Chapter 6: Orthogonality

In this chapter we generalise the notion of "perpendicular" vectors (meaning: at right angles) from $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$.

This notion can be generalised to arbitrary vector spaces, but that is way beyond the scope of this module.

Section 6.1: Definition

If $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$

then $\underline{x}^T \underline{y} = (x_1, \dots x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i$

is called the scalar product (or dot product, written $\underline{x} \cdot \underline{y}$) of $\underline{x}$ and $\underline{y}$.

(N.B. We identify the 1×1 matrix (z) with the scalar z.)

Example 6.2 $\underline{x} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$, $\underline{y} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$

$$\underline{x}.\underline{y} = 2\times 4 + (-3)\times 5 + 1\times 6 = 8 - 15 + 6 = -1$$

$$\underline{y}.\underline{x} = 4\times 2 + 5\times(-3) + 6\times 1 = -1.$$

Theorem 6.3 If $\underline{x}, \underline{y}, \underline{z} \in \mathbb{R}^n$, & $\alpha \in \mathbb{R}$, then

- $\underline{x}.\underline{y} = \underline{y}.\underline{x}$
- $(\underline{x}+\underline{y}).\underline{z} = \underline{x}.\underline{z} + \underline{y}.\underline{z}$
- $(\alpha\underline{x}).\underline{y} = \alpha(\underline{x}.\underline{y}) = \underline{x}.(\alpha\underline{y})$
- $\underline{x}.\underline{x} \geqslant 0$
- $\underline{x}.\underline{x} = 0$ only if $\underline{x} = \underline{0}$.

Proof

- $\underline{x}\cdot\underline{y} = \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i = \underline{y}\cdot\underline{x}$

  (if $\underline{x} = (x_1, \dots, x_n)^T$,
  $\underline{y} = (y_1, \dots, y_n)^T$,
  & $\underline{z} = (z_1, \dots z_n)^T$.)

- $$(\underline{x}+\underline{y})\cdot(\underline{z}) = \sum_{i=1}^{n} (x_i + y_i) z_i$$
  $$= \sum_{i=1}^{n} (x_i z_i + y_i z_i)$$
  $$= \sum_{i=1}^{n} x_i z_i + \sum_{i=1}^{n} y_i z_i$$
  $$= \underline{x}\cdot\underline{z} + \underline{y}\cdot\underline{z}$$

- $\underline{x}\cdot\underline{x} = \sum_{i=1}^{n} x_i x_i = \sum_{i=1}^{n} (x_i)^2 \geq 0$

- If $\underline{x}\cdot\underline{x} = 0$, then $\sum_{i=1}^{n} x_i^2 = 0$

  $\Rightarrow$ every $x_i = 0$

  i.e. $\underline{x} = (0, \dots, 0)^T = \underline{0}$.

Terminology 6.4 If $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$,

the *length* (or *norm*) of $\underline{x}$ is

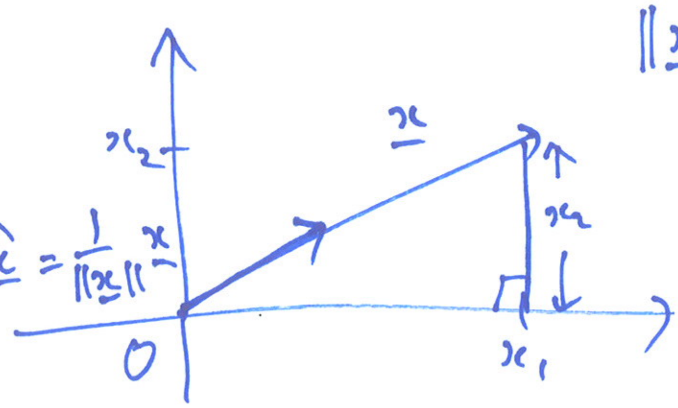
$$\|\underline{x}\| = \sqrt{\underline{x}.\underline{x}}$$
$$= \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

If $\|\underline{x}\| = 1$, then $\underline{x}$ is called a *unit vector*.

This agrees with what you know about $\mathbb{R}^2$ and $\mathbb{R}^3$.

$$\|\underline{x}\|^2 = x_1^2 + x_2^2$$
(Pythagoras)

unit vector $\hat{\underline{x}} = \frac{1}{\|\underline{x}\|}\underline{x}$



Since $\|\alpha\underline{x}\| = |\alpha| . \|\underline{x}\|$

we have $\frac{1}{\|\underline{x}\|}.\underline{x}$ ($= \underline{y}$, say) is a unit vector

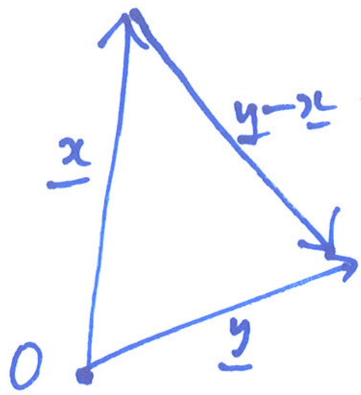
[$\|\alpha\underline{x}\| = \sqrt{\alpha^2 \sum x_i^2} = \sqrt{\alpha^2}\,\|\underline{x}\|$
$= |\alpha| . \|\underline{x}\|$.]

