Module 2 — Mathematical Background

Dr. Ahmad F. Taha

CE 4240/5240 — Intro to Infrastructure Systems Engineering

Email: ahmad.taha@vanderbilt.edu

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

VANDERBILT UNIVERSITY

August 29, 2024

Domain of a function

- Notations: scalar $x \in \mathbb{R}$ or vector $x \in \mathbb{R}^n$; functions and mappings
- **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, $\ln x$), $\mathrm{dom}(f) = \mathbb{R}_{++}$
- 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 $f(x)$:

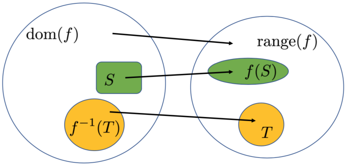
$$\mathrm{range}(f) = \{f(x) \,|\, x \in \mathrm{dom}(f)\}$$

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 $n \times n$):

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

Examples

Linear, affine, and quadratic functions

- Example on linear and quadratic functions
- A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *linear* if it is written in the form

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

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

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

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

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

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

  - $\nabla f(x) = Px + q$; $\nabla^2 f(x) = P$
  - 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

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

Linear and quadratic approximations of a function

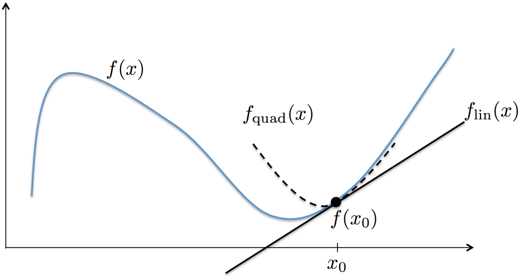
Consider a function $f(x) : \mathbb{R}^n \to \mathbb{R}$

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

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

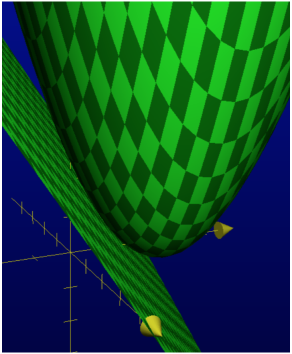
- Gradient vector evaluated at $x_1 = 0.5$, $x_2 = 0.5$

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

Example: Linear approximation

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

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



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 $\sup S = 1$

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

Examples

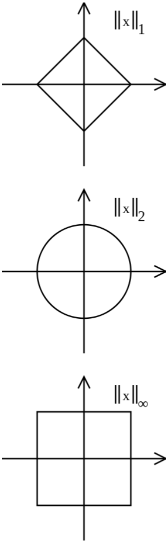


Figure: Unit circles $||x||_p = 1$ in different p norms.

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.

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: $\|A\|_1 = \max\limits_{1\leq j\leq n}\sum_{i=1}^{m}|a_{ij}|$: 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 $\|A\|_2 = \sigma_{\max}(A) \leq \|A\|_{\mathrm{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}$
- Infinity-norm: $\|A\|_\infty = \max\limits_{1\leq i\leq m}\sum_{j=1}^{n}|a_{ij}|$: maximum absolute row sum of the matrix
- Max-norm: $\|A\|_{\max} = \max_{ij}|a_{ij}|$
- Nuclear-norm: $\|A\|_* = \operatorname{trace}\left(\sqrt{A^*A}\right) = \sum_{i=1}^{\min\{m,n\}}\sigma_i(A)$

Norm bounds

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

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

Example: $A = \begin{bmatrix} 3 & 5 & 7 \\ 2 & 6 & 4 \\ 0 & 2 & 8 \end{bmatrix}$, then

- $\|A\|_1 = \max(|-3| + 2 + 0; 5 + 6 + 2; 7 + 4 + 8) = \max(5, 13, 19) = 19,$
- $\|A\|_\infty = \max(|-3| + 5 + 7; 2 + 6 + 4; 0 + 2 + 8) = \max(15, 12, 10) = 15.$

Eigenvalues and Eigenvectors

- Evalues/vectors are **only defined for square**[1] **matrices**
- For a matrix $A \in \mathbb{R}^{n\times n}$, we always have n evalues/evectors
  - Some of these evalues might be distinct, real, repeated, imaginary
  - To find evalues(A), solve this equation ($\boldsymbol{I}_n$: identity matrix of size n)

$$\det(\lambda \boldsymbol{I}_n - A) = 0 \text{ or } \det(A - \lambda \boldsymbol{I}_n) = 0 \Rightarrow a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n = 0$$

- **Example**: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$
- **Eigenvectors**: A number λ and a non-zero vector v satisfying

$$Av = \lambda v \Rightarrow (A - \lambda \boldsymbol{I}_n)v = 0$$

  are called an eigenvalue and an eigenvector of A
  - λ is an eigenvalue of an $n \times n$-matrix A if and only if $\lambda \boldsymbol{I}_n - A$ is not invertible, which is equivalent to $\det(\mathsf{A} - \lambda \mathsf{I}_\mathsf{n}) = 0$.

