boldsymbol{b} \in \mathcal{R}(\boldsymbol{A}^\top)$
  - Since $\mathcal{R}(\boldsymbol{A}^\top \boldsymbol{A}) \subseteq \mathcal{R}(\boldsymbol{A}^\top)$ and $rank(\boldsymbol{A}^\top \boldsymbol{A}) = rank(\boldsymbol{A}^\top)$, we have $\mathcal{R}(\boldsymbol{A}^\top \boldsymbol{A}) = \mathcal{R}(\boldsymbol{A}^\top)$
  - That is, $\boldsymbol{A}^\top \boldsymbol{b} \in \mathcal{R}(\boldsymbol{A}^\top \boldsymbol{A})$
- $\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{b}$ has exactly one solution iff $\boldsymbol{A}^\top \boldsymbol{A}$ is invertible
  - $\boldsymbol{A}^\top \boldsymbol{A}$ is symmetric, therefore diagonalizable
  - $\boldsymbol{A}^\top \boldsymbol{A}$ is invertible iff all its eigenvalues are nonzero

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - **Positive Definite Matrices and Quadratic Forms****
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Before We Start...

Caution!

This subsection requires the knowledge of matrix calculus.

Positive Definite Matrices (1/2)

Definition (Definite Matrices)

A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is called ***positive definite*** (resp., positive semidefinite/negative definitive/negative semidefinite) iff for any $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$, we have $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ (resp., $\geqslant 0 / < 0 / \leqslant 0$)

- There is no loss of generality if we assume $\boldsymbol{A}$ is symmetric
  - As $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top (\frac{1}{2}\boldsymbol{A} + \frac{1}{2}\boldsymbol{A}^\top)\boldsymbol{x}$ and the matrix $\frac{1}{2}\boldsymbol{A} + \frac{1}{2}\boldsymbol{A}^\top$ is always symmetric [Proof]

Theorem

A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is positive definite (or semidefinite) iff all eigenvalues of $\boldsymbol{A}$ are positive (or nonnegative).

Positive Definite Matrices (2/2)

Proof.

Let $\boldsymbol{T}$ be an orthogonal matrix whose column are eigenvectors of $\boldsymbol{A}$. For any matrix, let $\boldsymbol{y} = \boldsymbol{T}^{-1}\boldsymbol{x} = \boldsymbol{T}^\top \boldsymbol{x}$. We have $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}^\top \boldsymbol{T}^\top \boldsymbol{A}\boldsymbol{T}\boldsymbol{y} = \sum_{i=1}^{n} \lambda_i y_i^2$, and the proof follows. □

- What does positive definite mean anyway?
- Before we start, define the ***graph*** of a function $f : \mathcal{V} \to \mathbb{R}$, $\mathcal{V} \subseteq \mathbb{R}^n$, to be the set $\{[\boldsymbol{x}^\top, f(\boldsymbol{x})]^\top : \boldsymbol{x} \in \mathcal{V}\}$

Principle Minors (1/2)

- A ***minor*** of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the determinant of a matrix obtained by deleting some row and column of $\boldsymbol{A}$
- The ***principle minors*** of $\boldsymbol{A}$ are $det(\boldsymbol{A})$ and $n-1$ minors obtained by successively deleting some row and column of $\boldsymbol{A}$
- The ***leading principle minors*** of $\boldsymbol{A}$ are $det(\boldsymbol{A})$ and $n-1$ minors obtained by successively deleting the last row and column of $\boldsymbol{A}$

Principle Minors (2/2)

- There is a simple way to check if a matrix is positive definite:

Theorem

A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *is positive definite iff its leading principle minors are positive.*

Proof.

Since $\boldsymbol{A}$ is symmetric, it is diagonalizable. We have
$det(\boldsymbol{A}) = det(\boldsymbol{T}^{-1}\boldsymbol{D}\boldsymbol{T}) = det(\boldsymbol{T})^{-1}det(\boldsymbol{D})det(\boldsymbol{T}) = det(\boldsymbol{D}) = \prod_{i=1}^{n} \lambda_i$
and any minor of $\boldsymbol{A}$ equals to the multiplication of remaining eigenvalues.
Therefore, $\boldsymbol{A}$ is positive definite $\Leftrightarrow \lambda_i > 0$ for all $1 \leqslant i \leqslant n \Leftrightarrow$ the leading principle minors of $\boldsymbol{A}$ are positive. □

- Direction $\Leftarrow$ is ***not*** true in the semidefinite cases: $\boldsymbol{A}$ is positive semidefinite iff all principle minors (not only the leading principle minors) are nonnegative

Quadratic Forms (1/2)

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is ***quadratic*** iff it can be written as: $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} + c$ (the scalar coefficients do not matter)
  - $\boldsymbol{A}$ is symmetric, and f is said to be a ***quadratic form*** if $\boldsymbol{b} = \boldsymbol{0}$ and $c = 0$
- Our best intuition of a definite matrix is the shape of its corresponding quadratic form in a graph:

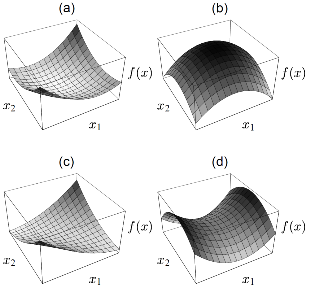


Figure : Quodratic form for a) positive definite; b) negative definite; c) positive definite but singular; d) indefinite matrix.

