

# Linear Algebra

## Chapter 7: Distance and Approximation

University of Seoul  
School of Computer Science  
Minho Kim

# Table of contents

Introduction: Taxicab Geometry

Inner Product Spaces

Norms and Distance Functions

Least Squares Approximation

The Singular Value Decomposition

Applications

# Outline

Introduction: Taxicab Geometry

Inner Product Spaces

Norms and Distance Functions

Least Squares Approximation

The Singular Value Decomposition

Applications

# Outline

Introduction: Taxicab Geometry

**Inner Product Spaces**

Norms and Distance Functions

Least Squares Approximation

The Singular Value Decomposition

Applications

# Outline

Introduction: Taxicab Geometry

Inner Product Spaces

**Norms and Distance Functions**

Least Squares Approximation

The Singular Value Decomposition

Applications

# Outline

Introduction: Taxicab Geometry

Inner Product Spaces

Norms and Distance Functions

**Least Squares Approximation**

The Singular Value Decomposition

Applications

# Best Approximation

- ▶ Which vector in a subspace  $W$  best approximates (or is closest to) the vector  $v$  outside  $W$ ?

## Definition: The Best Approximation

If  $W$  is a subspace of a normed linear space  $V$  (e.g.,  $\mathbb{R}^n$ ) and if  $v$  is a vector in  $V$ , then the **best approximation to  $v$  in  $W$**  is the vector  $\bar{v}$  in  $W$  such that

$$\|v - \bar{v}\| < \|v - w\|$$

for every vector  $w$  in  $W$  different from  $\bar{v}$ .

- ▶ “shortest distance”, “perpendicular distance”

## Best Approximation (cont'd)

### Theorem 7.8: The Best Approximation Theorem

If  $W$  is a finite-dimensional subspace of an inner product space  $V$  (e.g.,  $\mathbb{R}^n$ ) and if  $\mathbf{v}$  is a vector in  $V$ , then  $\text{proj}_W(\mathbf{v})$  is the best approximation to  $\mathbf{v}$  in  $W$ .

- ▶ What is the **error**?  $\|\mathbf{v} - \text{proj}_W(\mathbf{v})\|$  (distance from  $\mathbf{v}$  to  $W$ )

# Least Squares Approximation

- ▶ Given data points, (usually obtained from experiments) which function best approximates them? “best fit”
- ▶ How to minimize the **error**? (See Figure 7.13 (a) on p.587)

$$\rightarrow \text{error vector } e := \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- ▶ How to define the error?

$$\text{“}p\text{-norm” } \|e\|_p := \left( \sum_{i=1}^n |\epsilon_i|^p \right)^{\frac{1}{p}}$$

- ▶  $\|e\|_1 := \sum_{i=1}^n |\epsilon_i|$
- ▶  $\|e\|_2 := \sqrt{\sum_{i=1}^n |\epsilon_i|^2} = \|e\|$
- ▶  $\|e\|_\infty := \max(|\epsilon_1|, \dots, |\epsilon_n|)$

- ▶ Least square approximation: Which function best approximates the data points minimizing the **least squares error**  $\|e\|$ ?

## Least Squares Approximation (cont'd)

### Definition

If  $A$  is an  $m \times n$  matrix and  $\mathbf{b}$  is in  $\mathbb{R}^m$ , a **least squares solution** of  $A\mathbf{x} = \mathbf{b}$  is a vector  $\bar{\mathbf{x}}$  in  $\mathbb{R}^n$  such that

$$\|\mathbf{b} - A\bar{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all  $\mathbf{x}$  in  $\mathbb{R}^n$ .

## Solving Least Squares Problem

- ▶  $A\bar{\mathbf{x}} \in \text{col}(A)$ , therefore, the solution is the closest vector in  $\text{col}(A)$  to  $\mathbf{b}$ .
- ▶ By the “Best Approximation Theorem”,  $A\bar{\mathbf{x}} = \text{proj}_{\text{col}(A)}(\mathbf{b})$ .
- ▶  $\mathbf{b} - A\bar{\mathbf{x}} = \mathbf{b} - \text{proj}_{\text{col}(A)}(\mathbf{b}) = \text{perp}_{\text{col}(A)}(\mathbf{b})$  is orthogonal to  $\text{col}(A)$ 
  - $\Rightarrow \mathbf{b} - A\bar{\mathbf{x}}$  is orthogonal to all the columns of  $A$ .
  - $\Rightarrow A^T(\mathbf{b} - A\bar{\mathbf{x}}) = \mathbf{0}$
  - $\Rightarrow A^T A\bar{\mathbf{x}} = A^T \mathbf{b}$ : A system of **normal equations** for  $\bar{\mathbf{x}}$

