Module 0 — Mathematical Background

Dr. Ahmad F. Taha

CE 5999-02 Special Topics — Intro to Optimization

Email: ahmad.taha@vanderbilt.edu

Webpage: http://lab.vanderbilt.edu/taha

VANDERBILT UNIVERSITY

January 12, 2023

Domain of a function

- **Domain** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set of values of x such that $f(x)$ is defined (that is, $f(x)$ is a number in $\mathbb{R}$)

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n \mid -\infty < f(x) < \infty\}$$

  - $f(x) = a^T x, \mathrm{dom}(f) = \mathbb{R}^n$
  - $f(x) = \log x$ (the natural logarithm, $\ln x$), $\mathrm{dom}(f) = \mathbb{R}_{++}$
- In convex optimization, it is convenient to consider functions that take the value ∞ or $-\infty$
- More generally, we may have a function $f : \mathbb{R}^n \to \mathbb{R}^k$ that returns a vector in $\mathbb{R}^k$
  - The domain is the set of all x such that every entry of $f(x)$ is a number (not ∞ or $-\infty$ or undefined)

Range; image and inverse image

Consider a function $f : \mathbb{R}^n \to \mathbb{R}^k$.

Range is the set of all possible function values $f(x)$:

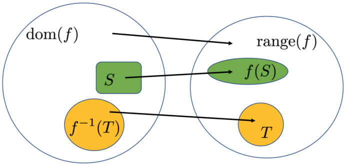
$$\text{range}(f) = \{f(x) \,|\, x \in \text{dom}(f)\}$$

The **image** of a set $S \subset \mathbb{R}^n$ under f is defined as

$$f(S) = \{f(x) \,|\, x \in S\} \subset \mathbb{R}^k$$

The **inverse image** of a set $T \subset \mathbb{R}^k$ under f is defined as

$$f^{-1}(T) = \{x \in \mathbb{R}^n \,|\, f(x) \in T\} \subset \mathbb{R}^n$$



Gradient vector and Hessian matrix

Consider a function $f(x) : \mathbb{R}^n \to \mathbb{R}$

Gradient vector of $f(x)$ at point y (column vector of length n):

$$\nabla f(y) = \begin{bmatrix} \frac{\partial f(y)}{\partial x_1} \\ \vdots \\ \frac{\partial f(y)}{\partial x_n} \end{bmatrix}$$

Hessian matrix of $f(x)$ at point y (symmetric matrix $n \times n$):

$$\nabla^2 f(y) = \begin{bmatrix} \frac{\partial^2 f(y)}{\partial^2 x_1} & \dots & \frac{\partial^2 f(y)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(y)}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f(y)}{\partial^2 x_n} \end{bmatrix}$$

Linear, affine, and quadratic functions

- A function $f:\mathbb{R}^n\rightarrow\mathbb{R}$ is called *linear* if it is written in the form

$$f(x)=a^Tx$$

  - $\nabla f(x)=a\,;\nabla^2 f(x)=0$
  - For scalar x: $(ax)'=a$ and $(ax)''=0$

- A function $f:\mathbb{R}^n\rightarrow\mathbb{R}$ is called *affine* if it is written in the form

$$f(x)=a^Tx+b$$

- A function $f:\mathbb{R}^n\rightarrow\mathbb{R}$ is called *quadratic* if it is written in the form

$$f(x)=\frac{1}{2}x^TPx+q^Tx+r$$

  - $\nabla f(x)=Px+q\,;\nabla^2 f(x)=P$
  - For scalar x: $\left(\frac{1}{2}Px^2+qx+r\right)'=Px+q$, $\left(\frac{1}{2}Px^2+qx+r\right)''=P$

Gradient of linear function

- Consider the column vectors $a \in \mathbb{R}^n$, $x \in \mathbb{R}^n$

$$f(x) = a^T x = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

- Take partial derivative with respect to x_k $(k = 1, \ldots, n)$

$$\frac{\partial (a^T x)}{\partial x_k} = a_k$$

- Organize everything in a vector

$$\nabla (a^T x) = \begin{bmatrix} \frac{\partial (a^T x)}{\partial x_1} \\ \vdots \\ \frac{\partial (a^T x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a$$

Gradient of quadratic function $\frac{1}{2}x^T Px$

- Matrix P has entries p_{ij} with $p_{ij} = p_{ji}$ for $i \neq j$ (symmetric matrix)

$$\frac{1}{2}x^T Px = \frac{1}{2}\sum_{i=1}^{n} p_{ii}x_i^2 + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} p_{ij}x_i x_j$$