Distances are just lengths :

Terminology 6.6 If $\underline{x}, \underline{y} \in \mathbb{R}^n$, the *distance between* $\underline{x}$ and $\underline{y}$ is

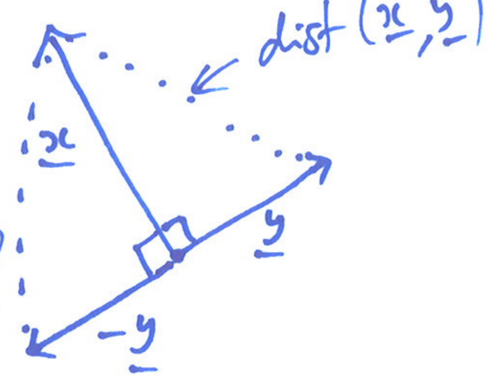
$$\text{dist}(\underline{x}, \underline{y}) = \|\underline{x} - \underline{y}\| = \|\underline{y} - \underline{x}\|$$

Example In $\mathbb{R}^2$



Example In $\mathbb{R}^2$, $\underline{x}$ and $\underline{y}$ are perpendicular *if and only if*

$$\underbrace{\|\underline{x} + \underline{y}\|}_{\text{dist}(\underline{x}, -\underline{y})} = \overbrace{\|\underline{x} - \underline{y}\|}^{\text{dist}(\underline{x}, \underline{y})}$$



Essentially, we use the same property as a *definition* of perpendicular vectors in $\mathbb{R}^n$.

This translate into $\underline{x} \cdot \underline{y} = 0$,

for the following reasons:

$$\mathrm{dist}(\underline{x}, \underline{y})^2 = \|\underline{x} - \underline{y}\|^2$$
$$= (\underline{x} - \underline{y}) \cdot (\underline{x} - \underline{y})$$
$$= \underline{x} \cdot \underline{x} - \underline{y} \cdot \underline{x} - \underline{x} \cdot \underline{y} + \underline{y} \cdot \underline{y}$$
$$= \underline{x} \cdot \underline{x} - 2\underline{x} \cdot \underline{y} + \underline{y} \cdot \underline{y}$$

$$\mathrm{dist}(\underline{x}, -\underline{y})^2 = \|\underline{x} + \underline{y}\|^2$$
$$= \underline{x} \cdot \underline{x} + 2\underline{x} \cdot \underline{y} + \underline{y} \cdot \underline{y}$$

So $\mathrm{dist}(\underline{x}, \underline{y}) = \mathrm{dist}(\underline{x}, -\underline{y})$ if & only if

$$-2\underline{x} \cdot \underline{y} = +2\underline{x} \cdot \underline{y}$$

if & only if $\underline{x} \cdot \underline{y} = 0$.

Definition 6.7 Two vectors $\underline{x}, \underline{y} \in \mathbb{R}^n$ are *perpendicular*, or *orthogonal*, to each other if $\underline{x} \cdot \underline{y} = 0$.

Theorem 6.8 $\underline{x}, \underline{y}$ are orthogonal if and only if

$$\|\underline{x} + \underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2$$

(*Pythagoras's theorem*!)

Proof $\|\underline{x} + \underline{y}\|^2 = \|\underline{x}\|^2 + 2\underline{x} \cdot \underline{y} + \|\underline{y}\|^2$ as above

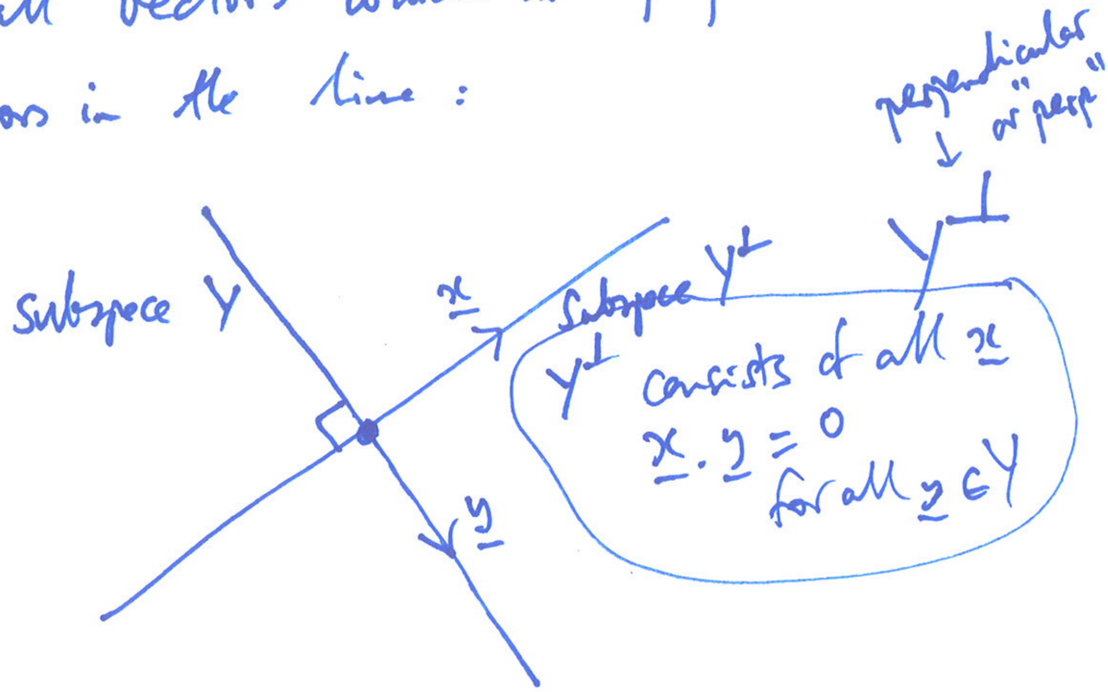
this $= \|\underline{x}\|^2 + \|\underline{y}\|^2$

