Orthogonality

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu

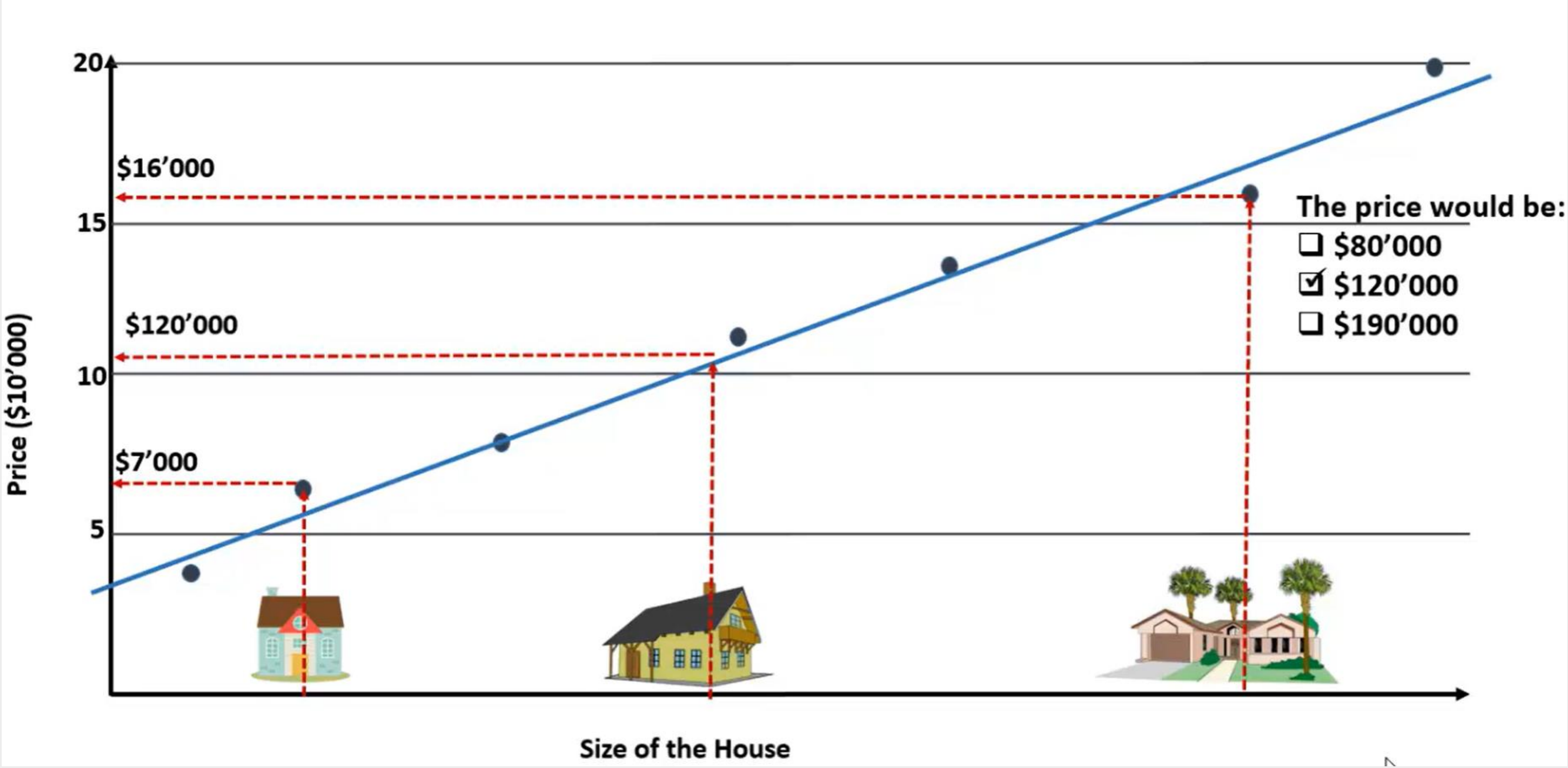
Maryam Ramezani maryam.ramezani@sharif.edu

