Numerical Optimization

Mathematical Background (I)

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Sets

Definition

A *set* is a collection of objects satisfying certain property P.

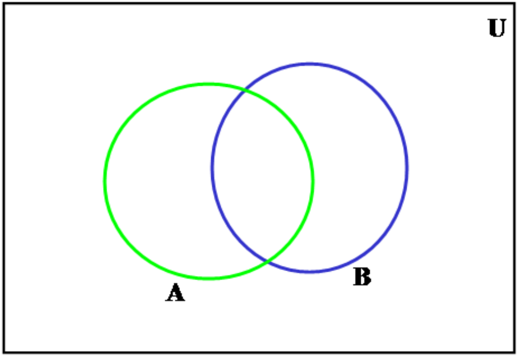
Examples:

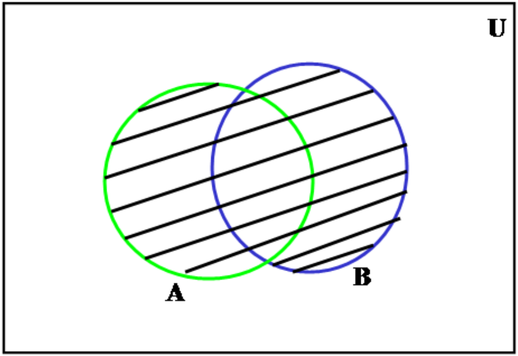
- A set of natural numbers, $\{1, 2, 3, \ldots\}$
- $\{x \in \mathbb{R} : 1 \leq x \leq 3\}$

Note: A set not containing any object is called the *empty* set and is denoted by ϕ.

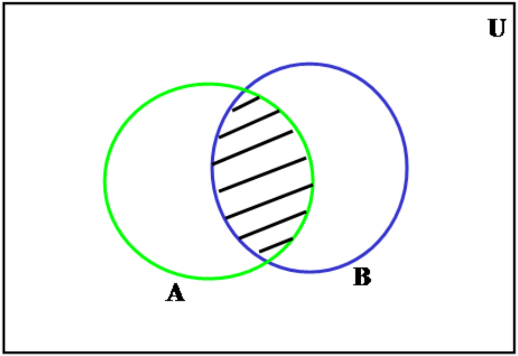
Let A and B be two sets.

- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- Difference: $A \backslash B = \{x : x \in A \text{ and } x \notin B\}$


U
A
B

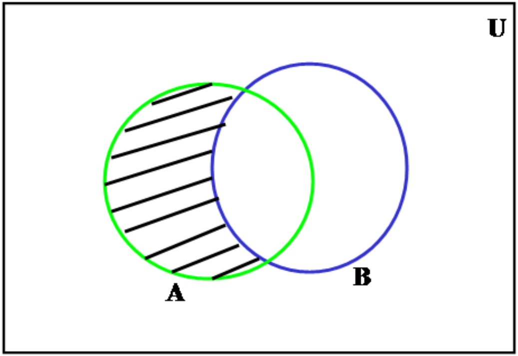


Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$

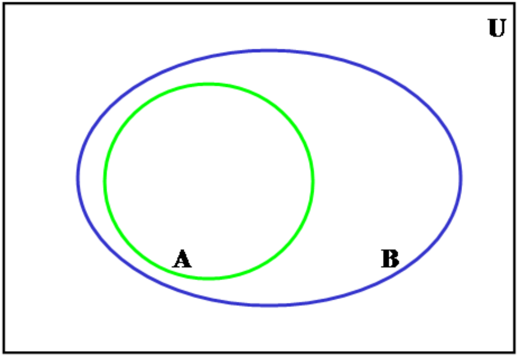


Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$

If the intersection of two sets is empty, we say that the sets are *disjoint*. That is, for two disjoint sets A and B, $A \cap B = \phi$.



Difference: $A \backslash B = \{x : x \in A \text{ and } x \notin B\}$



Let A and B be two sets. If A is a subset of B, that is, every member of A is also a member of B, we write $A \subseteq B$. Further, if A is a subset of B and there exists $y \in B$ such that $y \notin A$, then we write $A \subset B$.

Supremum and Infimum of a set

Definition

A set A of real numbers is said to be *bounded above*, if there is a real number y such that $x \leq y$ for every $x \in A$. The smallest possible real number y satisfying $x \leq y$ for every $x \in A$ is called the *least upper bound* or *supremum* of A and is denoted by $\sup\{x : x \in A\}$.

