ECE240 Introduction to Linear Dynamical Systems
Lecture 2: Vectors and Matrices

Dr. Yu Zhang
ECE Department, UC Santa Cruz

Fall 2025

Outline

1. Overview: Linear Algebra for Dynamical Systems
2. Vectors and Vector Norm
   - Linear Functions and Gradient
3. Matrices
   - Matrix-Matrix and Matrix-Vector Multiplications
   - Hessian, Jacobian, and Taylor Expansion
   - Matrix Inverse
   - Trace, Determinant, and Matrix Norm

Why Linear Algebra for Dynamical Systems (1/2)?

- State-space models are expressed using matrices:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad y(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t).$$

- Linear algebra concepts directly map to system properties:
  - Eigenvalues $\lambda_i(\mathbf{A}) \;\rightarrow\;$ stability and oscillatory modes.
  - Rank/Null space $\;\rightarrow\;$ controllability and observability.
  - Positive definite matrices $\;\rightarrow\;$ Lyapunov stability tests.
  - Matrix exponentials $\;\rightarrow\;$ closed-form solution of $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$.
  - SVD $\;\rightarrow\;$ balanced truncation and reduced-order models.

Why Linear Algebra for Dynamical Systems (2/2)?

- In $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$:
  - Range/Null space $\rightarrow$ reachable/steady-state subspaces.
  - Rank $\rightarrow$ dimension of controllable/observable space.
  - Eigenvalues $\rightarrow$ system modes.
  - Left eigenvectors $\rightarrow$ adjoint dynamics, dual systems.
- These concepts reappear in:
  - Controllability matrix $\mathcal{C} = [\mathbf{B}, \mathbf{AB}, \dots, \mathbf{A}^{n-1}\mathbf{B}]$.
  - Observability matrix $\mathcal{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix}$.

Vector

- Notation for vectors: $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} \in \mathbb{R}^N, \ \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} \in \mathbb{C}^N$
- A *linear function* of $\mathbf{x}$ can be expressed as an inner product:

$$f(\mathbf{x}) = \sum_{i=1}^{N} a_i x_i = \mathbf{a}^\top \mathbf{x},$$

  where $(\cdot)^\top$ denotes transpose, i.e., $\mathbf{a}^\top = [a_1 \ \cdots \ a_N]$.
- The *gradient* of a multivariate function is the vector of partial derivatives:

$$\nabla f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1} \ \cdots \ \frac{\partial f}{\partial x_N} \right]^\top.$$

- Q: Show that $\nabla f(\mathbf{x}) = \mathbf{a}$.
- The gradient is an expansion of the derivative concept. The derivative of a function implies how fast the function grows at a point. The gradient tells us in which direction the function grows the fastest, and how fast it grows in that direction.

Vector Norms

A vector norm is a real-valued function $\|\cdot\| : \mathbb{C}^N \mapsto \mathbb{R}_+$ satisfying the following three properties for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, and $a \in \mathbb{C}$:

1. *Positive definiteness:* $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
2. *Absolute homogeneity (scaling):* $\|a\mathbf{x}\| = |a| \cdot \|\mathbf{x}\|$
3. *Subadditivity (triangle inequality):* $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

If the property of positive definiteness is relaxed, we have a *seminorm*.

- Q: Define $0^0 = 0$, and $\|\mathbf{x}\|_0 = |x_1|^0 + |x_2|^0 + \cdots + |x_N|^0$. Is $\|\mathbf{x}\|_0$ a norm?
- Q: Prove $|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\|$.

ℓ_p Norms

For $\mathbf{x} \in \mathbb{C}^N$ and $p \geq 1$, $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}}$ is a vector norm. Commonly-used norms:

- $p = 2$: square root of the sum of squares $\|\mathbf{x}\|_2 = \sqrt{\sum_i |x_i|^2}$.
  The Euclidean norm reflects the ordinary distance from the origin to the point.
- $p = 1$: sum-absolute-value $\|\mathbf{x}\|_1 = \sum_i |x_i|$
- $p = \infty$: max-absolute-value $\|\mathbf{x}\|_\infty = \lim_{p \to \infty} \|\mathbf{x}\|_p = \max_i |x_i|$

1. Q: Find $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, and $\|\mathbf{x}\|_\infty$ for $\mathbf{x} = [1 + j,\ 2,\ -1]^\top$, where $j := \sqrt{-1}$.
2. Q: Show all three norms are equal for $\mathbf{x} = [c,\ 0,\ \cdots,\ 0]^\top$ for any $c \in \mathbb{R}$.