Overview

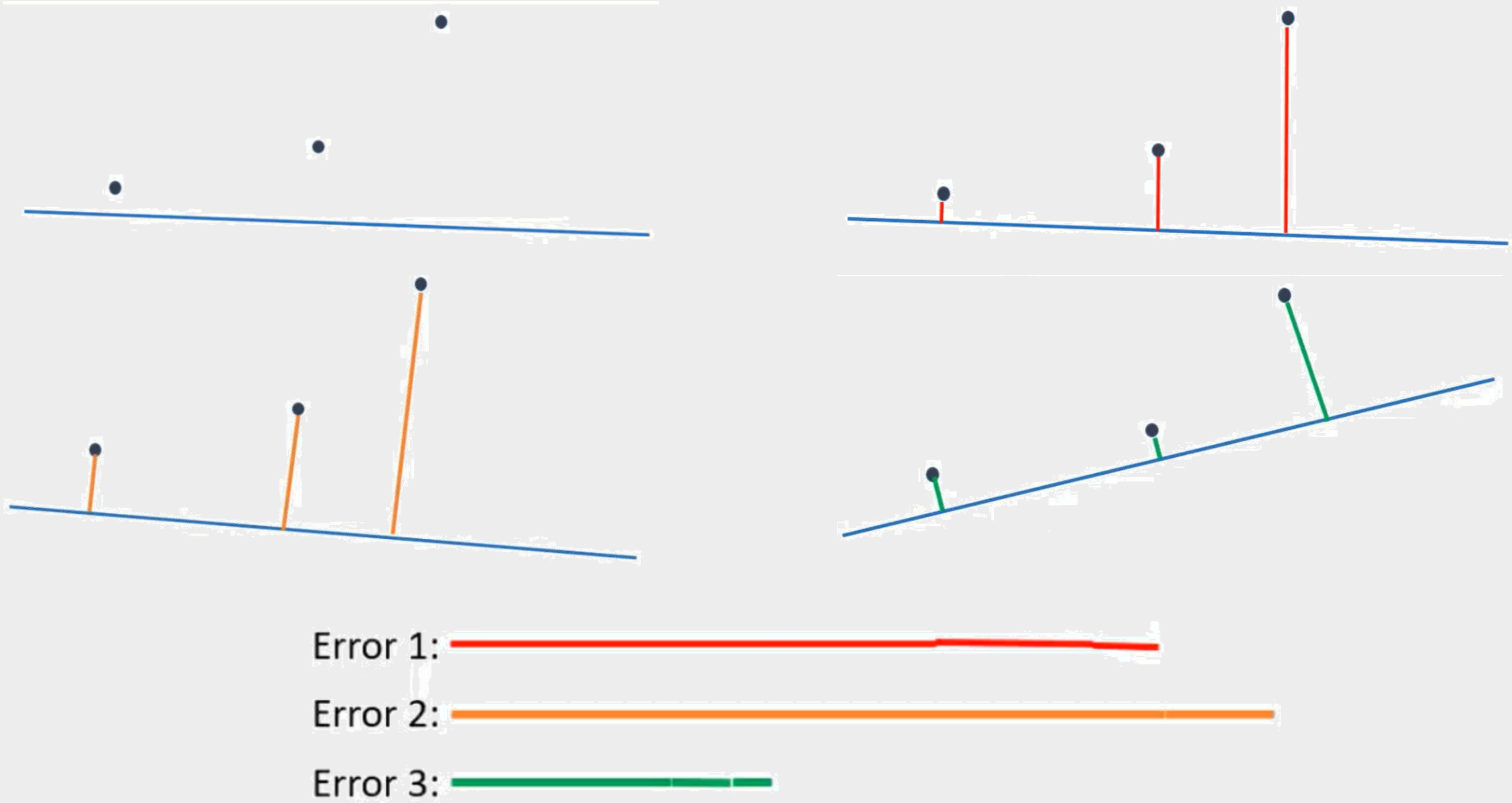
- Introduction
- Orthogonality
- Gram-Schmidt Algorithm
- Orthogonal Complements
- Projection