- Similarly, one can define *greatest lower bound* or *infimum*, $\inf\{x : x \in A\}$.

Example: Consider the set, $A = \{x : 1 \leq x < 3\}$

- $\sup\{x : x \in A\} = 3 (\notin A)$
- $\inf\{x : x \in A\} = 1 (\in A)$

Vector Space

A nonempty set S is called a *vector space* if

1. For any $\mathbf{x}, \mathbf{y} \in S$, $\mathbf{x} + \mathbf{y}$ is defined and is in S. Further,

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} \quad \text{(commutativity)}$$
$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z} \quad \text{(associativity)}$$

2. There exists an element in S, $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$.
3. For any $\mathbf{x} \in S$, there exists $\mathbf{y} \in S$ such that $\mathbf{x} + \mathbf{y} = \mathbf{0}$.
4. For any $\mathbf{x} \in S$ and $\alpha \in \mathbb{R}$, $\alpha\mathbf{x}$ is defined and is in S. Further, $\mathbf{1x} = \mathbf{x}$ for every $\mathbf{x}$.
5. For any $\mathbf{x}, \mathbf{y} \in S$ and $\alpha, \beta \in \mathbb{R}$,

$$\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$$
$$(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$$
$$\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$$

Elements in S are called *vectors*

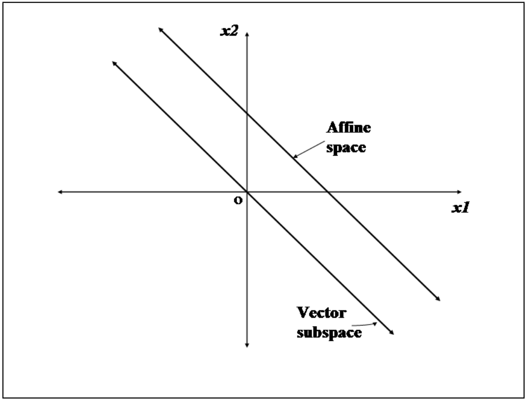
Notations

- $\mathbb{R}$: Vector space of real numbers
- $\mathbb{R}^n$: Vector space of real $n \times 1$ vectors
- n-vector $\mathbf{x}$ is an array of n scalars, $x_1, x_2, \ldots, x_n$
- $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$
- $\mathbf{x} \in \mathbb{R}^n, x_i \in \mathbb{R} \;\forall\; i$
- $\mathbf{x}^T = (x_1, x_2, \ldots, x_n)$
- $\mathbf{0}^T = (0, 0, \ldots, 0)$
- $\mathbf{1}^T = (1, 1, \ldots, 1)$ (We also use $\mathbf{e}$ to denote this vector)

Definition

If S and T are vector spaces such that $S \subseteq T$, then S is called a *subspace* of T.

Question: What are all possible subspaces of $\mathbb{R}^2$?



Spanning Set

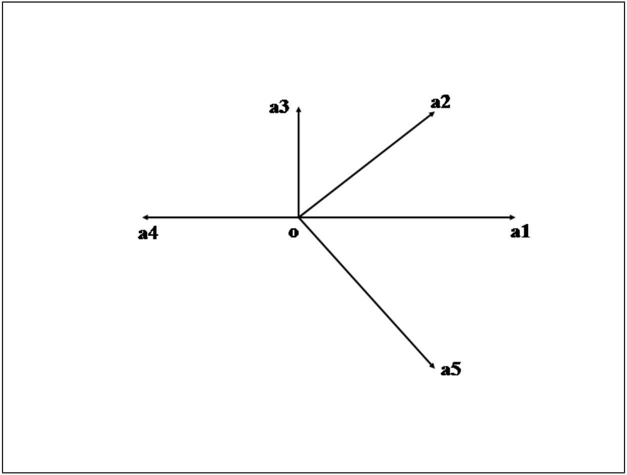
Definition

A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is said to *span* the vector space S if any vector $\mathbf{x} \in S$ can be represented as

$$\mathbf{x} = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i$$

for some real coefficients $\alpha_i,\ i = 1, \ldots, k$.