if and only if $\underline{x} \cdot \underline{y} = 0$.

Done.

Section 6.2: Orthogonal complements

Example In $\mathbb{R}^2$, a line through the origin is a subspace of dimension 1: it has a perpendicular subspace consisting of all vectors which are perpendicular to vectors in the line:



We generalise this to $\mathbb{R}^n$:

Terminology 6.9 If Y is a subspace of $\mathbb{R}^n$, then:

- a vector $\underline{x} \in \mathbb{R}$ is <u>orthogonal to Y</u> if it is orthogonal to every vector $\underline{y} \in Y$.
- the set of <u>all</u> such $\underline{x}$ is called the <u>orthogonal complement of Y</u>, written $Y^\perp$ (called <u>Y perpendicular</u>, or <u>Y perp</u>).

Formally:

$$Y^\perp = \{ \underline{x} \in \mathbb{R}^n \mid \underline{x} \cdot \underline{y} = 0 \text{ for } \underline{\text{all}}\ \underline{y} \in Y \}.$$

Example In $\mathbb{R}^3$, let $Y = \{ \alpha (1, -3, 2)^T : \alpha \in \mathbb{R} \}$

Then $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in Y^\perp$ iff $\underline{x} \cdot \alpha \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} = 0$ for all α

iff $(x_1 \ x_2 \ x_3) \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} = 0$

iff $x_1 - 3x_2 + 2x_3 = 0$.

iff $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3\beta - 2\gamma \\ \beta \\ \gamma \end{pmatrix} \quad \beta, \gamma \in \mathbb{R}$.

$\Rightarrow Y^\perp = \left\{ \begin{pmatrix} 3\beta - 2\gamma \\ \beta \\ \gamma \end{pmatrix} : \beta, \gamma \in \mathbb{R} \right\}$ has dimension 2.

Theorem 6.11 If Y is a subspace of $\mathbb{R}^n$ then

- $Y^\perp$ is a subspace of $\mathbb{R}^n$
- If $\mathcal{S}$ is a spanning set for Y,

then $\underline{x} \in Y^\perp \iff \underline{x}.\underline{s} = 0$ for every $\underline{s} \in \mathcal{S}$.

($\Rightarrow$ is obvious)

Proof

- If $\underline{x}_1, \underline{x}_2 \in Y^\perp$ then $\underline{x}_1 . \underline{y} = 0$ for all $\underline{y} \in Y$

and $\underline{x}_2 . \underline{y} = 0$ for all $\underline{y} \in Y$

$\Rightarrow (\underline{x}_1 + \underline{x}_2) \cdot \underline{y} = \underline{x}_1 \cdot \underline{y} + \underline{x}_2 \cdot \underline{y} = 0 + 0 = 0$ for all $\underline{y} \in Y$

$\Rightarrow \underline{x}_1 + \underline{x}_2 \in Y^\perp$.

If α is a scalar & $\underline{x} \in Y^\perp$

then $\underline{x}.\underline{y} = 0 \quad \forall \underline{y} \in Y$

$\Rightarrow (\alpha \underline{x}) . \underline{y} = \alpha(\underline{x} . \underline{y}) = \alpha 0 = 0$ for all $\underline{y} \in Y$

$\Rightarrow \alpha \underline{x} \in Y^\perp$.