[1] A square matrix has equal number of rows and columns.

Matrix Inverse

- Inverse of a generic 2by2 matrix:

$$\boxed{A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}$$

  – Notice that $A^{-1}A = AA^{-1} = I_2$

- Inverse of a generic 3by3 matrix:

$$\boxed{A^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}^T = \frac{1}{\det(A)} \begin{bmatrix} A & D & G \\ B & E & H \\ C & F & I \end{bmatrix}}$$

$$\begin{array}{lll} A = (ei - fh) & D = -(bi - ch) & G = (bf - ce) \\ B = -(di - fg) & E = (ai - cg) & H = -(af - cd) \\ C = (dh - eg) & F = -(ah - bg) & I = (ae - bd) \end{array}$$

$$\boxed{\det(A) = aA + bB + cC.}$$

  – Notice that $A^{-1}A = AA^{-1} = I_3$

Linear Algebra — Example 1

- Find the eigenvalues, eigenvectors, and inverse of matrix

$$A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$$

  - Eigenvalues: $\lambda_{1,2} = 5, -2$; evectors: $v_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top, v_2 = \begin{bmatrix} -\frac{4}{3} & 1 \end{bmatrix}^\top$
  - Inverse: $A^{-1} = -\frac{1}{10} \begin{bmatrix} 2 & -4 \\ -3 & 1 \end{bmatrix}$

- Write A in the matrix **diagonal transformation,** i.e., $A = TDT^{-1}$ where D is the diagonal matrix containing the eigenvalues of A:

$$\boxed{A = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^{-1}}$$

  - **Only valid for matrices with distinct, real eigenvalues**

Rank of a Matrix

- Rank of a matrix: $\mathrm{rank}(A)$ is equal to the number of linearly independent rows or columns
  - **Example 1:** $\mathrm{rank}\left(\begin{bmatrix} 1 & 1 & 0 & 2 \\ -1 & -1 & 0 & -2 \end{bmatrix}\right)$=?
  - **Example 2:** $\mathrm{rank}\left(\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}\right)$ =?
- Rank computation: reduce the matrix to a simpler form, generally row echelon form, by elementary row operations

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \to 2r_1 + r_2 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 3 & 5 & 0 \end{bmatrix} \to -3r_1 + r_3 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{bmatrix}$$

$$\to r_2 + r_3 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \to -2r_2 + r_1 \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \mathrm{rank}(A) = 2$$

Null Space of a Matrix

- The Null Space of any matrix A is the subspace $\mathcal{K}$ defined as follows:

$$\mathrm{N}(A) = \mathrm{Null}(A) = \ker(A) = \{x \in \mathcal{K} | Ax = \mathbf{0}\}$$

- $\mathrm{Null}(A)$ has the following three properties:
  - $\mathrm{Null}(A)$ always contains the zero vector, since $A0 = \mathbf{0}$
  - If $x \in \mathrm{Null}(A)$ and $y \in \mathrm{Null}(A)$, then $x + y \in \mathrm{Null}(A)$
  - If $x \in \mathrm{Null}(A)$ and c is a scalar, then $cx \in \mathrm{Null}(A)$
- **Example:** Find $\mathrm{N}(A)$

$$A = \begin{bmatrix} 2 & 3 & 5 \\ -4 & 2 & 3 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 3 & 5 \\ -4 & 2 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \left[\begin{array}{ccc|c} 2 & 3 & 5 & 0 \\ -4 & 2 & 3 & 0 \end{array}\right] \Rightarrow$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 1/16 & 0 \\ 0 & 1 & 13/8 & 0 \end{array}\right] \Rightarrow a = -\frac{1}{16}c, b = -\frac{13}{8}c \Rightarrow \boxed{\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \alpha \begin{bmatrix} -1/16 \\ -13/8 \\ 1 \end{bmatrix} = \tilde{\alpha} \begin{bmatrix} -1 \\ -26 \\ 16 \end{bmatrix}}$$

