1 : Theory of Matrices, Vector spaces and Linear Transformations

IT5506 – Mathematics for Computing II

Level III - Semester 5

© 2022 e-Learning Centre, UCSC

Intended Learning Outcomes

At the end of this lesson, you will be able to;

- Check whether an operation is an inner product.
- Calculate the norm of a vector and the angle between two vectors.
- Determine whether two vectors in an inner product space are orthogonal.
- Compute the scalar and vector projection of one vector onto another.
- Check whether a basis is orthogonal and/or orthonormal.
- Find an orthonormal basis of a subspace.
- Find least squares approximations for a system of equations.
- Determine whether a matrix is orthogonal.

© 2022 e-Learning Centre, UCSC

List of sub topics

1.8 Orthogonality (2 hours)

- 1.8.1 Inner (dot) product in real vector spaces with examples and its properties.
- 1.8.2 Orthogonal Vectors and Subspaces
- 1.8.3 Projections onto Lines
- 1.8.4 Orthogonal Bases and Gram-Schmidt orthogonalization process and the QR- decomposition.
- 1.8.5 Least square solution of a non-consistent linear system and the orthogonal projections.

© 2022 e-Learning Centre, UCSC

1.8.1 Real Inner/Dot Product in a Vector Space

Definition: A real inner product space is a real vector space V equipped with an operation that assigns to any pair of vectors $u, v \in V$ a real number $\langle u, v \rangle$, called the inner product of u and v. This operation must satisfy the following properties:

1. $\langle u, v \rangle = \langle v, u \rangle$, for all $u, v \in V$. (symmetry)
2. $\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$ for all $u, v, w \in V$, $\alpha, \beta \in \mathbb{R}$. (linearity)
3. $\langle u, u \rangle \geq 0,$ for all $u \in V$, and $\langle u, u \rangle = 0$ if and only if $u = 0$.

Definition (Dot product in $\mathbb{R}^n$):

Let $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be two vectors in $\mathbb{R}^n$. The dot product of u and v is a real number denoted by $\boldsymbol{u} \cdot \boldsymbol{v}$ and defined as

$$u \cdot v = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}^T \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

n

© 2022 e-Learning Centre, UCSC

$\mathbb{R}^{\boldsymbol{n}}$ with the dot product is an inner product space

Example 1.8.1.1: Define a mapping $\langle .\,,\ .\rangle: \mathbb{R}^n \ \rightarrow \mathbb{R}^n$ by $\langle u, v\rangle = u \cdot v, \forall\, \mathrm{u}, \mathrm{v} \in \mathbb{R}^n$.

We shall show that $\mathbb{R}^n$ with this mapping (dot product) is an inner product.

Solution: Let $u, v, w \in \mathbb{R}^n$, and $\alpha, \beta \in \mathbb{R}$

- $\langle u, v\rangle = u \cdot v = u^T v = (u^T v)^T = v^T u = v \cdot u = \langle v, u\rangle$.

  (Since $u^T v$ is a real number, $u^T v = (u^T v)^T$)

- $\langle u, \alpha v + \beta w\rangle = u^T(\alpha v + \beta w) = u^T(\alpha v) + u^T(\beta v)$

  $$= \alpha u^T v + \beta u^T v$$

  $$= \alpha\langle u, v\rangle + \beta\langle u, w\rangle.$$

- Since $\langle u, u\rangle = u \cdot u = u_1^2 + u_2^2 + \cdots + u_n^2$, $\langle u, u\rangle \geq 0$, and $\langle u, u\rangle = 0 \Leftrightarrow \mathrm{u} = \mathbf{0}.$

Hence**,** $\mathbb{R}^{\boldsymbol{n}}$ with the dot product is an inner product space

Note: Even though, there exist other inner product operations on $\mathbb{R}^n$ besides the dot product, in this course, we are going to consider only the dot product as the inner product.

© 2022 e-Learning Centre, UCSC

Properties of the Dot Product

Theorem 1.8.1.1: Let u, v, w be three vectors in $\mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$. Then

1. $u \cdot v = v \cdot u$
2. $(\alpha u + \beta v) \cdot w = \alpha(u \cdot w) + \beta(u \cdot w)$
3. $u \cdot (\alpha v + \beta w) = \alpha(u \cdot v) + \beta(u \cdot w)$.

These identities can be easily proved using the definition of dot product and are left as exercise.

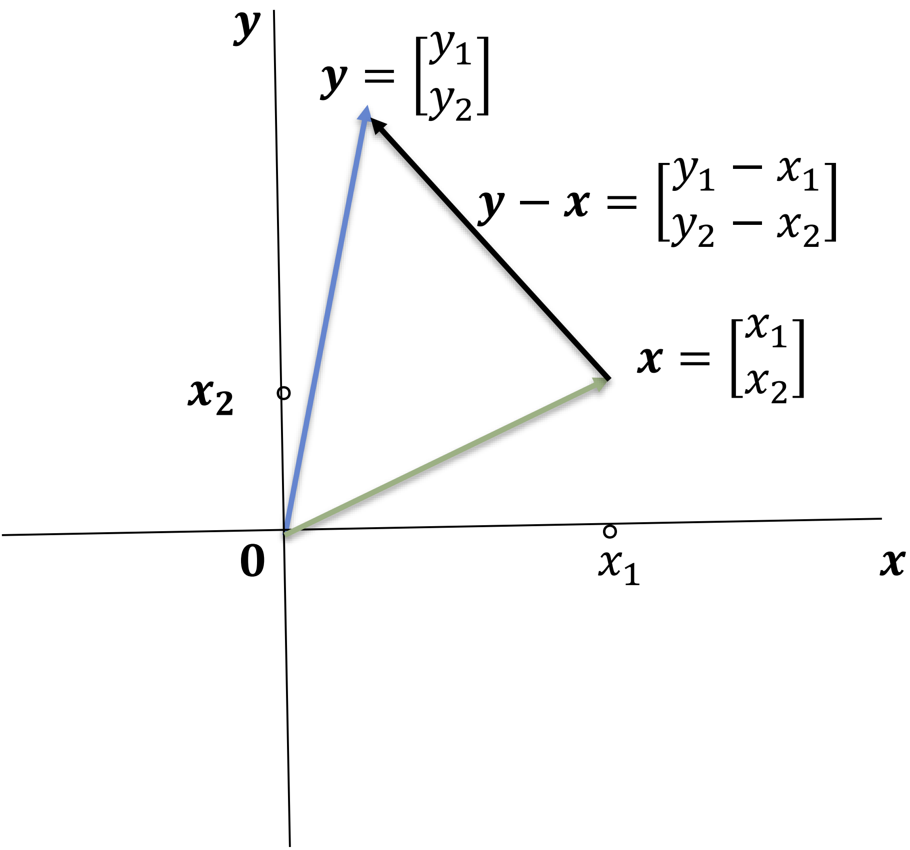
© 2022 e-Learning Centre, UCSC

Length of a vector in $\mathbb{R}^2$

The **length** (Euclidian length) or **norm** of a vector x is denoted $\| x \|$ and is defined as

$\| x \| = (x_1^2 + x_2^2)^{\frac{1}{2}}$.