- If $\underline{x}.\underline{s} = 0$ for every $\underline{s} \in \mathcal{S}$

then every $\underline{y} \in Y$ can be written as a linear combination $\underline{y} = \sum_{j=1}^{k} \alpha_j \underline{s}_j$

where $\alpha_j \in \mathbb{R}$ and $\underline{s}_j \in \mathcal{S}$

$$\Rightarrow \underline{x} \cdot \underline{y} = \underline{x} \cdot \left(\sum_{j=1}^{k} \alpha_j \underline{s}_j \right)$$

$$= \sum_{j=1}^{k} \alpha_j \, \underbrace{(\underline{x} \cdot \underline{s}_j)}_{= 0 \text{ by assumption}}$$

$$= 0$$

This is true for <u>every</u> $\underline{y} \in Y$:

therefore $\underline{x} \in Y^{\perp}$ <u>Done</u>

<u>Remark</u> It's most efficient if the spanning set S is a <u>basis</u> of Y.

Theorem 6.12

If $A \in \mathbb{R}^{m\times n}$ then

$$N(A) = \text{col}(A^T)^{\perp}$$

(and similarly $N(A^T) = \text{col}(A)^{\perp}$.)

Proof $\quad \underline{x} \in N(A) \iff A\underline{x} = \underline{0} \quad$ (defn.)

$$\iff \underline{x}^T A^T = \underline{0}^T = (0,\dots,0)$$

$$\iff \underline{x}^T \underline{y} = 0 \text{ for every column } \underline{y} \text{ of } A^T.$$

$\underline{y}$ runs through a basis for $\text{col}(A^T)$

$$\iff \underline{x}\cdot\underline{y} = 0$$

$$\iff \underline{x} \in \text{col}(A^T)^{\perp}$$

Example $\quad A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 0 & -2 \\ 1 & 1 & -3 \end{pmatrix}$

- $A^T = \begin{pmatrix} 1 & 2 & 1 \\ -2 & 0 & 1 \\ 3 & -2 & -3 \end{pmatrix} \Rightarrow \text{col}(A^T) = \left\{\alpha\begin{pmatrix}1\\-2\\3\end{pmatrix} + \beta\begin{pmatrix}2\\0\\-2\end{pmatrix} + \gamma\begin{pmatrix}1\\1\\-3\end{pmatrix}\right\}$

- $\Rightarrow \cancel{\text{col}(A^T)^{\perp} = \left\{\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} : x_1\begin{pmatrix}1\\-2\\3\end{pmatrix} + x_2\begin{pmatrix}2\\0\\-2\end{pmatrix} + x_3\begin{pmatrix}1\\1\\-3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\right\}}$

- $N(A) = \left\{\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} : \begin{pmatrix} 1 & -2 & 3 \\ 2 & 0 & -2 \\ 1 & 1 & -3 \end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\right\}$

$$\bullet\ \mathrm{Col}(A^T)^\perp = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = 0 \right.$$

$$\left. \& \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} = 0 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -3 \end{pmatrix} \right\}.$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : \begin{array}{l} x_1 - 2x_2 + 3x_3 = 0 \\ 2x_1 \quad\quad - 2x_3 = 0 \\ x_1 + x_2 - 3x_3 = 0 \end{array} \right\}.$$

$$= N(A)$$

Section 6.3: Orthogonal sets

<u>Example</u> The standard basis vectors e_i in $\mathbb{R}^n$ are all <u>orthogonal to each other</u>.

We generalise this idea:

<u>Definition 6.13</u> A set $\{\underline{u}_1, \dots, \underline{u}_k\}$ of vectors in $\mathbb{R}^n$ is called an <u>orthogonal set</u> if

$$\underline{u}_i \cdot \underline{u}_j = 0 \quad \text{for all } i \neq j.$$

<u>Example 6.14</u>

If $\underline{u}_1 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$, $\underline{u}_2 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$, $\underline{u}_3 = \begin{pmatrix} -1 \\ -4 \\ 7 \end{pmatrix}$

then $\{\underline{u}_1, \underline{u}_2, \underline{u}_3\}$ is an orthogonal set.

<u>Proof</u>:
$$\underline{u}_1 \cdot \underline{u}_2 = -3 + 2 + 1 = 0$$
$$\underline{u}_1 \cdot \underline{u}_3 = -3 - 4 + 7 = 0$$
$$\underline{u}_2 \cdot \underline{u}_3 = 1 - 8 + 7 = 0.$$
<u>Done</u>.

Remarkable Result 6.15

If $\{\underline{u}_1, \ldots, \underline{u}_k\}$ is an orthogonal set of **non-zero** vectors, then $\underline{u}_1, \ldots, \underline{u}_k$ are linearly independent.

Proof. If $c_1 \underline{u}_1 + c_2 \underline{u}_2 + \cdots + c_k \underline{u}_k = \underline{0}$
for $c_1, \ldots, c_k \in \mathbb{R}$,

then $(c_1 \underline{u}_1 + \cdots + c_k \underline{u}_k) \cdot \underline{u}_1 = \underline{0} \cdot \underline{u}_1 = 0$

$$= c_1 (\underline{u}_1 \cdot \underline{u}_1) + c_2 \underbrace{\cancel{(\underline{u}_2 \cdot \underline{u}_1)}}_{=0} + \cdots + c_k \underbrace{\cancel{(\underline{u}_k \cdot \underline{u}_1)}}_{=0}$$

since $\underline{u}_i \cdot \underline{u}_j = 0$ if $i \neq j$

$$\Rightarrow c_1 \underbrace{(\underline{u}_1 \cdot \underline{u}_1)}_{= \|\underline{u}_1\|^2 \neq 0} = 0$$

$= \|\underline{u}_1\|^2 \neq 0$ since $\underline{u}_1 \neq \underline{0}$.

