Lecture 0: Review of Calculus and Linear Algebra

CS4017: Introduction to Optimization

Join the group

群聊：CS4017 Intro to Optimization

该二维码7天内(3月3日前)有效，重新进入将更新

Outline

Vector norm

A **norm** $\|\cdot\|$ is a function from $\mathbb{R}^n \to \mathbb{R}$ that satisfies the following properties:

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

- used to measure the length of a vector $\mathbf{x}$, *i.e.*, $\|\mathbf{x}\|$
- the distance between two vectors $\mathbf{x}, \mathbf{y}$ can be measured by $\|\mathbf{x} - \mathbf{y}\|$

Examples

- p-**norm** $(p \geq 1)$ or Hölder norm: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$
- **2-norm** or Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} = (\mathbf{x}^\top \mathbf{x})^{1/2}$
- **1-norm** or Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- ∞-**norm** or maximum norm or sup-norm: $\|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$

Usual notation: ℓ_p-norm, ℓ_2-norm, ℓ_1-norm, and ℓ_∞-norm, respectively.

The following is not a norm but it is denoted as such:

- **0-"norm"** or ℓ_0-"norm": $\|\mathbf{x}\|_0 = \sum_{i=1}^n 1_{\{x_i \neq 0\}}$ = number of non-zero elements of $\mathbf{x}$

Matrix norm

Matrix norms are usually induced by vector norms. For $\mathbf{A} \in \mathbb{R}^{m \times n}$

- p-**norm**: $\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$
- **2-norm**: $\|\mathbf{A}\|_2 = \lambda_{\max}(\mathbf{A}^\top \mathbf{A})^{\frac{1}{2}} = \sigma_{\max}(\mathbf{A})$, *i.e.*, the largest singular value of $\mathbf{A}$
- **1-norm**: $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|$, *i.e.*, the maximum column absolute sum
- ∞-**norm**: $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$, *i.e.*, the maximum row absolute sum

The following norms are **not** induced by vector norms:

- **Frobenius norm**: $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$
- **nuclear norm**: $\|\mathbf{A}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})$

Inner product and angle

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

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

- The inner product of two matrices $X, Y \in \mathbb{R}^{m \times n}$ is defined as

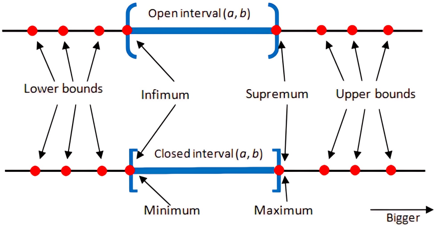
$$\langle X, Y \rangle = \mathbf{tr}\left(X^\top Y\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} y_{ij}$$

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

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

Supremum and infimum

- $C \subset \mathbb{R}$ is a set of real numbers.
  - The smallest upper bound of C: **supremum**, denoted $\sup C$.
  - The largest lower bound of C: **infimum**, denoted $\inf C$.
- If C is unbounded from above, then $\sup C = \infty$; if C is unbounded from below, then $\inf C = -\infty$
- Let $\emptyset$ denote the empty set, then we define $\sup \emptyset = -\infty$, $\inf \emptyset = \infty$
- If $\sup C$ or $\inf C \in C$, we say the supremum/infimum of C is attained or achieved, *i.e.*, it is the maximum/minimum of the set



Open and closed sets

- An element $\mathbf{x} \in C \subseteq \mathbb{R}^n$ is called an **interior point** of C if there exists an $\epsilon > 0$ for which

$$\{\mathbf{y} \mid \|\mathbf{y} - \mathbf{x}\| \leq \epsilon\} \subseteq C,$$

  i.e., there exists a ball centered at $\mathbf{x}$ that lies entirely in C.
- The set of all interior points of C is called the **interior** of C and is denoted as **int** C.
- A set is **open** if **int** $C = C$, *i.e.*, every point in C is an interior point.
- A set is **closed** if its complement $\mathbb{R}^n \setminus C := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \notin C\}$ is open.
- The **closure** of a set C is defined as

$$\textbf{cl } C = \mathbb{R}^n \setminus \textbf{int } (\mathbb{R}^n \setminus C).$$

  A point $\mathbf{x}$ is in the closure of C if for every $\epsilon > 0$, there is a $\mathbf{y} \in C$ with $\|\mathbf{x} - \mathbf{y}\| \leq \epsilon$.