Norm Inequalities (1/2)

1. The value of ℓ_p norm is non-increasing in p: $\|\mathbf{x}\|_a \leq \|\mathbf{x}\|_b$ for $a \geq b \geq 1$
2. Norm inequality chain: $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{N}\|\mathbf{x}\|_2 \leq N\|\mathbf{x}\|_\infty$
3. Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent iff there exist two positive reals C_1 and C_2 such that $C_1\|\mathbf{x}\|_a \leq \|\mathbf{x}\|_b \leq C_2\|\mathbf{x}\|_a$ for all $\mathbf{x} \in \mathbb{C}^N$.
4. **Theorem**: Any two norms in a finite-dimensional space are equivalent.

Norm Inequalities (2/2)

- **Inner product**:
  1. If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, then $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_k x_k \bar{y}_k = \mathbf{y}^H \mathbf{x} = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$, where $\bar{y}_k$ is the complex conjugate of y_k.
  2. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, then $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_k x_k y_k = \mathbf{x}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{x} = \langle \mathbf{y}, \mathbf{x} \rangle$.
  3. $|\langle \mathbf{x}, \mathbf{y} \rangle| = |\sum_k x_k \bar{y}_k| \leq \sum_k |x_k \bar{y}_k| = \sum_k |x_k y_k|$.
- **Cauchy-Schwarz inequality**:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2,$$

  where the equality holds iff $\mathbf{x} = c \cdot \mathbf{y}$ for some constant c.
- **Hölder's inequality**:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sum_k |x_k y_k| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \quad \text{for} \quad \frac{1}{p} + \frac{1}{q} = 1 \quad (p \geq 1)$$

Matrix

- Notation for matrices: $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{bmatrix} \in \mathbb{R}^{M \times N}$.
- Matrix-vector product $\mathbf{b} = \mathbf{A}\mathbf{x}$: $b_i = \sum_{j=1}^{N} A_{ij} x_j$ for $i = 1, \ldots, M$.
- If $\mathbf{A}_{:,j}$ denotes the j-th column of $\mathbf{A}$, we have

$$\mathbf{b} = \mathbf{A}\mathbf{x} = \sum_{j=1}^{N} x_j \mathbf{A}_{:,j} = x_1 \mathbf{A}_{:,1} + \ldots + x_N \mathbf{A}_{:,N}$$

- Matrix-matrix product: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$. Then,

$$\mathbf{C} = \mathbf{A}\mathbf{B} \in \mathbb{R}^{m \times p}, \text{ where } c_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}.$$

- Matrix multiplication is not commutative: the order of multiplication matters!

Matrix $\times$ Matrix: Four Ways

- Multiplying $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ can be viewed in 4 ways:

(MM1) **Row-Column inner products:** Each entry c_{ij} is the dot product of row i of A and column j of B.

(MM2) **Column combination:** Each column of $C = AB$ is a linear combination of columns of A, weighted by entries of B.

(MM3) **Row combination:** Each row of C is a linear combination of rows of B, weighted by entries of A.

(MM4) **Outer products:** AB is the sum of n rank-1 matrices:

$$AB = \sum_{k=1}^{n} a_k b_k^T$$

where a_k is column k of A and $b_k^\top$ is row k of B.

Block Matrix Multiplications

- Block matrices are widely used in system modeling.
- **Block** $\times$ **vector**:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Ax + By \\ Cx + Dy \end{bmatrix}$$

- **Block** $\times$ **block** (conformable sizes):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

- **Special cases**:
  - Upper triangular blocks: $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$.
  - Block-diagonal matrices: $\mathrm{diag}(A, D)$.

  Useful for interconnected subsystems and Kronecker structures.

Block Row Matrix × Vector

- Consider a block row matrix with k blocks:

$$\mathbf{M} = \begin{bmatrix} A_1 & A_2 & \cdots & A_k \end{bmatrix}, \quad \mathbf{x} = [x_1, x_2, \ldots, x_k]^\top$$

- The product is simply

$$\mathbf{Mx} = A_1 x_1 + A_2 x_2 + \cdots + A_k x_k.$$

- Examples:
  - If each A_i is $m \times n_i$ and $x_i \in \mathbb{R}^{n_i}$, the result is in $\mathbb{R}^m$.
  - For matrix $\mathcal{C} = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}$, multiplying by stacked inputs $[u_0, \ldots, u_{n-1}]^\top$ gives

$$Bu_0 + ABu_1 + \cdots + A^{n-1}Bu_{n-1},$$

which appears in controllability analysis.