- Take partial derivative with respect to x_k ($k = 1, \ldots, n$)
- The double summation has two parts that will give nonzero derivative: one for $i = k$ (the sum has $p_{kj}x_k x_j$); and another for all $i \neq k$, where we pick up a term $p_{ik}x_i x_k$

$$\frac{\partial}{\partial x_k}\left(\frac{1}{2}x^T Px\right) = p_{kk}x_k + \frac{1}{2}\sum_{j=1, j\neq k}^{n} p_{kj}x_j + \frac{1}{2}\sum_{i=1, i\neq k}^{n} p_{ik}x_i$$

$$= \frac{1}{2}\sum_{j=1}^{n} p_{kj}x_j + \frac{1}{2}\sum_{i=1}^{n} p_{ik}x_i = \frac{1}{2}\sum_{j=1}^{n} p_{kj}x_j + \frac{1}{2}\sum_{j=1}^{n} p_{jk}x_j \quad \text{[rename } i \text{ to } j\text{]}$$

$$= \sum_{j=1}^{n} p_{kj}x_j \quad [p_{kj} = p_{jk} \text{ for symmetric matrix}]$$

Gradient and Hessian of quadratic function $\frac{1}{2}x^TPx$

- Organize the partial derivatives in vector form

$$\nabla\left(\frac{1}{2}x^TPx\right) = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{1}{2}x^TPx\right) \\ \vdots \\ \frac{\partial}{\partial x_n}\left(\frac{1}{2}x^TPx\right) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n p_{1j}x_j \\ \vdots \\ \sum_{j=1}^n p_{nj}x_j \end{bmatrix} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \dots & \dots & \dots \\ p_{n1} & \dots & p_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = Px$$

- Second partial derivative

$$\frac{\partial^2}{\partial x_i \partial x_k}\left(\frac{1}{2}x^TPx\right) = \frac{\partial}{\partial x_i}\sum_{j=1}^n p_{kj}x_j = p_{ki} = p_{ik}$$

- Organize everything in matrix form

$$\nabla^2\left(\frac{1}{2}x^TPx\right) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1}\left(\frac{1}{2}x^TPx\right) & \dots & \frac{\partial^2}{\partial x_1 \partial x_n}\left(\frac{1}{2}x^TPx\right) \\ \dots & \dots & \dots \\ \frac{\partial^2}{\partial x_n \partial x_1}\left(\frac{1}{2}x^TPx\right) & \dots & \frac{\partial^2}{\partial x_n \partial x_n}\left(\frac{1}{2}x^TPx\right) \end{bmatrix} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \dots & \dots & \dots \\ p_{n1} & \dots & p_{nn} \end{bmatrix} = P$$

Linear and quadratic approximations of a function

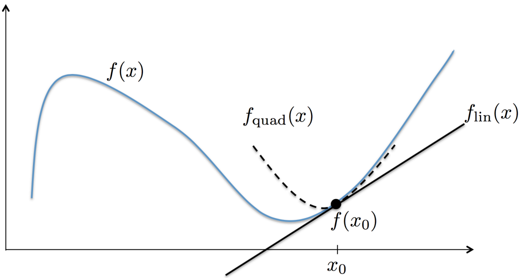
Consider a function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

Linear approximation (1st-order Taylor approximation) of $f(x)$ around point x_0

$$f_{\text{lin}}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0)$$

Quadratic approximation (2nd-order Taylor approximation) of $f(x)$ around point x_0

$$f_{\text{quad}}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$



Example: Gradient computation

- Find the linear approximation of $f(x_1, x_2) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}(x_2 - 2)^2$ at $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$
- Quadratic function

$$f(x_1, x_2) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}(x_2 - 2)^2 = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1 - 2x_2 + 2.5$$

$$= \frac{1}{2}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2.5 = \frac{1}{2}x^T P x + q^T x + r$$

- Gradient vector

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f(x_1, x_2)}{x_1} \\ \frac{\partial f(x_1, x_2)}{x_2} \end{bmatrix} = \begin{bmatrix} x_1 - 1 \\ x_2 - 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \end{bmatrix} = Px + q$$

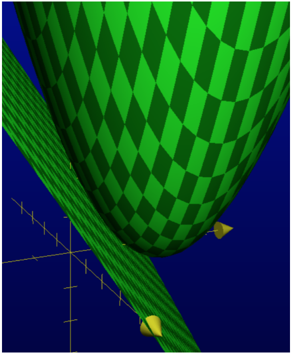
- Gradient vector evaluated at $x_1 = 0.5$, $x_2 = 0.5$

$$\nabla f(0.5, 0.5) = \begin{bmatrix} \frac{\partial f}{x_1} \\ \frac{\partial f}{x_2} \end{bmatrix} = \begin{bmatrix} -0.5 \\ -1.5 \end{bmatrix}$$