$\Rightarrow c_1 = 0$.

& similarly $c_j = 0$ for every j.

Hence $\underline{u}_1, \ldots \underline{u}_k$ are linearly independent.

<u>Terminology 6.16</u> An orthogonal <u>set</u> of <u>non-zero</u> vectors, which also <u>spans</u> a subspace H of $\mathbb{R}^n$, is therefore a <u>basis</u> for H, and is called an <u>orthogonal basis</u> for H.

<u>Example</u> $\{\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n\}$ form an orthogonal basis for $\mathbb{R}^n$.

- Computing coordinate of a vector with respect to an <u>orthogonal</u> basis is much easier than the general case (which requires Gaussian elimination):

<u>Theorem 6.17</u> Suppose $\{\underline{u}_1, \ldots, \underline{u}_k\}$ is an orthogonal basis for a subspace H of $\mathbb{R}^n$, & let $\underline{y} \in H$.

Then $\underline{y} = c_1\underline{u}_1 + c_2\underline{u}_2 + \cdots + c_k\underline{u}_k$

where $$c_j = \frac{\underline{y} \cdot \underline{u}_j}{\underline{u}_j \cdot \underline{u}_j}$$

Proof

$$\underline{y}\cdot\underline{u}_j = \sum_{i=1}^{k} c_i \underbrace{\underline{u}_i\cdot\underline{u}_j}_{=0 \text{ except when } i=j}$$

$$= c_j\, \underline{u}_j\cdot\underline{u}_j$$

& $\underline{u}_j \neq \underline{0} \Rightarrow \underline{u}_j\cdot\underline{u}_j \neq 0$

$$\Rightarrow c_j = \frac{\underline{y}\cdot\underline{u}_j}{\underline{u}_j\cdot\underline{u}_j} \quad \text{as required.}$$

Example 6.18

Express $\underline{y} = \begin{pmatrix} 6 \\ 1 \\ -8 \end{pmatrix}$ in terms of the orthogonal basis $\{\underline{u}_1, \underline{u}_2, \underline{u}_3\}$ of $\mathbb{R}^3$ given in Example 6.14.

Solution

$\underline{u}_1 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \underline{y} \cdot \underline{u}_1 = 18 + 1 - 8 = 11$

& $\underline{u}_1 \cdot \underline{u}_1 = 3^2 + 1^2 + 1^2 = 11$

$\underline{u}_2 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \Rightarrow \underline{y} \cdot \underline{u}_2 = -6 + 2 - 8 = -12$

$\underline{u}_2 \cdot \underline{u}_2 = (-1)^2 + 2^2 + 1^2 = 6$

$\underline{u}_3 = \begin{pmatrix} -1 \\ -4 \\ 7 \end{pmatrix} \Rightarrow \underline{y} \cdot \underline{u}_3 = -6 - 4 - 56 = -66$

$\underline{u}_3 \cdot \underline{u}_3 = 1^2 + 4^2 + 7^2 = 66.$

So $\underline{y} = c_1 \underline{u}_1 + c_2 \underline{u}_2 + c_3 \underline{u}_3$

where $c_1 = \dfrac{\underline{y} \cdot \underline{u}_1}{\underline{u}_1 \cdot \underline{u}_1} = \dfrac{11}{11} = 1,$

& $c_2 = \dfrac{-12}{6} = -2$

& $c_3 = \dfrac{-66}{66} = -1$

i.e. $\underline{y} = \underline{u}_1 - 2\underline{u}_2 - \underline{u}_3$

Check: $\begin{pmatrix} 6 \\ +1 \\ -8 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - 2\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} -1 \\ -4 \\ 7 \end{pmatrix}$ ✓.

Section 6.4 : Orthonormal sets

Given an <u>orthogonal</u> set $\{\underline{u}_1, \dots, \underline{u}_k\}$ we can <u>normalize</u> each vector $\underline{u}_j$ to a unit vector $\hat{\underline{u}}_j = \frac{1}{\|\underline{u}_j\|}\underline{u}_j$. The resulting set $\{\hat{\underline{u}}_1, \dots, \hat{\underline{u}}_k\}$ is called <u>orthonormal</u>.

<u>Terminology 6.19</u> $\{\underline{v}_1, \dots, \underline{v}_k\} \subseteq \mathbb{R}^n$ is an <u>orthonormal set</u> if $\underline{v}_i \cdot \underline{v}_j = \delta_{ij}$

where $$\delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j. \end{cases}$$

An orthonormal set which is a basis of H is called an <u>orthonormal basis</u> of H.