$\| x \|^2 = x_1^2 + x_2^2 = x \cdot x = x^T x$.



The **length** or **distance** between two vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ is denoted by $\| y - x \|$ and is defined as

$\| \boldsymbol{y} - \boldsymbol{x} \|^2 = (y_1 - x_1)^2 + (y_2 - x_2)^2$.

$$\| \boldsymbol{y} - \boldsymbol{x} \|^2 = (\boldsymbol{y} - \boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x})$$

$$= (\boldsymbol{y} - \boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}).$$

© 2022 e-Learning Centre, UCSC

Length of a vector in $\mathbb{R}^{\boldsymbol{n}}$

Definition: Let $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ be a vector in $\mathbb{R}^{\boldsymbol{n}}$. Then the length of u or the **norm** of u is denoted by $\| u \|$ and is defined as $\| u \| = (u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}$.

Properties of length: Let $u \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$. Then

- $\| u \| \geq 0$;
- $\| u \| = 0$ if and only if $u = \mathbf{0}$.
- $\| \alpha\, u \| = |\alpha| \, \| u \|$.
- $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = \boldsymbol{u} \cdot \boldsymbol{u} = \| u \|^{\mathbf{2}}$.

Definition (Unit Vector): A vector $u \in \mathbb{R}^{\boldsymbol{n}}$ is called a unit vector if it has length 1, that is, if $\| u \| = 1$.

© 2022 e-Learning Centre, UCSC

Normalizing a vector

Example 1.8.1.2: Consider the vector $\mathrm{v} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$ in $\mathbb{R}^3$, Find the unit vector u that has the same direction as v. (u is known as the normalizing vector of v)

Solution: We have $\| v \| = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{1 + 4 + 9} = \sqrt{14}$. Hence,

$$u = \frac{v}{\| v \|} = \frac{1}{\sqrt{14}} v = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{14} \\ -2/\sqrt{14} \\ 3/\sqrt{14} \end{bmatrix}.$$

Exercise: Let $\mathrm{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathrm{v} = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} \in \mathbb{R}^3$. Find $\| u + v \|$ and $\| u - v \|$.

© 2022 e-Learning Centre, UCSC

The Cauchy-Schwarz Inequality in $\mathbb{R}^n$

Theorem 1.8.1.2: Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. Then $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \| \boldsymbol{u} \| \ \| \boldsymbol{v} \|$. Furthermore equality is obtained if and only if one of $\boldsymbol{u}$ or $\boldsymbol{v}$ is a scalar multiple of the other.

Proof: If $\boldsymbol{u} = \boldsymbol{0}$, then $|\boldsymbol{u} \cdot \boldsymbol{v}| = 0$ and $\| \boldsymbol{u} \| \ \| \boldsymbol{v} \| = 0$. Hence, the theorem follows.

Therefore, we may assume that $u \neq \boldsymbol{0}$. Define a function $f\colon \mathbb{R} \to \mathbb{R}$ by

$$f(t) = (t\boldsymbol{u} + \boldsymbol{v}) \cdot (t\boldsymbol{u} + \boldsymbol{v}), \forall t \in \mathbb{R}.$$

Then by the properties of dot product, we know that $f(t) \geq 0$ for all t $\in \mathbb{R}$ and

$$\begin{aligned} f(t) &= tu \cdot (tu + v) + \mathrm{v} \cdot (tu + v) \\ &= t^2\,(u \cdot u) + t\,(u \cdot v) + v \cdot tu + v \cdot v \\ &= \| u \|^2 \ t^2 + 2(u \cdot v)t + \| v \|^2. \end{aligned}$$

This is a quadratic function in t, and it has one or zero roots if and only if $\big(2(u.v)\big)^2 - 4 \| u \|^2 \| v \|^2 \leq 0$. Which is equivalent to

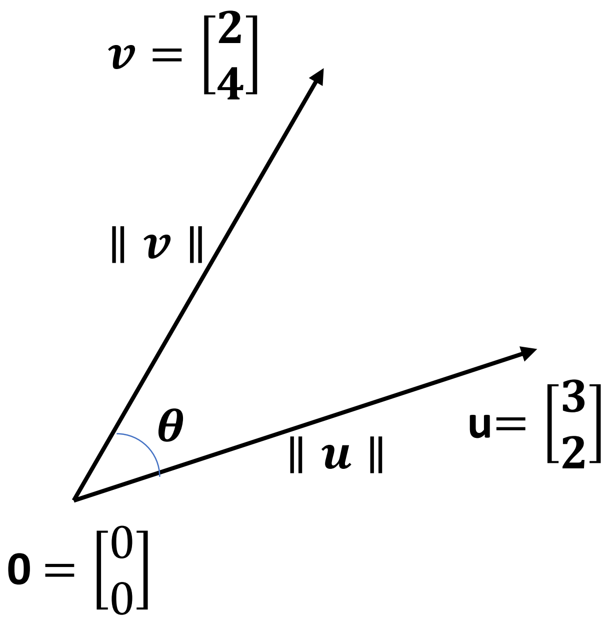
$$|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \| \boldsymbol{u} \| \ \| \boldsymbol{v} \|.$$

Note that if $u = \alpha v$ for some $\alpha \in \mathbb{R}$, then $|u \cdot v| = |(\alpha v) \cdot v| = |\alpha| \ \| \boldsymbol{v} \|^2$, and $\| \boldsymbol{u} \| \ \| \boldsymbol{v} \| = \| \boldsymbol{\alpha v} \| \ \| \boldsymbol{v} \| = |\alpha| \ \| \boldsymbol{v} \|^2$.