Example: Linear approximation

Linear approximation around $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

$$\begin{aligned} f_{\text{lin}}(x_1, x_2) &= f(0.5, 0.5) + \nabla f(0.5, 0.5)^T \begin{bmatrix} x_1 - 0.5 \\ x_2 - 0.5 \end{bmatrix} \\ &= 1.25 + \begin{bmatrix} -0.5 \\ -1.5 \end{bmatrix}^T \begin{bmatrix} x_1 - 0.5 \\ x_2 - 0.5 \end{bmatrix} \\ &= 1.25 - 0.5(x_1 - 0.5) - 1.5(x_2 - 0.5) \\ &= -0.5x_1 - 1.5x_2 + 2.25 \end{aligned}$$



Interior points and open sets

The **ball** of radius $\epsilon > 0$ around x is $\{y | \|y - x\| \leq \epsilon\}$.

The notion of a ball is used to define *open sets* and the *interior of a set*.

Consider a set S and a point $x \in S$. The point x is called an **interior point** of S if there is a ball around x that is contained entirely in S, or in other words, if there is an $\epsilon > 0$ such that $\{y \mid \|y - x\| \leq \epsilon\} \subset S$.

The set of all interior points of S is called the **interior** of S and is denoted by $\text{int}(S)$.

A set $S \subset \mathbb{R}^n$ is **open** if for every $x \in S$ there is an $\epsilon > 0$ such that
$\{y \mid \|y - x\| \leq \epsilon\} \subset S$.
In other words, a set is open if it contains only its interior points.

Closed sets

A set is **closed** if and only if its complement is open.

There are other ways to define closed sets:

Consider a set $S \subset \mathbb{R}^n$. A point $x \in \mathbb{R}^n$ is called **limit of the set** S if there is a sequence of points in S converging to x.

A set is closed if it contains all its limits.

Boundary and closure

Consider a set $S \subset \mathbb{R}^n$.

A point x is called a **boundary point** of S if every ball around x, however small, contains a point in S and a point not in S.

The boundary points of a set form the **boundary** of the set, which is denoted as $\text{bd}(S)$.

A set is closed if it contains its boundary. A set is open if it contains no boundary points.

The **closure** of a set is defined as the set together with its boundary points.

$$\text{cl}(S) = S \cup \text{bd}(S)$$

A set may be open, closed, both, or neither.

Least upper bound (supremum)

- Consider a set on the real line, $S \subset \mathbb{R}$
- A number $a \in \mathbb{R}$ is an **upper bound** of S if $x \leq a$ for all $x \in S$.
- The **least upper bound** of a set S, denoted by $\sup S$, is defined by two properties:
  1. it is an upper bound of S
  2. it is less than all other upper bounds of S
- There are three cases:
  1. S is empty, and then, $\sup S = -\infty$
  2. S is unbounded above (e.g., $S = [5, \infty)$), and then, $\sup S = \infty$
  3. S is bounded above, and then $\sup S$ is a number

Greatest lower bound (infimum)

- Consider a set on the real line, $S \subset \mathbb{R}$
- A number $b \in \mathbb{R}$ is a **lower bound** of S if $x \geq b$ for all $x \in S$.
- The **greatest lower bound** of a set S, denoted by $\inf S$, is defined by two properties:
  1. it is a lower bound of S
  2. it is greater than all other lower bounds of S
- There are three cases:
  1. S is empty, and then, $\inf S = \infty$
  2. S is unbounded below (e.g., $S = (-\infty, -1]$), and then, $\inf S = -\infty$
  3. S is bounded below, and then $\inf S$ is a number

$\max$ and $\min$ vs. $\sup$ and $\inf$

- Consider a set on the real line, $S \subset \mathbb{R}$
- The **maximum** of S is a number $x \in S$ such that for all $y \in S$, it holds that $y \leq x$
- The thing to note: $\max$ of a set must be a number ***in the set***
- $\sup S$ and $\inf S$ are useful because they are always defined, even when $\max$ and $\min$ are not defined
- For example, if $S = (0, 1)$, then $\max S$ is not defined, but $\sup S = 1$
- If S is closed, then $\max$ and $\sup$ are the same (likewise, $\min$ and $\inf$)

Optimal value of an optimization problem

We denote the optimal value of an optimization problem as follows

$$f^* = \text{minimize } f_0(x)$$
$$\text{subj. to } f_i(x) \leq 0, i = 1, \ldots, m$$
$$h_i(x) = 0, i = 1, \ldots, p.$$

We mean that we have to find the image of the feasible set

$$\mathcal{F} = \{x \in \mathbb{R}^n \,|\, f_i(x) \leq 0, i = 1, \ldots, m; h_i(x) = 0, i = 1, \ldots, p\}$$

under f_0,

$$S = \{f_0(x) \,|\, x \in \mathcal{F}\}$$

and then, f^* is the greatest lower bound of that set: $f^* = \inf S$.