Introduction

Least Squares Error Correction

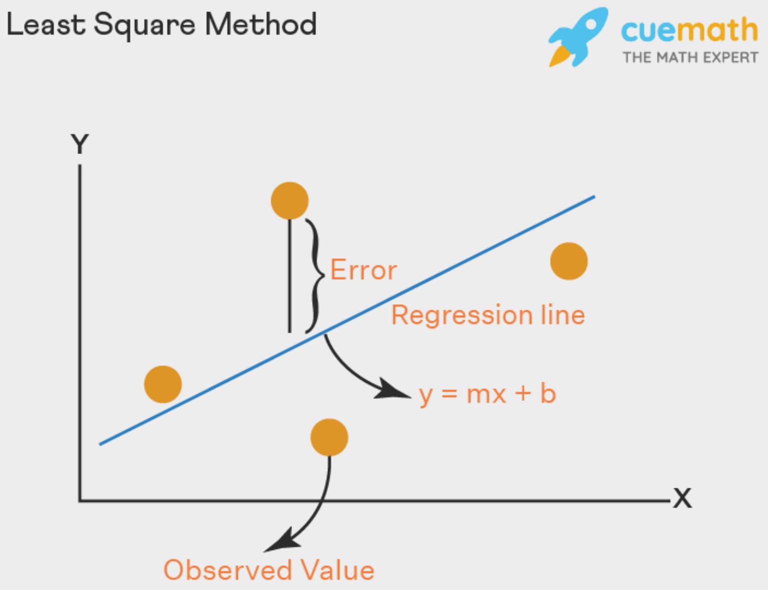


Least Squares Error Correction



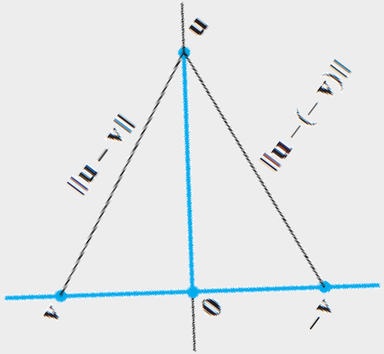
- Objective: $\hat{y}=\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n + b$

$$\min ||y - \hat{y}||$$



Orthogonality

- Geometry



- Algebra

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = \mathbf{0}$.

Suppose V is an inner product space.
Two vectors $\mathbf{v}, \mathbf{w} \in V$ are called **orthogonal** if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$

Orthogonal Sets

- A set of vectors $\{a_1, \dots, a_k\}$ in R^n is orthogonal set if each pair of distinct vectors is orthogonal (mutually orthogonal vectors).

Definition

A basis B of an inner product space V is called an **orthonormal basis** of V if

a) $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{v} \neq \boldsymbol{w} \in B$, and (mutual orthogonality)

b) $\|\mathbf{v}\| = 1$ for all $\mathbf{v} \in B$. (normalization)

- set of n-vectors $a_1, \dots, a_k$ are *(mutually) orthogonal* if $a_i \perp a_j$ for $i \neq j$
- They are *normalized* if $\|a_i\| = 1$ for $i = 1, \dots, k$
- They are *orthonormal* if both hold
- Can be expressed using inner products as

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Orthogonal Sets

Example

- Zero vector is orthogonal to every vector in vector space V
- The standard basis of $\mathbb{R}^n$ or $\mathbb{C}^n$ is an orthogonal set with respect to the standard inner product.

Orthogonal Sets

Theorem

If $S = \{a_1, \dots, a_k\}$ is an orthogonal set of nonzero vectors in R^n, then S is linearly independent and is a basis for the subspace spanned by S.

Proof

If k = n, then prove that S is a basis for R^n

Corollary

- A simple way to check if an n−vector y is a linear combination of the orthonormal vectors $a_1, \dots, a_k$, if and only if:

$$y = (a_1^T y)a_1 + \ \dots + (a_k^T y)a_k$$

- For orthogonal vectors $a_1, \dots, a_k$:

$$y = c_1 a_1 + \cdots + c_k a_k$$

$$c_j = \frac{y.a_j}{a_j.a_j}$$

Orthonormal vectors

Independence-dimension inequality

If the n-vectors $a_1, \ldots, a_k$ are linearly independent, then $k \leq n$.

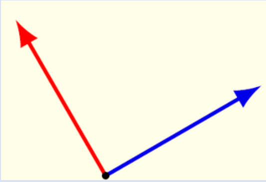
- Orthonormal sets of vectors are linearly independent
- By independence–dimension inequality, must have $k \leq n$
- When $k = n, a_1, \ldots, a_n$ are an *orthonormal basis*

Example

- Standard unit n-vectors $e_1, \ldots, e_n$
- The 3-vectors

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \qquad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

- The 2-vectors shown below



- The standard basis in $P_n(x)\ [-1,1]$ (be the set of real-valued polynomials of degree at most n.)

Orthogonal Subsets

Orthogonal Subspaces

Definition

- Two subspaces W_1 and W_2 of the same space V are orthogonal, denoted by $W_1 \perp W_2$, if and only if each vector $w_1 \in W_1$ is orthogonal to each vector $w_2 \in W_2$ for all w_1, w_2 in W_1, W_2 respectively:
$$< w_1, w_2 > = 0$$

Orthogonal Complements

Orthogonal Complements

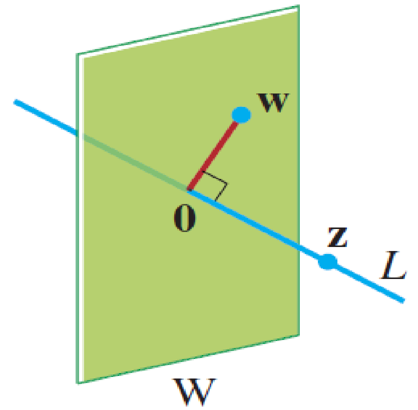
Definition

- If a vector z is orthogonal to every vector in a subspace W of R^n, then z is said to be orthogonal to W.
- **The set of all vectors z that are orthogonal** to W is called the orthogonal complement of W and is denoted by $W^{\perp}$

Example

W be a plane through the origin in $\mathbb{R}^3$.