Quadratic Forms (2/2)

- Why $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} + c$ is a paraboloid when $\boldsymbol{A}$ is positive definite?
- Since $\boldsymbol{A}$ is symmetric, we have $f'(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{A}^\top) - \boldsymbol{b}^\top = \boldsymbol{x}^\top \boldsymbol{A} - \boldsymbol{b}^\top$
  - This implies that the solution to $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \mathbf{0}$, say $\boldsymbol{x}^*$, is a stationary point of f
- We can rewrite $f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}^* + (\boldsymbol{x}-\boldsymbol{x}^*))^\top \boldsymbol{A}(\boldsymbol{x}^* + (\boldsymbol{x}-\boldsymbol{x}^*)) - \boldsymbol{b}^\top(\boldsymbol{x}^* + (\boldsymbol{x}-\boldsymbol{x}^*)) + c = \cdots = f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^\top \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^*)$ [Proof]

Outline

1. **Linear Algebra**
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. **Geometry**
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. **Point Set Topology****
   - Topological Spaces
   - Manifolds

Matrix Norms

- The set of matrices $\mathbb{R}^{m\times n}$ can be viewed as a vector space $\mathbb{R}^{mn}$
- How to define a norm in this space?

Definition (Matrix Norm)

A function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ is called the ***matrix norm*** if it satisfies:
a) $\|\boldsymbol{A}\| \geqslant 0, \forall \boldsymbol{A} \in \mathbb{R}^{m\times n}$ and the equality holds iff $\boldsymbol{A} = \boldsymbol{O}$ (positivity);
b) $\|r\boldsymbol{A}\| = |r| \|\boldsymbol{A}\|, \forall \boldsymbol{A} \in \mathbb{R}^{m\times n}, r \in \mathbb{R}$ (homogeneity);
c) $\|\boldsymbol{A}+\boldsymbol{B}\| \leqslant \|\boldsymbol{A}\| + \|\boldsymbol{B}\|, \forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ (triangle inequality).
For our purpose, we consider only the ***sub-multiplicative norm*** that satisfies an additional property for square matrices:
d) $\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\| \|\boldsymbol{B}\|, \forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$.

Frobenius Norms

- A common matrix norm is the ***Frobenius norm***:
  $\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$
  - Equivalent to the Euclidean norm in $\mathbb{R}^{mn}$
  - Is a sub-multiplicative norm [Proof]
- The Frobenius norm is unitarily invariant
  - Given an unitary (or orthogonal) matrix $\boldsymbol{U}$,
    $\|\boldsymbol{U}\boldsymbol{A}\|_F = \|\boldsymbol{U}\boldsymbol{a}_1\|_2 + \cdots + \|\boldsymbol{U}\boldsymbol{a}_1\|_2 = \|\boldsymbol{a}_1\|_2 + \cdots + \|\boldsymbol{a}_1\|_2 = \|\boldsymbol{A}\|_F$
- If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric, then
  $\|\boldsymbol{A}\|_F = \left\|\boldsymbol{U}^\top \boldsymbol{D}\boldsymbol{U}\right\|_F = \|\boldsymbol{D}\|_F = \sqrt{\sum_{i=1}^{n}\lambda_i^2}$

Low Rank Approximation

Theorem

Given a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *and* $k < rank(\boldsymbol{A})$, *the solution to the problem*

$$\arg_{\boldsymbol{M}} \min \|\boldsymbol{A} - \boldsymbol{M}\|_F$$
$$\textit{subject to } rank(\boldsymbol{M}) = k$$

is $\boldsymbol{M} = \boldsymbol{U}\widetilde{\boldsymbol{D}}\boldsymbol{U}^\top$, *where the columns of* $\boldsymbol{U}$ *are the eigenvectors of* $\boldsymbol{A}$ *and* $\widetilde{\boldsymbol{D}}$ *is a diagonal matrix containing only the* k *largest eigenvalues of* $\boldsymbol{A}$ *(with others being* 0*).*

Proof.

