Ch. 5 Orthogonality

5.1 Scalar product in $\mathbb{R}^n$

Consider two vectors $\begin{cases} x = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n \\ y = (y_1, y_2, \cdots, y_n)^T \in \mathbb{R}^n \end{cases}$

Def. The scalar product of x and y is

$$x^T y = (x_1 \ x_2 \ \cdots \ x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Ex. 1 If $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $y = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$, then the scalar product of x and y is:

$$x^T y = 1 \times (-1) + 2 \times 0 + 3 \times 2 = 5$$

Def. The Euclidean length (or the 2-norm) of $x = (x_1, x_2, \cdots, x_n)^T$ is defined by

$$\|x\| = \sqrt{x^T x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Def. If $x, y \in \mathbb{R}^n$, the distance between the vectors is $\|y - x\|$.

Ex. 2 If $x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $y = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, then the distance between x and y is:

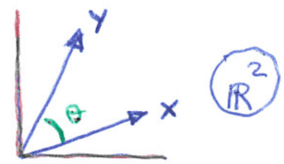
$$\|y - x\| = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} = \sqrt{(-1-1)^2 + (0-2)^2} = \sqrt{4+4} = \sqrt{8}.$$

We also have:

$$\|x\| = \sqrt{x_1^2 + x_2^2} = \sqrt{1^2 + 2^2} = \sqrt{5}$$

$$\|y\| = \sqrt{y_1^2 + y_2^2} = \sqrt{(-1)^2 + 0^2} = 1$$

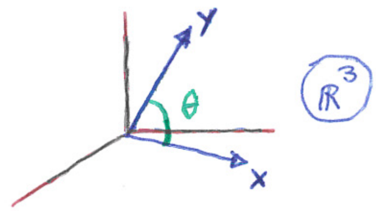
$$x^T y = 1 \times (-1) + 2 \times 0 = -1$$



Theorem Let $x, y \in \mathbb{R}^2$ or $\mathbb{R}^3$ be two non-zero vectors.

Let θ be the angle between x and y.

Then $$\cos\theta = \frac{x^T y}{\|x\| \cdot \|y\|}, \quad 0 \le \theta \le \pi.$$



EX.3 If $x = \begin{pmatrix}1\\2\end{pmatrix}$ and $y = \begin{pmatrix}3\\1\end{pmatrix}$, find the angle θ between them.

$x^T y = 1\times 3 + 2\times 1 = 5$

$\|x\| = \sqrt{1^2 + 2^2} = \sqrt{5}$

$\|y\| = \sqrt{3^2 + 1^2} = \sqrt{10}$

$$\Rightarrow \cos\theta = \frac{x^T y}{\|x\| \cdot \|y\|} = \frac{5}{\sqrt{5}\cdot\sqrt{10}} = \frac{\sqrt{5}}{\sqrt{10}} = \sqrt{\frac{5}{10}} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}$$

$$\Rightarrow \theta = 45^\circ \quad \text{or} \quad \theta = \frac{\pi}{4}$$

Corollary (Cauchy-Schwarz inequality) If $x, y \in \mathbb{R}^2$ or $\mathbb{R}^3$, then

$$|x^T y| \le \|x\| \cdot \|y\|.$$

the equality hold iff one of the vectors is $\mathbf{0}$ or one vector is the multiple of the other.

EX.4 Verify that Cauchy-Schwarz inequality hold for vectors in EX.3.

$5 = |x^T y| \le \|x\| \cdot \|y\| = \sqrt{50}$ $\quad (5^2 \le (\sqrt{50})^2 \Rightarrow 5 \le \sqrt{50})$

Def. the vectors $x, y \in \mathbb{R}^2$ or $\mathbb{R}^3$ are said to be orthogonal if $x^T y = 0$.

In this case $\cos\theta = 0 \Rightarrow \theta = \frac{\pi}{2}$. We write $x \perp y$.

<u>Def.</u> Any vector $u \in \mathbb{R}^n$ with length 1 is called a unit vector.

For a unit vector $u = (u_1, u_2, \cdots, u_n)^T \in \mathbb{R}^n$, we therefore have:

$$\|u\| = \sqrt{u_1^2 + \cdots + u_n^2} = 1$$

For example $e_1 = (1, 0, \cdots, 0)^T \in \mathbb{R}^n$ is a unit vector.

Given a non-zero vector $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$, we can find a unit vector corresponding to $\mathbf{x}$ by <u>normalizing</u> $\mathbf{x}$ (divide $\mathbf{x}$ by its length):

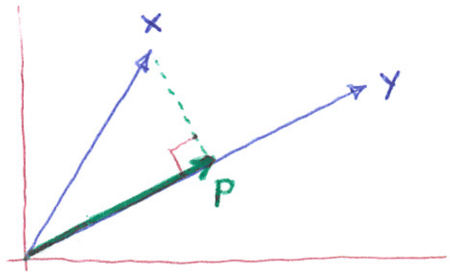
$$u = \frac{1}{\|\mathbf{x}\|} \cdot \mathbf{x}$$

Then it is clear that: $\|u\| = \frac{1}{\|\mathbf{x}\|} \cdot \|\mathbf{x}\| = 1$

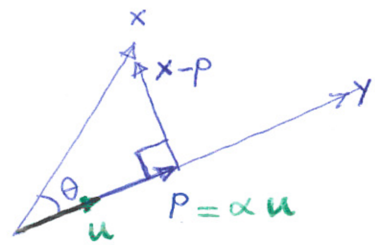
<u>Remark</u> Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ or $\mathbb{R}^3$. Let $u = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ and $\mathbf{v} = \frac{\mathbf{y}}{\|\mathbf{y}\|}$. Then,

$$\cos\theta = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} = u^T \mathbf{v}.$$

Orthogonal Projection



The projection vector p of x onto y is obtained by drawing a line from x orthogonal to y; see the figure.



To find p, we notice that $x - p \perp p$.