<u>Example</u> $\{\underline{e}_1, \dots, \underline{e}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ because $\underline{e}_i \cdot \underline{e}_i = 1$.

Example 6.21 If $\underline{u}_1 = \begin{pmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}$, $\underline{u}_2 = \begin{pmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$, $\underline{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ then $\{\underline{u}_1, \underline{u}_2, \underline{u}_3\}$ is an orthonormal basis of $\mathbb{R}^3$.

Why?

$$\underline{u}_1 \cdot \underline{u}_1 = (2^2 + 1^2 + 1^2)/6 = 1$$

$$\underline{u}_2 \cdot \underline{u}_2 = 1$$

$$\underline{u}_3 \cdot \underline{u}_3 = 1$$

$$\underline{u}_1 \cdot \underline{u}_2 = \frac{1}{\sqrt{6}\sqrt{3}}(-2 + 1 + 1) = 0$$

$$\underline{u}_1 \cdot \underline{u}_3 = \frac{1}{\sqrt{6}\sqrt{2}}(0 - 1 + 1) = 0$$

$$\underline{u}_2 \cdot \underline{u}_3 = \frac{1}{\sqrt{3}\sqrt{2}}(0 - 1 + 1) = 0.$$

Theorem 6.22 Let $U \in \mathbb{R}^{m \times n}$.

The columns of U are orthonormal $\iff$

$$U^T U = I_n.$$

Proof The (i,j) entry of $U^T U$ is

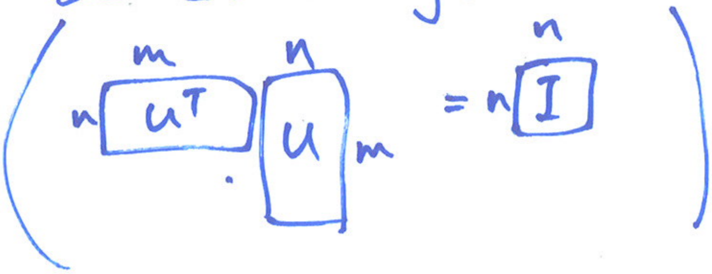
$$\underline{x}^T \underline{y} \quad \text{where } \underline{x}^T = i\text{'th row of } U^T$$
$$\& \quad \underline{y} = j\text{'th column of } U.$$
$$(\underline{x} = i\text{'th column of } U)$$

$$= \underline{x} \cdot \underline{y}$$

If the columns of U are orthonormal

then $\underline{x} \cdot \underline{y} = \begin{cases} 0 & \text{if } \underline{x} \neq \underline{y} \\ 1 & \text{if } \underline{x} = \underline{y}. \end{cases} \Rightarrow U^T U = I_n$

& conversely. Done.



Remarkable consequence (for square matrices)

If $U \in \mathbb{R}^{n \times n}$ & the columns of U are orthonormal, then the rows of U are orthonormal.

Proof If columns of U are orthonormal, then

$U^T U = I \Rightarrow U^T = U^{-1}$ (because U is square)

$\Rightarrow U U^T = I \Rightarrow$ columns of U^T are orthonormal

$\Rightarrow$ rows of U are orthonormal.

Theorem 6.23 If the columns of $U \left(\in \mathbb{R}^{m \times n}\right)$ are orthonormal, then

- $(U\underline{x}) \cdot (U\underline{y}) = \underline{x} \cdot \underline{y}$

(Hence
- $\| U\underline{x} \| = \|x\|$,
- $(U\underline{x}) \cdot (U\underline{y}) = 0 \iff \underline{x} \cdot \underline{y} = 0$.)

Proof $(U\underline{x}) \underset{\substack{\uparrow \\ \text{dot product}}}{\cdot} (U\underline{y}) = (U\underline{x})^T (U\underline{y})$ (matrix multiplication)

$= (\underline{x}^T U^T)(U\underline{y})$

$= \underline{x}^T (U^T U) \underline{y}$

$= \underline{x}^T I \underline{y}$

$= \underline{x}^T \underline{y}$

$= \underline{x} \cdot \underline{y}$ Done.

Terminology 6.24

If Q is a square matrix and $Q^T Q = I$, then Q is called an orthogonal matrix.

(**Warning**: it's easy to confuse this meaning of the word "orthogonal" with other meanings we have already used.)