We only give an intuitive proof here. Since $\boldsymbol{A}$ is symmetric, we have $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$. Recall that the Frobenius norm is unitarily invariant, we have an equivalent objective: $\arg_{\boldsymbol{M}} \min \left\|\boldsymbol{D} - \boldsymbol{U}^\top\boldsymbol{M}\boldsymbol{U}\right\|_F$. Since $\boldsymbol{D}$ is diagonal, $\boldsymbol{U}^\top\boldsymbol{M}\boldsymbol{U}$ should be diagonal too to minimize the objective, implying that $\boldsymbol{M} = \boldsymbol{U}\widetilde{\boldsymbol{D}}\boldsymbol{U}^\top$ for some diagonal matrix $\widetilde{\boldsymbol{D}}$. Let λ_i and $\widetilde{d}_i$ be the ith diagonal element of $\boldsymbol{D}$ and $\widetilde{\boldsymbol{D}}$ respectively, we have $\left\|\boldsymbol{D} - \boldsymbol{U}\boldsymbol{M}\boldsymbol{U}^\top\right\|_F = \sqrt{\sum_{i=1}^n (\lambda_i - \widetilde{d}_i)^2}$. Since $rank(\boldsymbol{M}) = k$, only k of the $\widetilde{d}_i$s can be nonzero. Therefore, $\boldsymbol{M}$ is the best approximation when these nonzero $\widetilde{d}_i$s are the k largest eigenvalues of $\boldsymbol{A}$. □

Induced Norms (1/2)

- We can define another type of matrix norms based on vector norms
- Let $\|\cdot\|_{(m)}$ and $\|\cdot\|_{(n)}$ be two vector norms, we define the ***induced norm*** for $\mathbb{R}^{m\times n}$ as: $\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|_{(n)}=1} \|\boldsymbol{Ax}\|_{(m)}, \forall \boldsymbol{A} \in \mathbb{R}^{m\times n}$
  - We say that a matrix norm $\|\cdot\|$ is ***induced by*** (or ***compatible with***) the vector norms $\|\cdot\|_{(m)}$ and $\|\cdot\|_{(n)}$ if for all $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\|\boldsymbol{Ax}\|_{(m)} \leqslant \|\boldsymbol{A}\| \|\boldsymbol{x}\|_{(n)}$
  - The induced norm is a sub-multiplicative norm [Homework]

Induced Norms (2/2)

Theorem

Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the matrix norm $\|\boldsymbol{A}\|$ induced by the Euclidean norm equals $\sqrt{\lambda_{\max}}$, where $\lambda_{\max}$ is the largest eigenvalue of the matrix $\boldsymbol{A}^\top \boldsymbol{A}$.

Proof.

Since $\boldsymbol{A}^\top \boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, from our previous discussions we know that $\boldsymbol{A}^\top \boldsymbol{A}$ is diagonalizable. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be its eigenvalues and $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ be the orthonormal set of eigenvectors corresponding to these eigenvalues[a]. Consider an arbitrary $\boldsymbol{x}$, $\|\boldsymbol{x}\|_{(2)} = 1$, we have

$\boldsymbol{x} = c_1 \boldsymbol{x}_1 + \cdots + c_n \boldsymbol{x}_n$ and $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = c_1^2 + \cdots + c_n^2 = 1$. Furthermore,
$\|\boldsymbol{A}\boldsymbol{x}\|_{(2)}^2 = \left\langle \boldsymbol{x}, \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} \right\rangle = \langle c_1 \boldsymbol{x}_1 + \cdots + c_n \boldsymbol{x}_n, c_1 \lambda_1 \boldsymbol{x}_1 + \cdots + c_n \lambda_n \boldsymbol{x}_n \rangle =$
$\lambda_1 c_1^2 + \cdots + \lambda_n c_n^2 \leqslant \lambda_1 (c_1^2 + \cdots + c_n^2) = \lambda_1$, implying that $\|\boldsymbol{A}\boldsymbol{x}\|_{(2)} \leqslant \sqrt{\lambda_1}$.
Note the maximum of $\|\boldsymbol{A}\boldsymbol{x}\|_{(2)}$ is attainable when $\boldsymbol{x} = \boldsymbol{x}_1$. Therefore,
$\|\boldsymbol{A}\| = \sqrt{\lambda_1} = \sqrt{\lambda_{\max}}$. □

[a] Actually, $\boldsymbol{A}^\top \boldsymbol{A}$ is positive semidefinite, as $\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle \geqslant 0, \forall \boldsymbol{x} \in \mathbb{R}^n$. So $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$.

Rayleigh's Quotient

- Applying the similar argument above, we have:

Theorem (Rayleigh's Quotient)

Given a symmetric matrix $\boldsymbol{P} \in \mathbb{R}^{n \times n}$*, then* $\forall \boldsymbol{x} \in \mathbb{R}^n$*,*

$$\lambda_{\min} \leqslant \frac{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} \leqslant \lambda_{\max},$$

where $\lambda_{\min}$ *and* $\lambda_{\max}$ *are the smallest and largest eigenvalues of* $\boldsymbol{P}$ *respectively.*

- $\frac{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} = \lambda_i$ when $\boldsymbol{x}$ is the corresponding eigenvector of λ_i

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Matrix Exponential

Caution!