Open and closed sets

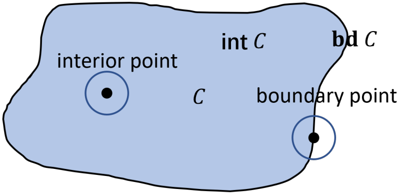
- The boundary of the set C is defined as

$$\textbf{bd } C = \textbf{cl } C \setminus \textbf{int } C.$$

- A boundary point $\mathbf{x}$ (*i.e.*, $\mathbf{x} \in \textbf{bd } C$) satisfies: for all $\epsilon > 0$, there exist $\mathbf{y} \in C$ and $\mathbf{z} \notin C$ with

$$\|\mathbf{y} - \mathbf{x}\|_2 \leq \epsilon, \qquad \|\mathbf{z} - \mathbf{x}\|_2 \leq \epsilon.$$

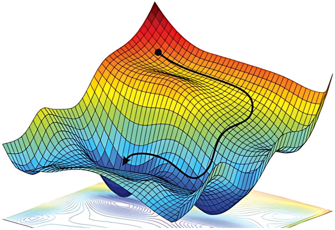
- C is **closed** if it contains its boundary, *i.e.*, $\textbf{bd } C \subseteq C$. It is **open** if it contains no boundary points, *i.e.*, $C \bigcap \textbf{bd } C = \emptyset$.



Gradient

For $f : \mathbb{R}^n \to \mathbb{R}$, the gradient is the vector (we use column vector as the convention) of partial derivatives:

$$\nabla f(\mathbf{x})_i = \frac{\partial f(\mathbf{x})}{\partial x_i}, \quad i = 1, \ldots, n.$$



Gradient rules for vectors

Suppose $\mathbf{a}, \mathbf{A}$ are not functions of $\mathbf{x}$, and $\mathbf{u} = \mathbf{u}(\mathbf{x}), \mathbf{v} = \mathbf{v}(\mathbf{x})$ are vectors, then

- $\frac{\partial \mathbf{a}}{\partial \mathbf{x}} = \mathbf{0}$
- $\frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{I}$
- $\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$
- $\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$
- $\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
- $\frac{\partial (\mathbf{u}\cdot\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \cdot \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \cdot \mathbf{u}$
  (Chain rule)
- $\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$
- $\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$

Find more in `https://en.wikipedia.org/wiki/Matrix_calculus` (the *denominator layout*).

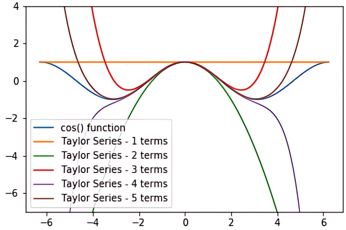
Taylor expansion

The second-order Taylor expansion for $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x}_0$ can be written as

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \nabla f(\mathbf{x})^\top (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \nabla^2 f(\mathbf{x})(\mathbf{x} - \mathbf{x}_0),$$

where $\nabla^2 f(\mathbf{x})$ is the $n \times n$ Hessian matrix of f at $\mathbf{x}$, given by

$$\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}.$$



Outline

Rank

The **rank** of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted by $\mathbf{rank}\,\mathbf{A}$, is defined as

- the maximum number of linearly independent rows of $\mathbf{A}$
  $=$ the maximum number of linearly independent columns of $\mathbf{A}$.

Key properties:

- $\mathbf{rank}\,\mathbf{A} = \mathbf{rank}\,\mathbf{A}^\top$
- $\mathbf{rank}\,\mathbf{A} \leq \min\{m, n\}$
- $\mathbf{rank}\,k\mathbf{A} = \mathbf{rank}\,\mathbf{A}$ for $k \neq 0$
- $\mathbf{rank}\,\mathbf{A} = 0$ iff $\mathbf{A} = \mathbf{0}$
- $\mathbf{rank}\,\mathbf{A} = \mathbf{rank}\,\mathbf{A}\mathbf{A}^\top = \mathbf{rank}\,\mathbf{A}^\top\mathbf{A}$
- $\mathbf{rank}(\mathbf{A} + \mathbf{B}) \leq \mathbf{rank}\,\mathbf{A} + \mathbf{rank}\,\mathbf{B}$
- $\mathbf{rank}\,\mathbf{A}\mathbf{B} \leq \min\{\mathbf{rank}\,\mathbf{A}, \mathbf{rank}\,\mathbf{B}\}$
- $\mathbf{rank}\,\mathbf{A}\mathbf{B} \geq \mathbf{rank}\,\mathbf{A} + \mathbf{rank}\,\mathbf{B} - n$