Let $u = \dfrac{y}{\|y\|}$ be the unit vector in the direction of y.

We want to find $\alpha \in \mathbb{R}$ s.t. $p = \alpha u$. (Note the p is in the directon of u)

We know that $x - \alpha u \perp \alpha u$. This means in the right triangle we have

$$\cos\theta = \frac{\|p\|}{\|x\|} = \frac{\|\alpha u\|}{\|x\|} = \frac{\alpha \cdot \|u\|}{\|x\|} = \frac{\alpha}{\|x\|}$$

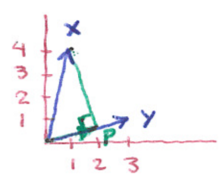
$$\Rightarrow \quad \alpha = \|x\| \cos\theta = \|x\| \cdot \frac{x^T y}{\|x\| \, \|y\|} = \frac{x^T y}{\|y\|}$$

Hence, $p = \alpha u = \dfrac{x^T y}{\|y\|} \cdot \dfrac{y}{\|y\|} = \dfrac{x^T y}{y^T y}\, y$

α is the scalar projection of x onto y: $\quad \alpha = \dfrac{x^T y}{\|y\|}$

p is the vector projection of x onto y: $\quad p = \dfrac{x^T y}{y^T y}\, y$

EX.5 Find the projection vector p of $x = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ onto $y = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$.



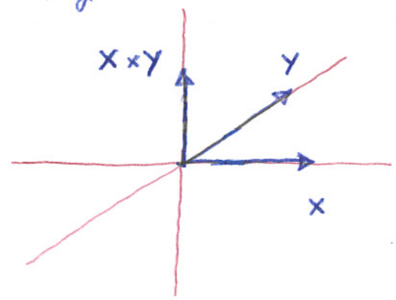
$$p = \frac{x^T y}{y^T y}\, y = \frac{1 \times 3 + 4 \times 1}{3^2 + 1^2} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \frac{7}{10} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 21/10 \\ 7/10 \end{pmatrix} = \begin{pmatrix} 2.1 \\ 0.7 \end{pmatrix}.$$

Cross product of two vectors

Given two vectors $x, y \in \mathbb{R}^3$, their cross product is defined as:

$$x \times y = \begin{pmatrix} x_2 y_3 - y_2 x_3 \\ y_1 x_3 - x_1 y_3 \\ x_1 y_2 - y_1 x_2 \end{pmatrix}$$

If x and y are linearly independent, then $x \times y$ is orthogonal to both x and y and therefore normal to the plane containing them.



If x and y have the same direction (not linearly independent) or if either one has zero length, then $x \times y = 0$.

For any vectors $x, y \in \mathbb{R}^3$, we have $\|x \times y\| = \|x\| \, \|y\| \sin\theta$,

where θ is the angle between x and y.

The concepts of angle between two vectors, Cauchy-Schwarz inequality, and orthogonality can be extended to $\mathbb{R}^n$:

For any vectors $x, y \in \mathbb{R}^n$, we have:

1) $\cos\theta = \dfrac{x^T y}{\|x\| \, \|y\|}$, $\quad 0 \leq \theta \leq \pi$

2) Cauchy-Schwarz inequality: $\quad -1 \leq \dfrac{x^T y}{\|x\| \, \|y\|} \leq 1$

3) if $x^T y = 0$, we say x and y are orthogonal: $x \perp y$.

Remark For $x, y \in \mathbb{R}^n$: $\|x+y\|^2 = (x+y)^T (x+y) = \|x\|^2 + 2x^T y + \|y\|^2$.

If $x \perp y \Rightarrow x^T y = 0 \Rightarrow \|x+y\|^2 = \|x\|^2 + \|y\|^2$ this is known as Pythagorean law.

Application to Statistics

We want to study how closely exam scores for a class correlate with scores on homework assignments. This may be useful to find out the level of difficulty.

1) We first collect data for 7 students:

	Scores (out of 200)		
student	HW assignments	Midterms	Final
1	198	200	196
2	160	165	165
3	158	158	133
4	150	165	91
5	175	182	151
6	134	135	101
7	152	136	80
Average	161	163	131

2) Next, we need to do some data processing:

A) Translate the scores to average zero.

In each column, we subtract the average score from each column and store the translated scores in a matrix:

$$X = \begin{bmatrix} 37 & 37 & 65 \\ -1 & 2 & 34 \\ -3 & -5 & 2 \\ -11 & 2 & -40 \\ 14 & 19 & 20 \\ -27 & -28 & -30 \\ -9 & -27 & -51 \end{bmatrix}$$

Remark 1 Each column vector of X has mean zero.

Remark 2 The entries of X represent the deviations from the average for each of the three sets of scores.

B) Normalize each column to make them unit vectors and store them in a matrix:

$$U = \begin{bmatrix} 0.74 & 0.65 & 0.62 \\ -0.02 & 0.03 & 0.33 \\ -0.06 & -0.09 & 0.02 \\ -0.22 & 0.03 & -0.38 \\ 0.28 & 0.33 & 0.19 \\ -0.54 & -0.49 & -0.29 \\ -0.18 & -0.47 & -0.49 \end{bmatrix}$$

If $X = [X_1 \ X_2 \ X_3]$

then $U = \left[\underbrace{\frac{X_1}{\|X_1\|}}_{u_1} \ \underbrace{\frac{X_2}{\|X_2\|}}_{u_2} \ \underbrace{\frac{X_3}{\|X_3\|}}_{u_3} \right]$