This subsection requires the knowledge of Taylor's theorem.

- Given a scalar x, by Taylor's theorem we have $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$
- Similarly, given a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we can define the ***matrix exponential*** as $e^{\boldsymbol{A}} = \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{k!} = \boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \cdots \in \mathbb{R}^{n \times n}$
  - $e^{\boldsymbol{O}} = \boldsymbol{I}$, $\left(e^{\boldsymbol{A}}\right)^\top = e^{\boldsymbol{A}^\top}$ [Proof]
- Unlike the scalar version, $e^{\boldsymbol{A}+\boldsymbol{B}} \neq e^{\boldsymbol{A}} e^{\boldsymbol{B}}$ unless $\boldsymbol{AB} = \boldsymbol{BA}$
  - If $\boldsymbol{A}$ and $\boldsymbol{B}$ commute, we can write $(\boldsymbol{A}+\boldsymbol{B})^k = \sum_{i=0}^{k} \binom{k}{i} \boldsymbol{A}^i \boldsymbol{B}^{k-i}$, so $\frac{(\boldsymbol{A}+\boldsymbol{B})^k}{k!} = \sum_{i=0}^{k} \frac{\boldsymbol{A}^i}{i!} \frac{\boldsymbol{B}^{k-i}}{(k-i)!}$, implying $e^{\boldsymbol{A}+\boldsymbol{B}} = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \frac{\boldsymbol{A}^i}{i!} \frac{\boldsymbol{B}^{k-i}}{(k-i)!} = \sum_{r=0}^{\infty} \frac{\boldsymbol{A}^r}{r!} \sum_{s=0}^{\infty} \frac{\boldsymbol{B}^s}{s!} = e^{\boldsymbol{A}} e^{\boldsymbol{B}}$
- If $\boldsymbol{A} = \boldsymbol{UDU}^{-1}$ is diagonalizable, we have $e^{\boldsymbol{A}} = \boldsymbol{U} e^{\boldsymbol{D}} \boldsymbol{U}^{-1}$, where $e^{\boldsymbol{D}}$ is a diagonal matrix whose the ith diagonal element equals to e^{λ_i} [Proof]

Matrix Logarithm

- The exponential $e^{\boldsymbol{A}}$ of an anitsymmetric (resp. antihermitian) matrix $\boldsymbol{A}$ is orthogonal (resp. unitary)
    - $(e^{\boldsymbol{A}})^\top e^{\boldsymbol{A}} = e^{\boldsymbol{A}^\top} e^{\boldsymbol{A}} = e^{-\boldsymbol{A}} e^{\boldsymbol{A}} = e^{\boldsymbol{O}} = \boldsymbol{I}$
- We call $\boldsymbol{B}$ the ***matrix logarithm*** of $\boldsymbol{A}$ iff $\boldsymbol{A} = e^{\boldsymbol{B}}$, denoted by $\ln \boldsymbol{A}$
- Not every matrix has a logarithm
- Nevertheless, if a matrix $\boldsymbol{A}$ is diagonalizable, we can easily find its logarithm
    - Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{-1}$, we have $\ln \boldsymbol{A} = \boldsymbol{U}(\ln \boldsymbol{D})\boldsymbol{U}^{-1}$, where $\ln \boldsymbol{D}$ is a diagonal matrix whose the ith diagonal element equals to $\ln \lambda_i$ [Proof]

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Affine Spaces (1/2)

- Recall that the linear variety is defined as $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}$ for some $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{y} \in \mathbb{R}^m$
  - If we can find a solution $\boldsymbol{x}_0$, then for any $\boldsymbol{v} \in \mathcal{N}(\boldsymbol{A})$, $\boldsymbol{x} = v + \boldsymbol{x}_0$ is also a solution
  - A linear variety is a "translated nullspace"
- Geometry discusses the properties of “shapes” in a vector space
  - Since these shapes may not pass through the origin, they lie in the “translated subspaces”

Affine Spaces (2/2)

Definition (Affine Space)

Given a vector space $\mathcal{V}$, a set of points $\mathcal{A}$ is called the ***affine space*** iff there exists a map $\mathcal{A} \times \mathcal{V} \to \mathcal{A}$, denoted by $a + \boldsymbol{v}$ for all $a \in \mathcal{A}$ and $\boldsymbol{v} \in \mathcal{V}$, with the following properties:

a) For all $a \in \mathcal{A}$, $a + \boldsymbol{0} = a$;

b) For all $\boldsymbol{a} \in \mathcal{A}$ and $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$, $(a + \boldsymbol{v}) + \boldsymbol{w} = a + (\boldsymbol{v} + \boldsymbol{w})$;

c) For any $a, b \in \mathcal{A}$ there exists a unique $\boldsymbol{v} \in \mathcal{V}$ such that $a = b + \boldsymbol{v}$.