Therefore, we could write the optimization problem as

$$f^* = \inf\{f_0(x) \,|\, f_i(x) \leq 0, i = 1, \ldots, m; h_i(x) = 0, i = 1, \ldots, p\}.$$

Vector Norms

The **norm** of a vector $x \in \mathbb{R}^n$ is a function $\|.\| : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following three properties for all $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, and $a \in \mathbb{R}$.

1. *Positive definiteness:* $\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x = 0$
2. *Scaling:* $\|ax\| = |a|\|x\|$
3. *Triangle inequality:* $\|x + y\| \leq \|x\| + \|y\|$

- We will use the notation $\|.\|$ for any norm satisfying the previous properties
- We will use the notation $\|.\|_p$ for a specific norm, to be defined shortly
- The norm is also called the **length** of the vector

Distance

The **distance**, or **metric**, between two points in $\mathbb{R}^n$ is defined as

$$d(x, y) = \|x - y\|.$$

The distance satisfies the following three properties.

1. *Positive definiteness:* $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$
2. *Symmetry:* $d(x, y) = d(y, x)$
3. *Triangle inequality:* $d(x, z) \leq d(x, y) + d(y, z)$

ℓ_p norms

For $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $p \geq 1$, the ℓ_p norm is defined as

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

1. $p = 2$: Euclidean norm $\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^n x_i^2}$
   - Inner product $x^T x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \ldots + x_n^2 = \|x\|_2^2$
2. $p = 1$: sum-abs-values $\|x\|_1 = \sum_i |x_i|$
3. $p = \infty$: max-abs-value $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$

Cauchy-Schwartz and Hölder inequalities

Consider vectors $x = [x_1, \ldots, x_n]^T$ and $y = [y_1, \ldots, y_n]^T$.

Cauchy-Schwartz inequality:

$$|x^T y| \leq \|x\|_2 \|y\|_2$$

Equality holds if and only if the vectors x and y are linearly dependent, that is, if there are $a \in \mathbb{R}$ and $b \in \mathbb{R}$, not both zero, such that $ax + by = 0$.

Hölder's inequality: If $\frac{1}{p} + \frac{1}{q} = 1$ with $p \geq 1$ ($p = 1$ gives $q = \infty$), then

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \|y\|_q$$

Equality holds if and only if there are $a \geq 0$ and $b \geq 0$, not both zero, such that $a|x_i|^p = b|y_i|^q$ for all i.

Dual norms

Consider a norm $\|.\|$ on $\mathbb{R}^n$.

The dual norm is defined as

$$\|z\|_* = \sup\{z^T x \mid \|x\| \leq 1\}$$

For example, the dual of the Euclidean norm is the Euclidean norm:

$$\sup\{z^T x \mid \|x\|_2 \leq 1\} = \|z\|_2$$

Proof: Use Cauchy-Schwartz inequality

The dual of ℓ_p norm is the ℓ_q norm, where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof: Use Hölder's inequality

Matrix Norms

Matrix norm is a function $\|\cdot\| : K^{m\times n} \to \mathbb{R}$ that must satisfy the following properties for all scalars α and matrices $A, B \in K^{m\times n}$

- $\|A\| \geq 0$ (positive-valued)
- $\|A\| = 0 \iff A = 0_{m,n}$ (definite)
- $\|\alpha A\| = |\alpha|\|A\|$ (absolutely homogeneous)
- $\|A + B\| \leq \|A\| + \|B\|$ (sub-additive or satisfying the triangle inequality)
- $\|AB\| \leq \|A\|\|B\|$ (sub-multiplicative property—super useful)

Specific Matrix Norms

- Frobenius-norm:
$$\|A\|_{\mathsf{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\operatorname{trace}\left(A^*A\right)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)}$$
- 1-norm: $\|A\|_1 = \max\limits_{1\leq j\leq n} \sum\limits_{i=1}^{m} |a_{ij}|$: max absolute column sum of the matrix
- 2-norm: $\|A\|_2 = \sqrt{\lambda_{\max}\left(A^*A\right)} = \sigma_{\max}(A)$ where $\sigma_{\max}(A)$ is the largest singular value of matrix A and A^* denotes the conjugate transpose
  - Note that $\|A\|_2 = \sigma_{\max}(A) \leq \|A\|_{\mathrm{F}} = \left(\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}$
- Infinity-norm: $\|A\|_\infty = \max\limits_{1\leq i\leq m} \sum\limits_{j=1}^{n} |a_{ij}|$: maximum absolute row sum of the matrix
- Max-norm: $\|A\|_{\max} = \max_{ij} |a_{ij}|$
- Nuclear-norm: $\|A\|_* = \operatorname{trace}\left(\sqrt{A^*A}\right) = \sum\limits_{i=1}^{\min\{m,n\}} \sigma_i(A)$