Therefore, if u is a scalar multiple of v, then $|\boldsymbol{u} \cdot \boldsymbol{v}| = \| \boldsymbol{u} \| \ \| \boldsymbol{v} \|$.

© 2022 e-Learning Centre, UCSC

Interpretation of Cauchy-Schwarz Inequality in $\mathbb{R}^2$



$\| u \| = \sqrt{3^2 + 2^2} = \sqrt{13}.$

$\| v \| = \sqrt{2^2 + (4)^2} = \sqrt{20}.$

$|u \cdot v| = |(3 \times 2) + (2 \times 4)| = |14| = 14.$

Not that,

$|u \cdot v| = 14 < \sqrt{13} \times \sqrt{20} = \sqrt{260} = \| u \| \| v \| .$

We knew from A/L Mathematics that,

$$u \cdot v = \| u \| \| v \| \cos\theta \, .$$

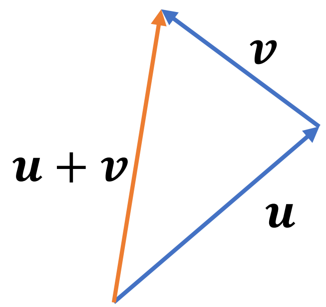
$$\Rightarrow \quad |u \cdot v| = \| u \| \| v \| \; |\cos\theta| \; \leq \; \| u \| \| v \| \quad (\because |\cos\theta| \leq 1).$$

© 2022 e-Learning Centre, UCSC

Triangle Inequality in $\mathbb{R}^n$

Theorem 1.8.1.3: For any $u, v \in \mathbb{R}^n$, $\| \boldsymbol{u} + \boldsymbol{v} \| \leq \| \boldsymbol{u} \| + \| \boldsymbol{v} \|$.

Proof: Let $u, v \in \mathbb{R}^n$.



$\| \boldsymbol{u} + \boldsymbol{v} \|^2 = (\boldsymbol{u} + \boldsymbol{v}) \cdot (\boldsymbol{u} + \boldsymbol{v})$

$= \boldsymbol{u} \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{v}$

$= \| \boldsymbol{u} \|^2 + 2\,(u \cdot v) + \| \boldsymbol{v} \|^2$

$\leq \| \boldsymbol{u} \|^2 + 2\,|\boldsymbol{u} \cdot \boldsymbol{v}| + \| \boldsymbol{v} \|^2$ $\quad(\because\ u \cdot v \leq |u \cdot v|\,)$

$\leq \| \boldsymbol{u} \|^2 + 2 \| \boldsymbol{u} \| \, \| \boldsymbol{v} \| + \| \boldsymbol{v} \|^2$ $\quad(\because$ Cauchy–Schwarz Inequality)

$= (\| \boldsymbol{u} \| + \| \boldsymbol{v} \|)^2.$

Therefore, $\| \boldsymbol{u} + \boldsymbol{v} \|^2 \leq (\| \boldsymbol{u} \| + \| \boldsymbol{v} \|)^2$. By taking square roots of both sides,
We obtain $\| \boldsymbol{u} + \boldsymbol{v} \| \leq \| \boldsymbol{u} \| + \| \boldsymbol{v} \|$.

© 2022 e-Learning Centre, UCSC

Exercises

1. Let u, v be two vectors in $\mathbb{R}^n$. Using the properties of dot product prove the following:
   - $(a \cdot b) = \frac{1}{4}(\| \boldsymbol{u} + \boldsymbol{v} \|^2 - \| \boldsymbol{u} - \boldsymbol{v} \|^2)$
   - $\| \boldsymbol{u} + \boldsymbol{v} \|^2 + \| \boldsymbol{u} - \boldsymbol{v} \|^2 = 2 \| \boldsymbol{u} \|^2 + 2 \| \boldsymbol{v} \|^2$.

2. Find the angle between the vectors $u = \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}$ and $v = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$.

3. Determine whether the formula $\langle u, v \rangle = u^T A v$ for all $u, v \in \mathbb{R}^2$ determines an inner product on $\mathbb{R}^2$, where $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

© 2022 e-Learning Centre, UCSC

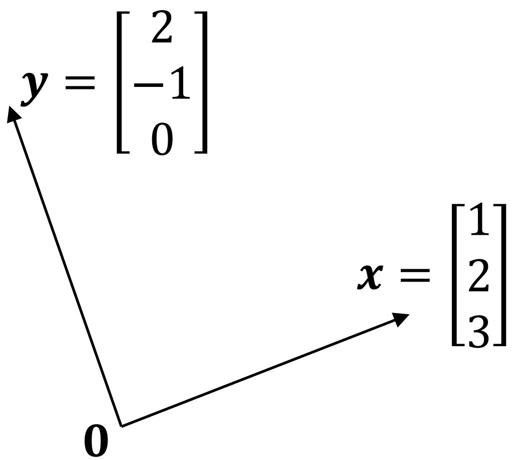
1.8.2 Orthogonal vectors and Subspaces in $\mathbb{R}^n$

Definition: Two vectors $\boldsymbol{x}, \boldsymbol{y}$ of a vector space $(\mathbb{R}^n)$ are said to be orthogonal (perpendicular) to each other if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.

We also write $\boldsymbol{x} \perp \boldsymbol{y}$ to indicate that x and y are orthogonal.

Note: zero (**0**) vector is orthogonal to every other vector x in the vector space, since $x \cdot \mathbf{0} = 0$.

Example: Determine whether two vectors $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ are orthogonal.



$\boldsymbol{x} \cdot \boldsymbol{y} = 1 \times 2 + 2 \times (-1) + 3 \times 0 = 0.$

Hence, $\boldsymbol{x} \perp \boldsymbol{y}$.

Also verify that $\| x + y \|^2 = \| x \|^2 + \| y \|^2 = 19$.

© 2022 e-Learning Centre, UCSC

Orthogonal Complement

Definition: Let S be a subset of $\mathbb{R}^n$. The orthogonal complement of S is the set

$$S^{\perp} = \{x \in \mathbb{R}^n : \ x \cdot w = 0 \text{ for all } w \in S\}.$$

Theorem 1.8.2.1: If S is any subset of $\mathbb{R}^n$, then $S^{\perp}$ is a subspace of $\mathbb{R}^n$.