- Property c) can be written as $a - b = \boldsymbol{v}$
- Intuitively, an affine space is a "translated vector space" where the origin is undefined

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Line Segments

Definition (Line Segment)

Given two points x and y in an affine space, the set $\{x + \delta(y - x) : \delta \in [0,1]\}$ is called the ***line segment*** between x and y.

- A line segment is a "shape" in the affine space where x and y lie
- Note there is no reason why x and y cannot be vectors
  - If points are vectors, they can be summed directly to get a new point (vector)
  - A line segment between two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ can be defined alternatively as the ***convex combination*** of $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e., $\{(1-\delta)\boldsymbol{x} + \delta\boldsymbol{y} \in \mathbb{R}^n : \delta \in [0,1]\}$
- We focus on the vector points from now on

Curves

Definition (Curve)

Let $\mathcal{I}$ be an interval of real numbers. A ***curve*** is a continuous function $\gamma : \mathcal{I} \to \mathbb{R}^n$. We also say that the curveγ is ***parametrized*** by the continous function.

- E.g., let $\mathcal{I} = [0, 2\pi]$, we can define a circle (a closed curve) γ in $\mathbb{R}^2$ parametrized by $\gamma(t) = [\cos(t), \sin(t)]^\top, \forall t \in \mathcal{I}$
- A curve is called the ***plane curve*** when $n = 2$ and ***space curve*** when $n = 3$

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Hyperplanes (1/2)

Definition (Hyperplane)

Given $y \in \mathbb{R}$ and $\boldsymbol{a} \in R^n$ where $\boldsymbol{a} \neq \boldsymbol{0}$, the set $H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = y\}$ is called the ***hyperplane*** of $\mathbb{R}^n$.

- A hyperplane is an affine space translated from the subspace $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = 0\}$ of $\mathbb{R}^n$
  - Since the dimension of the subspace is always $n-1$, we say that the hyperplane always has dimension $n-1$
- A hyperplane H divides $\mathbb{R}^n$ into the ***positive half-space*** $H_+ = \{\boldsymbol{x} \in \mathbb{R}^n : a_1x_1 + \cdots + a_nx_n \geqslant 0\}$ and ***negative half-space*** $H_- = \{\boldsymbol{x} \in \mathbb{R}^n : a_1x_1 + \cdots + a_nx_n \leqslant 0\}$
  - Both H_+ and H_- are subspaces of $\mathbb{R}^n$ [Proof]

Hyperplanes (2/2)

- For any $\boldsymbol{x}_1, x2 \in H$, the vector $\boldsymbol{a}$ is orthogonal to $\boldsymbol{x}_1 - \boldsymbol{x}_2$ and is called the ***normal*** of H
  - As $\langle \boldsymbol{a}, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle = \boldsymbol{a}^\top \boldsymbol{x}_1 - \boldsymbol{a}^\top \boldsymbol{x}_2 = y - y = 0$
- If a linear variety $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}$ has dimension less than n (i.e., $\boldsymbol{A} \neq \boldsymbol{O}$), then it is the intersection of a finite number of hyperplanes

Outline

1 Linear Algebra
- Vector Spaces, Linear Transformations, and Matrices
- Matrices
- Eigenvalues and Eigenvectors
- Inner Products and Norms
- Positive Definite Matrices and Quadratic Forms**
- Matrix Norms
- Matrix Exponential and Logarithm**

2 Geometry
- Affine Spaces
- Line Segments and Curves
- Hyperplanes
- Convex Sets
- Neighborhoods

3 Point Set Topology**
- Topological Spaces
- Manifolds

Convex Sets (1/2)

- So far we have seen many sets, e.g., vector spaces, subspaces, affine spaces, shapes (line segments and sets consisting of a single point), etc.

Definition (Convex Set)

A set Θ of points is ***convex*** iff for any $\boldsymbol{u}, \boldsymbol{w} \in \Theta$, we have $(1-\delta)\boldsymbol{u} + \delta\boldsymbol{v} \in \Theta, \forall\delta \in (0,1)$.

- Why "convex?"

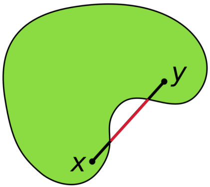
Convex Sets (1/2)

- So far we have seen many sets, e.g., vector spaces, subspaces, affine spaces, shapes (line segments and sets consisting of a single point), etc.

Definition (Convex Set)

A set Θ of points is ***convex*** iff for any $\boldsymbol{u}, \boldsymbol{w} \in \Theta$, we have $(1-\delta)\boldsymbol{u}+\delta\boldsymbol{v} \in \Theta, \forall\delta \in (0,1)$.