Norm bounds

For matrix $A \in \mathbb{R}^{m \times n}$ of rank r, the following inequalities hold

- $\|A\|_2 \leq \|A\|_F \leq \sqrt{r}\|A\|_2$
- $\|A\|_F \leq \|A\|_* \leq \sqrt{r}\|A\|_F$
- $\|A\|_{\max} \leq \|A\|_2 \leq \sqrt{mn}\|A\|_{\max}$
- $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty$
- $\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1$
- $\|A\|_2 \leq \sqrt{\|A\|_1\|A\|_\infty}$

Interpretation of matrix-vector multiplication: Column version

Consider a matrix $A \in \mathbb{R}^{m\times n}$ with columns $a_i \in \mathbb{R}^m$, $i = 1, \ldots, n$:

$$A = [a_1 \cdots a_n].$$

The matrix-vector multiplication Ax, where $x \in \mathbb{R}^n$, can be interpreted as a linear combination of the columns of A, weighted by the entries of x:

$$Ax = x_1 a_1 + \ldots + x_n a_n$$

Interpretation of matrix-vector multiplication: Row version

Consider a matrix $B \in \mathbb{R}^{p\times n}$ with rows b_i^T, $i = 1, \ldots, p$, where $b_i \in \mathbb{R}^n$:

$$B = \begin{bmatrix} b_1^T \\ \vdots \\ b_p^T \end{bmatrix}.$$

The matrix-vector multiplication Bx, where $x \in \mathbb{R}^n$, can be interpreted as a column vector where each entry is the inner product between a row of B and x:

$$Bx = \begin{bmatrix} b_1^T x \\ \vdots \\ b_p^T x \end{bmatrix}.$$

Interpretation of matrix-matrix multiplication: Outer product

- Consider a matrix $A \in \mathbb{R}^{m \times p}$ with columns $a_i \in \mathbb{R}^m$, $i = 1, \ldots, p$:

$$A = [a_1 \cdots a_p]$$

- Consider a matrix $B \in \mathbb{R}^{p \times n}$ with rows b_i^T, $i = 1, \ldots, p$, where $b_i \in \mathbb{R}^n$:

$$B = \begin{bmatrix} b_1^T \\ \vdots \\ b_p^T \end{bmatrix}$$

- The matrix-matrix product AB can be interpreted as a sum of matrices, where each matrix is the outer product between one column of A and one row of B:

$$AB = \sum_{i=1}^{p} a_i b_i^T$$

- Example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 5 & 6 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 7 & 8 \end{bmatrix}$$

Symmetric matrices

- Consider the $n \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

- Matrix A is symmetric when

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & & & \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix} = A^T$$

Positive/negative definite and indefinite matrices

- Consider a **symmetric** $n \times n$ matrix A
- In the following definitions, x is a vector $n \times 1$ ($x \in \mathbb{R}^n$)

	Definition	Eigenvalues
Positive semidefinite	$x^T Ax \geq 0$ **for all** x	all are ≥ 0
Positive definite	$x^T Ax > 0$ **for all** $x \neq 0$	all are > 0
Negative semidefinite	$x^T Ax \leq 0$ **for all** x	all are ≤ 0
Negative definite	$x^T Ax < 0$ **for all** $x \neq 0$	all are < 0
Indefinite	$x^T Ax > 0$ **for some** x **and** $x^T Ax < 0$ **for some other** x	some are > 0 **and** some are < 0

Submatrices and minors

- For an $n \times n$ matrix A, any matrix created by taking diagonal entries of A together with the corresponding off-diagonal entries is called a **principal submatrix** of A, and its determinant is called **principal minor**
- The upper left $k \times k$ corner (for $k = 1, 2, \ldots, n$) of A is called a **leading principal submatrix** of A, and its determinant is called **leading principal minor**

$$A = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix}$$

Leading principal submatrices

$$a, \begin{bmatrix} a & b \\ e & f \end{bmatrix}, \begin{bmatrix} a & b & c \\ e & f & g \\ i & j & k \end{bmatrix}, \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix}$$