Rank

The **rank** of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted by $\mathbf{rank}\,\mathbf{A}$, is defined as

- the maximum number of linearly independent rows of $\mathbf{A}$
  $=$ the maximum number of linearly independent columns of $\mathbf{A}$.

We say that

- $\mathbf{A}$ has full column rank if $m \geq n, \mathbf{rank}\,\mathbf{A} = n$
- $\mathbf{A}$ has full row rank if $m \leq n, \mathbf{rank}\,\mathbf{A} = m$
- $\mathbf{A}$ has full rank if $\mathbf{rank}\,\mathbf{A} = \min\{m, n\}$
- Otherwise it is rank deficient

Range and null space

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

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

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

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

We also have the row space $\mathcal{R}(\mathbf{A}^\top)$ and the left null space $\mathcal{N}(\mathbf{A}^\top)$

Orthogonal complement

Definition

The **orthogonal complement** of a subspace $\mathcal{V}$ is

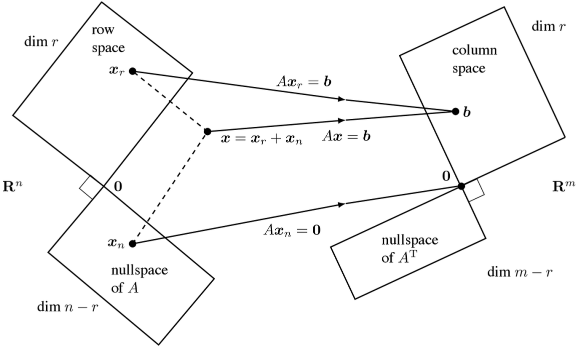
$$\mathcal{V}^\perp = \{\mathbf{x} \mid \mathbf{z}^\top \mathbf{x} = 0 \text{ for all } \mathbf{z} \in \mathcal{V}\}.$$

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

- $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^\top)^\perp$
  - $\mathcal{N}(\mathbf{A}) \oplus \mathcal{R}(\mathbf{A}^\top) = \mathbb{R}^n$
- $\mathcal{R}(\mathbf{A}) = \mathcal{N}(\mathbf{A}^\top)^\perp$
  - $\mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^\top) = \mathbb{R}^m$

$\oplus$ refers to the *orthogonal direct sum*.

The 4 subspaces



Find more in
`https://blog.csdn.net/Insomnia_X/article/details/125930326`

Eigenvalues

Definition

Given a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ if $\mathbf{v} \neq \mathbf{0}$ satisfies

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}, \qquad \text{for some } \lambda \in \mathbb{C}, \mathbf{v} \in \mathbb{C}^n$$

Then we call

- λ an **eigenvalue** of $\mathbf{A}$
- $\mathbf{v}$ an **eigenvector** of $\mathbf{A}$ associated with λ
- $(\mathbf{v}, \lambda)$ an eigen-pair of $\mathbf{A}$

Eigenvalues

Some properties:

- $(\alpha\mathbf{v}, \lambda)$ is also an eigen-pair for any $\alpha \in \mathbb{C}, \alpha \neq 0$
- $\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \Leftrightarrow \lambda \in \{\lambda_1, \ldots, \lambda_n\}$
- $\det(\mathbf{A}) = \Pi_{i=1}^{n} \lambda_i$
- $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$
- Eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$

Eigen-decomposition

A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be **diagonalizable**, or admit an **eigen-decomposition**, if there exists a nonsingular $\mathbf{V} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{\Lambda} = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$. For $\mathbf{V}$ (column-wise concatenation of eigenvectors), we have

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \Leftrightarrow \mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda} \Leftrightarrow \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

Necessary and sufficient condition for the existence of eigen-decomposition

A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ admits an eigen-decomposition if and only if there exists n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbf{A}$.