- Why "convex?"Any line segment cannot have portions that fall outside of the convex set



Convex Sets (2/2)

- Examples: $\mathbb{R}^n$, a half-space, a hyperplane, a linear variety, a line or line segment, a set of a single point, etc.
- Convex subsets of $\mathbb{R}^n$ have the following properties [Homework]:
  - Given a convex set Θ and $\beta \in \mathbb{R}$, the set $\beta\Theta = \{\boldsymbol{x} : \boldsymbol{x} = \beta\boldsymbol{v}, \boldsymbol{v} \in \Theta\}$ is convex
  - Given a convex sets Θ_1 and Θ_2, the set $\Theta_1 + \Theta_2 = \{\boldsymbol{x} : \boldsymbol{x} = \boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{v}_1 \in \Theta_1, \boldsymbol{v}_2 \in \Theta_2\}$ is convex
  - The intersection of convex sets is convex
- A point $\boldsymbol{x} \in \Theta$ is called an ***extreme point*** of Θ iff there are no two distinct points $\boldsymbol{u}, \boldsymbol{v} \in \Theta$ such that $\boldsymbol{x} = (1-\delta)\boldsymbol{u} + \delta\boldsymbol{v}$ for some $\delta \in (0,1)$
  - E.g., vertices (i.e., corners) of a polyhedron or endpoints of a line segment

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Neighborhoods (1/2)

Definition (Neighborhood)

A ***neighborhood*** of a point $\mathbf{x} \in \mathbb{R}^n$ is the set $\{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| < \varepsilon\}$, where ε is some positive real number.

- A point $\mathbf{x}$ in a set S is said to be an ***interior point*** of S iff S contains some neighborhood of $\mathbf{x}$
- A point $\mathbf{x}$ is said to be a ***boundary point*** of S iff every neighborhood of $\mathbf{x}$ contains a point in S and a point not in S
  - $\mathbf{x}$ may or may not be an element of S
  - The set of all boundary points of S is called the ***boundary*** of S
- An ***open set*** S contains a neighborhood of each of its points (i.e., contains only interior points)
  - Given $a, b \in \mathbb{R}$, the sets (a, b) and $\{[a, b]^\top : a^2 + 5b^2 < 1\}$ are open
- A set S is said to be ***closed*** if its complement $\mathbb{R}^n \backslash S$ is open (or intuitively, if it contains the boundary)
  - $[a, b]$ is closed

Neighborhoods (1/2)

Definition (Neighborhood)

A ***neighborhood*** of a point $\boldsymbol{x} \in \mathbb{R}^n$ is the set $\{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\}$, where ε is some positive real number.

- A point $\boldsymbol{x}$ in a set S is said to be an ***interior point*** of S iff S contains some neighborhood of $\boldsymbol{x}$
- A point $\boldsymbol{x}$ is said to be a ***boundary point*** of S iff every neighborhood of $\boldsymbol{x}$ contains a point in S and a point not in S
  - $\boldsymbol{x}$ may or may not be an element of S
  - The set of all boundary points of S is called the ***boundary*** of S
- An ***open set*** S contains a neighborhood of each of its points (i.e., contains only interior points)
  - Given $a, b \in \mathbb{R}$, the sets (a, b) and $\{[a, b]^\top : a^2 + 5b^2 < 1\}$ are open
- A set S is said to be ***closed*** if its complement $\mathbb{R}^n \backslash S$ is open (or intuitively, if it contains the boundary)
  - $[a, b]$ is closed

Neighborhoods (2/2)

- A set S that can be contained in a ball of finite radius is said to be ***bounded***
  - That is, for any point $\boldsymbol{x} \in S$, there exists some positive real number $r \in \mathbb{R}$ such that $\|\boldsymbol{x}\| < r$
- A set S is ***compact*** iff it is both closed and bounded
  - Given $a, b \in \mathbb{R}$. Is (a, b) compact?
  - How about $[a, b]$?

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Point Set Topology

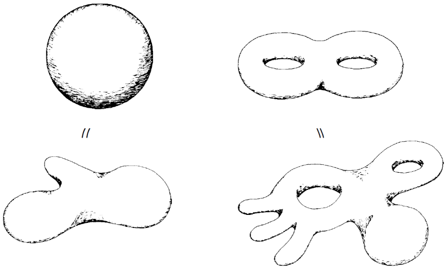
- Line segments, curves, surfaces, hyperplanes are basically sets of points
- Point set topology treat these sets as "spaces" and discusses their properties

Caution!

This section requires the knowledge of function continuity and limit.

Geometry vs. Topology

- Imagine that a shape is made by rubber
  - It can be "deformed" (e.g., either rotated, sheared, flipped, scaled etc. by linear by transformations; or bended, stretched, twisted etc. by nonlinear functions)
  - But not teared, or cut and then glued
- ***Geometry*** discusses the properties (e.g., volume, curvature, distance, angle, etc.) of shapes that are changed as they are deformed
- ***Topology*** discusses the shapes' nature which is unaffected by deformation



Topological Properties

- Examples of the topological properties?