Observation 6.25

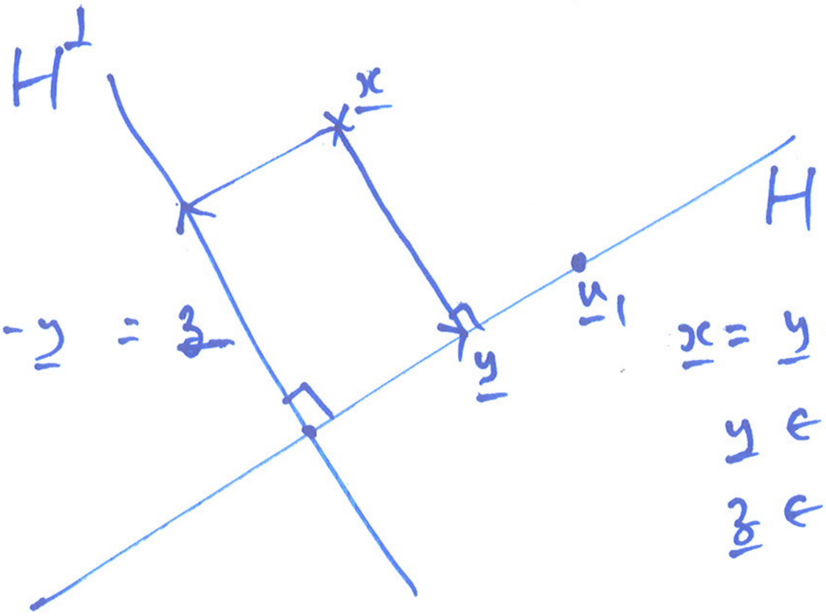
If Q $(\in \mathbb{R}^{n \times n})$ is an orthogonal matrix, then

- Q is invertible and $Q^{-1} = Q^T$.
- if $\{\underline{v}_1, \dots, \underline{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$, then so is $\{Q\underline{v}_1, \dots, Q\underline{v}_n\}$.

Proof $Q\underline{v}_i \cdot Q\underline{v}_j = \underline{v}_i \cdot \underline{v}_j$ by 6.23

$= \delta_{ij}$ Done.

Section 6.5: Orthogonal projections



$\underline{x} - \underline{y} = \underline{z}$

$\underline{x} = \underline{y} + \underline{z}$

$\underline{y} \in H$

$\underline{z} \in H^\perp$.

$$\underline{y} = \left(\frac{\underline{x}\cdot\underline{u}_1}{\underline{u}_1\cdot\underline{u}_1}\right)\underline{u}_1 \quad \text{and} \quad \underline{z} = \underline{x} - \underline{y}.$$

$$\begin{aligned}\underline{z}\cdot\underline{u}_1 &= \underline{x}\cdot\underline{u}_1 - \underline{y}\cdot\underline{u}_1 \\ &= \underline{x}\cdot\underline{u}_1 - \left(\frac{\underline{x}\cdot\underline{u}_1}{\underline{u}_1\cdot\underline{u}_1}\right)\underline{u}_1\cdot\underline{u}_1 \\ &= \underline{x}\cdot\underline{u}_1 - \underline{x}\cdot\underline{u}_1 = 0.\end{aligned}$$

So $\underline{z} \in H^\perp$.

More generally, in $\mathbb{R}^n$ we have:

<u>Theorem 6.26</u> (Orthogonal Decomposition Theorem)

If H is a subspace of $\mathbb{R}^n$, and $\underline{x} \in \mathbb{R}^n$ then $\underline{x}$ can be uniquely expressed as

$$\underline{x} = \underline{y} + \underline{z} \quad \text{with} \quad \underline{y} \in H, \ \underline{z} \in H^\perp.$$

Moreover, if $\{\underline{u}_1, \ldots, \underline{u}_k\}$ is an orthogonal basis for H, then

$$\underline{y} = \left(\frac{\underline{x} \cdot \underline{u}_1}{\underline{u}_1 \cdot \underline{u}_1}\right)\underline{u}_1 + \cdots + \left(\frac{\underline{x} \cdot \underline{u}_k}{\underline{u}_k \cdot \underline{u}_k}\right)\underline{u}_k .$$

<u>Proof</u> • $\underline{y}$ is a linear combination of $\underline{u}_1, \ldots, \underline{u}_k$

$\Rightarrow \underline{y} \in H.$

• To check that $\underline{x} - \underline{y} \in H^\perp$ we need to show $(\underline{x} - \underline{y}) \cdot \underline{u}_i = 0$ for each i:

Now $$\underline{y} \cdot \underline{u}_i = \left(\frac{\underline{x} \cdot \underline{u}_1}{\underline{u}_1 \cdot \underline{u}_1}\right)\underbrace{\underline{u}_1 \cdot \underline{u}_i}_{=0 \text{ unless } i=1} + \cdots$$

$$= \left(\frac{\underline{x} \cdot \underline{u}_i}{\underline{u}_i \cdot \underline{u}_i}\right)\underline{u}_i \cdot \underline{u}_i + 0\text{'s}$$

$$= \underline{x} \cdot \underline{u}_i$$

$\Rightarrow (\underline{x} - \underline{y}) \cdot \underline{u}_i = 0$ as required.