Singular value decomposition (SVD)

Theorem

Given any $\mathbf{A} \in \mathbb{R}^{m\times n}$, there exists a 3-tuple $(\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}) \in \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times n}$ such that

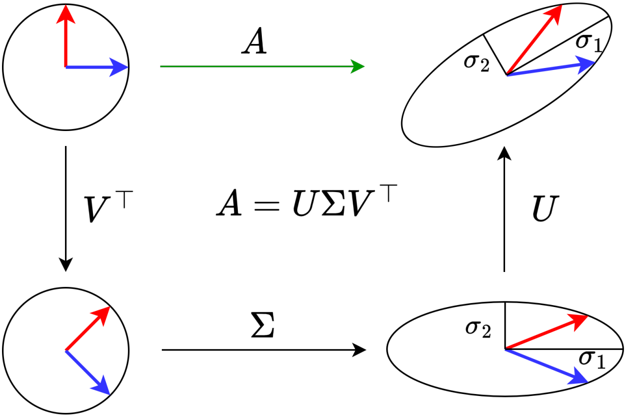
$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top,$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, and $\boldsymbol{\Sigma}$ takes the form

$$[\Sigma]_{ij} = \begin{cases} \sigma_i, & i = j \\ 0, & i \neq j \end{cases}, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0,\ p = \min\{m, n\}.$$

- σ_i is called the ith **singular value**, $\sigma_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}^\top\mathbf{A})}$.
- $\mathbf{u}_i$ and $\mathbf{v}_i$ are called the ith left and right singular vectors, respectively
  - $\mathbf{u}_i^\top\mathbf{A} = \sigma_i\mathbf{v}_i^\top$
  - $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$

Singular value decomposition (SVD)



Reduced SVD

Let r be the number of nonzero singular values, and

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$$

We have

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^\top \\ \mathbf{V}_2^\top \end{bmatrix} = \mathbf{U}_1 \tilde{\mathbf{\Sigma}} \mathbf{V}_1^\top$$

where

- $\mathbf{U}_1 = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{R}^{m \times r}$
- $\tilde{\mathbf{\Sigma}} = \mathbf{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r \times r}$
- $\mathbf{V}_1 = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \in \mathbb{R}^{n \times r}$
- outer-product form: $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$

Read more about SVD in
`https://www.cnblogs.com/pinard/p/6251584.html`.

Pseudo-inverse

The **pseudo-inverse** (or Moore-Penrose inverse) of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\mathbf{A}^\dagger = \mathbf{V}_1 \tilde{\mathbf{\Sigma}}^{-1} \mathbf{U}_1^\top \in \mathbb{R}^{n \times m}.$$

- $\mathbf{A}^\dagger$ always exists and is unique
- outer-product form: $\mathbf{A}^\dagger = \sum_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^\top$

In general $\mathbf{A}\mathbf{A}^\dagger \neq \mathbf{I}$ and $\mathbf{A}^\dagger\mathbf{A} \neq \mathbf{I}$.

Find more in `https://www.cnblogs.com/bigmonkey/p/12070331.html`.

Pseudo-inverse

Special cases of pseudo-inverse

- $\mathbf{A}$ has full-column rank
  - $\mathbf{A}^\dagger = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$
  - $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}$ (left inverse)

- $\mathbf{A}$ has full-row rank
  - $\mathbf{A}^\dagger = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^{-1}$
  - $\mathbf{A} \mathbf{A}^\dagger = \mathbf{I}$ (right inverse)

- $\mathbf{A}$ is square and has full rank
  - $\mathbf{A}^\dagger = \mathbf{A}^{-1}$

Positive semidefinite

A matrix $\mathbf{A} \in \mathbb{S}^n$ ($\mathbb{S}^n$ denotes the set of symmetric matrices; they have real eigenvalues) is said to be

- **positive semidefinite (PSD)**, $\mathbf{A} \succeq \mathbf{0}$: if $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
- **positive definite (PD)**, $\mathbf{A} \succ \mathbf{0}$: if $\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
- **negative (semi)definite** when $-\mathbf{A}$ is positive (semi)definite
- **indefinite**, $\mathbf{A} \not\succeq \mathbf{0}$: if neither $\mathbf{A}$ and $-\mathbf{A}$ are PSD