Linear Algebra — Example 2

- Find the determinant, rank, and null-space set of this matrix:

$$B = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

  - $\det(B) = 0$
  - $\text{rank}(B) = 2$
  - null$(B) = \alpha \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}, \forall\ \alpha \in \mathbb{R}$
- Is there a relationship between the determinant and the rank of a matrix?
  - Yes! Matrix drops rank if determinant = zero $\Rightarrow$ 1 zero evalue
- True or False?
  - $AB = BA$ for all A and B—**FALSE!**
  - A and B are invertible $\rightarrow (A+B)$ is invertible—**FALSE!**

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

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

Definiteness

- We need a way to figure out if vector-valued functions $f(x)$ are always positive or negative...why?
- For example, consider $f(x) = x_1^2 + x_2^2 + 2x_1x_2$. Is this function always positive?
- Well $f(x) = (x_1 + x_2)^2$ so this function is always non-negative
- But how do we assert that large-scale quadratic forms are positive/negative/etc..?
- Need to learn about *definiteness*
- A quadratic function $f(x) \in \mathbb{R}^n \to \mathbb{R}$ is **definite** if the zero vector is the only value where the form is zero (i.e., $f(0) = 0$)
- A quadratic function is positive definite if it is definite and $f(x) \geq 0$ for all x
- A quadratic function is negative definite if it is definite and $f(x) \leq 0$ for all x

Positive/negative definite and indefinite matrices

- Consider a **symmetric** $n \times n$ matrix A
- In the following definitions, x is a vector $n \times 1$ ($x \in \mathbb{R}^n$)
- A quadratic form is **definite** if the zero vector is the only place the form is zero (i.e., $f(0) = 0$)

	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

Example

- Is this matrix positive definite? $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$
- For what values of b is $A = \begin{bmatrix} 2 & -1 & b \\ -1 & 2 & -1 \\ b & -1 & 2 \end{bmatrix}$ positive definite?

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

More on rank of a matrix

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

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

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

- Matrix A^TA is $n \times n$, and we have that $\mathrm{rank}(A^TA) = \mathrm{rank}(A) = n$
- Therefore, A^TA is full rank and cannot have any zero eigenvalues
- Likewise, if $\mathrm{rank}(A) = m$, 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

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 ($\|q_i\|_2 = 1$)

Singular value decomposition (SVD)

A matrix $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = r$ 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:

  $\sigma_{\max}(A) = \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$
- $\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots \sigma_r)$ 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,

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

- $A = U_1 \Sigma_1 V_1^T$ 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 $\sigma_i(A) = \sqrt{\lambda_i(A^2)} = |\lambda_i(A)|$. And remember that

$$\|A\|_2 = \sqrt{\lambda_{\max}\left(A^* A\right)} = \sigma_{\max}(A)$$

Example

Find the SVD of $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \in \mathbb{R}^{2\times 3}, \quad m = 2, n = 3, r = 2.$

- First, compute $AA^T = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$.
- The find the evalues of AA^T:

$$\det(\lambda I - AA^T) = (\lambda - 25)(\lambda - 9) = 0, \quad \sigma_1 = \sqrt{25} = 5\sigma_2 = \sqrt{9} = 3, \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$$

- Need to construct V now by finding evectors associated with $\lambda = 25$ and $\lambda = 9$.
- For $\lambda = 25$, we perform elementary row operations on $A^T A - 25I$ and get $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} v_1 = 0$. Then, $v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

Example (Cont'd)

- For $\lambda = 9$, we compute $A^T A - 9I$ and get

$$\begin{bmatrix} 1 & 0 & \frac{-1}{4} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix} v_2 = 0, \quad v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix}$$

- To construct v_3, we find nullspace $\mathcal{N}(A)$ and get

$$v_3 = \begin{bmatrix} \frac{-2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

- Hence $V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{-2}{3} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{18}} & \frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & \frac{1}{3} \end{bmatrix}$.

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

$$Ax = b$$

- This linear system of equations is one of the most solved problems
- Special case: $m = n$, i.e., A is a square matrix
  - If A is invertible (or non-zero determinant, or not rank-deficient), then $x = A^{-1}b$ is the unique solution
  - If If A is *not* invertible, no unique solution
- General case ($m \neq n$), we need to create new matrix inverse
  - That is, the *Moore-Penrose inverse* of A defined as $A^\dagger$
  - If A has linearly independent columns ($m > n$) and hence $A^T A$ is invertible, then

$$\textit{// left inverse:} \quad A^\dagger = \left(A^T A\right)^{-1} A^T \quad \Rightarrow \quad \boxed{x = A^\dagger b}$$