$L = W^{\perp}$and $W = L^{\perp}$



Orthogonal Complements

Theorem

$W^{\perp}$ is a subspace of $\mathbb{R}^n$.

Theorem

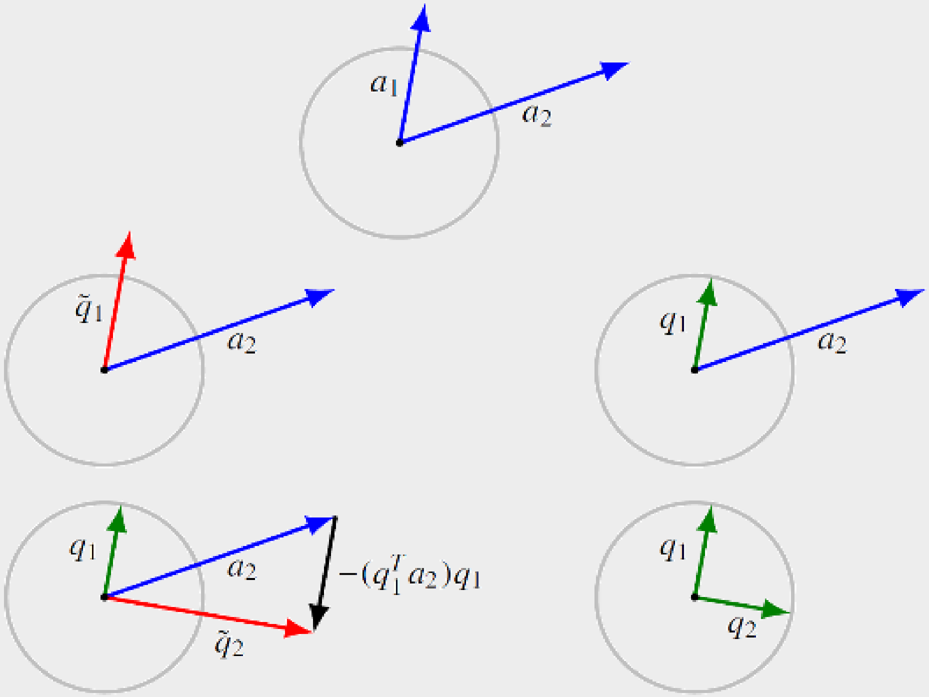
$W^{\perp} \cap W = \{\mathbf{0}\}$.

Important

We emphasize that W_1 and W_2 can be orthogonal without being complements.
$W_1 = span((1,0,0))$ and $W_2 = span((0,1,0))$.

Gram–Schmidt Algorithm

- Find orthonormal basis for span $\{a_1, a_2, \ldots, a_k\}$
- Geometry:



Gram–Schmidt (orthogonalization) algorithm

- Find orthonormal basis for span $\{a_1, a_2, \dots, a_k\}$
- Algebra:

1) $q1 = \frac{a_1}{\|a_1\|}$

2) $\widetilde{q_2} = a_2 - (q_1^T a_2)q_1 \rightarrow q_2 = \frac{\widetilde{q_2}}{\|\widetilde{q_2}\|}$

3) $\widetilde{q_3} = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2 \rightarrow q_3 = \frac{\widetilde{q_3}}{\|\widetilde{q_3}\|}$

.

.

.

k) $\widetilde{q_k} = a_k - (q_1^T a_k)q_1 - \cdots - \left(q_{k-1}^T a_k\right)q_{k-1} \rightarrow q_k = \frac{\widetilde{q_k}}{\|\widetilde{q_k}\|}$

Gram–Schmidt (orthogonalization) algorithm

Example

Find orthogonal set for $a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathrm{c} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

- Why $\{q_1, q_2, \dots, q_k\}$ is a orthonormal basis for span $\{a_1, a_2, \dots, a_k\}$?