Proof: Since **0** is orthogonal to all vectors, $\mathbf{0} \in S^{\perp}$.

Assume that $u, v \in S^{\perp}$. We have to show that $u + v \in S^{\perp}$.

Let $x \in S$. Then we have $(u + v) \cdot x = u \cdot x + u \cdot y = 0 + 0 = 0$.

That is; $(u + v) \cdot x = 0$, for all $x \in S$.

Therefore, $u + v \in S^{\perp}$.

Similarly, we can show that for any $v \in S^{\perp}$ and $\alpha \in \mathbb{R}$, we have $\alpha v \in S^{\perp}$.

Therefore, $S^{\perp}$ is a subspace of $\mathbb{R}^n$.

© 2022 e-Learning Centre, UCSC

Orthogonal and Orthonormal sets of vectors in $\mathbb{R}^n$

Definition: A set of vectors $\mathrm{S} = \{x_1, x_2, \cdots, x_k\}$ in $\mathbb{R}^n$ $(k \leq n)$ is said to be an **orthogonal** set if the vectors are non-zero and pairwise orthogonal, i.e., $x_i \neq 0$ for all i, and $x_i \perp x_j$ for all $i \neq j$.

Moreover, the set of vectors $\mathrm{U} = \{u_1, u_2, \cdots, u_k\}$ in $\mathbb{R}^n$ is called **orthonormal** if it is orthogonal and $\| u_i \| = 1$ for each i.

Theorem 1.8.2.2: Let $\mathrm{S} = \{x_1, x_2, \cdots, x_k\} \subseteq \mathbb{R}^n$ $(k \leq n)$ is Orthogonal. Then S is linearly independent.

Proof: Suppose $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = \mathbf{0}$ for $\alpha_1, \alpha_2, \cdots, \alpha_k \in \mathbb{R}$.

Let $i \in \{1, 2, \cdots, k\}$ be arbitrary.

$$0 = (\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k) \cdot x_i = \alpha_1 x_1 \cdot x_i + \cdots + \alpha_i x_i \cdot x_i + \cdots + \alpha_k x_k \cdot x_k$$

$$= \alpha_i x_i \cdot x_i \quad \left(\because\ x_j \cdot x_i = 0, for\ i \neq j\right).$$

That is, $\alpha_i x_i \cdot x_i = \alpha_i \| x_i \|^2 = 0$. Since $x_i \neq \mathbf{0}$, we have $\alpha_i = 0$.

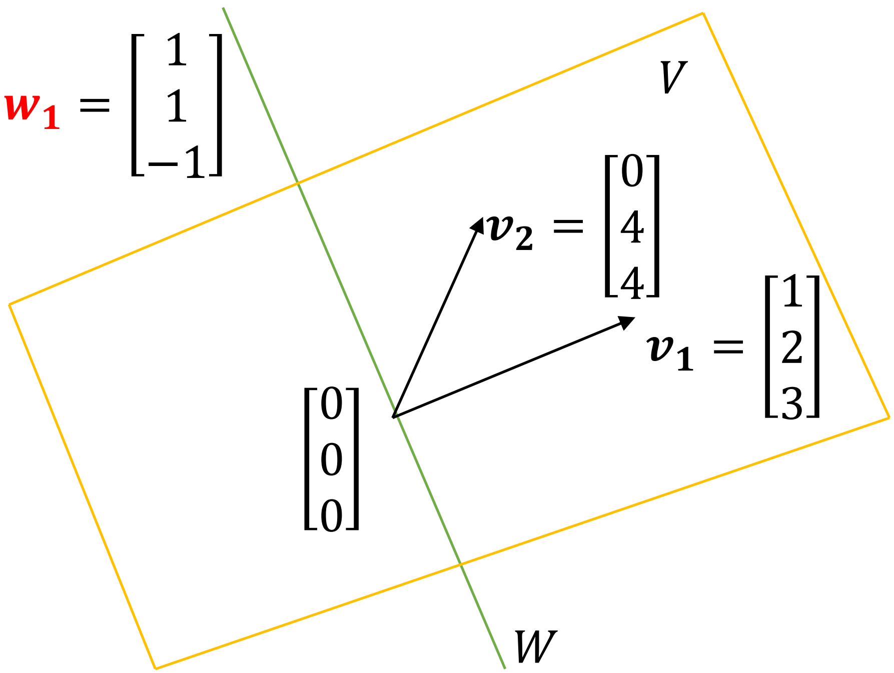
Hence, $\alpha_i = 0, \forall i$. Therefore, $x_1, x_2, \cdots, x_k$ are linearly independent.

© 2022 e-Learning Centre, UCSC

Orthogonal Subspaces

Definition: Two subspaces V and W of the same vector space $\mathbb{R}^n$ are orthogonal if and only if every vector v in V is orthogonal to every vector w in W.

$$V \perp W \iff v \perp w, \forall v \in V \,\&\, \forall w \in W \iff v \cdot w = 0, \forall v \in V \,\&\, \forall w \in W.$$



Example:

Let $V = \left\langle \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix} \right\} \right\rangle$. Then V is a 2 dimensional subspace of $\mathbb{R}^3$.

Let W $= \left\langle \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\} \right\rangle$. Then W is a 1 dimensional subspace of $\mathbb{R}^3$.

Since every vector v in V is orthogonal to every vector w in W, $V \perp W$

© 2022 e-Learning Centre, UCSC

Fundamental Theorem of Orthogonality

Theorem 1.8.2.3 : The row space of a real matrix $A = \left[a_{ij}\right]_{m\times n}$ is orthogonal to the null space of A (in $\mathbb{R}^n$). The column space of A is orthogonal to the null space of A^T (in $\mathbb{R}^m$).

Proof: Let $x \in N(A)$.

$\Rightarrow Ax = 0$. That is; $\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_n} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$.

$\Rightarrow \quad (\text{row 1 of } A) \cdot x = 0, (\text{row 2 of } A) \cdot x = 0, \cdots, (\text{row m of } A) \cdot x = 0.$

$\Rightarrow \quad x \perp (\text{row 1 of } A), \quad x \perp (\text{row 2 of } A), \cdots, \quad x \perp (\text{row n of } A).$

$$\Rightarrow \quad x \perp Span\left(\left\{\begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \end{bmatrix}, \begin{bmatrix} a_{21} \\ \vdots \\ a_{2n} \end{bmatrix}, \cdots, \begin{bmatrix} a_{m1} \\ \vdots \\ a_{mn} \end{bmatrix}\right\}\right) = Row\ space(A) = C(A^T)$$

Since $x \in N(A)$ was arbitrary, we have $N(A) \perp Row\ space(A) = C(A^T)$.

© 2022 e-Learning Centre, UCSC

Proof Continued

Let $y \in N(A^T)$.

$\Rightarrow A^T y = [\mathbf{0}]_{n\times 1}$.

$\Rightarrow y^T A = [\mathbf{0}]_{1\times n}$.

$$\Rightarrow \quad \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}$$

$\Rightarrow \quad y \cdot (\text{column 1 of } A) = 0,\ y \cdot (\text{column 2 of } A) = 0, \cdots,\ y \cdot (\text{column n of } A) = 0.$

$\Rightarrow \quad y \perp (\text{column 1 of } A), y \perp (\text{column 2 of } A), \cdots,\ y \perp (\text{column n of } A).$

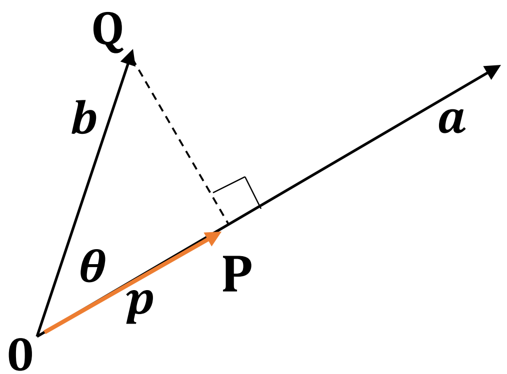
$$\Rightarrow \quad y \perp Span\left(\left\{\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}, \cdots, \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}\right\}\right) \Rightarrow \quad y \perp C(A).$$

Since $y \in N(A^T)$ was arbitrary, we have $N(A^T) \perp C(A)$.

© 2022 e-Learning Centre, UCSC

1.8.3 Projections Onto Lines

It is sometimes important to find the component of a vector in a particular direction. Consider the following picture:



The distance from 0 to P (measured positively in the direction of $\boldsymbol{a}$) is called the component of $\boldsymbol{b}$ in the direction of $\boldsymbol{a}$.

The vector $\overrightarrow{OP} = \boldsymbol{p}$ is called the projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$, and is denoted $proj_{\boldsymbol{a}}(\boldsymbol{b})$.

$OP = \| \boldsymbol{b} \| \cos \boldsymbol{\theta}$

On the other hand, we have $\boldsymbol{a} \cdot \boldsymbol{b} = \| \boldsymbol{a} \| \| \boldsymbol{b} \| \cos \boldsymbol{\theta}$.

Therefore,

$\| \boldsymbol{b} \| \cos \boldsymbol{\theta} = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\|}$.

$$proj_{\boldsymbol{a}}(\boldsymbol{b}) = \overrightarrow{OP} = |OP| \frac{\boldsymbol{a}}{\| \boldsymbol{a} \|} = \| \boldsymbol{b} \| \cos \boldsymbol{\theta} \frac{\boldsymbol{a}}{\| a \|} = \| \boldsymbol{a} \| \| \boldsymbol{b} \| \cos \boldsymbol{\theta} \frac{\boldsymbol{a}}{\| \boldsymbol{a} \|^2}$$

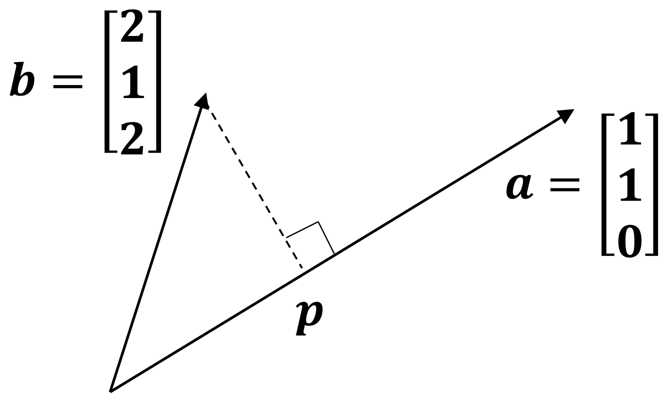
$$= (\boldsymbol{a} \cdot \boldsymbol{b}) \frac{\boldsymbol{a}}{\| \boldsymbol{a} \|^2}$$

$$= \frac{\boldsymbol{a}^T \boldsymbol{b}}{\boldsymbol{a}^T \boldsymbol{a}} \boldsymbol{a}.$$

© 2022 e-Learning Centre, UCSC

Projections Onto Lines

Example 1.8.3.1 : Find the projection of $\boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ onto $\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.



$$proj_{\boldsymbol{a}}(\boldsymbol{b}) = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\ \boldsymbol{a} = \left(\frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$= \left(\frac{3}{2}\right)\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \\ 0 \end{bmatrix}.$$

Note: $proj_{\boldsymbol{a}}(\boldsymbol{b}) = \left(\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\right)\ \boldsymbol{a} = \boldsymbol{a}\left(\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\right) = \left(\frac{\boldsymbol{a}\ \boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}\right)\boldsymbol{b}$ $\quad \left(\because \left(\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\right) \text{ is a real number}\right).$

The matrix $P = \left(\frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}\right) = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is known as the projection matrix, and the linear transformation $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$, defined by $Tx = Px, \forall x \in \mathbb{R}^3$, is known as the projection mapping.

© 2022 e-Learning Centre, UCSC

1.8.4 Orthogonal Bases, Gram-Schmidt Orthogonalization Process and the QR- Decomposition

Definition: A basis $\mathcal{B} = \{b_1, b_2, \cdots, b_n\}$ of $\mathbb{R}^n$ is said to be an orthogonal basis if $\mathcal{B}$ is orthogonal. If, moreover, $\| b_i \| = 1$ for each i, we say that $\mathcal{B}$ is an orthonormal basis for $\mathbb{R}^n$.

Example:

- $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right\}$ is a basis for $\mathbb{R}^2$, but this is not orthogonal.
- $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$ is an orthogonal basis for $\mathbb{R}^2$, but this not orthonormal.
- $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$ is an orthonormal basis for $\mathbb{R}^2$ - known as standard basis for $\mathbb{R}^2$.
- $\left\{\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}, \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}\right\}$ is an orthonormal basis for $\mathbb{R}^2$ for all $\theta \in \mathbb{R}$.

© 2022 e-Learning Centre, UCSC

Gram-Schmidt Orthogonalization Process

The Gram-Schmidt orthogonalization procedure is a method for turning any basis of a finite dimensional vector space into an orthogonal/orthonormal one.

Theorem 1.8.4.1: Let V be a finite dimensional vector space. Let $\{a_1, a_2, \cdots, a_n\}$ be a linearly independent subset of V. Then the Gram-Schmidt orthogonalization process uses the vectors $a_1, a_2, \cdots, a_n$ to construct orthonormal vectors $q_1, q_2, \cdots, q_n$ such that $\langle\{a_1, a_2, \cdots, a_i\}\rangle = \langle\{q_1, q_2, \cdots, q_i\}\rangle$, $\forall i = 1,2, \cdots n$.

There is no problem with q_1: it can go in the direction of a_1. We divide by the length, so that $q_1 = \frac{a_1}{\|a_1\|}$ is a unit vector.

We will show this results for the cases $n = 2 \,\&\, 3$.

© 2022 e-Learning Centre, UCSC

Gram-Schmidt Orthogonalization Process $(n = 2)$

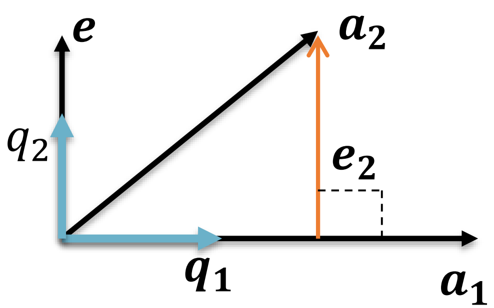
Let a_1, a_2 be two linearly independent vectors in a finite dimensional vector space.

Let $q_1 = \frac{a_1}{\|a_1\|}$. Then $\| q_1 \| = 1$, and $\langle\{a_1\}\rangle = \langle\{q_1\}\rangle$.

Projection of $\boldsymbol{a_2}$ onto the line through $\boldsymbol{a_1}$ is $(q_1 \cdot a_2)q_1 = (q_1^T a_2)q_1$.

Let $\boldsymbol{e_2} = a_2 - (q_1^T a_2)q_1$. Then $\boldsymbol{e_2} \perp q_1$ because $q_1 \cdot \boldsymbol{e_2} = q_1 \cdot (a_2 - (q_1^T a_2)q_1)$

$$= q_1 \cdot a_2 - (q_1 \cdot a_2)(q_1 \cdot q_1) = 0.$$



Let $q_2 = \frac{e_2}{\|e\|}$. Then $\| q_2 \| = 1$, $q_2 \perp q_1$, and $\langle\{a_1, a_2\}\rangle = \langle\{q_1, q_2\}\rangle$.

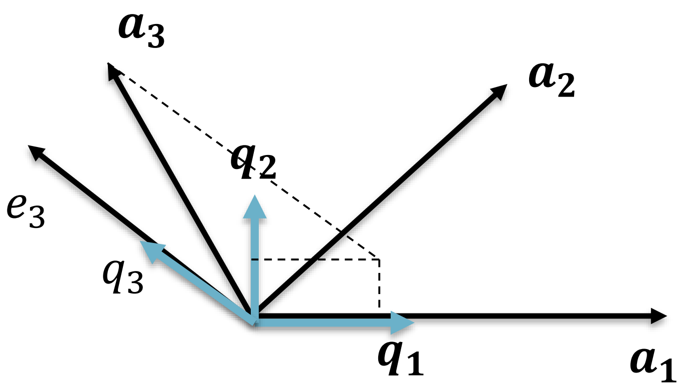
That is; the subspace spanned by the basis a_1, a_2 and the subspace spanned by the orthonormal basis q_1, q_2 are the same.

© 2022 e-Learning Centre, UCSC

Gram-Schmidt Orthogonalization Process $(n = 3)$

Let a_1, a_2, a_3 be three linearly independent vectors in a finite dimensional vector space.

a_3 can not be in the plane of q_1 and q_2, which is the plane of a_1 and a_2. However, it may have a component in that plane, and that has to be subtracted.



Let $e_3 = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2$.

This is the one idea of the whole Gram-Schmidt process, to subtract from every new vector its components in the directions that are already settled. That idea is used over and over again.

Let $q_3 = \frac{e_3}{\|e_3\|}$. Then $\| q_3 \| = 1$, and it can be easily shown that $q_3 \perp q_1$, $q_3 \perp q_2$, and $\langle\{a_1, a_2, a_3\}\rangle = \langle\{q_1, q_2, q_3\}\rangle$.

The Gram-Schmidt process starts with independent vectors $a_1, a_2, \cdots, a_n$ and ends with orthonormal vectors $q_1, q_2, \cdots, q_n$. At step j, it subtracts from a_j its components in the directions $q_1, q_2, \cdots, q_{j-1}$ that are already settled:

$e_j = a_j - (q_1^T a_j)q_1 - (q_2^T a_j)q_2 - \cdots - (q_{j-1}^T a_j)q_{j-1}$, and $q_j = \frac{e_j}{\|e_j\|}$.

© 2022 e-Learning Centre, UCSC

Gram-Schmidt orthogonalization process $(n = 3)$

The QR- decomposition: Relation ships between given basis vectors $a_i's$ and the equivalent orthonormal basis $q_i's$ when $n = 3$:
Let A, Q be the matrices containing a_1, a_2, a_3 and q_1, q_2, q_3 as its column vectors respectively. We have,
$a_1 = (q_1^T a_1)q_1$
$a_2 = (q_1^T a_2)q_1 + (q_2^T a_2)q_2$
$a_3 = (q_1^T a_3)q_1 + (q_2^T a_3)q_2 + (q_3^T a_3)q_3.$

$$\Rightarrow \quad \begin{bmatrix} \vdots & \vdots & \vdots \\ a_1 & a_2 & a_3 \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots \\ q_1 & q_2 & q_3 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} q_1^T a_1 & q_1^T a_2 & q_1^T a_3 \\ 0 & q_2^T a_2 & q_2^T a_3 \\ 0 & 0 & q_3^T a_3 \end{bmatrix}$$

$\Rightarrow \quad A = QR.$

R is upper triangular because of the way Gram-Schmidt process was done. The first vectors a_1 and q_1 fell on the same line. Then q_1, q_2 were in the same plane as a_1, a_2. The third vectors a_3 and q_3 were not involved until step 3.