Block Column Matrix × Vector (1/2)

- Consider a block column matrix with k blocks:

$$\mathbf{M} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix}, \quad \mathbf{x} \in \mathbb{R}^n.$$

- The product is

$$\mathbf{M}\mathbf{x} = \begin{bmatrix} A_1\mathbf{x} \\ A_2\mathbf{x} \\ \vdots \\ A_k\mathbf{x} \end{bmatrix}.$$

Block Column Matrix $\times$ Vector (2/2)

Examples:

- Each $\mathbf{A}_i$ is $m_i \times n$; then result is stacked in $\mathbb{R}^{m_1+m_2+\cdots+m_k}$.
- Observability matrix:

$$\mathcal{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix}, \qquad \mathcal{O}\mathbf{x}_0 = \begin{bmatrix} \mathbf{Cx}_0 \\ \mathbf{CAx}_0 \\ \vdots \\ \mathbf{CA}^{n-1}\mathbf{x}_0 \end{bmatrix},$$

which collects the output sequence from the initial state $\mathbf{x}_0$.

Quadratic functions

- A *quadratic function* of $\mathbf{x}$ can be expressed as

$$
\begin{aligned}
f_2(\mathbf{x}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}x_i x_j \\
&= \sum_{i=1}^{N} x_i \left(\sum_{j=1}^{N} A_{ij}x_j\right) \\
&= \sum_{i=1}^{N} x_i [\mathbf{A}\mathbf{x}]_i = \mathbf{x}^\top \mathbf{A}\mathbf{x}
\end{aligned}
$$

- Q: Express $x_1^2 - 2x_1x_2 + 2x_2$ as $\mathbf{x}^\top\mathbf{A}\mathbf{x} + \mathbf{b}^\top\mathbf{x}$.
- Q: Show that $\nabla f_2(\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$.

Hessian and Jacobian matrices

- *Hessian*: the matrix of second-order partial derivatives of $f : \mathbb{R}^N \mapsto \mathbb{R}$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

  Q: For symmetric $\mathbf{A}$, show that $\nabla^2 f_2(\mathbf{x}) = 2\mathbf{A}$.

- *Jacobian*: matrix whose rows are the gradients of $\mathbf{f} : \mathbb{R}^N \mapsto \mathbb{R}^M$

$$\mathbf{J}_{M \times N} = \frac{d\mathbf{f}}{d\mathbf{x}} = \begin{bmatrix} \nabla f_1(\mathbf{x})^\top \\ \vdots \\ \nabla f_M(\mathbf{x})^\top \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1} & \cdots & \frac{\partial f_M}{\partial x_N} \end{bmatrix}$$

  or, component-wise: $\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}$

Taylor's series expansion and mean value theorem

- Univariate function:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \mathcal{O}\left((x - x_0)^2\right)$$

- Multivariate function:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\nabla f(\mathbf{x}_0))^\top(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \nabla^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \mathcal{O}\left(\|\mathbf{x} - \mathbf{x}_0\|^2\right)$$

- Mean value theorem (MVT): There exist y and z between x and x_0 such that

$$f(x) = f(x_0) + f(y)'(x - x_0)$$

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(z)}{2}(x - x_0)^2$$

MVT can be generalized to multivariate functions.

Special square matrices

- *Symmetric matrix*: $\mathbf{A} = \mathbf{A}^\top$ for $\mathbf{A} \in \mathbb{R}^{N \times N}$
- *Hermitian matrix*: $\mathbf{B} = \mathbf{B}^H$ for $\mathbf{B} \in \mathbb{C}^{N \times N}$
- *Orthonormal matrix*: $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top\mathbf{A} = \mathbf{I}$ for $\mathbf{A} \in \mathbb{R}^{N \times N}$
- *Unitary matrix*: $\mathbf{B}\mathbf{B}^H = \mathbf{B}^H\mathbf{B} = \mathbf{I}$ for $\mathbf{B} \in \mathbb{C}^{N \times N}$

Matrix Inverse: Definition

Definition

For a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, an *inverse matrix* $\mathbf{A}^{-1}$ is a matrix such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n,$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

- Not all matrices are invertible.
- If $\mathbf{A}^{-1}$ exists, $\mathbf{A}$ is called *nonsingular* or *invertible*.
- If no inverse exists, $\mathbf{A}$ is called *singular*.