# The Least Squares Theorem

## Theorem 7.9: The Least Squares Theorem

Let  $A$  be an  $m \times n$  matrix and let  $\mathbf{b}$  be in  $\mathbb{R}^m$ . Then  $A\mathbf{x} = \mathbf{b}$  always has at least one least squares solution  $\bar{\mathbf{x}}$ . Moreover,

- $\bar{\mathbf{x}}$  is a least squares solution of  $A\mathbf{x} = \mathbf{b}$  if and only if  $\bar{\mathbf{x}}$  is a solution of the normal equations  $A^T A \bar{\mathbf{x}} = A^T \mathbf{b}$ .
- $A$  has linearly independent columns if and only if  $A^T A$  is invertible. In this case, the least squares solution of  $A\mathbf{x} = \mathbf{b}$  is unique and is given by

$$\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

- Least squares error:  $\|\mathbf{e}\| = \|\mathbf{b} - A\bar{\mathbf{x}}\|$

## Least Squares via the $QR$ Factorization

When  $A \in \mathbb{R}^{m \times n}$  and  $\text{rank}(A) = n$ ,

$$\begin{aligned}A^T A \bar{x} &= A^T \mathbf{b} \\ \Rightarrow (QR)^T QR \bar{x} &= (QR)^T \mathbf{b} && (A = QR) \\ \Rightarrow R^T Q^T QR \bar{x} &= R^T Q^T \mathbf{b} \\ \Rightarrow R^T R \bar{x} &= R^T Q^T \mathbf{b} && (Q^T = Q^{-1}) \\ \Rightarrow \bar{x} &= R^{-1} Q^T \mathbf{b} && (R \text{ is invertible})\end{aligned}$$

### Theorem 7.10

Let  $A$  be an  $m \times n$  matrix with linearly independent columns and let  $\mathbf{b}$  be in  $\mathbb{R}^m$ . If  $A = QR$  is a  $QR$  factorization of  $A$ , then the unique least squares solution  $\bar{x}$  of  $Ax = \mathbf{b}$  is

$$\bar{x} = R^{-1} Q^T \mathbf{b}$$

► Example 7.30 (p.592)

→ Don't compute  $R^{-1}$ , but solve  $R\bar{x} = Q^T \mathbf{b}$ .

# Orthogonal Projection Revisited

## Theorem 7.11

Let  $W$  be a subspace of  $\mathbb{R}^m$  and let  $A$  be an  $m \times n$  matrix whose columns form a basis for  $W$ . If  $\mathbf{v}$  is any vector in  $\mathbb{R}^n$ , then the orthogonal projection of  $\mathbf{v}$  onto  $W$  is the vector

$$\text{proj}_W(\mathbf{v}) = A(A^T A)^{-1} A^T \mathbf{v}$$

The linear transformation  $P : \mathbb{R}^m \rightarrow \mathbb{R}^n$  that projects  $\mathbb{R}^m$  onto  $W$  has  $A(A^T A)^{-1} A^T$  as its standard matrix.

Proof:

1. By the Least Squares Theorem, the unique least squares solution to  $A\mathbf{x} = \mathbf{v}$  is

$$\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{v}$$

2. And since  $A\bar{\mathbf{x}} = \text{proj}_{\text{col}(A)}(\mathbf{v}) = \text{proj}_W(\mathbf{v})$
3. Therefore,

$$\text{proj}_W(\mathbf{v}) = A((A^T A)^{-1} A^T \mathbf{v}) = (A(A^T A)^{-1} A^T) \mathbf{v}$$

## Pseudoinverse of a Matrix

- ▶  $\mathbf{x} = A^{-1}\mathbf{b}$  is the unique solution of  $A\mathbf{x} = \mathbf{b}$
- ▶  $\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$  is the unique least squares solution of  $A\mathbf{x} = \mathbf{b}$

→  $(A^T A)^{-1} A^T$  plays the role of an “inverse of  $A$ ”

Definition: Pseudoinverse

If  $A$  is a matrix with linearly independent columns, then the **pseudoinverse** of  $A$  is the matrix  $A^+$  defined by

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

- ▶ What if  $A$  is a square matrix?

## Pseudoinverse of a Matrix (cont'd)

- ▶ Which properties do they have?

### Theorem 7.12

Let  $A$  be a matrix with linearly independent columns. Then the pseudoinverse  $A^+$  of  $A$  satisfies the following properties, called the **Penrose conditions** for  $A$ :

- $AA^+A = A$
- $A^+AA^+ = A^+$
- $AA^+$  and  $A^+A$  are symmetric.

# Outline

Introduction: Taxicab Geometry

Inner Product Spaces

Norms and Distance Functions

Least Squares Approximation

**The Singular Value Decomposition**

Applications

# Outline

Introduction: Taxicab Geometry

Inner Product Spaces

Norms and Distance Functions

Least Squares Approximation

The Singular Value Decomposition

**Applications**