Properties

Let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$, then

- $\mathbf{A} \succeq \mathbf{0}, \alpha \geq 0$ $(\mathbf{A} \succ \mathbf{0}, \alpha > 0) \Rightarrow \alpha\mathbf{A} \succeq \mathbf{0}$ $(\alpha\mathbf{A} \succ \mathbf{0})$
- $\mathbf{A}, \mathbf{B} \succeq \mathbf{0}$ $(\mathbf{A} \succeq \mathbf{0}, \mathbf{B} \succ \mathbf{0}) \Rightarrow \mathbf{A} + \mathbf{B} \succeq \mathbf{0}$ $(\mathbf{A} + \mathbf{B} \succ \mathbf{0})$
- $\mathbf{A} \succ \mathbf{0} \Leftrightarrow \mathbf{A}^{-1} \succ \mathbf{0}$
- $\mathbf{A} \succeq \mathbf{0}$ $(\mathbf{A} \succ \mathbf{0}) \Leftrightarrow \lambda_i \geq 0$ $(\lambda_i > 0)$ for $i = 1, \ldots, n$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues
- $\mathbf{A} \succeq \mathbf{0}$ $(\mathbf{A} \succ \mathbf{0}) \Rightarrow \mathrm{tr}(\mathbf{A}) \geq 0$ and $\det(\mathbf{A}) \geq 0$ $\big(\mathrm{tr}(\mathbf{A}) > 0$ and $\det(\mathbf{A}) > 0\big)$
- $\mathbf{A} \not\succeq \mathbf{0} \Leftrightarrow \lambda_i > 0$ for some i and $\lambda_i < 0$ for some i

Properties

Let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$, then

- $\mathbf{A} \succeq \mathbf{0} \Rightarrow a_{ii} \geq 0$ for all i
- $\mathbf{A} \succ \mathbf{0} \Rightarrow a_{ii} > 0$ for all i
- $\mathbf{A} \in \mathbb{S}^n$ is PSD iff it can be factored as $\mathbf{A} = \mathbf{C}^\top \mathbf{C}$ for some $\mathbf{C} \in \mathbb{R}^{m \times n}$

Definition

$\mathbf{A} \succeq \mathbf{B}$ $(\mathbf{A} \succ \mathbf{B}, \mathbf{A} \not\succeq \mathbf{B})$ means that $\mathbf{A} - \mathbf{B}$ is PSD (PD, indefinite)

Find more in `https://en.wikipedia.org/wiki/Definite_matrix`.

Determinant

The **determinant** is defined for a square matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ (`https://en.wikipedia.org/wiki/Determinant`). Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times m}$, then some key properties are as follows:

- $\mathbf{A}$ is nonsingular / invertible iff $\det(\mathbf{A}) \neq 0$
- $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$
- $\det(\mathbf{A}^\top) = \det(\mathbf{A})$
- $\det(\alpha \mathbf{A}) = \alpha^m \det(\mathbf{A})$ for any $\alpha \in \mathbb{R}$
- $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$ for any nonsingular $\mathbf{A}$
- if $\mathbf{A}$ is triangular, then $\det(\mathbf{A}) = \Pi_{i=1}^m a_{ii}$
- $\det(\mathbf{A}) = \Pi_{i=1}^n \lambda_i$

Trace

Let $\mathbf{A} \in \mathbb{R}^{m \times m}$, the **trace** of $\mathbf{A}$ is defined as $\text{tr}(\mathbf{A}) := \sum_{i=1}^{n} a_{ii}$. Main properties:

- $\text{tr}(\alpha \mathbf{A}) = \alpha \text{tr}(\mathbf{A})$
- $\text{tr}(\mathbf{A}^\top) = \text{tr}(\mathbf{A})$
- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$
  - $\text{tr}(\mathbf{b}\mathbf{a}^\top) = \mathbf{a}^\top \mathbf{b}$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$
  - $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB})$ (cyclic permutation)
  - $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CBA}) = \text{tr}(\mathbf{ACB})$ if $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{S}^m$ (any permutation is allowed if all matrices are symmetric)
- $\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$

Thank you

Any questions?