Matrix Inverse: Conditions

When does $\mathbf{A}^{-1}$ exist?

A square matrix $\mathbf{A}$ is invertible if and only if:

- $\det(\mathbf{A}) \neq 0$
- $\mathrm{rank}(\mathbf{A}) = n$
- Its columns (and rows) are linearly independent

Key Properties

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

Matrix Inverse: Example

2x2 Case

For

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

if $\det(\mathbf{A}) = ad - bc \neq 0$, then

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

- Easy to compute for 2×2.
- For larger matrices: use Gaussian elimination or decomposition (LU, QR).

Pseudo-Inverse of a Matrix

Motivation: For non-square or singular matrices, the inverse $\mathbf{A}^{-1}$ does not exist. The **Moore–Penrose pseudo-inverse** $\mathbf{A}^{\dagger}$ generalizes the concept of matrix inverse.

Definition: The pseudo-inverse $\mathbf{A}^{\dagger}$ of a real matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the unique matrix $\mathbf{A}^{\dagger} \in \mathbb{R}^{n \times m}$ satisfying:

$$
\begin{aligned}
&1.\ \mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}, \\
&2.\ \mathbf{A}^{\dagger}\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{A}^{\dagger}, \\
&3.\ (\mathbf{A}\mathbf{A}^{\dagger})^{\top} = \mathbf{A}\mathbf{A}^{\dagger}, \\
&4.\ (\mathbf{A}^{\dagger}\mathbf{A})^{\top} = \mathbf{A}^{\dagger}\mathbf{A}.
\end{aligned}
$$

Special cases:

- If $\mathbf{A}$ is invertible and square, $\mathbf{A}^{\dagger} = \mathbf{A}^{-1}$.
- If $\mathbf{A}$ has full column rank $(m > n)$, then $\mathbf{A}^{\dagger} = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}$.
- If $\mathbf{A}$ has full row rank $(m < n)$, then $\mathbf{A}^{\dagger} = \mathbf{A}^{\top}(\mathbf{A}\mathbf{A}^{\top})^{-1}$.

Geometric Interpretation

Key idea: $\mathbf{A}^\dagger$ provides the **minimum-norm least squares solution** to an over- or under-determined system.

For $\mathbf{A}\mathbf{x} = \mathbf{b}$:

$$\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$$

- If $\mathbf{A}$ is tall $(m > n)$: $\mathbf{A}^\dagger$ finds the best-fit $\mathbf{x}$ minimizing residuals.
- If $\mathbf{A}$ is wide $(m < n)$: $\mathbf{A}^\dagger$ finds the *minimum-norm* $\mathbf{x}$ satisfying $\mathbf{A}\mathbf{x} = \mathbf{b}$ (if consistent).

Interpretation: $\mathbf{A}^\dagger$ projects $\mathbf{b}$ orthogonally onto Range$(\mathbf{A})$ and computes the solution x^* in the least-squares sense.

Computation via SVD

Let the singular value decomposition (SVD) of $\mathbf{A}$ be:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top, \quad \mathbf{U} \in \mathbb{R}^{m\times m},\ \mathbf{V} \in \mathbb{R}^{n\times n}$$

where $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$ and $r = \operatorname{rank}(\mathbf{A})$.

Then,

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

where $\boldsymbol{\Sigma}^+$ is obtained by taking the reciprocal of nonzero singular values:

$$\boldsymbol{\Sigma}^+ = \operatorname{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0\right)$$

Benefits:

- Works for any $\mathbf{A}$ (square, rectangular, or singular).
- Provides a numerically stable way to compute $\mathbf{A}^\dagger$.

Example: Pseudo-Inverse in Least Squares

Consider $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$

Since $\mathbf{A}$ is not square and not full rank:

$$\mathbf{A}^{\dagger} = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

Then, the least squares solution:

$$\mathbf{x}^{*} = \mathbf{A}^{\dagger}\mathbf{b} = \frac{1}{4}\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Interpretation: $\mathbf{A}^{\dagger}$ automatically handles rank deficiency and gives the minimum-norm least squares solution.

Trace

- For square matrices: $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^\top) = \sum_{i=1}^n A_{ii} = \sum_{i=1}^n \lambda_i(\mathbf{A})$ (sum of all eigenvalues of $\mathbf{A}$)
- Inner product of two matrices: $\mathrm{tr}(\mathbf{A}^\top \mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A}^\top) = \sum_{i,j} A_{ij} B_{ij}$
- Cyclic property: $\mathrm{tr}(\mathbf{ABCD}) = \mathrm{tr}(\mathbf{BCDA}) = \mathrm{tr}(\mathbf{CDAB}) = \mathrm{tr}(\mathbf{DABC})$
- $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$, $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$
- Generally, $\mathrm{tr}(\mathbf{AB}) \neq \mathrm{tr}(\mathbf{A}) \times \mathrm{tr}(\mathbf{B})$.