Topological Properties

- Examples of the topological properties? Loosely speaking,
  - Dimension (number of element in a basis)
  - Compactness
  - Connectedness
  - Separation (we will see this later when talking about the Hausdorff spaces)
- Properties of a topological space are described using the ***open sets***

Topological Spaces

Definition (Topological Space)

Given a set of point X. Let $\mathcal{T}$ be a set of subsets of X. Then $(X,\mathcal{T})$ is called a ***topological space*** iff

a) Both $\emptyset$ and X are in $\mathcal{T}$;

b) Any union of arbitrary (possibly infinitely) many elements of $\mathcal{T}$ is an element of $\mathcal{T}$;

c) Any intersection of finitely many elements of $\mathcal{T}$ is an element of $\mathcal{T}$.

We call $\mathcal{T}$ a ***topology*** on X, and the sets in $\mathcal{T}$ are called the ***open sets***.

- When $X = \mathbb{R}^n$, our previous definition of an open set (i.e., a set containing an ε-ball around each its point) is just a special case here
  - The collection of those open sets is called the ***standard topology*** on $\mathbb{R}^n$
  - We can define different topologies on $\mathbb{R}^n$ such as the cofinite topology: $\mathcal{T} = \{X \backslash A : A = X \text{ or } A \text{ is finite}\}$

Neighborhood

Definition (Neighborhood)

A ***neighborhood*** (or specifically, ***open neighborhood***) of a point p in a topological space $(X, \mathcal{T})$ is an open set in $\mathcal{T}$ containing p.

- Our previous definition of a neighborhood (i.e., an ε-ball) is a special case
    - An ε-ball is itself an open set (with a particular shape)

Sequences and Limits

Definition (Limit of a Sequence)

In a topological space $(X, \mathcal{T})$, a point $p^* \in X$ is called the ***limit*** of a sequence of points $\{p^{(k)}\}_{k \in \mathbb{N}}$ in X iff for every neighborhood S of p^*, there exists $K \in \mathbb{N}$, such that $p^{(k)} \in S$ for all $k > K$.

- The limit of a sequence may not be unique, as the neighborhoods of points may not be separable
  - Consider two points p and q in the cofinite topological space on $\mathbb{R}$, any neighborhood of p (e.g., $\mathbb{R} \backslash \{q\}$) and q (e.g., $\mathbb{R} \backslash \{p\}$) must overlap

Separation

- An important topological property is that whether two points are separable:

Definition (Hausdorff Space)

A topological space $(X, \mathcal{T})$ is ***Hausdorff*** iff given any two points p and q in X, if there exists a neighborhood U of p and V of q respectively such that $U \cap V = \emptyset$.

- Every sequence $\{p^{(k)}\}_k$ has a unique limit p^* in the Hausdorff space, and we write $\lim_{k\to\infty} p^{(k)} = p^*$
- We can then perform calculus in the Hausdorff spaces

Function Continuity (1/2)

Definition (Continuity)

A function $f : X \to Y$ between two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ is ***continuous*** iff given any open set $U \in \mathcal{T}_Y$, the inverse image $f^{-1}(U) = \{x \in X : f(x) \in U\}$ is open.

- How does this related with our previous definition of continuity?
  - Recall that a function f is continuous at a iff $\lim_{x \to a} f(x) = f(a)$; that is, given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all x, $\|x - a\| < \delta$, we have $\|f(x) - f(a)\| < \varepsilon$

Function Continuity (2/2)

Theorem

Let $\boldsymbol{f}:\mathbb{R}^n \to \mathbb{R}^m$ be a function between two standard topological spaces $(\mathbb{R}^n, \mathcal{T}_n)$ and $(\mathbb{R}^m, \mathcal{T}_m)$. For any $\boldsymbol{a} \in \mathbb{R}^n$, $\lim_{\boldsymbol{x}\to\boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{a})$ iff for any open set $U \in \mathcal{T}_m$, $f^{-1}(U)$ is open.

Proof.

$\Rightarrow$ If $f^{-1}(U) = \emptyset$ we are done since the empty set is always open. Otherwise, consider any point $\boldsymbol{a} \in f^{-1}(U)$. Since U is open, there exists $\varepsilon > 0$ such that the set $\{\boldsymbol{y} \in \mathbb{R}^m : \|\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{a})\| < \varepsilon\}$ is contained in U. By definition of $\lim_{\boldsymbol{x}\to\boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{a})$, there exists $\delta > 0$ such that the set $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{a}\| < \delta\}$ is contained in $f^{-1}(U)$. Since for any point $\boldsymbol{a}$, its neighborhood is contained in $f^{-1}(U)$. $f^{-1}(U)$ is an open set.
$\Leftarrow$ Given any $\varepsilon > 0$, define $U = \{\boldsymbol{y} \in \mathbb{R}^m : \|\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{a})\| < \varepsilon\}$. Since $f^{-1}(U)$ is an open set and $\boldsymbol{a} \in f^{-1}(U)$, there exists $\delta > 0$ such that $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{a}\| < \delta\}$ is contained in $f^{-1}(U)$, implying that if $\|\boldsymbol{x} - \boldsymbol{a}\| < \delta$ then $\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{a})\| < \varepsilon$. □