  - Else if A has linearly independent rows ($m < n$), then

$$\textit{// right inverse:} \quad A^\dagger = A^T \left(A A^T\right)^{-1} \quad \Rightarrow \quad \boxed{x = A^\dagger b}$$

First order ODE

- *Dynamic systems*, things that move in space and time, play a great role in our understanding of the universe
- Examples: climate change models, energy, water, transportation, GPS, epidemiology, zoology, and 100 other critical applications
- Role of dynamic systems in the pandemic, climate change, systems biology, infrastructure, etc...
- Many ways to model dynamic systems (model-driven, model-free, data-driven, etc...)
- Arguably, best way to model dynamic systems is via differential equations
- Ability to predict the future, plan, operate, control, foresee the unforeseen
- We'll have an entire 4–5 lectures on this topic, but let's cover the basic math fundamentals

ODEs 101

- Let's start with the first differential equation (or initial value problem IVP):

$$\text{Autonomous-IVP}: \quad \dot{x}(t) = \alpha x(t), \quad x(t_0) = x(0) = \text{given/known}$$

- Example: leaky tank/battery, mosquitos when it's raining
- Here, $x \in \mathbb{R}$, i.e., it's a scalar not a vector
- How do you solve this differential equation for $x(t)$?
- What happens for $x(t)$ as $t \to \infty$?
- What role does α play in the *evolution* of $x(t)$?
- What happens if we have a control input and the IVP is replaced by:

$$\text{Controlled-IVP}: \quad \dot{x}(t) = \alpha x(t) + \beta u(t), \quad x(t_0) = x(0) = \text{given/known}, \quad u(t) \text{ is given for all}$$

ODEs 101 (Cont'd)

- Solution of the Autonomous-IVP:

$$x(t) = e^{\alpha t} x(0)$$

- Solution of the Controlled-IVP:

$$x(t) = e^{\alpha t} x(0) + \beta \int_0^t e^{\alpha(t-\tau)} u(\tau) d\tau$$

- In this class, we are not interested in simple scalar ODEs, but more so in network or infrastructure of ODEs
- That is, $x \in \mathbb{R}^n$ and instead of the scalar ODE, we have something more general

$$\dot{x}(t) = Ax(t) + Bu(t)$$

- Multi-input, multi-output dynamic system modeling an infrastructure, a network, a system, etc...
- Examples

Dynamic Models in Nature

- Predator-prey equations are 1st order non-linear, ODEs
- Describe the dynamics of biological systems where 2 species interact
- One species as a predator and the other as a prey
- Populations change through time according to these equations:

$$\dot{x}(t) = \alpha x(t) - \beta x(t)z(t)$$

$$\dot{z}(t) = \delta x(t)z(t) - \gamma y(t)$$

– $x(t)$: # of preys (e.g., rabbits)

– $z(t)$: # of predators (e.g., foxes)

– $\dot{x}(t), \dot{z}(t)$: growth rates of the 2 species—function of time, t

– $\alpha, \beta, \gamma, \delta$: +ve real parameters depicting the interaction of the species

Mathematical Model

$$\dot{x}(t) = \alpha x(t) - \beta x(t)z(t)$$

$$\dot{z}(t) = \delta x(t)z(t) - \gamma z(t)$$

- Prey's population grows exponentially $(\alpha x(t))$—why?
- Rate of predation is assumed to be proportional to the rate at which the predators and the prey meet ($\beta x(t)z(t)$)
- If either $x(t)$ or $z(t)$ is zero then there can be no predation
- $\delta x(t)z(t)$ represents the growth of the predator population
- No prey $\Rightarrow$ no food for the predator $\Rightarrow$ $z(t)$ decays
- Is there an equilibrium? What is it?

Dynamics in Epidemiology

- **Epidemiology:** The branch of medicine that deals with the incidence, distribution, and possible control of diseases and other factors relating to health
- In the past 10 years, mathematicians, biologists, and physicists studied mathematical models of epidemics
- Why is that important?
- Various models focus on different things:
  - SIR Model: **S** for the number susceptible, **I** for the number of infectious, and **R** for the number recovered
  - SIS Model: Infections like cold and influenza, do not possess lasting immunity
  - SEIR: **E** for exposed
  - MSIR: **M** stands for maternally-derived immunity
  - SEIS and many, many more