Example : The vectors,
$a1 = (1,0)^T, a2 = (1,1)^T; a3 = (0,1)^T, a4 = (-1,0)^T$ and $a5 = (1,-1)^T$ span $\mathbb{R}^2$



Linear Independence

Definition

A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is said to *linearly independent* if

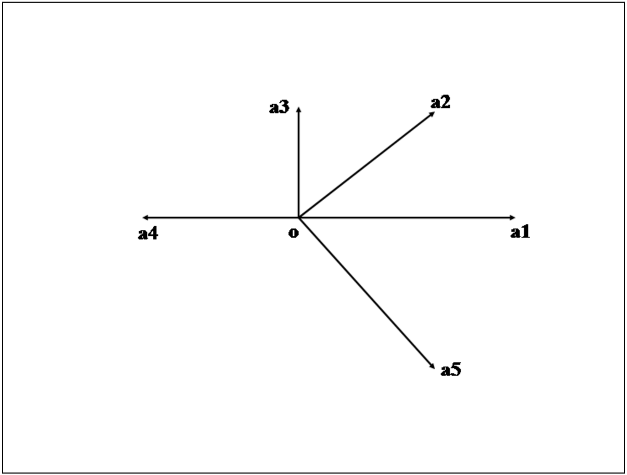
$$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i = 0 \Rightarrow \alpha_i = 0 \ \ \forall\, i.$$

Otherwise, they are linearly dependent and one of them is a linear combination of the others.

Mathematical Background

Example : In $\mathbb{R}^2$,

- $a1 = (1, 0)$ and $a2 = (1, 1)$ are linearly independent.
- $a1 = (1, 0)$ and $a4 = (-1, 0)$ are linearly dependent.



Basis

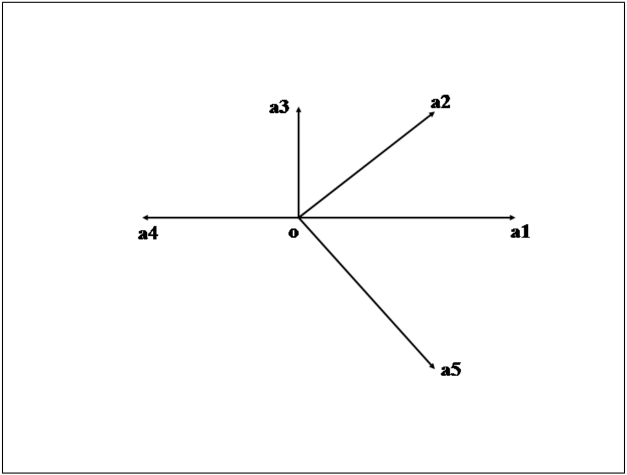
Definition

A set of vectors is said to be a *basis* for the vector space S if it is linearly independent and spans S.

Mathematical Background

Example : For $\mathbb{R}^2$,

- $a1 = (1, 0)$ and $a2 = (1, 1)$ form a basis
- $a1 = (1, 0)$ and $a3 = (0, 1)$ form a basis



- A vector space does not have a unique basis.
- If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is a basis for S, then any $\mathbf{x} \in S$ can be *uniquely* represented using $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$.
- Any two bases of a vector space have the same cardinality.
- The dimension of the vector space S is the cardinality of a basis of S.
- The dimension of the vector space $\mathbb{R}^n$ is n.
- Let $\mathbf{e}_i$ denote an n-dimensional vector whose i-th element is 1 and the remaining elements are 0's. Then, the set $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ forms a *standard* basis for $\mathbb{R}^n$.
- A basis for the vector space S is a maximal independent set of vectors which spans the space S.
- A basis for the vector space S is a minimal spanning set of vectors which spans the space S.

Functions

Definition

A function f from a set A to a set B is a rule that assigns to each x in A a unique element $f(x)$ in B. This function can be represented by

$$f : A \longrightarrow B.$$

Note:

- A: *Domain* of f
- $\{y \in B : (\exists x)[y = f(x)]\}$: *Range* of f
- *Range* of $f \subseteq B$

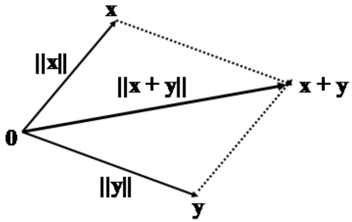
Examples:

- $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = x^2$
- $f : (-1, 1) \longrightarrow \mathbb{R}$ defined as $f(x) = \frac{1}{|x|-1}$

Definition

A *norm* on $\mathbb{R}^n$ is a real-valued function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ which obeys

- $\|\mathbf{x}\| \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$,
- $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for every $\mathbf{x} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, and
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for every $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$.