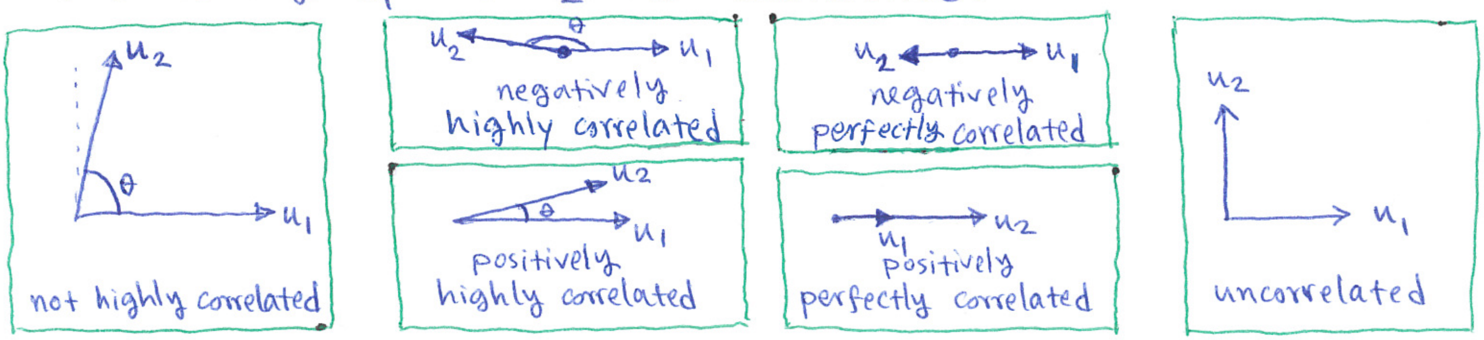
3) Statistical analysis:

To compare two sets of scores, we compute the cosine of the angle between their corresponding column vectors. A cosine value near 1 indicates that the two sets of scores are highly correlated.

For two vectors x_1 and x_2: $$\cos\theta = \frac{x_1^T x_2}{\|x_1\| \cdot \|x_2\|} = u_1^T u_2 \approx 0.92.$$

A perfect correlation of 1 would correspond to the case where the two sets of translated normalized vectors are proportional, that is $u_2 = \alpha u_1$.

A correlation of zero would correspond to the case $u_1 \perp u_2$. In this case we say u_1 and u_2 are uncorrelated.



We can compute the correlation between each pair of scores, we compute the matrix
$$C = U^T U = \begin{pmatrix} 1 & 0.92 & 0.83 \\ 0.92 & 1 & 0.83 \\ 0.83 & 0.83 & 1 \end{pmatrix}$$

The matrix C is referred to as a correlation matrix.

For instance, | correlation between u_1 and u_3 is 0.83.
correlation between u_1 and u_2 is 0.92.

This means there is a higher correlation between HW scores & midterm than between HW scores & final.

5.2 Orthogonal subspaces

Let $A \in \mathbb{R}^{m \times n}$ with the null space $N(A) = \{ x \in \mathbb{R}^n \mid Ax = 0 \}$.

Consider a vector $x = (x_1, \cdots, x_n)^T \in N(A) \Rightarrow Ax = 0$

$\Rightarrow$ we have $a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n = 0$ for $i = 1, \cdots, m$.

Hence, x is orthogonal to any i-th row of A.

Hence, x is orthogonal to any i-th column of A^T.

Hence, x is orthogonal to any linear combination of the column vectors of A^T.

Hence, x is orthogonal to any vector y in the column space of A^T: $x^T y = 0$

Thus each vector in $N(A)$ is orthogonal to every vector in $C(A^T)$.

We say $N(A)$ is orthogonal to $C(A^T)$. $\Rightarrow$ we write $N(A) \perp C(A^T)$.

Def. Two subspaces X and Y of $\mathbb{R}^n$ are said to be orthogonal if

$$x^T y = 0, \quad \forall x \in X, \ \forall y \in Y.$$

we write $X \perp Y$

Ex. 6 $X =$ subspace of $\mathbb{R}^3$ spanned by $\{e_1, e_2\}$.

$Y =$ subspace of $\mathbb{R}^3$ spanned by e_3.

If $x \in X$, it must be of the form $x = (x_1, x_2, 0)^T$.

If $y \in Y$, it must be of the form $y = (0, 0, y_3)^T$.

Then, $x^T y = x_1 \times 0 + x_2 \times 0 + 0 \times y_3 = 0 \Rightarrow X \perp Y$.

Def. Let X be a subspace of $\mathbb{R}^n$.

The orthogonal complement of X, denoted by $X^\perp$, is defined as:

$$X^\perp = \{ \mathbf{z} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{z} = 0, \ \forall\, \mathbf{x} \in X \}.$$

Ex. 7 Consider $X = \text{span}(\mathbf{e}_1)$ and $Y = \text{span}(\mathbf{e}_2)$ two subspaces of $\mathbb{R}^3$.

(1) Are X and Y orthogonal?

(2) Are X and Y orthogonal complements?

1) $$\begin{cases} \text{If } \mathbf{x} \in X = \text{span}(\mathbf{e}_1) \Rightarrow \mathbf{x} = (x_1, 0, 0)^T \\ \text{If } \mathbf{y} \in Y = \text{span}(\mathbf{e}_2) \Rightarrow \mathbf{y} = (0, y_2, 0)^T \end{cases} \Rightarrow \mathbf{x}^T\mathbf{y} = x_1 \times 0 + 0 \times y_2 + 0 \times 0 = 0$$

$$\Downarrow$$

$$X \perp Y.$$

2) $$\begin{aligned} X^\perp &= \{ \mathbf{z} \in \mathbb{R}^3 \mid \mathbf{x}^T\mathbf{z} = 0, \ \forall\, \mathbf{x} = (x_1, 0, 0)^T \} \\ &= \{ \mathbf{z} \in \mathbb{R}^3 \mid x_1 z_1 + 0 \times z_2 + 0 \times z_3 = 0 \} \\ &= \{ \mathbf{z} \in \mathbb{R}^3 \mid x_1 z_1 = 0, \ \forall\, x_1 \in \mathbb{R} \} \\ &= \{ \mathbf{z} \in \mathbb{R}^3 \mid z_1 = 0 \} = \{ \mathbf{z} = (0, z_2, z_3) \} = \text{span}(\mathbf{e}_2, \mathbf{e}_3) \neq Y \end{aligned}$$

$$\begin{aligned} Y^\perp &= \{ \mathbf{z} \in \mathbb{R}^3 \mid \mathbf{y}^T\mathbf{z} = 0, \ \forall\, \mathbf{y} = (0, y_2, 0)^T \} \\ &= \cdots = \text{span}(\mathbf{e}_1, \mathbf{e}_2) \neq X \end{aligned}$$

Hence, X and Y are not orthogonal complements.

Remarks. $\boxed{1}$ If $X, Y \subset \mathbb{R}^n$ and $X \perp Y \Rightarrow X \cap Y = \{\mathbf{0}\}$

proof. If $\mathbf{x} \in X \cap Y$ and $X \perp Y$, then $\mathbf{x}^T\mathbf{x} = 0 \Rightarrow \|\mathbf{x}\|^2 = 0 \Rightarrow \mathbf{x} = \mathbf{0}$.

$\boxed{2}$ If $X \subset \mathbb{R}^n$, then $X^\perp \subset \mathbb{R}^n$.

Fundamental Subspaces

Consider $A \in \mathbb{R}^{m\times n}$.

vector $b \in \mathbb{R}^m$ is in the column space of A iff $b = Ax$ for some $x \in \mathbb{R}^n$.

If we think of A as a linear transformation $L(x) = Ax : \mathbb{R}^n \to \mathbb{R}^m$, then the column space of A is the same as the range of A.

$$\text{range}(A) = \{ b \in \mathbb{R}^m \mid b = Ax,\ x \in \mathbb{R}^n \} = \underset{\text{column space}}{C(A)}$$

$$\text{range}(A^T) = \{ c \in \mathbb{R}^n \mid c = A^T y,\ y \in \mathbb{R}^m \} = C(A^T)$$

$C(A)$ is the same as $R(A^T)$, except that $C(A)$ contains $m\times 1$ vectors and $R(A^T)$ contains $1\times m$ vectors. Hence, $b \in C(A)$ iff $b^T \in R(A^T)$.

Similarly, $c \in C(A^T)$ iff $c^T \in R(A)$.

Recall (frome page 8) that $C(A^T) \perp N(A)$. Similarly, $C(A) \perp N(A^T)$.
The following theorem states that $N(A)$ is actually the orthogonal complement of $C(A^T)$.

<u>theorem</u> (Fundamental Subspaces Theorem)
If $A \in \mathbb{R}^{m\times n}$, then $N(A) = C(A^T)^\perp$ and $N(A^T) = C(A)^\perp$.

<u>proof</u>. 1) Since $C(A^T) \perp N(A) \Rightarrow \boxed{N(A) \subset C(A^T)^\perp}$

2) if $x \in C(A^T)^\perp \Rightarrow x \perp$ columns of $A^T \Rightarrow Ax = 0 \Rightarrow x \in N(A) \Rightarrow \boxed{C(A^T)^\perp \subset N(A)}$

Hence, $\boxed{N(A) = C(A^T)^\perp}$. Similarly, we can show $\boxed{N(A^T) = C(A)^\perp}$.

<u>EX-8</u> Consider $A = \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}$. $\Rightarrow A^T = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$

$C(A) = \text{span}((1,2)^T)$ $\qquad$ $C(A^T) = \text{span}(e_1)$

$R(A) = \text{span}(e_1)$ $\qquad$ $R(A^T) = \text{span}((1,2))$

$\text{rang}(A) = \{ b \in \mathbb{R}^2 \mid b = Ax,\ x = (x_1, x_2)^T \in \mathbb{R}^2 \} = \{ b \in \mathbb{R}^2 \mid b = (x_1, 2x_1)^T = x_1(1,2)^T,\ x_1 \in \mathbb{R} \} = \text{span}((1,2)^T)$

$N(A) = \{ x \in \mathbb{R}^2 \mid Ax = 0 \} = \{ x \in \mathbb{R}^2 \mid x_1 = 0 \} = \text{span}(e_2)$

$N(A^T) = \{ x \in \mathbb{R}^2 \mid A^T x = 0 \} = \{ x \in \mathbb{R}^2 \mid x_1 + 2x_2 = 0 \} = \text{span}((-2,1)^T)$

$$\begin{cases} C(A) = \mathrm{span}\left(\begin{pmatrix}1\\2\end{pmatrix}\right) \\ N(A^T) = \mathrm{span}\left(\begin{pmatrix}-2\\1\end{pmatrix}\right) \end{cases}$$

Since $\begin{pmatrix}1\\2\end{pmatrix} \perp \begin{pmatrix}-2\\1\end{pmatrix} \Rightarrow$ every vector in $C(A) \perp$ every vector in $N(A^T)$

$\Downarrow$

$C(A) \perp N(A^T)$

Moreover, $C(A)^\perp = \{ \mathbf{z} \in \mathbb{R}^2 \mid \mathbf{x}^T \mathbf{z} = 0, \ \forall\, \mathbf{x} \in C(A) \}$

$= \{ \mathbf{z} \in \mathbb{R}^2 \mid (\alpha, 2\alpha)\begin{pmatrix}z_1\\z_2\end{pmatrix} = \alpha(z_1 + 2z_2) = 0 \} = \mathrm{span}\left((-2,1)^T\right) = N(A^T)$

$$\begin{cases} C(A^T) = \mathrm{span}(e_1) \\ N(A) = \mathrm{span}(e_2) \end{cases}$$

Since $e_1 \perp e_2 \Rightarrow$ every vector in $C(A^T) \perp$ every vector in $N(A)$

$\Downarrow$

$C(A^T) \perp N(A)$

Moreover, $C(A^T)^\perp = \{ \mathbf{z} \in \mathbb{R}^2 \mid \mathbf{x}^T \mathbf{z} = 0, \ \forall\, \mathbf{x} \in C(A^T) \}$

$= \{ \mathbf{z} \in \mathbb{R}^2 \mid (\alpha, 0)\begin{pmatrix}z_1\\z_2\end{pmatrix} = \alpha z_1 = 0 \} = \mathrm{span}(e_2) = N(A)$.