Other principal submatrices

$$a, f, k, p, \begin{bmatrix} a & c \\ i & k \end{bmatrix}, \begin{bmatrix} a & d \\ m & p \end{bmatrix}, \begin{bmatrix} f & g \\ j & k \end{bmatrix}, \ldots, \begin{bmatrix} a & c & d \\ i & k & l \\ m & o & p \end{bmatrix}, \ldots$$

Criterion on definiteness using minors

- A symmetric matrix is positive definite if and only if all its leading principal minors are positive
- A symmetric matrix is positive semidefinite if and only if all its principal minors are nonnegative
- A symmetric matrix is negative definite if and only if all its leading principal minors of odd order (1×1, 3×3, ...) are negative and all its leading principal minors of even order (2×2, 4×4, ...) are positive
- A symmetric matrix is negative semidefinite if and only if all its principal minors of odd order are nonpositive and all its principal minors of even order are nonnegative

A couple of reminders about the inverse matrix

Suppose matrix A is invertible

- The transpose and the inverse can be taken in any order, so sometimes we write the result as A^{-T}

$$(A^{-1})^T = (A^T)^{-1} = A^{-T}$$

- If the eigenvalues of symmetric matrix A are λ_i, then the eigenvalues of A^{-1} are $\frac{1}{\lambda_i}$
- So if A is positive definite, then A^{-1} is also positive definite
- If A is negative definite, then A^{-1} is also negative definite

Rank of a matrix

Consider a matrix $A \in \mathbb{R}^{m\times n}$.

$$\begin{aligned}\operatorname{rank}(A) &= \text{maximum number of linearly independent rows of } A\\ &= \text{maximum number of linearly independent columns of } A\\ &= \text{number of nonzero eigenvalues of } A\text{, if } A \text{ is symmetric}\\ &= \text{number of (positive) singular values of } A \text{ (more on this later)}\\ &= \operatorname{rank}(A^TA) = \operatorname{rank}(AA^T)\end{aligned}$$

- For a rectangular matrix $A \in \mathbb{R}^{m\times n}$, we have that $\operatorname{rank}(A) \leq m$ and $\operatorname{rank}(A) \leq n$
- A square matrix $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $\operatorname{rank}(A) = n$ (full rank)

Positive semidefiniteness of A^TA and AA^T

Consider a rectangular matrix $A \in \mathbb{R}^{m\times n}$

Matrix A^TA is symmetric and positive semidefinite, for any $A \in \mathbb{R}^{m\times n}$

- Indeed, for any $z \in \mathbb{R}^n$, it holds that

$$z^T(A^TA)z = (Az)^T(Az) = \|Az\|_2^2 \geq 0$$

- Likewise, AA^T is symmetric and positive semidefinite for any A

If $\text{rank}(A) = n$, then A^TA is positive definite

- Matrix A^TA is $n \times n$, and we have that $\text{rank}(A^TA) = \text{rank}(A) = n$
- Therefore, A^TA is full rank and cannot have any zero eigenvalues
- Already proved that A^TA is positive semidefinite, so A^TA must be positive definite
- Likewise, if $\text{rank}(A) = m$, then AA^T is positive definite

If A is square and full-rank, then A^TA and AA^T are positive definite

Cholesky decomposition

- For a **positive definite** matrix A, the Cholesky decomposition is given by

$$A = LL^T$$

  where L is a lower triangular matrix which is unique

- Matlab implements the Cholesky factorization with the command `R=chol(A)` and returns an upper triangular matrix R such that

$$A = R^T R$$

- If you want a lower triangular matrix, the command is `R=chol(A,'lower')`
- The Cholesky factorization $A = LL^T$ is applicable to **positive semidefinite** matrices as well (L is not unique in this case)
- Matlab's command does not work in this case; to compute L use the algorithm provided in Golub and Van Loan, *Matrix Computations*, Sec. 4.2.8

Range and nullspace of a matrix

Consider matrices $A = [a_1 \ldots a_n] \in \mathbb{R}^{m \times n}$ and $B = \begin{bmatrix} b_1^T \\ \vdots \\ b_p^T \end{bmatrix} \in \mathbb{R}^{p \times n}$

The range of A is defined as the set of all vectors that can be obtained as multiplication of A and some vector:

$$\begin{aligned} \text{range}(A) = \mathcal{R}(A) &= \{Ax \mid x \in \mathbb{R}^n\} \\ &= \{x_1 a_1 + \ldots + x_n a_n \mid x_i \in \mathbb{R}\} = \text{span}(a_1, \ldots, a_n) \end{aligned}$$

Therefore, the system $Ax = b$ has a solution if and only if $b \in \text{range}(A)$

The nullspace of B is defined as the set of all vectors whose multiplication with B gives the zero vector:

$$\text{nullspace}(B) = \mathcal{N}(B) = \{x \mid Bx = 0\} = \{x \mid b_1^T x = 0, \ldots, b_p^T x = 0\}$$

(solution set of **homogeneous** system of linear equations)

Symmetric matrices and eigenvalue-eigenvector properties

Set of symmetric $n \times n$ matrices: $\mathbb{S}^n$.