Mathematical Background

Let $\mathbf{x} \in \mathbb{R}^n$.
Some popular norms:

- L_2 or Euclidean norm

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} (x_i)^2 \right)^{\frac{1}{2}}$$

- L_1 norm

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

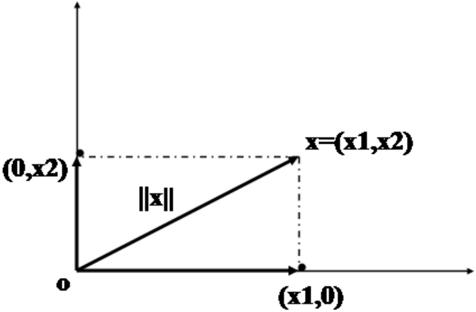
- L_∞ norm

$$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

Mathematical Background

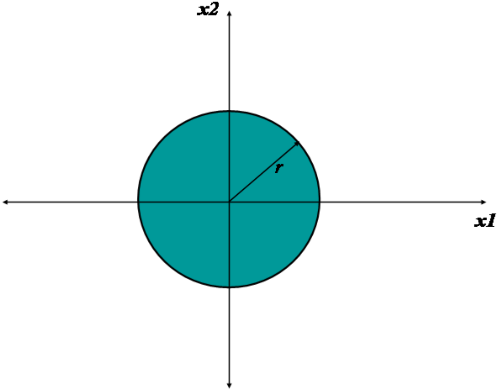
Illustration of L_2 norm:

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} (x_i)^2\right)^{\frac{1}{2}}$$

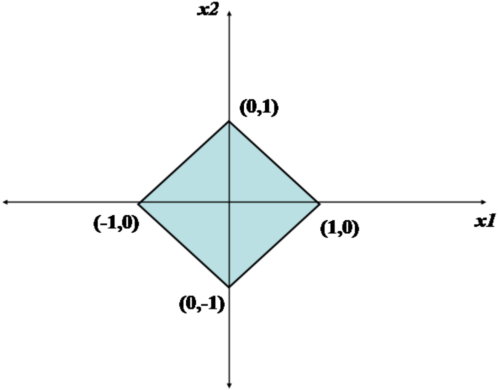


Mathematical Background

- $S = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq r\}$

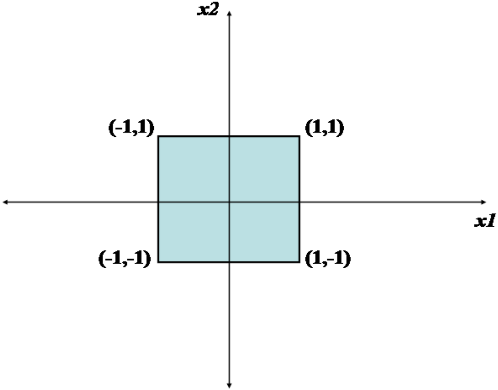


- $S = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_1 \leq 1\}$



Mathematical Background

- $S = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_\infty \leq 1\}$



- In general, the class of L_p $(1 \leq p < \infty)$ vector norms is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

- *Question*: Does the convergence of a particular optimization algorithm depend on what norm its stopping criterion used?

Result

If $\|\cdot\|_p$ and $\|\cdot\|_q$ are any two norms on $\mathbb{R}^n$, then there exist positive constants α and β such that

$$\alpha\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_q \leq \beta\|\mathbf{x}\|_p$$

for any $\mathbf{x} \in \mathbb{R}^n$.

Inner Product

Definition

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{x} \neq 0 \neq \mathbf{y}$. The *inner* or *dot* product of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathbf{x} \cdot \mathbf{y} \equiv \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i \cdot y_i = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos \theta$$

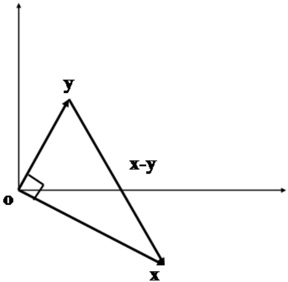
where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

Note:

- $\mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$.
- $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$
- $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ (Cauchy-Schwartz inequality)

Orthogonality

- Suppose $\mathbf{x}$ and $\mathbf{y}$ are perpendicular to each other.



Using Pythagoras formula,

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{y}\|^2,$$

which gives $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}^T\mathbf{y}$. That is, $\mathbf{x}^T\mathbf{y} = 0$

Orthogonality

Definition

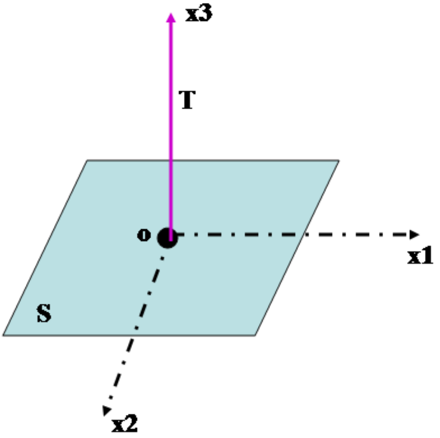
Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$. $\mathbf{x}$ and $\mathbf{y}$ are said to *perpendicular* or *orthogonal* to each other if $\mathbf{x}^T\mathbf{y} = 0$.

Definition

- Two subspaces S and T of the same vector space $\mathbb{R}^n$ are orthogonal if every vector $\mathbf{x} \in S$ is orthogonal to every vector $\mathbf{y} \in T$, i.e. $\mathbf{x}^T\mathbf{y} = 0 \ \forall \mathbf{x} \in S, \mathbf{y} \in T$.

Definition

Given a subspace S of $\mathbb{R}^n$, the space of all vectors orthogonal to S is called the *orthogonal complement* of S.

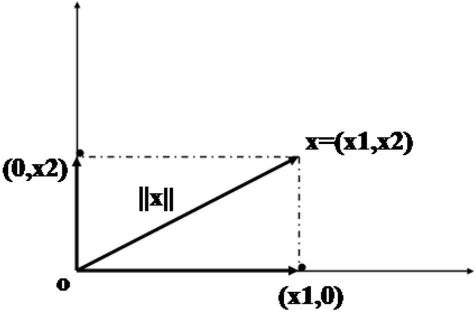


Mutual Orthogonality

Definition

Vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ are said to be *mutually orthogonal* if $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ for all $i \neq j$. If, in addition, $\|\mathbf{x}_i\| = 1$ for every i, the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is said to be *orthonormal*.

Mutual Orthogonality



- Is the set of mutually orthogonal vectors linearly independent?

Result

If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are mutually orthogonal nonzero vectors, then they are linearly independent.

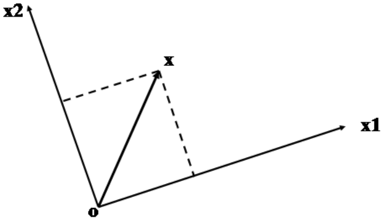
We need to show that

$$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i = 0 \Rightarrow \alpha_i = 0 \ \ \forall\, i.$$

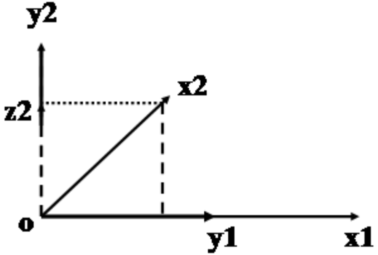
Proof.

Let $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \ldots + \alpha_k \mathbf{x}_k = 0$.
Therefore, $(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \ldots + \alpha_k \mathbf{x}_k)^T \mathbf{x}_1 = 0$, or,
$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i^T \mathbf{x}_1 = 0$.
This gives $\alpha_1 \mathbf{x}_1^T \mathbf{x}_1 = 0$ which implies $\alpha_1 = 0$.
Similarly we can show that each α_i is zero.
Therefore, the mutually orthogonal vectors are linearly independent. $\square$

Suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthonormal.
Given any vector $\mathbf{x}$, we can write $\mathbf{x} = (\mathbf{x}^T\mathbf{x}_1)\mathbf{x}_1 + (\mathbf{x}^T\mathbf{x}_2)\mathbf{x}_2$.
We require orthonormality of given set of vectors.



Question: How to produce an orthonormal basis starting with a given basis $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$?



Gram-Schmidt Procedure

- Given $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$, a basis in $\mathbb{R}^3$
- To produce an orthonormal basis $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$.
- Without loss of generality, set $\mathbf{y}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$
- Consider $\mathbf{x}_2$ and remove its component in the $\mathbf{y}_1$ direction.

$$\mathbf{z}_2 = \mathbf{x}_2 - (\mathbf{x}_2^T \mathbf{y}_1)\mathbf{y}_1$$

- $\mathbf{z}_2$ is orthogonal to $\mathbf{y}_1$
- Set $\mathbf{y}_2 = \frac{\mathbf{z}_2}{\|\mathbf{z}_2\|}$
- Start with $\mathbf{x}_3$ and remove its components in the $\mathbf{y}_1$ and $\mathbf{y}_2$ directions.

$$\mathbf{z}_3 = \mathbf{x}_3 - (\mathbf{x}_3^T \mathbf{y}_1)\mathbf{y}_1 - (\mathbf{x}_3^T \mathbf{y}_2)\mathbf{y}_2$$

- $\mathbf{z}_3$ is orthogonal to $\mathbf{y}_1$ and $\mathbf{y}_2$
- Set $\mathbf{y}_3 = \frac{\mathbf{z}_3}{\|\mathbf{z}_3\|}$
- Easy to extend this procedure to a basis in $\mathbb{R}^n$

Matrices

- $\mathbf{A} \in \mathbb{R}^{m \times n}$. $\mathbf{A}$ is a matrix of size $m \times n$.

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix}$$

- A_{ij} denotes (i,j)-element of $\mathbf{A}$.
- $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ where $\mathbf{a}_i \in \mathbb{R}^m,\ i = 1, \dots, n$
- The *transpose* of $\mathbf{A}$, denoted by $\mathbf{A}^T$ is the $n \times m$ matrix whose (i,j)-element is A_{ji}.
- $\mathbf{A}^T = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{pmatrix}$

Matrices

- Diagonal Matrix: A square matrix Λ such that $\Lambda_{ij} = 0,\ i \neq j$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

- Identity Matrix ($\mathbf{I}$): A diagonal matrix such that $I_{ii} = 1\ \forall\ i$

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

- Lower Triangular Matrix ($\mathbf{L}$) : A square matrix such that $L_{ij} = 0, i < j$

Matrices

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

Definition

The subspace of $\mathbb{R}^m$, spanned by the column vectors of $\mathbf{A}$ is called the *column space* of $\mathbf{A}$. The subspace of $\mathbb{R}^n$, spanned by the row vectors of $\mathbf{A}$ is called the *row space* of $\mathbf{A}$

Definition

Column Rank : The dimension of the column space
Row Rank : The dimension of the row space

Definition

The column rank of a matrix $\mathbf{A}$ equals its row rank, and this common value is called the *rank* of $\mathbf{A}$.

- Let $\mathbf{A} = \left(\begin{smallmatrix} 1 & 3 & -2 & 4 \\ -1 & -3 & 1 & -2 \end{smallmatrix}\right)$. $\text{rank}(\mathbf{A}) = 2$
- The rank of a matrix is 0 if and only of it is a zero matrix.
- Matrices with the smallest rank - Rank one matrices
  Example:

$$\begin{pmatrix} 3 & 1 & -1 \\ -3 & -1 & 1 \\ 6 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 1 & -1 \end{pmatrix} = \mathbf{u}\mathbf{v}^T$$

- Every matrix of rank one has the simplest form, $\mathbf{A} = \mathbf{u}\mathbf{v}^T$.

Matrices

Definition

A square matrix $\mathbf{A}$ is said to be *invertible* if there exists a matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. There is at most one such $\mathbf{B}$ and is denoted by $\mathbf{A}^{-1}$.

Easy to verify that,

- $$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \text{ if } (ad-bc) \neq 0.$$
- $$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^{-1} = \begin{pmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{pmatrix} \quad \text{if} \quad \lambda_1, \lambda_2 \neq 0.$$

Matrices

A product of invertible matrices is invertible and

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

- We denote the determinant of a matrix $\mathbf{A}$ by $\det(\mathbf{A})$.

If $\det(\mathbf{A}) \neq 0$, then $\mathbf{A}$ is invertible.

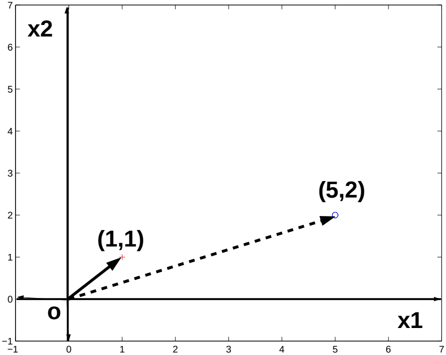
- The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if

$$\begin{aligned} \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} &\neq 0 \\ \text{i.e.} \quad ad - bc &\neq 0 \end{aligned}$$

- The matrix $\mathbf{Q}$ is orthogonal if $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

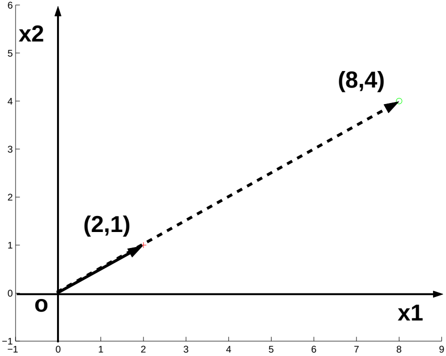
Matrix-vector multiplication, Ax

- $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- $\mathbf{Ax} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$



Matrix-vector multiplication, Ax

- $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.
- $\mathbf{Ax} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} = 4\begin{pmatrix} 2 \\ 1 \end{pmatrix} = 4\mathbf{x}$



Eigenvalues and Eigenvectors

Definition

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. The *eigenvalues* and *eigenvectors* of $\mathbf{A}$ are the real or complex scalars λ and n-dimensional vectors $\mathbf{x}$ such that

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}, \ \ \mathbf{x} \neq \mathbf{0}.$$

- $\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \Rightarrow (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$
- λ is an eigenvalue of $\mathbf{A}$ if and only if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \qquad (\text{characteristic equation of } \mathbf{A})$$

- This equation has n roots and are called the eigenvalues of $\mathbf{A}$.

Eigenvalues and Eigenvectors

- Let $\mathbf{A} = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}$.
- Characteristic equation:

$$\det \begin{pmatrix} 4-\lambda & -5 \\ 2 & -3-\lambda \end{pmatrix} = 0$$
$$\Rightarrow (\lambda^2 - \lambda - 2) = 0$$
$$\Rightarrow \lambda = 2 \text{ or } \lambda = -1$$

- $\lambda_1 = 2, (\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_1 = \mathbf{0}$ gives $\mathbf{x}_1$ to be a multiple of $(5, 2)^T$.
- $\lambda_2 = -1, (\mathbf{A} - \lambda_2 \mathbf{I})\mathbf{x}_2 = \mathbf{0}$ gives $\mathbf{x}_2$ to be a multiple of $(1, 1)^T$.
- Eigenvalues of $\mathbf{A}$: 2 and -1
- The corresponding eigenvectors of $\mathbf{A}$: $(5, 2)^T$ and$(1, 1)^T$

Symmetric Matrices

Definition

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. The matrix $\mathbf{A}$ is said to be *symmetric* if $\mathbf{A}^T = \mathbf{A}$.

- Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric. Then,
  - $\mathbf{A}$ has n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, and
  - a corresponding set of eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ can be chosen to be orthonormal.
  - $\mathbf{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is an orthogonal matrix ($\mathbf{S}^{-1} = \mathbf{S}^T$).
  - $\mathbf{S}^T\mathbf{A}\mathbf{S} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} = \Lambda$

Quadratic Form

- Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a symmetric matrix
- Consider $f(x) = \mathbf{x}^T\mathbf{A}\mathbf{x}$, a *pure quadratic form*

$\mathbf{A}$ is said to be	if
positive definite (pd)	$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for every nonzero $\mathbf{x} \in \mathbb{R}^n$
positive semi-definite (psd)	$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$
negative definite (nd)	$\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$ for every nonzero $\mathbf{x} \in \mathbb{R}^n$
negative semi-definite (nsd)	$\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$ for every $\mathbf{x} \in \mathbb{R}^n$
indefinite	$\mathbf{A}$ is neither positive definite nor negative definite

- *Question*: How to numerically check the positive definiteness of $\mathbf{A}$?

Quadratic Form

- Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a symmetric matrix
- Consider $f(x) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, a *pure quadratic form*
- Eigenvalues of $\mathbf{A}$: $\lambda_1, \lambda_2, \ldots, \lambda_n$
- Orthonormal Eigenvectors of $\mathbf{A}$: $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$
- $\mathbf{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$

$$\begin{aligned} \mathbf{x}^T \mathbf{A} \mathbf{x} &= \mathbf{x}^T \mathbf{S} \Lambda \mathbf{S}^T \mathbf{x} \\ &= \mathbf{y}^T \Lambda \mathbf{y} \\ &= \sum_{i=1}^{n} \lambda_i y_i^2 \end{aligned}$$

Therefore, $\lambda_i > 0 \quad \forall\, i \quad \Rightarrow \quad \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$

To prove that $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0 \Rightarrow$ Every eigen value of $\mathbf{A}$ is positive.

- Given, $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for every $\mathbf{x} \neq 0$
- Therefore, $\mathbf{x}_i^T\mathbf{A}\mathbf{x}_i > 0$ for every eigen vector $\mathbf{x}_i$
- That is, $\lambda_i\mathbf{x}_i^T\mathbf{x}_i > 0$ for every eigen vector $\mathbf{x}_i$
- Thus, $\lambda_i > 0$ for every eigen vector $\mathbf{x}_i$.

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric. Then,

$\mathbf{A}$ is said to be	if and only if, all the eigenvalues of $\mathbf{A}$ are
positive definite (pd)	positive
positive semi-definite (psd)	non-negative
negative definite (nd)	negative
negative semi-definite (nsd)	non-positive

- $\mathbf{A}$ is indefinite if and only if, it has both positive and negative eigenvalues.

Some other ways of checking positive definiteness

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric.

- Sylvester's criterion: $\mathbf{A}$ is positive definite if all its leading principal minors are positive.

$$\begin{pmatrix} a & b & c \\ b & e & f \\ c & f & g \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & e & f \\ c & f & g \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & e & f \\ c & f & g \end{pmatrix}$$

- $\mathbf{A}$ is positive definite if there exists a unique lower triangular matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ with positive diagonal components such that $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ (Cholesky Decomposition).

Examples

- $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$ is *positive definite*

  (The eigenvalues are $2-\sqrt{2}, 2+\sqrt{2}$ and 2).
- $\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$ is *positive semi-definite*
- $\begin{pmatrix} 1 & -2 & 4 \\ -2 & 2 & 0 \\ 4 & 0 & -7 \end{pmatrix}$ is *indefinite*

Solution of $\mathbf{Ax} = \mathbf{b}$

- Let $\mathbf{A} \in \mathbb{R}^{n \times n}$, symmetric and positive definite
- Solution of $\mathbf{Ax} = \mathbf{b}$ is $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$
- Instead, use Cholesky decomposition of $\mathbf{A}$, $\mathbf{A} = \mathbf{L}\mathbf{L}^T$
- The given system of equations is $\mathbf{L}\mathbf{L}^T\mathbf{x} = \mathbf{b}$
- Solve the *triangular* system, $\mathbf{Ly} = \mathbf{b}$ using *forward substitution* to get $\mathbf{y}$.
- Solve the *triangular* $\mathbf{L}^T\mathbf{x} = \mathbf{y}$ using *backward substitution* to get $\mathbf{x}^*$.
- Cholesky decomposition is a *numerically stable* procedure

Solution of $\mathbf{Ax} = \mathbf{b}$

- $\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 4 \\ -4 \end{pmatrix}$
- Cholesky decomposition of $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ gives

$$\mathbf{L} = \begin{pmatrix} 1.4142 & 0 & 0 \\ -0.7071 & 1.2247 & 0 \\ 0 & -0.8165 & 1.1547 \end{pmatrix}$$

- Solution of $\mathbf{Ly} = \mathbf{b}$ gives $\mathbf{y} = \begin{pmatrix} 0 \\ 3.2660 \\ -1.1547 \end{pmatrix}$
- Solution of $\mathbf{L}^T\mathbf{x} = \mathbf{y}$ results in

$$\mathbf{x}^* = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

Some References

- Strang G., Linear Algebra and Its Applications, Thomson-Brooks/Cole (2006).
- Golub G. H. and Van Loan C. F., Matrix Computations, The Johns Hopkins University Press (1996), Hindustan Book Agency (India) (2007).