Every m by n matrix with independent columns can be factored into $A = QR$. The columns of Q are orthonormal, and R is upper triangular and invertible. When $m = n$ and all matrices are square, Q becomes an orthogonal matrix.

© 2022 e-Learning Centre, UCSC

Gram-Schmidt orthogonalization process

Example 1.8.4.1: $a_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}, a_2 = \begin{bmatrix}1\\0\\0\end{bmatrix}, a_3 = \begin{bmatrix}2\\1\\0\end{bmatrix}$ are given. Find orthonormal vectors q_1, q_2, q_3 such that $\langle\{a_1, a_2, a_3\}\rangle = \langle\{q_1, q_2, q_3\}\rangle$.

Solution: $q_1 = \frac{a_1}{\|a_1\|} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$.

$\boldsymbol{e_2} = a_2 - (q_1^T a_2)q_1 = \begin{bmatrix}1\\0\\0\end{bmatrix} - \left(\frac{1}{\sqrt{2}}\right)\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}$. Hence, $q_2 = \frac{\boldsymbol{e_2}}{\|\boldsymbol{e_2}\|} = \left(\frac{1}{\sqrt{2}}\right)\begin{bmatrix}1\\0\\-1\end{bmatrix}$.

$\boldsymbol{e_3} = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2 = \begin{bmatrix}2\\1\\0\end{bmatrix} - \left(\sqrt{2}\right)\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix} - \left(\sqrt{2}\right)\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix} = \begin{bmatrix}0\\1\\0\end{bmatrix}$.

$q_3 = \frac{\boldsymbol{e_3}}{\|\boldsymbol{e_3}\|} = \begin{bmatrix}0\\1\\0\end{bmatrix}$. Also $A = \begin{bmatrix}1 & 1 & 2\\0 & 0 & 1\\1 & 0 & 0\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\0 & 0 & 1\\\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{bmatrix}\begin{bmatrix}\sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2}\\0 & \frac{1}{\sqrt{2}} & \sqrt{2}\\0 & 0 & 1\end{bmatrix} = QR$.

© 2022 e-Learning Centre, UCSC

Orthogonal Matrices

Definition: Let $q_1, q_2, \cdots, q_n$ be orthonormal vectors in $\mathbb{R}^n$ (Orthonormal basis for $\mathbb{R}^n$). The square matrix Q with $q_1, q_2, \cdots, q_n$ as its columns is called **orthogonal matrix**.

Proposition 1.8.4.2: If Q is a square orthogonal matrix, then $Q^TQ = I$.

Proof:

$$\begin{bmatrix} \cdots & q_1^T & \cdots \\ \cdots & q_2^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & q_n^T & \cdots \end{bmatrix}_{n\times n} \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ q_1 & q_2 & \cdots & q_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = I_n \quad \left(\because \begin{array}{c} q_i^T\, q_i = 1, \forall i \\ q_i^T\, q_j = 0, \forall i \neq j \end{array} \right).$$

Hence, $\boldsymbol{Q^TQ = I}$ and $Q^{-1} = Q^T$.

© 2022 e-Learning Centre, UCSC

Examples of Orthogonal Matrices

1. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & cos\theta \end{bmatrix}$ for any $\theta \in \mathbb{R}$.

2. $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$

© 2022 e-Learning Centre, UCSC

Properties of Orthogonal Matrices

Proposition 1.8.4.3: Let Q be an orthogonal Matrix of order n. Then

1. Q preserves lengths ($\| Qx \| = \| x \|, \forall x \in \mathbb{R}^n$)

2. Q preserves inner products and angles ($(Qx)^T(Qy) = x^T y, \forall x, y \in \mathbb{R}^n$).

Proof:

1. $\| Qx \|^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \| x \|^2 \quad (\because Q^T Q = I)$

   Hence, $\| Qx \| = \| x \|, \forall x \in \mathbb{R}^n$.

2. $(Qx)^T(Qy) = x^T Q^T Q y = x^T y \quad (\because Q^T Q = I)$

   Hence, $(Qx)^T(Qy) = x^T y, \forall x, y \in \mathbb{R}^n$.

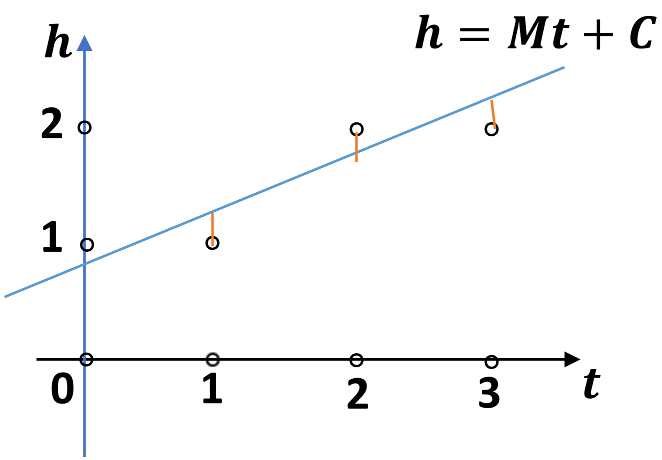
© 2022 e-Learning Centre, UCSC

1.8.5 Least Squares Approximations by a line

Consider the following system of linear equations $Ax = b$:

$$\begin{aligned} M + C &= 1 \\ 2M + C &= 2 \\ 3M + C &= 2 \end{aligned} \qquad \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} M \\ C \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

This system is inconsistent. We would like to find the best possible $\hat{x} = \begin{bmatrix} \widehat{M} \\ \hat{C} \end{bmatrix}$ such that error$\| b - A\hat{x} \|$ is minimal. That is, we would like to find the best line $h = \widehat{M}\mathrm{t} + \hat{C}$ that is very close to all data points.



Least Square Error E

$$= (M + C - 1)^2 + (2M + C - 2)^2 + (3M + C - 2)^2$$

$$\frac{\partial E}{\partial C} = 2(M + C - 1) + 2(2M + C - 2) + 2(3M + C - 2)$$

$$\frac{\partial E}{\partial M} = 2(M + C - 1) + 4(2M + C - 2) + 6(3M + C - 2)$$

$$\frac{\partial E}{\partial C} = 0 \Rightarrow 6M + 3C = 5 \ \& \ \frac{\partial E}{\partial M} = 0 \Rightarrow 14M + 6C = 11$$

Hence, the least square error is obtained at $M = {}^{1}/_{2}$ and $C = {}^{2}/_{3}$. The best line is

$$h = {}^{1}/_{2}\, t + {}^{2}/_{3}.$$

© 2022 e-Learning Centre, UCSC

Least Squares Approximation Problem

Problem: Given a (possibly inconsistent) system of linear equations $Ax = b$, find x such that $\| Ax - b \|$ is as small as possible. We call such a vector x a least squares approximation for the system of linear equations.

Let $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Then

$$Ax - b = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}\, x_n - b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}\, x_n - b_m \end{bmatrix},$$ and therefore,

$$\| Ax - b \|^2 = (a_{11}x_1 + \cdots + a_{1n}\, x_n - b_1)^2 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}\, x_n - b_m)^2.$$

Therefore, minimizing $\| Ax - b \|$ is the same as minimizing the sum of the squares of the errors of all the equations.

We note that $\| Ax - b \| = 0$ if and only if $Ax - b = \mathbf{0}$. Therefore, if the system of equations $Ax = b$ is consistent, then its least squares approximations are exactly the solutions of the system of equations in the usual sense.

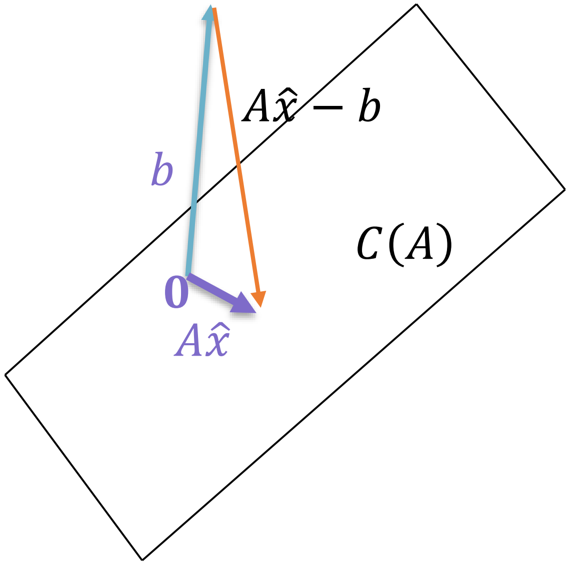
© 2022 e-Learning Centre, UCSC

Solution of the least squares approximation problem

Theorem 1.8.5.1: A vector $\hat{x}$ is a least squares approximation of the system of equations $Ax = b$ if and only if $A^T A\hat{x} = A^T b$.

Proof: Let $a_1, a_2, \cdots, a_n$ be the columns of the matrix A. Then the column space of A is $C(A) = Span\,(\{a_1, a_2, \cdots, a_n\})$. If $x = [x_1 \quad \cdots \quad x_n]^T$ is any column vector, then by the definition of matrix multiplication, we have $Ax = x_1 a_1 + \cdots + x_n a_n$.

Therefore, the equation $Ax = b$ has a solution if and only if $b \in C(A)$.
For $\hat{x}$ to be a least squares approximation, we want $\| A\hat{x} - b \|$ to be as small as possible. This means that we are looking for the element of $C(A)$ that is closest to b. we know that this happens when $(A\hat{x} - b) \perp C(A)$.



$(A\hat{x} - b) \perp C(A) \;\Leftrightarrow (A\hat{x} - b) \perp a_i$ for each i

$\Leftrightarrow a_i \cdot (A\hat{x} - b) = a_i{}^T(A\hat{x} - b) = 0$ for each i

$\Leftrightarrow A^T(A\hat{x} - b) = 0$

$\Leftrightarrow A^T A\hat{x} = A^T b$. (This is called as Normal Equation.)

© 2022 e-Learning Centre, UCSC

Least Squares Approximation

Example 1.8.5.1: Find the least squares approximation for the system of equations

$2x_1 + 2x_2 + 2x_3 = 1$

$x_1 - \ x_2 \ - x_3 = -2$

$-x_1 - \ x_2 + 2\,x_3 = 4$

$2x_1 + 2\,x_2 \ - x_3 = -8.$

Solution: $Ax = b$, where $A = \begin{bmatrix} 2 & 2 & 2 \\ 1 & -1 & -1 \\ -1 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$, $\quad b = \begin{bmatrix} 1 \\ -2 \\ 4 \\ -8 \end{bmatrix}$, and $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

By Theorem 1.8.5.1, the least squares approximation $\hat{x}$ is given by the solution of the system of equations $A^T A\hat{x} = A^T b$.

$$A^T A = \begin{bmatrix} 2 & 1 & -1 & 2 \\ 2 & -1 & -1 & 2 \\ 2 & -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 2 \\ 1 & -1 & -1 \\ -1 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 10 & 8 & -1 \\ 8 & 10 & 1 \\ -1 & 1 & 10 \end{bmatrix}.$$

© 2022 e-Learning Centre, UCSC

Least Squares Approximation

Solution of example 1.8.5.1 continue:

$$A^T b = \begin{bmatrix} 2 & 1 & -1 & 2 \\ 2 & -1 & -1 & 2 \\ 2 & -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 4 \\ -8 \end{bmatrix} = \begin{bmatrix} -20 \\ -16 \\ 20 \end{bmatrix}.$$

Therefore, the least squares approximation, we solve the system of equations:

$$\begin{bmatrix} 10 & 8 & -1 \\ 8 & 10 & 1 \\ -1 & 1 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -20 \\ -16 \\ 20 \end{bmatrix}.$$

The unique solution of this system is $x_1 = -1, x_2 = -1$ and $x_3 = 2$.

Note: We can double-check this answer by checking whether $Ax - b$ is orthogonal to every column vector of A.

© 2022 e-Learning Centre, UCSC

Exercise

- Find the least squares approximation for the system of equations:

$$x_1 + 2x_2 + 2x_3 = 5$$

$$x_1 + x_2 - x_3 = 11$$

$$x_1 + 2x_2 - x_3 = -18$$

$$2x_1 - x_2 + 2x_3 = 0.$$

© 2022 e-Learning Centre, UCSC