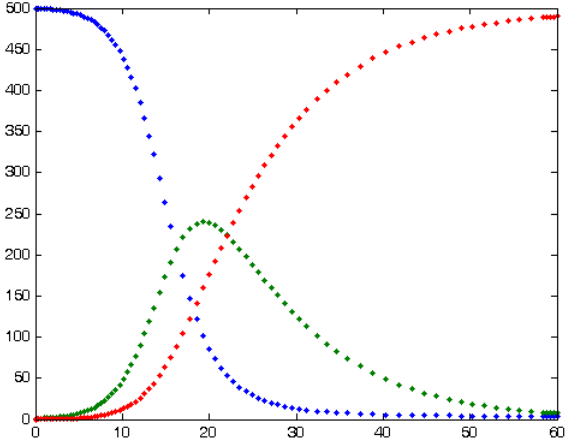
SIR Model

- Here, we present the dynamic model for the SIR model
- We take flu as an example of the SIR model
- Define variable $S(t), I(t), R(t)$ representing the number of people in each category at time t. The SIR model can be written as

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta IS}{N}\\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I\\ \frac{dR}{dt} &= \gamma I.\end{aligned}$$

- N is the total number of people, with $S(t) + I(t) + R(t) = N$
- The force of infection F can be written as $F = \beta I/N$
- β is the contact rate, and γ is the transition rate (rate of recovery)

So who do these quantities vary?



Blue represents **Susceptible**, Green represents **Infected**, and Red represents the **Recovered** population.

Dynamic Systems and Infras

- We will spend 4–5 classes on modeling infrastructure
- But we shall all see that many infra models will be written as

$$\dot{x}(t) = Ax(t) + Bu(t)$$

  where $x(t) \in \mathbb{R}^n$ is the *state* of the system and $u(t) \in \mathbb{R}^m$ is the *control input* of the system (more on that in the next slides)

- Examples
- Discussion on scale, number of states, inputs, etc..

Dynamic System Models

- Consider a multi-input, multi-output (MIMO) dynamic system with
  - n **internal states** $(x_1(t), x_2(t), \ldots, x_n(t))$
  - m **control inputs** $(u_1(t), u_2(t), \ldots, u_m(t))$
  - p **measurement outputs** $(y_1(t), y_2(t), \ldots, y_p(t))$
- An infrastructure modeler/engineer comes and gives you the relationship between $x(t), u(t)$, and $y(t)$ as

$$\begin{aligned} \dot{x}_i(t) &= \sum_{j=1}^{n} a_{ij} x_j(t) + \sum_{j=1}^{m} b_{ij} u_k(t), \quad \forall i = 1, 2, \ldots, n \\ y_l(t) &= \sum_{j=1}^{n} c_{lj} x_j(t) + \sum_{j=1}^{m} d_{lj} u_j(t), \quad \forall l = 1, 2, \ldots, p \\ & \quad x_i(0), \quad \forall i = 1, 2, \ldots, n \;\; \text{are all given} \end{aligned}$$

  - Constants $a_{ij}, b_{ij}, c_{lj}, d_{lj}$: all given, model parameters of the dynamic system, can be any real number; initial conditions $x_i(0)$ are given
- Almost all dynamic systems can be written as above (more on modeling specific infras in Module 4)

State-Space Representations

- **State-space** (SS) theory: representing the system by a **vector-form first order ODE**:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x_{\text{initial}} = x_{t_0}, \tag{1}$$

$$y(t) = Cx(t) + Du(t), \tag{2}$$

  - $y(t) \in \mathbb{R}^p$: output-vector and A, B, C, D are constant matrices from parameters $a_{ij}, b_{ij}, c_{lj}, d_{lj}$
  - $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{p \times n}, D \in \mathbb{R}^{p \times m}$
- The above two equations represent a relationship between the input and output of the system via the internal system states
- The above 2 equations are nothing but a matrix-vector first ODEs
- These equations elegantly model infrastructure systems
- What about uncertainty? Nonlinearity?

Explicit Analytical Solution

- How do we solve this large-scale ODE?
- For linear dynamic systems, solution is actually analytically obtained
- Very elegant too, extends the case of scalar ODEs (like $\dot{x}(t) = \alpha x(t)$):

$$\boxed{x(t) = e^{At}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau}$$

- This means that the sensor data (output) can be predicted via plugging $x(t)$ into $y(t) = Cx(t) + Du(t)$
- Why is this important???
- How do you implement this? A high-level picture

Solving via Matlab/Discretization

- You can still apply the analytical solution, but for large-scale systems you might into issues (due to integration and matrix exponential)
- Alternatively, you can go to Matlab and use `ode45`

Questions And Suggestions?

Thank You!

Please visit

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

IFF you want to know more ☺