A symmetric $n \times n$ matrix always has n real eigenvalues.

1. Eigenvalues may be repeated though (not a problem).
2. A general square but nonsymmetric matrix may have complex eigenvalues.

A symmetric $n \times n$ matrix always has n linearly independent eigenvectors $(q_1, \ldots, q_n)$.

- In particular, the eigenvectors can be chosen so they are *orthonormal*.
  1. This means that they are mutually orthogonal ($q_i^T q_j = 0$ for $i \neq j$)
  2. ...and they have unit norm ($\|q_i\|_2 = 1$)

Implications of eigenvalue-eigenvector properties for symmetric matrices

Let λ_i, $i = 1, \ldots, n$, be the (real) eigenvalues of $A \in \mathbb{S}^n$.

Let q_i, $i = 1, \ldots, n$, be the corresponding n *orthonormal* eigenvectors.

1. The set of eigenvectors forms a basis of $\mathbb{R}^n$. That is, *for any* $x \in \mathbb{R}^n$, we can write

$$x = \mu_1 q_1 + \mu_2 q_2 + \ldots + \mu_n q_n$$

   for some coefficients $\mu_i \in \mathbb{R}$.

2. Eigendecomposition of matrix A:

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^T = Q \Lambda Q^T$$

   where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and Q is an orthonormal matrix (that is, $Q^T Q = Q Q^T = I$), with the eigenvectors of A as columns:

$$Q = [q_1 \ldots q_n].$$

Bound on quadratic forms

Suppose we order the n (real) eigenvalues of $A \in \mathbb{S}^n$ as follows without loss of generality:

$$\lambda_{\max}(A) = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} \geq \lambda_n = \lambda_{\min}(A).$$

1. For any $x \in \mathbb{R}^n$, it holds that

$$\lambda_{\min}(A)\|x\|_2^2 \leq x^T Ax \leq \lambda_{\max}(A)\|x\|_2^2$$

   Proof: Write $x = \sum_{i=1}^n c_i q_i$, for some coefficients c_i, and use the orthogonality of the eigenvectors.

2. The right inequality holds as equality if you choose $x = tq_1$ for any $t \geq 0$.

Note that the property holds regardless of the signs of the eigenvalues.

A formula for the maximum eigenvalue

Consider the previous two properties when x is a unit-norm vector u, that is $\|u\|_2 = 1$: (equivalently, set $u = x/\|x\|_2$). We have that

1. $u^T A u \leq \lambda_{\max}(A)$, for any $u \in \mathbb{R}^n$ such that $\|u\|_2 = 1$
2. $u^T A u = \lambda_{\max}(A)$, if $u = q_1$ (unit-norm eigenvector corresponding to $\lambda_{\max}$)

By definition of the $\sup$, we have that

$$\lambda_{\max}(A) = \sup_{\|u\|_2=1} u^T A u.$$

Singular value decomposition (SVD)

A matrix $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$ can always be written as

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \underbrace{U_1}_{m \times r} \underbrace{\Sigma_1}_{r \times r} \underbrace{V_1^T}_{r \times n} = \underbrace{U}_{m \times m} \underbrace{\Sigma}_{m \times n} \underbrace{V^T}_{n \times n}$$

where

- we have r *singular values*, which are all positive:

  $\sigma_{\max}(A) = \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$
- $\Sigma_1 = \text{diag}(\sigma_1, \ldots \sigma_r)$ and $\Sigma = \begin{bmatrix} \Sigma_1 & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix}$
- $U_1 \in \mathbb{R}^{m \times r}$ has columns called *left singular vectors* of A; and $U_1^T U_1 = I_r$
- $V_1 \in \mathbb{R}^{n \times r}$ has columns called *right singular vectors* of A; and $V_1^T V_1 = I_r$
- U and V are orthonormal ($m \times m$ and $n \times n$, respectively) and in addition,

$$U = [\ \underbrace{U_1}_{\text{basis of } \mathcal{R}(A)} \quad \underbrace{U_2}_{\text{basis of } \mathcal{N}(A^T)}\ ] \qquad V = [\ \underbrace{V_1}_{\text{basis of } \mathcal{R}(A^T)} \quad \underbrace{V_2}_{\text{basis of } \mathcal{N}(A)}\ ]$$

- $A = U_1 \Sigma_1 V_1^T$ is sometimes called "thin" SVD

Relationship between SVD and eigendecomposition

Suppose that $A \in \mathbb{R}^{m \times n}$.

$$A^T A = V_1 \Sigma_1 U_1^T U_1 \Sigma_1 V_1^T = V_1 \Sigma_1^2 V_1^T = V \Sigma^2 V^T$$

Since V is orthonormal, we conclude that $V\Sigma^2 V^T$ is the eigendecomposition of $A^T A$.