<u>Theorem</u> If S is a subspace of $\mathbb{R}^n$, then $\dim(S) + \dim(S^\perp) = n$.

If $\{x_1, \cdots, x_r\}$ is a basis for S and $\{x_{r+1}, \cdots, x_n\}$ is a basis for $S^\perp$, then $\{x_1, \cdots, x_n\}$ is a basis for $\mathbb{R}^n$.

<u>Def.</u> Let U and V be subspaces of a vector space W.

if $\forall\, w \in W$ can be written uniquely as $u + v$, where $u \in U$ and $v \in V$, then we say W is a <u>direct sum</u> of U and V, and we write $W = U \oplus V$.

<u>Theorem⊕.</u> If S is a subspace of $\mathbb{R}^n$, then $\mathbb{R}^n = S \oplus S^\perp$.

<u>Theorem</u> if S is a subspace of $\mathbb{R}^n$, then $(S^\perp)^\perp = S$.

Ex. 9 Let $A \in \mathbb{R}^{m \times n}$. Then $N(A)$ and $C(A^T)$ are subspaces of $\mathbb{R}^n$.

We know that $C(A^T)$ and $N(A)$ are orthogonal complements in $\mathbb{R}^n$

Hence, by theorem $\oplus$, $C(A^T) \oplus N(A) = \mathbb{R}^n$.

Ex. 10 Let $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$.

(1) Find $N(A)$ and $C(A^T)$.

(2) Verify that $N(A) \oplus C(A^T) = \mathbb{R}^3$.

1) $N(A) = \{ x \in \mathbb{R}^3 \mid Ax = 0 \} = \{ x \in \mathbb{R}^3 \mid \begin{pmatrix} 2x_1 \\ 3x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \} = \{ (0,0,x_3), x_3 \in \mathbb{R} \}$

$\boxed{N(A) = \text{span}(e_3)}$

$A^T = \begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}$

To find $C(A^T)$, we reduce A^T to RE form and find the columns corresponding to leading 1's: $A^T = \begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow C(A^T) = \text{span}\left(\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} \right)$

$\Rightarrow \boxed{C(A^T) = \text{span}(e_1, e_2)}$

3) $N(A) \oplus C(A^T) = \text{span}(e_1, e_2, e_3) = \mathbb{R}^3$. O.K.

5.3 least Squares Problems

Motivation

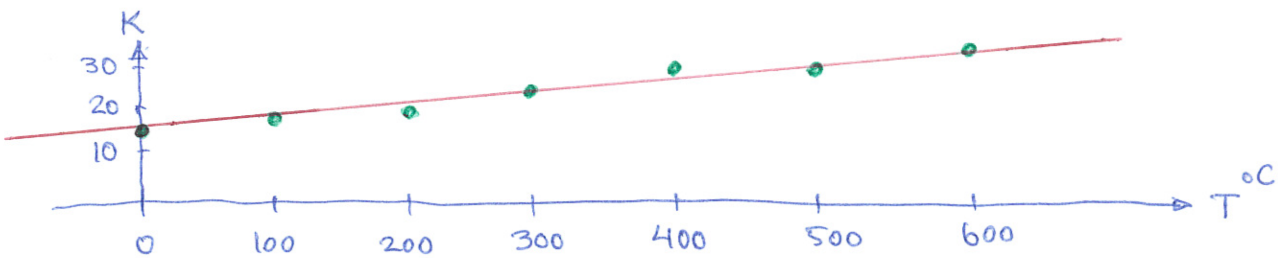
Assume that we want to derive a formula for the thermal conductivity of stainless steel K as a function of temperature T:

$$K = K(T)$$

We first measure K at different temperatures T between $0^{\circ C}$ to $600^{\circ C}$.

we collect the measured data:

i	1	2	3	4	5	6	7
T_i	0	100	200	300	400	500	600
K_i	14.5	17	17.6	19	22	23.1	24.2



We then try to find a function that best fits the data. For example we can try to find a line that fits the 7 data points:

$$\boxed{K = \alpha + \beta . T}$$

See the red line in figure above.

We can then use the formula to find K at different T values.

This problem is called a least squares problems, and the method for finding α and β is called the method of least squares.

Now, let us consider a general case with m data points:

i	1	2	3	...	m
x_i	x_1	x_2	x_3	...	x_m
y_i	y_1	y_2	y_3	...	y_m

We wish to find a linear function $y = \alpha x + \beta$ that best fits the data

We will require that the linear function passes through all data points.

this means, we want: $y_i = \alpha + \beta x_i$, $i = 1, 2, \cdots, m$

this gives us m equations with 2 unknowns (α, β):

$$\begin{cases} \alpha + x_1 \cdot \beta = y_1 \\ \alpha + x_2 \beta = y_2 \\ \vdots \qquad\qquad \vdots \\ \alpha + x_m \beta = y_m \end{cases} \xrightarrow{\text{in matrix form}} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

We therefore obtain an overdetermined system of equations if $m \geq 3$.

Let's write the system as $A\mathbf{x} = \mathbf{b}$

where $A \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^{n \times 1}$, $\mathbf{b} \in \mathbb{R}^{m \times 1}$, with $m > n$.

Such systems are usually inconsistent. So, in general, we cannot find a vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$ for which $A\mathbf{x} = \mathbf{b}$. Instead, we try to find a vector $\mathbf{x}$ for which $A\mathbf{x}$ is as close as possible to $\mathbf{b}$.

Hence, we want $\mathbf{r} = A\mathbf{x} - \mathbf{b}$, which is called residual, be as small as possible. We then try to find $\mathbf{x}$ that makes the Euclidean length of $\mathbf{r}$ as small as possible.