Homeomorphism

Definition (Homeomorphism)

Two topological spaces $(X, \mathfrak{T}_X)$ and $(Y, \mathfrak{T}_Y)$ are ***homeomorphic*** (or ***topological isomorphic***) if there exists a function $f : X \to Y$ such that:
a) f is a bijection (i.e., one-to-one and onto);
b) f is an open map (i.e., for any open set $U \subseteq X$, $\{f(x) : x \in U\} \subseteq Y$ is open);
c) f is continuous.

- Intuitively, two homeomorphic spaces are "the same" from the topological point of view
  - All topological properties are preserved
- Is $(-1, 1)$ homeomorphic to $\mathbb{R}$?

Homeomorphism

Definition (Homeomorphism)

Two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ are ***homeomorphic*** (or ***topological isomorphic***) if there exists a function $f : X \to Y$ such that:
a) f is a bijection (i.e., one-to-one and onto);
b) f is an open map (i.e., for any open set $U \subseteq X$, $\{f(x) : x \in U\} \subseteq Y$ is open);
c) f is continuous.

- Intuitively, two homeomorphic spaces are "the same" from the topological point of view
  - All topological properties are preserved
- Is $(-1, 1)$ homeomorphic to $\mathbb{R}$?
  - Yes, as we can define $f : (-1, 1) \to \mathbb{R}$, $f(x) = \tan(\frac{\pi}{2}x)$
- Also, $\{[x_1, x_2, x_3]^\top \in \mathbb{R}^3 : x_3 = x_1 + x_2\}$ is homeomorphic to $\mathbb{R}^2$
  - We say the function $\boldsymbol{f} : \mathbb{R}^2 \to \mathbb{R}^3$, $\boldsymbol{f}(x_1, x_2) = (x_1, x_2, x_1 + x_2)$, ***embeds*** $\mathbb{R}^2$ into $\mathbb{R}^3$

Outline

1. Linear Algebra
   - Vector Spaces, Linear Transformations, and Matrices
   - Matrices
   - Eigenvalues and Eigenvectors
   - Inner Products and Norms
   - Positive Definite Matrices and Quadratic Forms**
   - Matrix Norms
   - Matrix Exponential and Logarithm**
2. Geometry
   - Affine Spaces
   - Line Segments and Curves
   - Hyperplanes
   - Convex Sets
   - Neighborhoods
3. Point Set Topology**
   - Topological Spaces
   - Manifolds

Manifolds (1/2)

- Many complex shapes in the real world have a simple shape when we look at a just tiny portion of them

Definition (Manifold)

A ***manifold*** $(M, \mathcal{T})$ of dimension k embedded in $\mathbb{R}^n$ is a Hausdorff space such that for any point $p \in M \subseteq \mathbb{R}^n$, there exists a small neighborhood of p which is homeomorphic to $\mathbb{R}^k$.

- Curves and surfaces are examples of manifolds of dimension 1 and 2 respectively
- The mapping between the local neighborhoods and $\mathbb{R}^k$ need not be linear
  - Consider a unit circle $M = \{[x_1, x_2]^\top : x_1^2 + x_2^2 = 1\}$ in $\mathbb{R}^2$, any point $\boldsymbol{p}$ lies in at least one of the 4 open sets $M_{top} = \{[x_1, x_2]^\top \in M : x_2 > 0\}$, $M_{right} = \{[x_1, x_2]^\top \in M : x_1 > 0\}$, M_{bottom}, and M_{left}
  - Each of these sets is homeomorphic to $\mathbb{R}^k$ (e.g., we can define $f_{top}(x_1, x_2) = \tan(\frac{\pi}{2} x_1)$)

Manifolds (2/2)

- When we say a shape looks like a "donut" in a 3-dimensional space we are looking at its ***extrinsic*** properties from the 3-dimensional space
- Manifold provides an ***intrinsic*** pint of view of a shape
  - All topological properties of a tiny portion of a manifold is the same with those of the Euclidean space
- Generally, a manifold can be constructed by "patching" the overlapping local neighborhoods (e.g., M_{top}, M_{right}, M_{bottom}, and M_{left})
- The invertible mappings (e.g., f_{top}, f_{right}, f_{bottom}, and f_{left}) between these neighborhoods and $\mathbb{R}^k$ are called ***charts***
- A specific collection of charts which covers a manifold is called the ***atlas***
  - An atlas is not unique as we can use different combinations of charts to cover a manifold