Therefore,

1. $\lambda_i(A^T A) = \sigma_i^2(A)$, $i = 1, \ldots, r$; and $\lambda_i(A^T A) = 0$, $i = r+1, \ldots, n$
2. V contains the eigenvectors of $A^T A$

Considering AA^T, we likewise have that

1. $\lambda_i(AA^T) = \sigma_i^2(A)$, $i = 1, \ldots, r$; and $\lambda_i(AA^T) = 0$, $i = r+1, \ldots, m$
2. U contains the eigenvectors of AA^T

When A is symmetric, then $\sigma_i(A) = \sqrt{\lambda_i(A^2)} = |\lambda_i(A)|$

Pseudoinverse (Moore-Penrose inverse)

For any matrix $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r$, the pseudoinverse $A^\dagger \in \mathbb{R}^{n\times m}$ is defined using the SVD as

$$A^\dagger = V_1\Sigma_1^{-1}U_1^T = V\Sigma^\dagger U^T$$

where $\Sigma^\dagger = \begin{bmatrix} \Sigma_1^{-1} & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix}$.

The pseudoinverse $A^\dagger = B$ is the unique matrix B that simultaneously satisfies the following properties:

$$ABA = A \qquad (AB)^T = AB$$

$$BAB = B \qquad (BA)^T = BA$$

Bounds using $\sigma_{\max}$

1. For any $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, it holds that

$$\|Ax\|_2 \leq \sigma_{\max}(A)\|x\|_2$$

   The bound is achieved when $x = tv_1$ (right sing. vector corresponding to $\sigma_{\max}(A)$)
   Proof: $\|Ax\|_2^2 = x^T A^T Ax \leq \lambda_{\max}(A^T A)\|x\|_2^2 = \sigma^2_{\max}(A)\|x\|_2^2$

2. For any $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$, it holds that

$$y^T Ax \leq \sigma_{\max}(A)\|x\|_2\|y\|_2$$

Formulas for $\sigma_{\max}$ and the spectral norm

Based on the previous two inequalities, we obtain that

$$\sigma_{\max}(A) = \sup_{\|u\|_2=1} \|Au\|_2 = \sup_{x\neq 0, y\neq 0} \frac{y^T A x}{\|x\|_2 \|y\|_2}$$

The maximum singular value is a norm on $\mathbb{R}^{m\times n}$, called *spectral norm* or ℓ_2-norm, and is sometimes denoted by $\rho(A)$:

$$\rho(A) = \|A\|_2 = \sigma_{\max}(A)$$

(it can be verified that it satisfies the 3 properties of norm)

Special cases:

- When $A \in \mathbb{S}^n$ (i.e., A is symmetric), $\rho(A) = \max_{i=1,\ldots,n} |\lambda_i(A)|$
- When $A \in \mathbb{R}^{m\times 1}$ (i.e., A is a vector), $\rho(A)$ becomes the Euclidean norm (ℓ_2-norm)

Explicit solution of a linear system of equations

With $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, consider the system of equations

$$Ax = b$$

The general solution is written in the form

$$x = x_p + v$$

where

- x_p is any solution if $Ax = b$;
- v is any vector $v \in \mathrm{nullspace}(A)$

Schur complement

Consider the symmetric block matrix $X \in \mathbb{S}^n$ (so $A = A^T$, $C = C^T$)

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

- Suppose A is invertible: $S = C - B^T A^{-1} B$ is the **Schur complement of A in X**
- Then the following hold:
  1. $X \succ 0 \iff A \succ 0$ and $S \succ 0$
  2. If $A \succ 0$, then $X \succeq 0 \Leftrightarrow S \succeq 0$
- Suppose C is invertible: $S = A - BC^{-1}B^T$ is the **Schur complement of C in X**
- Then the following hold:
  1. $X \succ 0 \iff C \succ 0$ and $S \succ 0$
  2. If $C \succ 0$, then $X \succeq 0 \Leftrightarrow S \succeq 0$

Questions And Suggestions?

Thank You!

Please visit

https://lab.vanderbilt.edu/taha/

IFF you want to know more ☺