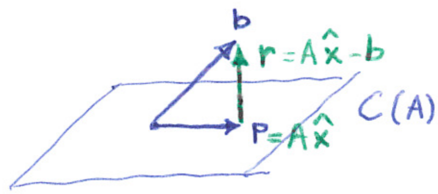
So, we wish to find $\mathbf{x}$ such that $\|\mathbf{r}\|$ is as small as possible.

Minimization of $\|\mathbf{r}\|$ is equivalent to minimizing $\|\mathbf{r}\|^2$.

A vector $\hat{\mathbf{x}}$ that minimizes $\|A\mathbf{x}-\mathbf{b}\|^2$, is called the least squares solution of the system $A\mathbf{x}=\mathbf{b}$.

How to find $\hat{\mathbf{x}}$?

Let $\mathbf{p} = A\hat{\mathbf{x}}$. We know that $\mathbf{p} \in C(A)$.



We want $\mathbf{p}$ be a vector in $C(A)$ that is closest to $\mathbf{b}$.

The shortest vector from $\mathbf{b}$ to $C(A)$ is the one which is orthogonal to $C(A)$.

Hence, we want $\mathbf{r} \in C(A)^{\perp} = N(A^T) \implies 0 = A^T\mathbf{r} = A^T(A\hat{\mathbf{x}} - \mathbf{b})$

$$\implies \boxed{A^TA\hat{\mathbf{x}} = A^T b}$$

This is an $n\times n$ system, since $A^TA \in \mathbb{R}^{n\times n}$.

<u>theorem</u> If $A \in \mathbb{R}^{m\times n}$ with rank n, the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$ have a unique solution $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$, and $\hat{\mathbf{x}}$ is the unique least squares solution of the system $A\mathbf{x}=\mathbf{b}$.

<u>EX. 11</u> Find the least squares solution to the system $\begin{cases} x_1 + x_2 = 3 \\ -2x_1 + 3x_2 = 1 \\ 2x_1 - x_2 = 2 \end{cases}$

We write the system in form $A\mathbf{x}=\mathbf{b}$, where $A = \begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$.

The normal equations are: $A^TA\mathbf{x} = A\mathbf{b} \Rightarrow \begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix} \Rightarrow \begin{cases} x_1 = \frac{83}{50} \\ x_2 = \frac{71}{50} \end{cases}$

EX. 12 Consider the following 4 data points:

i	1	2	3	4
x_i	0	1	2	3
y_i	3	2	4	4

1) Find the best least squares fit by a linear function.

2) Find the best least squares fit by a quadratic function.

---

1) Consider a linear function $y = \alpha + \beta x$. We want:

$y_i = \alpha + \beta x_i$, $i = 1, 2, 3, 4$

Hence, we obtain the system
$$\begin{cases} \alpha + x_1 \beta = y_1 \\ \alpha + x_2 \beta = y_2 \\ \alpha + x_3 \beta = y_3 \\ \alpha + x_4 \beta = y_4 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{A} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{\mathbf{a}} = \underbrace{\begin{bmatrix} 3 \\ 2 \\ 4 \\ 4 \end{bmatrix}}_{\mathbf{b}}$$

The least squares solution is obtaind by:

$$A^T A \, \mathbf{a} = A^T \mathbf{b} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 4 \\ 4 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 13 \\ 22 \end{bmatrix} \Rightarrow \begin{cases} \alpha = 2.5 \\ \beta = 0.5 \end{cases} \Rightarrow$$

The linear least squares fit is $\boxed{y = 2.5 + 0.5x}$

2) Consider a quadratic function $y = \alpha + \beta x + \gamma x^2$.

We want $y_i = \alpha + \beta x_i + \gamma x_i^2$, $i = 1, 2, 3, 4$.

We obtain the system:

$$\begin{cases} \alpha + x_1 \beta + x_1^2 \gamma = y_1 \\ \alpha + x_2 \beta + x_2^2 \gamma = y_2 \\ \alpha + x_3 \beta + x_3^2 \gamma = y_3 \\ \alpha + x_4 \beta + x_4^2 \gamma = y_3 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}}_{a} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{b}$$

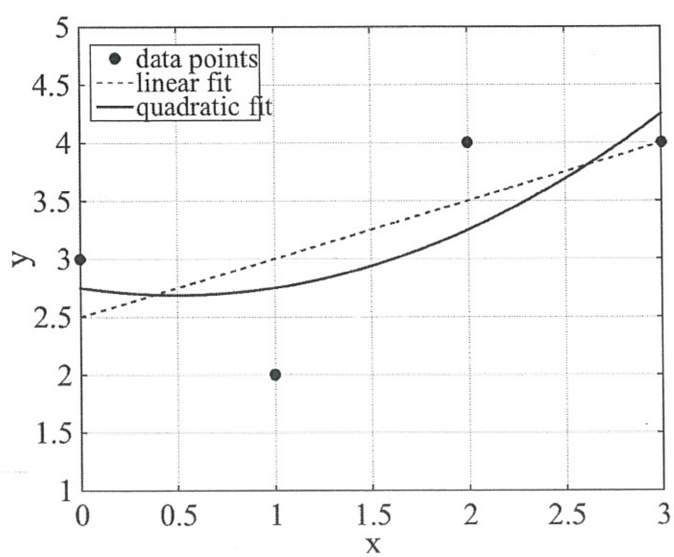
the least squares solution is obtained by

$$A^T A a = A^T b \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 4 \\ 4 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 13 \\ 22 \\ 54 \end{bmatrix} \Rightarrow \begin{cases} \alpha = 2.75 \\ \beta = -0.25 \\ \gamma = 0.25 \end{cases}$$

The quadratic least squares fit is $\boxed{y = 2.75 - 0.25\,x + 0.25\,x^2}$

The data points and the two least squares fits are shown below.



5.4 Inner Product Spaces

Recall:

The Scalar product in $\mathbb{R}^n$:

For $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$, $\quad \mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$.

Based on scalar product in $\mathbb{R}^n$, we defined the following:

1) orthogonality: $\quad \mathbf{x} \perp \mathbf{y} \Leftrightarrow \mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = 0$

2) Euclidean length (norm) of $\mathbf{x}$: $\quad \|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$

3) angle between $\mathbf{x}$ and $\mathbf{y}$: $\quad \cos\theta = \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$

Scalar products in $\mathbb{R}^n$ are very useful. For instance, we could solve least squares problems by scalar products in $\mathbb{R}^n$.

Scalar products are not only defined in $\mathbb{R}^n$. We can generalize scalar products to other vector spaces. We call them inner products. This is the topic of this section.

Def. An inner product on a vector space V is an operation on V that to each pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V assigns a real number $\langle \mathbf{x}, \mathbf{y} \rangle$ satisfying the following conditions:

(1) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality iff $\mathbf{x} = 0$

(2) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle \quad \forall\, \mathbf{x}, \mathbf{y} \in V$

(3) $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle \quad \forall\, \mathbf{x}, \mathbf{y}, \mathbf{z} \in V, \ \forall\, \alpha, \beta \in \mathbb{R}$

Def. A vector space V equipped with an inner product is called an inner product space.

Consider $V = \mathbb{R}^n$. The standard inner product in $\mathbb{R}^n$ is the scalar product:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i , \quad \text{where } x, y \in \mathbb{R}^n .$$

This is not the only possible/available inner product though. We can define other inner products in $\mathbb{R}^n$. For example, we can consider a "weighted" scalar product as an inner product for $\mathbb{R}^n$:

$$\langle x, y \rangle = \sum_{i=1}^{n} w_i \cdot x_i y_i , \quad \text{where } w_i \geq 0 \text{ are weights.}$$

Consider $V = \mathbb{R}^{m \times n}$. An inner product for $\mathbb{R}^{m \times n}$ can be defined by:

$$\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij} , \quad \text{where } A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m \times n} .$$

Consider $V = C[a,b]$. An inner product on $C[a,b]$ may be defined by:

$$\langle f, g \rangle = \int_a^b f(x) g(x)\, dx , \quad \text{where } f, g \in C[a,b] .$$

Question Why $\langle f, g \rangle = \int_a^b f(x) g(x)\, dx$ is an inner product in $C[a,b]$?

Answer Because it satisfies the three conditions of an inner product.

For example, consider condition (1):

$$\langle f, f \rangle = \int_a^b f^2(x)\, dx \geq 0 \quad \text{with equality iff } f \equiv 0 .$$

We can also define a weighted integral as an inner product in $C[a,b]$:

$$\langle f, g \rangle = \int_a^b w(x) f(x) g(x)\, dx ,$$ where $w(x)$ is a positive continuous function on $[a,b]$, called a weight function.

Consider $V = P_n$. An inner product on P_n can be defined by:

$$\langle p, q \rangle = \sum_{i=1}^{n} p(x_i) \cdot q(x_i) ,$$ where $p, q \in P_n$ and $x_1, \dots, x_n$ are distinct real numbers.

Based on inner products, we can define the following:

1) the norm (or length) of a vector $\mathbf{v}$ in an inner product space V:

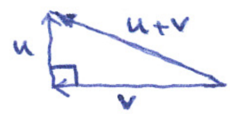
$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}, \quad \forall\, \mathbf{v} \in V.$$

2) orthogonality: $\mathbf{u} \perp \mathbf{v}$ if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, $\forall\, \mathbf{u}, \mathbf{v} \in V$.

<u>theorem</u> (Pythagorean Law) If $\mathbf{u} \perp \mathbf{v}$, then $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

<u>proof</u>. $\|\mathbf{u}+\mathbf{v}\|^2 = \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + 2\underbrace{\langle \mathbf{u}, \mathbf{v} \rangle}_{0} + \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

in $\mathbb{R}^2$, this is just the Pythagorean theorem:



3) Angle θ between $\mathbf{u}$ and $\mathbf{v}$: $\cos\theta = \dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$.

<u>Theorem</u> (Cauchy-Schwarz inequality) $|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\| \cdot \|\mathbf{v}\|$.
Equality holds iff $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

4) orthogonal projection: If $\mathbf{u}, \mathbf{v} \in V$ and $\mathbf{v} \neq 0$, then <u>vector projection</u> of $\mathbf{u}$ onto $\mathbf{v}$ is given by $\mathbf{p} = \alpha \dfrac{\mathbf{v}}{\|\mathbf{v}\|}$

wher $\alpha = \dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|}$ is the scalar projection of $\mathbf{u}$ onto $\mathbf{v}$.

<u>Note</u>: $\mathbf{u} - \mathbf{p} \perp \mathbf{p}$

EX. 13 Let $V = C[-1,1]$ with the inner product $\langle f,g\rangle = \int_{-1}^{1} f(x)\, g(x)\, dx$.

1) Show that the vectors 1 and x are orthogonal.

2) Show that the vectors 1 and x^2 are not orthogonal.

3) Determine the lengths of vectors 1, x, and x^2.

4) Verify the Pythagorean law for 1 and x.

---

1) $$\langle 1, x\rangle = \int_{-1}^{1} 1 \cdot x \, dx = \frac{1}{2}\left[x^2\right]_{x=-1}^{x=1} = \frac{1}{2}\left[(1)^2 - (-1)^2\right] = \frac{1}{2}(1-1) = 0.$$

2) $$\langle 1, x^2\rangle = \int_{-1}^{1} 1 \cdot x^2 \, dx = \frac{1}{3}\left[x^3\right]_{x=-1}^{x=1} = \frac{1}{3}\left[(1)^3 - (-1)^3\right] = \frac{1}{3}(1+1) = \frac{2}{3} \neq 0.$$

3) $$\langle 1, 1\rangle = \int_{-1}^{1} 1 \cdot 1 \, dx = 2 \Rightarrow \|1\| = \sqrt{\langle 1,1\rangle} = \sqrt{2}$$

$$\langle x, x\rangle = \int_{-1}^{1} x \cdot x \, dx = \frac{2}{3} \Rightarrow \|x\| = \sqrt{\langle x,x\rangle} = \sqrt{\frac{2}{3}}$$

$$\langle x^2, x^2\rangle = \int_{-1}^{1} x^2 \cdot x^2 \, dx = \frac{1}{5}\left[x^5\right]_{x=-1}^{x=1} = \frac{1}{5}(1+1) = \frac{2}{5} \Rightarrow \|x^2\| = \sqrt{\langle x^2, x^2\rangle} = \sqrt{\frac{2}{5}}$$

4) Since $1 \perp x$, the Pythagorean law must hold:

$$\begin{cases} \|1+x\|^2 = \langle 1+x, 1+x\rangle = \int_{-1}^{1} (1+x)^2 \, dx = \left[x + x^2 + \frac{1}{3}x^3\right]_{x=-1}^{x=1} = 1+1+\frac{1}{3}+1-1+\frac{1}{3} = \frac{8}{3} \\ \|1\|^2 + \|x\|^2 = 2 + \frac{2}{3} = \frac{8}{3} \end{cases}$$

Hence $\|1+x\|^2 = \|1\|^2 + \|x\|^2$.

Ex. 14 Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 3 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 1 \\ 3 & 0 \\ -3 & 4 \end{pmatrix}$, and consider the inner product $\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$ and the corresponding norm

$$\|A\|_F = \sqrt{\langle A, A \rangle} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

. This is called the Frobenius norm.

1) Determine $\langle A, B \rangle$.

2) Determine $\|A\|_F$.

3) Determine $\|B\|_F$.

4) Verify the Cauchy-Schwarz inequality.

---

1) $\langle A, B \rangle = 1 \times (-1) + 1 \times (1) + 1 \times 3 + 2 \times 0 + 3 \times (-3) + 3 \times 4 = 6$

2) $\|A\|_F = \sqrt{1^2 + 1^2 + 1^2 + 2^2 + 3^2 + 3^2} = 5$

3) $\|B\|_F = \sqrt{(-1)^2 + 1^2 + 3^2 + 0^2 + (-3)^2 + 4^2} = 6$

4) $\underbrace{|\langle A, B \rangle|}_{6} \le \underbrace{\|A\|_F}_{5} \underbrace{\|B\|_F}_{6}$ O.K. since $6 \le 30$.

More about norms

Norms can be defined independent of inner products. We have seen that we can define a norm based on an inner product, that is, we can consider the norm:

$$\|v\| = \sqrt{\langle v, v \rangle}, \qquad \forall\, v \in V.$$

In general, we can define norms in other ways.

Def. A vector space V is said to be a normed linear space if to each vector $v \in V$, a real number $\|v\|$, called the norm of v, is associated, satisfying:

1) $\|v\| \geq 0$ with equality iff $v = \mathbf{0}$.
2) $\|\alpha v\| = |\alpha|\,\|v\|, \quad \forall\, \alpha \in \mathbb{R},\ \forall\, v \in V.$
3) $\|u + v\| \leq \|u\| + \|v\|, \quad \forall\, u, v \in V$

↳ this is called the triangle inequality.

We can define norms in many ways and independent of inner products.

A few common norms:

$V = \mathbb{R}^n$: $\boxed{\|x\|_1 = \sum_{i=1}^{n} |x_i|}$ 1-norm, where $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$.

$V = \mathbb{R}^n$: $\boxed{\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|}$ infinity norm, where $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$.

$V = \mathbb{R}^n$: $\boxed{\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}}$ p-norm, where $p \geq 1$.

Remark. In the particular case when $p = 2$, we obtain the 2-norm:

$$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x, x \rangle}$$ which corresponds to the inner product $\langle x, x \rangle$.

Remark. If $p \neq 2$, the p-norm $\|\cdot\|_p$ does not correspond to any inner product.

If the norm is not derived from an inner product, the Pythagorean law will NOT hold necessarily.

EX.15 Consider $V = \mathbb{R}^2$ with the infinity norm. Verify that the Phythogorean law will not hold for $x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $y = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$.

$$\|x\|_\infty^2 + \|y\|_\infty^2 = [\max(1,2)]^2 + [\max(4,2)]^2 = 2^2 + 4^2 = 20$$

$$\|x+y\|_\infty^2 = [\max(|1-4|, |2+2|)]^2 = [\max(3,4)]^2 = 4^2 = 16$$

Hence, we have $\|x+y\|_\infty^2 < \|x\|_\infty^2 + \|y\|_\infty^2$.

Conclusion: Although $x \perp y$ in this example, Pythagorean law does not hold, since the infinity norm does not correspond to any inner product.

EX.16 Let $x = (4, -5, 3)^T$ be a vector in $\mathbb{R}^3$.

Compute $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$.

$$\|x\|_1 = |4| + |-5| + |3| = 4 + 5 + 3 = 12$$
$$\|x\|_2 = \sqrt{16 + 25 + 9} = 5\sqrt{2}$$
$$\|x\|_\infty = \max(|4|, |-5|, |3|) = \max(4,5,3) = 5$$

Def. Let x and y be two vectors in a normed linear space. The distance between x and y is the number $\|x - y\|$.

EX.17 Compute the distance between $x = (1,2)^T$ and $y = (3,1)^T$ in $\|\cdot\|_\infty$ and $\|\cdot\|_1$.

$$\|x - y\|_\infty = \max(|1-3|, |2-1|) = \max(2,1) = 2 \qquad \|x-y\|_1 = |1-3| + |2-1| = 2+1 = 3$$