Determinant

- The determinant of $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a scalar, denoted by $\det(\mathbf{A})$ or $|\mathbf{A}|$.
- **Leibniz formula**: $\det(\mathbf{A}) = \sum_{\tau\in S_n} \operatorname{sgn}(\tau) \prod_{i=1}^n a_{i\tau(i)} = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{\sigma(i)i}$. The sum involves all permutations. Each summand is a product of entries of the matrix multiplied with a sign, which returns $+1$ and -1 for even and odd permutations, respectively.
- **Laplace expansion**: $\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} M_{ij}$, where the *minor* M_{ij} is the determinant of the $(n-1)\times(n-1)$ matrix that results from $\mathbf{A}$ by removing its ith row and jth column.
- For a 2×2 matrix, $\det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.
- For a 3×3 matrix,

$$\det(\mathbf{A}) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}.$$

Geometric Interpretation of Determinant (1/2)[1]

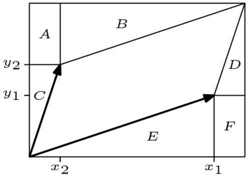

Figure: A parallelogram formed by two bold vectors. Let S(parallelogram) denote its area.

$$\begin{aligned}
S(\text{parallelogram}) &= S(\text{rectangle}) - S(A) - S(B) - \cdots - S(F) && (1)\\
&= (x_1 + x_2)(y_1 + y_2) - x_2 y_1 - \frac{x_1 y_1}{2} - \frac{x_2 y_2}{2} - \frac{x_2 y_2}{2} - \frac{x_1 y_1}{2} - x_2 y_1 && (2)\\
&= x_1 y_2 - x_2 y_1 && (3)\\
&= \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} && (4)
\end{aligned}$$

[1] https://mathresearch.utsa.edu/wiki/index.php?title=The_Geometric_Interpretation_of_the_Determinant

Geometric Interpretation of Determinant (2/2)[2]

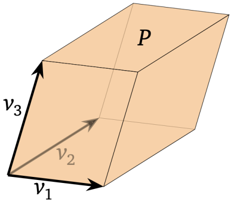

Figure: A parallelepiped P formed by three bold vectors. Let $\mathrm{vol}(P)$ denote its volume.

Theorem (Determinants and volumes)

Let $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \in \mathbb{R}^{n\times n}$, *let* P *be the parallelepiped determined by these vectors. Then,* $|\det(A)| = \mathrm{vol}(P)$.

[2] `https://textbooks.math.gatech.edu/ila/determinants-volumes.html`

Properties of Determinant[3]

- $\det(\mathbf{A}^\top) = \det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i(\mathbf{A})$ (product of all eigenvalues of $\mathbf{A}$)
- $\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$
- $\det(k\mathbf{A}) = k^n \det(\mathbf{A})$ for $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $k \in \mathbb{R}$.
- $\mathbf{A}$ is invertible $\Leftrightarrow \det(\mathbf{A}) \neq 0$.
- $\det(\mathbf{A}+\mathbf{B}) \neq \det(\mathbf{A}) + \det(\mathbf{B})$ in general
- $\det(\mathbf{AB}) = \det(\mathbf{A}) \times \det(\mathbf{B})$

[3] https://ocw.mit.edu/courses/18-06sc-linear-algebra-fall-2011/5dd3f8ec0a398fd74264fef3fd591f81_MIT18_06SCF11_Ses2.5sum.pdf

Effect of Row/Column Operations and Basis Change on $\det(\mathbf{A})$

Row/Column Operations

- Row/Column swap $\Rightarrow$ sign change: $\det(\text{swap}_i^j(\mathbf{A})) = -\det(\mathbf{A})$
- Row/Column scaling by $\alpha \neq 0 \Rightarrow$ determinant scales by α: $\det(\alpha\, \text{row}_i(\mathbf{A})) = \alpha \det(\mathbf{A})$
- Add multiple of one row/column to another $\Rightarrow$ unchanged: $\det(\text{row}_i + \beta\, \text{row}_j) = \det(\mathbf{A})$

Under Coordinate Transformation (Similarity Transform):

$$\boxed{\det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) = \det(\mathbf{A})}$$

Interpretation: The determinant is **basis invariant**: it depends only on the linear transformation itself, not on the coordinate system used to represent it.

Geometric meaning: Changing the basis reshapes the axes but does not change the volume scaling factor.

Determinant of Special Matrix

Matrix Type	**Formula for** $\det(\mathbf{A})$	**Key Idea**
Upper/Lower Triangular	$\prod_i a_{ii}$	Product of diagonal entries
Orthogonal / Unitary	$\lvert\det(\mathbf{A})\rvert = 1$	Rotation / reflection
Permutation	± 1	Sign $=$ parity of swaps
Block Triangular	$\det(\mathbf{A}) = \det(\mathbf{B})\det(\mathbf{C})$	Diagonal blocks multiply

Matrix Norms

A matrix norm is a real-valued function $\|\cdot\| : \mathbb{C}^{m\times n} \mapsto \mathbb{R}_+$ satisfying the following three properties for all $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m\times n}$, and $a \in \mathbb{C}$:

1. *Positive definiteness:* $\|\mathbf{A}\| \geq 0$, and $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}$
2. *Absolute homogeneity (scaling):* $\|a\mathbf{A}\| = |a| \cdot \|\mathbf{A}\|$
3. *Subadditivity (triangle inequality):* $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$
4. *Every norm can be rescaled to be sub-multiplicative:* $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$

Entrywise Matrix Norms

$$\boxed{\|\mathbf{A}\|_{p,p} = \|\text{vec}(\mathbf{A})\|_p = \left(\sum_{i,j} |A_{ij}|^p\right)^{1/p}}$$

The special case $p = 2$ is the Frobenius norm, and $p = \infty$ yields the maximum norm.

1. $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} |A_{ij}|^2} = \sqrt{\text{tr}\left(\mathbf{A}^H\mathbf{A}\right)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(\mathbf{A})}$,
2. $\|\mathbf{A}\|_{\max} = \max_{ij} |A_{ij}|$

Matrix Norms Induced by Vector Norms

$$\boxed{\|\mathbf{A}\|_{\alpha,\beta} = \sup\left\{\frac{\|\mathbf{A}\mathbf{x}\|_\beta}{\|\mathbf{x}\|_\alpha} : \mathbf{x} \in \mathbb{C}^n \text{ with } \mathbf{x} \neq \mathbf{0}\right\}}$$

If $\alpha = \beta = p$, then $\|\mathbf{A}\|_p = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$. Special cases:

1. $\|\mathbf{A}\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^m |A_{ij}|$: the maximum absolute *column sum* of $\mathbf{A}$.
2. $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^n |A_{ij}|$: the maximum absolute *row sum* of $\mathbf{A}$.
3. $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}\left(\mathbf{A}^H\mathbf{A}\right)} = \sigma_{\max}(\mathbf{A})$: the largest singular value of $\mathbf{A}$ (aka spectral norm).

Schatten Norms

Apply the ℓ_p norm to the vector of all singular values of a matrix.

$$\boxed{\|\mathbf{A}\|_p = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^p(\mathbf{A}) \right)^{\frac{1}{p}}}$$

Special cases:

1. $p = 1 \implies$ nuclear norm: $\|A\|_* = \mathrm{tr}\left(\sqrt{\mathbf{A}^H\mathbf{A}}\right) = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})$,
2. $p = 2 \implies$ Frobenius norm.
3. $p = \infty \implies$ spectral norm.

Matrix Norm Inequalities

Let $\|\mathbf{A}\|_p$ denote the norm induced by the vector ℓ_p-norm. For matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ of rank r, the following inequalities hold:

1. $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{r}\|\mathbf{A}\|_2$
2. $\|\mathbf{A}\|_F \leq \|\mathbf{A}\|_* \leq \sqrt{r}\|\mathbf{A}\|_F$
3. $\|\mathbf{A}\|_{\max} \leq \|\mathbf{A}\|_2 \leq \sqrt{mn}\|\mathbf{A}\|_{\max}$
4. $\frac{1}{\sqrt{n}}\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\|_2 \leq \sqrt{m}\|\mathbf{A}\|_\infty$
5. $\frac{1}{\sqrt{m}}\|\mathbf{A}\|_1 \leq \|\mathbf{A}\|_2 \leq \sqrt{n}\|\mathbf{A}\|_1$
6. $\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_1\|\mathbf{A}\|_\infty}$