  - $\{q_1, q_2, \dots, q_k\}$ are normalized.
  - $\{q_1, q_2, \dots, q_k\}$ is a orthogonal set
  - a_i is a linear combination of $\{q_1, q_2, \dots, q_i\}$

  $\Downarrow$

  span$\{q_1, q_2, \dots, q_k\}$ = span$\{a_1, a_2, \dots, a_k\}$

- q_i is a linear combination of $\{a_1, a_2, \dots, a_i\}$

Gram–Schmidt (orthogonalization) algorithm

- Given n–vectors $a_1, \dots, a_k$

  for $i = 1, \dots, k$

  1. Orthogonalization: $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - \cdots - (q_{i-1}^T a_i)q_{i-1}$
  2. Test for linear dependence: if $\tilde{q}_i = 0$, quit
  3. Normalization: $q_i = \dfrac{\tilde{q}_i}{\|\tilde{q}_i\|}$

Note

- If G–S does not stop early (in step 2), $a_1, \dots, a_k$ are linearly independent.
- If G–S stops early in iteration $i = j$, then a_j is a linear combination of $a_1, \dots, a_{j-1}$ (so $a_1, \dots, a_k$ are linearly dependent)

$$a_j = (q_1^T a_j)q_1 + \cdots + (q_{j-1}^T a_j)q_{j-1}$$

Complexity of Gram–Schmidt algorithm

- Gram–Schmidt algorithm gives us an explicit method for determining if a list of vectors is linearly dependent or independent.

- Given n-vectors $a_1, \dots, a_k$ $\quad O(nk^2)$

  for $i = 1, \dots, k$

  1. Orthogonalization: $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - \cdots - \left(q_{i-1}^T a_i\right)q_{i-1}$
  2. Test for linear dependence: if $\tilde{q}_i = 0$, quit
  3. Normalization: $q_i = \dfrac{\tilde{q}_i}{\|\tilde{q}_i\|}$

Orthonormal basis

Corollary

Every finite-dimensional inner product space has an orthonormal basis.

Existence of Orthonormal Bases

- Every finite-dimensional inner product space has an orthonormal basis.
- Since finite-dimensional inner product spaces (by definition) have a basis consisting of finitely many vectors, and the Gram-Schmidt process tells us how to convert that basis into an orthonormal basis, we now know that every finite-dimensional inner product space has an orthonormal basis.

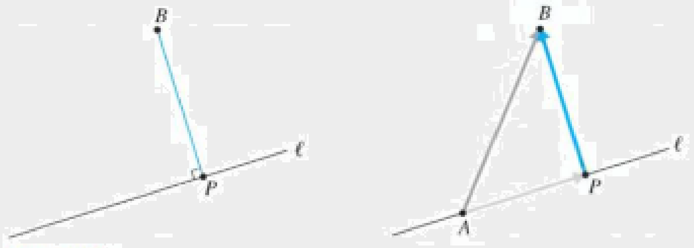
Example

Find an orthonormal basis for $P_2(x)$ *in* $[-1, 1]$ with respect to the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$$

Projection

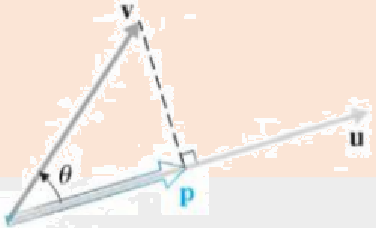
- Finding the distance from a point B to line l = Finding the length of line segment BP
- AP: projection of AB onto the line l



Definition

If **u** and **v** are vectors in $\mathbb{R}^n$ and $\mathbf{u} \neq \mathbf{0}$, then the **projection of v onto u** is the vector $proj_{\mathbf{u}}(\mathbf{v})$ defined by

$$proj_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right)\mathbf{u}$$



The projection of **v** onto **u**

Example

Write x as a linear combination of a_1, a_2, a_3?

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad a_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written **uniquely** in the form:

$\mathrm{proj}_W\, \mathbf{y}$.

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

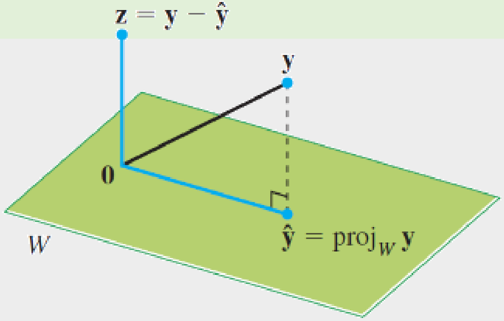
where $\hat{\mathbf{y}}$ is in W and **z is in $W^{\perp}$**. In fact, if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1.\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $z = \mathbf{y} - \hat{\mathbf{y}}$

Important

The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the particular basis used in (2).



The orthogonal projection of $\mathbf{y}$ onto W.

References

- Chapter 1: Advanced Linear and Matrix Algebra, Nathaniel Johnston
- Chapter 6: Linear Algebra David Cherney
- Linear Algebra and Optimization for Machine Learning
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares