

# CS229 Section: Linear Algebra

# Aditya Agrawal

Slides adapted from past CS229 teams

June 28, 2024

# Outline

- 1 Basic Concepts and Notation
- 2 Matrix Multiplication
- 3 Operations and Properties
- 4 Matrix Calculus

# Basic Concepts and Notation







# Outline

- 1 Basic Concepts and Notation
- 2 **Matrix Multiplication**
- 3 Operations and Properties
- 4 Matrix Calculus

## Vector-Vector Product

- *inner product* or *dot product*

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i.$$

- *outer product*

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}.$$



# Matrix-Vector Product

- If we write  $A$  by rows, then we can express  $Ax$  as,

$$y = Ax = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}.$$

# Matrix-Vector Product

- If we write  $A$  by columns, then we have:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \dots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a^1 \end{bmatrix} x_1 + \begin{bmatrix} a^2 \end{bmatrix} x_2 + \dots + \begin{bmatrix} a^n \end{bmatrix} x_n \quad (1)$$

$y$  is a **linear combination** of the *columns* of  $A$ .



# Matrix-Vector Product

It is also possible to multiply on the left by a row vector.

- expressing  $A$  in terms of rows we have:

$$y^T = x^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}$$

$$= x_1 \begin{bmatrix} - & a_1^T & - \end{bmatrix} + x_2 \begin{bmatrix} - & a_2^T & - \end{bmatrix} + \dots + x_m \begin{bmatrix} - & a_m^T & - \end{bmatrix}$$

$y^T$  is a linear combination of the *rows* of  $A$ .



## Matrix-Matrix Multiplication (different views)

2. As a sum of outer products

$$C = AB = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \dots & a^p \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & b_1^T & - \\ - & b_2^T & - \\ & \vdots & \\ - & b_p^T & - \end{bmatrix} = \sum_{i=1}^p a^i b_i^T .$$

## Matrix-Matrix Multiplication (different views)

3. As a set of matrix-vector products.

$$C = AB = A \begin{bmatrix} \begin{array}{c} | \\ b^1 \\ | \end{array} & \begin{array}{c} | \\ b^2 \\ | \end{array} & \dots & \begin{array}{c} | \\ b^n \\ | \end{array} \end{bmatrix} = \begin{bmatrix} \begin{array}{c} | \\ Ab^1 \\ | \end{array} & \begin{array}{c} | \\ Ab^2 \\ | \end{array} & \dots & \begin{array}{c} | \\ Ab^n \\ | \end{array} \end{bmatrix}. \quad (2)$$

Here the  $i$ th column of  $C$  is given by the matrix-vector product with the vector on the right,  $c_i = Ab_i$ . These matrix-vector products can in turn be interpreted using both viewpoints given in the previous subsection.

## Matrix-Matrix Multiplication (different views)

4. As a set of vector-matrix products.

$$C = AB = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \quad B = \begin{bmatrix} - & a_1^T B & - \\ - & a_2^T B & - \\ & \vdots & \\ - & a_m^T B & - \end{bmatrix}.$$





# Outline

- 1 Basic Concepts and Notation
- 2 Matrix Multiplication
- 3 Operations and Properties**
- 4 Matrix Calculus

# Operations and Properties



# Trace

The **trace** of a square matrix  $A \in \mathbb{R}^{n \times n}$ , denoted  $\text{tr}A$ , is the sum of diagonal elements in the matrix:

$$\text{tr}A = \sum_{i=1}^n A_{ii}.$$

The trace has the following properties:

- For  $A \in \mathbb{R}^{n \times n}$ ,  $\text{tr}A = \text{tr}A^T$ .
- For  $A, B \in \mathbb{R}^{n \times n}$ ,  $\text{tr}(A + B) = \text{tr}A + \text{tr}B$ .
- For  $A \in \mathbb{R}^{n \times n}$ ,  $t \in \mathbb{R}$ ,  $\text{tr}(tA) = t \text{tr}A$ .
- For  $A, B$  such that  $AB$  is square,  $\text{tr}AB = \text{tr}BA$ .
- For  $A, B, C$  such that  $ABC$  is square,  $\text{tr}ABC = \text{tr}BCA = \text{tr}CAB$ , and so on for the product of more matrices.





## Examples of Norms

In fact, all three norms presented so far are examples of the family of  $\ell_p$  norms, which are parameterized by a real number  $p \geq 1$ , and defined as

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



# Matrix Norms

Norms can also be defined for matrices, such as the Frobenius norm,

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\text{tr}(A^T A)}.$$

Many other norms exist, but they are beyond the scope of this review.

# Linear Independence

A set of vectors  $\{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^m$  is said to be **(linearly) dependent** if one vector belonging to the set *can* be represented as a linear combination of the remaining vectors; that is, if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values  $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$ ; otherwise, the vectors are **(linearly) independent**.

# Linear Independence

A set of vectors  $\{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^m$  is said to be **(linearly) dependent** if one vector belonging to the set *can* be represented as a linear combination of the remaining vectors; that is, if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values  $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$ ; otherwise, the vectors are **(linearly) independent**.

**Example:**

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

are linearly dependent because  $x_3 = -2x_1 + x_2$ .

# Rank of a Matrix

- The **column rank** of a matrix  $A \in \mathbb{R}^{m \times n}$  is the largest number of columns of  $A$  that constitute a linearly independent set.

# Rank of a Matrix

- The **column rank** of a matrix  $A \in \mathbb{R}^{m \times n}$  is the largest number of columns of  $A$  that constitute a linearly independent set.
- The **row rank** is the largest number of rows of  $A$  that constitute a linearly independent set.

# Rank of a Matrix

- The **column rank** of a matrix  $A \in \mathbb{R}^{m \times n}$  is the largest number of columns of  $A$  that constitute a linearly independent set.
- The **row rank** is the largest number of rows of  $A$  that constitute a linearly independent set.
- For any matrix  $A \in \mathbb{R}^{m \times n}$ , it turns out that the column rank of  $A$  is equal to the row rank of  $A$  (prove it yourself!), and so both quantities are referred to collectively as the **rank** of  $A$ , denoted as  $\text{rank}(A)$ .

## Properties of the Rank

- For  $A \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A) \leq \min(m, n)$ . If  $\text{rank}(A) = \min(m, n)$ , then  $A$  is said to be **full rank**.

## Properties of the Rank

- For  $A \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A) \leq \min(m, n)$ . If  $\text{rank}(A) = \min(m, n)$ , then  $A$  is said to be **full rank**.
- For  $A \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A) = \text{rank}(A^T)$ .







# The Inverse of a Square Matrix

- The **inverse** of a square matrix  $A \in \mathbb{R}^{n \times n}$  is denoted  $A^{-1}$ , and is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

# The Inverse of a Square Matrix

- The **inverse** of a square matrix  $A \in \mathbb{R}^{n \times n}$  is denoted  $A^{-1}$ , and is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

- We say that  $A$  is **invertible** or **non-singular** if  $A^{-1}$  exists and **non-invertible** or **singular** otherwise.

# The Inverse of a Square Matrix

- The **inverse** of a square matrix  $A \in \mathbb{R}^{n \times n}$  is denoted  $A^{-1}$ , and is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

- We say that  $A$  is **invertible** or **non-singular** if  $A^{-1}$  exists and **non-invertible** or **singular** otherwise.
- In order for a square matrix  $A$  to have an inverse  $A^{-1}$ , then  $A$  must be full rank.

# The Inverse of a Square Matrix

- The **inverse** of a square matrix  $A \in \mathbb{R}^{n \times n}$  is denoted  $A^{-1}$ , and is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

- We say that  $A$  is **invertible** or **non-singular** if  $A^{-1}$  exists and **non-invertible** or **singular** otherwise.
- In order for a square matrix  $A$  to have an inverse  $A^{-1}$ , then  $A$  must be full rank.
- Properties (Assuming  $A, B \in \mathbb{R}^{n \times n}$  are non-singular):
  - ▶  $(A^{-1})^{-1} = A$
  - ▶  $(AB)^{-1} = B^{-1}A^{-1}$
  - ▶  $(A^{-1})^T = (A^T)^{-1}$ . For this reason this matrix is often denoted  $A^{-T}$ .

# Orthogonal Matrices

- Two vectors  $x, y \in \mathbb{R}^n$  are **orthogonal** if  $x^T y = 0$ .
- A vector  $x \in \mathbb{R}^n$  is **normalized** if  $\|x\|_2 = 1$ .
- A square matrix  $U \in \mathbb{R}^{n \times n}$  is **orthogonal** if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being **orthonormal**).

# Orthogonal Matrices

- Two vectors  $x, y \in \mathbb{R}^n$  are **orthogonal** if  $x^T y = 0$ .
- A vector  $x \in \mathbb{R}^n$  is **normalized** if  $\|x\|_2 = 1$ .
- A square matrix  $U \in \mathbb{R}^{n \times n}$  is **orthogonal** if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being **orthonormal**).

- **Properties:**

- ▶ The inverse of an orthogonal matrix is its transpose.

$$U^T U = I = U U^T.$$



# Orthogonal Matrices

- Two vectors  $x, y \in \mathbb{R}^n$  are **orthogonal** if  $x^T y = 0$ .
- A vector  $x \in \mathbb{R}^n$  is **normalized** if  $\|x\|_2 = 1$ .
- A square matrix  $U \in \mathbb{R}^{n \times n}$  is **orthogonal** if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being **orthonormal**).

- **Properties:**

- ▶ The inverse of an orthogonal matrix is its transpose.

$$U^T U = I = U U^T.$$

- ▶ Operating on a vector with an orthogonal matrix will not change its Euclidean norm, i.e.,

$$\|Ux\|_2 = \|x\|_2$$

for any  $x \in \mathbb{R}^n$ ,  $U \in \mathbb{R}^{n \times n}$  orthogonal.

# Span and Projection

- The **span** of a set of vectors  $\{x_1, x_2, \dots, x_n\}$  is the set of all vectors that can be expressed as a linear combination of  $\{x_1, \dots, x_n\}$ . That is,

$$\text{span}(\{x_1, \dots, x_n\}) = \left\{ v : v = \sum_{i=1}^n \alpha_i x_i, \alpha_i \in \mathbb{R} \right\}.$$

# Span and Projection

- The **span** of a set of vectors  $\{x_1, x_2, \dots, x_n\}$  is the set of all vectors that can be expressed as a linear combination of  $\{x_1, \dots, x_n\}$ . That is,

$$\text{span}(\{x_1, \dots, x_n\}) = \left\{ v : v = \sum_{i=1}^n \alpha_i x_i, \alpha_i \in \mathbb{R} \right\}.$$

- The **projection** of a vector  $y \in \mathbb{R}^m$  onto the span of  $\{x_1, \dots, x_n\}$  is the vector  $v \in \text{span}(\{x_1, \dots, x_n\})$ , such that  $v$  is as close as possible to  $y$ , as measured by the Euclidean norm  $\|v - y\|_2$ .

$$\text{Proj}(y; \{x_1, \dots, x_n\}) = \operatorname{argmin}_{v \in \text{span}(\{x_1, \dots, x_n\})} \|y - v\|_2.$$

# Range

- The **range** or the column space of a matrix  $A \in \mathbb{R}^{m \times n}$ , denoted  $\mathcal{R}(A)$ , is the the span of the columns of  $A$ . In other words,

$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}.$$

# Range

- The **range** or the column space of a matrix  $A \in \mathbb{R}^{m \times n}$ , denoted  $\mathcal{R}(A)$ , is the the span of the columns of  $A$ . In other words,

$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}.$$

- Assuming  $A$  is full rank and  $n < m$ , the projection of a vector  $y \in \mathbb{R}^m$  onto the range of  $A$  is given by,

$$\text{Proj}(y; A) = \operatorname{argmin}_{v \in \mathcal{R}(A)} \|v - y\|_2.$$

# Null space

The **nullspace** of a matrix  $A \in \mathbb{R}^{m \times n}$ , denoted  $\mathcal{N}(A)$  is the set of all vectors that equal 0 when multiplied by  $A$ , i.e.,

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

# The Determinant

The **determinant** of a square matrix  $A \in \mathbb{R}^{n \times n}$ , is a function  $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ , and is denoted  $|A|$  or  $\det A$ .

Given a matrix

$$\begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix},$$

consider the set of points  $S \subset \mathbb{R}^n$  as follows:

$$S = \{v \in \mathbb{R}^n : v = \sum_{i=1}^n \alpha_i a_i \text{ where } 0 \leq \alpha_i \leq 1, i = 1, \dots, n\}.$$

The absolute value of the determinant of  $A$  is a measure of the “volume” of the set  $S$ .

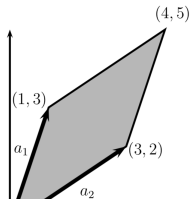
# The Determinant: Intuition

For example, consider the  $2 \times 2$  matrix,

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \quad (3)$$

Here, the rows of the matrix are

$$a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad a_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$





# The Determinant: Properties

Algebraically, the determinant satisfies the following three properties:

1. The determinant of the identity is 1,  $|I| = 1$ . (Geometrically, the volume of a unit hypercube is 1).

# The Determinant: Properties

Algebraically, the determinant satisfies the following three properties:

1. The determinant of the identity is 1,  $|I| = 1$ . (Geometrically, the volume of a unit hypercube is 1).
2. Given a matrix  $A \in \mathbb{R}^{n \times n}$ , if we multiply a single row in  $A$  by a scalar  $t \in \mathbb{R}$ , then the determinant of the new matrix is  $t|A|$ , (Geometrically, multiplying one of the sides of the set  $S$  by a factor  $t$  causes the volume to increase by a factor  $t$ .)

# The Determinant: Properties

Algebraically, the determinant satisfies the following three properties:

1. The determinant of the identity is 1,  $|I| = 1$ . (Geometrically, the volume of a unit hypercube is 1).
2. Given a matrix  $A \in \mathbb{R}^{n \times n}$ , if we multiply a single row in  $A$  by a scalar  $t \in \mathbb{R}$ , then the determinant of the new matrix is  $t|A|$ , (Geometrically, multiplying one of the sides of the set  $S$  by a factor  $t$  causes the volume to increase by a factor  $t$ .)
3. If we exchange any two rows  $a_i^T$  and  $a_j^T$  of  $A$ , then the determinant of the new matrix is  $-|A|$ , for example

# The Determinant: Properties

Algebraically, the determinant satisfies the following three properties:

1. The determinant of the identity is 1,  $|I| = 1$ . (Geometrically, the volume of a unit hypercube is 1).
2. Given a matrix  $A \in \mathbb{R}^{n \times n}$ , if we multiply a single row in  $A$  by a scalar  $t \in \mathbb{R}$ , then the determinant of the new matrix is  $t|A|$ , (Geometrically, multiplying one of the sides of the set  $S$  by a factor  $t$  causes the volume to increase by a factor  $t$ .)
3. If we exchange any two rows  $a_i^T$  and  $a_j^T$  of  $A$ , then the determinant of the new matrix is  $-|A|$ , for example

# The Determinant: Properties

Algebraically, the determinant satisfies the following three properties:

1. The determinant of the identity is 1,  $|I| = 1$ . (Geometrically, the volume of a unit hypercube is 1).
2. Given a matrix  $A \in \mathbb{R}^{n \times n}$ , if we multiply a single row in  $A$  by a scalar  $t \in \mathbb{R}$ , then the determinant of the new matrix is  $t|A|$ , (Geometrically, multiplying one of the sides of the set  $S$  by a factor  $t$  causes the volume to increase by a factor  $t$ .)
3. If we exchange any two rows  $a_i^T$  and  $a_j^T$  of  $A$ , then the determinant of the new matrix is  $-|A|$ , for example

In case you are wondering, it is not immediately obvious that a function satisfying the above three properties exists. In fact, though, such a function does exist, and is unique (which we will not prove here).

# The Determinant: Properties

- For  $A \in \mathbb{R}^{n \times n}$ ,  $|A| = |A^T|$ .
- For  $A, B \in \mathbb{R}^{n \times n}$ ,  $|AB| = |A||B|$ .
- For  $A \in \mathbb{R}^{n \times n}$ ,  $|A| = 0$  if and only if  $A$  is singular (i.e., non-invertible). (If  $A$  is singular then it does not have full rank, and hence its columns are linearly dependent. In this case, the set  $S$  corresponds to a “flat sheet” within the  $n$ -dimensional space and hence has zero volume.)
- For  $A \in \mathbb{R}^{n \times n}$  and  $A$  non-singular,  $|A^{-1}| = 1/|A|$ .

# The Determinant: Formula

Let  $A \in \mathbb{R}^{n \times n}$ ,  $A_{\setminus i, \setminus j} \in \mathbb{R}^{(n-1) \times (n-1)}$  be the *matrix* that results from deleting the  $i$ th row and  $j$ th column from  $A$ .

The general (recursive) formula for the determinant is

$$\begin{aligned} |A| &= \sum_{i=1}^n (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}| \quad (\text{for any } j \in 1, \dots, n) \\ &= \sum_{j=1}^n (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}| \quad (\text{for any } i \in 1, \dots, n) \end{aligned}$$

with the initial case that  $|A| = a_{11}$  for  $A \in \mathbb{R}^{1 \times 1}$ . If we were to expand this formula completely for  $A \in \mathbb{R}^{n \times n}$ , there would be a total of  $n!$  ( $n$  factorial) different terms. For this reason, we hardly ever explicitly write the complete equation of the determinant for matrices bigger than  $3 \times 3$ .

# The Determinant: Examples

However, the equations for determinants of matrices up to size  $3 \times 3$  are fairly common, and it is good to know them:

$$\begin{aligned} |[a_{11}]| &= a_{11} \\ \left| \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right| &= a_{11}a_{22} - a_{12}a_{21} \\ \left| \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \right| &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{aligned}$$



## Quadratic Forms

Given a square matrix  $A \in \mathbb{R}^{n \times n}$  and a vector  $x \in \mathbb{R}^n$ , the scalar value  $x^T A x$  is called a **quadratic form**. Written explicitly, we see that

$$x^T A x = \sum_{i=1}^n x_i (A x)_i = \sum_{i=1}^n x_i \left( \sum_{j=1}^n A_{ij} x_j \right) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j .$$

## Quadratic Forms

Given a square matrix  $A \in \mathbb{R}^{n \times n}$  and a vector  $x \in \mathbb{R}^n$ , the scalar value  $x^T A x$  is called a **quadratic form**. Written explicitly, we see that

$$x^T A x = \sum_{i=1}^n x_i (A x)_i = \sum_{i=1}^n x_i \left( \sum_{j=1}^n A_{ij} x_j \right) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j .$$

We often implicitly assume that the matrices appearing in a quadratic form are symmetric.

$$x^T A x = (x^T A x)^T = x^T A^T x = x^T \left( \frac{1}{2} A + \frac{1}{2} A^T \right) x,$$

# Positive Semidefinite Matrices

A symmetric matrix  $A \in \mathbb{S}^n$  is:

- **positive definite** (PD), denoted  $A \succ 0$  if for all non-zero vectors  $x \in \mathbb{R}^n$ ,  $x^T A x > 0$ .
- **positive semidefinite** (PSD), denoted  $A \succeq 0$  if for all vectors  $x^T A x \geq 0$ .
- **negative definite** (ND), denoted  $A \prec 0$  if for all non-zero  $x \in \mathbb{R}^n$ ,  $x^T A x < 0$ .
- **negative semidefinite** (NSD), denoted  $A \preceq 0$  ) if for all  $x \in \mathbb{R}^n$ ,  $x^T A x \leq 0$ .
- **indefinite**, if it is neither positive semidefinite nor negative semidefinite — i.e., if there exists  $x_1, x_2 \in \mathbb{R}^n$  such that  $x_1^T A x_1 > 0$  and  $x_2^T A x_2 < 0$ .

# Positive Semidefinite Matrices

- One important property of positive definite and negative definite matrices is that they are always full rank, and hence, invertible.
- Given any matrix  $A \in \mathbb{R}^{m \times n}$  (not necessarily symmetric or even square), the matrix  $G = A^T A$  (sometimes called a ***Gram matrix***) is always positive semidefinite. Further, if  $m \geq n$  and  $A$  is full rank, then  $G = A^T A$  is positive definite.

# Eigenvalues and Eigenvectors

Given a square matrix  $A \in \mathbb{R}^{n \times n}$ , we say that  $\lambda \in \mathbb{C}$  is an **eigenvalue** of  $A$  and  $x \in \mathbb{C}^n$  is the corresponding **eigenvector** if

$$Ax = \lambda x, \quad x \neq 0.$$

Intuitively, this definition means that multiplying  $A$  by the vector  $x$  results in a new vector that points in the same direction as  $x$ , but scaled by a factor  $\lambda$ .

# Eigenvalues and Eigenvectors

We can rewrite the equation above to state that  $(\lambda, x)$  is an eigenvalue-eigenvector pair of  $A$  if,

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

But  $(\lambda I - A)x = 0$  has a non-zero solution to  $x$  if and only if  $(\lambda I - A)$  has a non-empty nullspace, which is only the case if  $(\lambda I - A)$  is singular, i.e.,

$$|(\lambda I - A)| = 0.$$

We can now use the previous definition of the determinant to expand this expression  $|(\lambda I - A)|$  into a (very large) polynomial in  $\lambda$ , where  $\lambda$  will have degree  $n$ . It's often called the **characteristic polynomial** of the matrix  $A$ .

# Properties of eigenvalues and eigenvectors

- The trace of a  $A$  is equal to the sum of its eigenvalues,

$$\text{tr}A = \sum_{i=1}^n \lambda_i.$$

# Properties of eigenvalues and eigenvectors

- The trace of a  $A$  is equal to the sum of its eigenvalues,

$$\text{tr}A = \sum_{i=1}^n \lambda_i.$$

- The determinant of  $A$  is equal to the product of its eigenvalues,

$$|A| = \prod_{i=1}^n \lambda_i.$$



# Properties of eigenvalues and eigenvectors

- The trace of a  $A$  is equal to the sum of its eigenvalues,

$$\text{tr}A = \sum_{i=1}^n \lambda_i.$$

- The determinant of  $A$  is equal to the product of its eigenvalues,

$$|A| = \prod_{i=1}^n \lambda_i.$$

- The rank of  $A$  is equal to the number of non-zero eigenvalues of  $A$ .

# Properties of eigenvalues and eigenvectors

- The trace of a  $A$  is equal to the sum of its eigenvalues,

$$\text{tr}A = \sum_{i=1}^n \lambda_i.$$

- The determinant of  $A$  is equal to the product of its eigenvalues,

$$|A| = \prod_{i=1}^n \lambda_i.$$

- The rank of  $A$  is equal to the number of non-zero eigenvalues of  $A$ .
- Suppose  $A$  is non-singular with eigenvalue  $\lambda$  and an associated eigenvector  $x$ . Then  $1/\lambda$  is an eigenvalue of  $A^{-1}$  with an associated eigenvector  $x$ , i.e.,  $A^{-1}x = (1/\lambda)x$ .

# Properties of eigenvalues and eigenvectors

- The trace of a  $A$  is equal to the sum of its eigenvalues,

$$\text{tr}A = \sum_{i=1}^n \lambda_i.$$

- The determinant of  $A$  is equal to the product of its eigenvalues,

$$|A| = \prod_{i=1}^n \lambda_i.$$

- The rank of  $A$  is equal to the number of non-zero eigenvalues of  $A$ .
- Suppose  $A$  is non-singular with eigenvalue  $\lambda$  and an associated eigenvector  $x$ . Then  $1/\lambda$  is an eigenvalue of  $A^{-1}$  with an associated eigenvector  $x$ , i.e.,  $A^{-1}x = (1/\lambda)x$ .
- The eigenvalues of a diagonal matrix  $D = \text{diag}(d_1, \dots, d_n)$  are just the diagonal entries  $d_1, \dots, d_n$ .

# Eigenvalues and Eigenvectors of Symmetric Matrices

Throughout this section, let's assume that  $A$  is a symmetric real matrix (i.e.,  $A^T = A$ ). We have the following properties:

1. All eigenvalues of  $A$  are real numbers. We denote them by  $\lambda_1, \dots, \lambda_n$ .
2. There exists a set of eigenvectors  $u_1, \dots, u_n$  such that (i) for all  $i$ ,  $u_i$  is an eigenvector with eigenvalue  $\lambda_i$  and (ii)  $u_1, \dots, u_n$  are unit vectors and orthogonal to each other.

# New Representation for Symmetric Matrices

- Let  $U$  be the orthonormal matrix that contains  $u_i$ 's as columns:

$$U = \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix}$$

## New Representation for Symmetric Matrices

- Let  $U$  be the orthonormal matrix that contains  $u_i$ 's as columns:

$$U = \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

- Let  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  be the diagonal matrix that contains  $\lambda_1, \dots, \lambda_n$ .

$$AU = \begin{bmatrix} | & | & \cdots & | \\ Au_1 & Au_2 & \cdots & Au_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & \cdots & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix} = U \text{diag}(\lambda_1, \dots, \lambda_n) = U\Lambda$$

## New Representation for Symmetric Matrices

- Let  $U$  be the orthonormal matrix that contains  $u_i$ 's as columns:

$$U = \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

- Let  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  be the diagonal matrix that contains  $\lambda_1, \dots, \lambda_n$ .

$$AU = \begin{bmatrix} | & | & \cdots & | \\ Au_1 & Au_2 & \cdots & Au_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & \cdots & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix} = U \text{diag}(\lambda_1, \dots, \lambda_n) = U\Lambda$$

- Recalling that orthonormal matrix  $U$  satisfies that  $UU^T = I$ , we can diagonalize matrix  $A$ :

$$A = AUU^T = U\Lambda U^T \quad (4)$$

## Background: representing vector w.r.t. another basis

- Any orthonormal matrix  $U = \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix}$  defines a new basis of  $\mathbb{R}^n$ .
- For any vector  $x \in \mathbb{R}^n$  can be represented as a linear combination of  $u_1, \dots, u_n$  with coefficient  $\hat{x}_1, \dots, \hat{x}_n$ :

$$x = \hat{x}_1 u_1 + \cdots + \hat{x}_n u_n = U\hat{x}$$

- Indeed, such  $\hat{x}$  uniquely exists

$$x = U\hat{x} \Leftrightarrow U^T x = \hat{x}$$

In other words, the vector  $\hat{x} = U^T x$  can serve as another representation of the vector  $x$  w.r.t the basis defined by  $U$ .



## “Diagonalizing” matrix-vector multiplication

- Left-multiplying matrix  $A$  can be viewed as left-multiplying a diagonal matrix w.r.t the basis of the eigenvectors.
  - ▶ Suppose  $x$  is a vector and  $\hat{x}$  is its representation w.r.t to the basis of  $U$ .
  - ▶ Let  $z = Ax$  be the matrix-vector product.
  - ▶ the representation  $z$  w.r.t the basis of  $U$ :

$$\hat{z} = U^T z = U^T A x = U^T U \Lambda U^T x = \Lambda \hat{x} = \begin{bmatrix} \lambda_1 \hat{x}_1 \\ \lambda_2 \hat{x}_2 \\ \vdots \\ \lambda_n \hat{x}_n \end{bmatrix}$$

- We see that left-multiplying matrix  $A$  in the original space is equivalent to left-multiplying the diagonal matrix  $\Lambda$  w.r.t the new basis, which is merely scaling each coordinate by the corresponding eigenvalue.

## “Diagonalizing” matrix-vector multiplication

Under the new basis, multiplying a matrix multiple times becomes much simpler as well. For example, suppose  $q = AAAx$ .

$$\hat{q} = U^T q = U^T AAAx = U^T U \Lambda U^T U \Lambda U^T U \Lambda U^T x = \Lambda^3 \hat{x} = \begin{bmatrix} \lambda_1^3 \hat{x}_1 \\ \lambda_2^3 \hat{x}_2 \\ \vdots \\ \lambda_n^3 \hat{x}_n \end{bmatrix}$$

## “Diagonalizing” quadratic form

As a directly corollary, the quadratic form  $x^T A x$  can also be simplified under the new basis

$$x^T A x = x^T U \Lambda U^T x = \hat{x}^T \Lambda \hat{x} = \sum_{i=1}^n \lambda_i \hat{x}_i^2$$

(Recall that with the old representation,  $x^T A x = \sum_{i=1, j=1}^n x_i x_j A_{ij}$  involves a sum of  $n^2$  terms instead of  $n$  terms in the equation above.)



# Outline

- 1 Basic Concepts and Notation
- 2 Matrix Multiplication
- 3 Operations and Properties
- 4 Matrix Calculus

# Matrix Calculus











# The Hessian

Suppose that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a function that takes a vector in  $\mathbb{R}^n$  and returns a real number. Then the **Hessian** matrix with respect to  $x$ , written  $\nabla_x^2 f(x)$  or simply as  $H$  is the  $n \times n$  matrix of partial derivatives,

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Note that the Hessian is always symmetric, since

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







## Gradients of Quadratic Function

Now consider the quadratic function  $f(x) = x^T A x$  for  $A \in \mathbb{S}^n$ . Remember that

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j.$$

To take the partial derivative, we'll consider the terms including  $x_k$  and  $x_k^2$  factors separately:

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial x_k} &= \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j \\ &= \frac{\partial}{\partial x_k} \left[ \sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right] \\ &= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k \end{aligned}$$

## Gradients of Quadratic Function

Now consider the quadratic function  $f(x) = x^T A x$  for  $A \in \mathbb{S}^n$ . Remember that

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j.$$

To take the partial derivative, we'll consider the terms including  $x_k$  and  $x_k^2$  factors separately:

$$\begin{aligned}\frac{\partial f(x)}{\partial x_k} &= \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j \\ &= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k \\ &= \sum_{i=1}^n A_{ik} x_i + \sum_{j=1}^n A_{kj} x_j = 2 \sum_{i=1}^n A_{ki} x_i,\end{aligned}$$





- $\nabla_x b^T x = b$
- $\nabla_x^2 b^T x = 0$
- $\nabla_x x^T A x = 2Ax$  (if  $A$  symmetric)
- $\nabla_x^2 x^T A x = 2A$  (if  $A$  symmetric)