• <u>Uniqueness</u>: if $\underline{x} = \underline{y}_1 + \underline{z}_1$ ($\underline{y}_1 \in H$, $\underline{z}_1 \in H^\perp$)

and $\underline{x} = \underline{y}_2 + \underline{z}_2$ ($\underline{y}_2 \in H$, $\underline{z}_2 \in H^\perp$)

$$\Rightarrow \underline{y}_1 + \underline{z}_1 = \underline{y}_2 + \underline{z}_2$$

$$\Rightarrow \underbrace{\underline{y}_1 - \underline{y}_2}_{\in H} = \underbrace{\underline{z}_2 - \underline{z}_1}_{\in H^\perp} \in H \cap H^\perp.$$

& if $\underline{y}_1 - \underline{y}_2 = \sum_{i=1}^{k} c_i \underline{u}_i \in H \cap H^\perp$

then $\left(\sum_{i=1}^{k} c_i \underline{u}_i\right) \cdot \underline{u}_j = c_j \underline{u}_j \cdot \underline{u}_j = 0$

$\Rightarrow c_j = 0$ for all j.

$$\Rightarrow \underline{y}_1 - \underline{y}_2 = \underline{0} \quad \Rightarrow \underline{z}_1 - \underline{z}_2 = \underline{0}$$

$\Rightarrow \underline{y}_1 = \underline{y}_2$ and $\underline{z}_1 = \underline{z}_2$ <u>as required.</u>

<u>Terminology</u> $\underline{y}$ is the <u>orthogonal projection</u> of $\underline{x}$ onto H, written $\text{proj}_H(\underline{x})$.

Now $\text{proj}_H(\underline{x})$ is the closest point to $\underline{x}$ that lies in H:

Theorem 6.27 (Best Approximation Theorem)

If H is a subspace of $\mathbb{R}^n$, and $\underline{x} \in \mathbb{R}^n$ then $\text{proj}_H(\underline{x})$ is the closest point of H to $\underline{x}$ in the sense that:

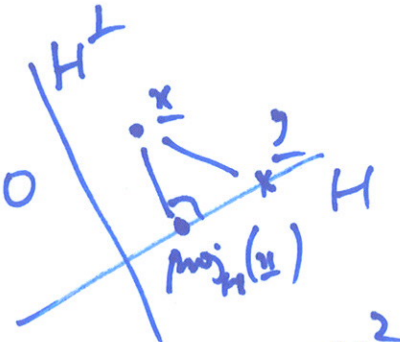
$$\| \underline{x} - \text{proj}_H(\underline{x}) \| \leq \| \underline{x} - \underline{y} \|$$

for all $\underline{y} \in H$.

Proof $\quad \underline{y} - \text{proj}_H(\underline{x}) \in H$

$$\left(\Rightarrow \underline{y} - \text{proj}_H(\underline{x}) = \sum_{i=1}^{k} a_i \underline{u}_i \right)$$

Now $\underline{x} - \text{proj}_H \underline{x} \in H^\perp$



$$\Rightarrow (\underline{x} - \text{proj}_H \underline{x}) . (\underline{y} - \text{proj}_H \underline{x}) = 0$$

$\Rightarrow$ by Pythagoras

$$\| \underline{x} - \underline{y} \|^2 = \| \underline{x} - \text{proj}_H(\underline{x}) \|^2 + \underbrace{\| \underline{y} - \text{proj}_H(\underline{x}) \|^2}_{\geq 0}$$

$$\geq \| \underline{x} - \text{proj}_H(\underline{x}) \|^2$$

$\Rightarrow \| \underline{x} - \underline{y} \| \geq \| \underline{x} - \text{proj}_H(\underline{x}) \|$ since lengths ≥ 0.

Done.

Theorem 6.28 If H is a subspace of $\mathbb{R}^n$, then

- $(H^\perp)^\perp = H$.
- $\dim H + \dim H^\perp = n$.

Proof Exercise 4 on CW 10.

Section 6.6 : Gram-Schmidt process

This is a process for obtaining an orthogonal basis for a subspace H of $\mathbb{R}^n$ (from an arbitrary basis).

<u>Idea</u>: Given $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k$ a basis for H,

set $\underline{v}_1 = \underline{x}_1$

Now $\underline{x}_2 = \mathrm{proj}_{\mathrm{Span}(\underline{v}_1)}(\underline{x}_2) + \begin{pmatrix}\text{something} \\ \perp \underline{v}_1\end{pmatrix}$

$$= \left(\frac{\underline{x}_2 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1}\right)\underline{v}_1 + \cdots$$

So set $\underline{v}_2 = \underline{x}_2 - \left(\dfrac{\underline{x}_2 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1}\right)\underline{v}_1$

then $\underline{v}_3 = \underline{x}_3 - \left(\dfrac{\underline{x}_3 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1}\right)\underline{v}_1 - \left(\dfrac{\underline{x}_3 \cdot \underline{v}_2}{\underline{v}_2 \cdot \underline{v}_2}\right)\underline{v}_2$

etc.

So we've arranged that each $\underline{v}_i$ is orthogonal to $\underline{v}_1, \dots, \underline{v}_{i-1}$.

Theorem 6.29 If $\{\underline{x}_1, \ldots, \underline{x}_k\}$ is a basis for H, a subspace of $\mathbb{R}^n$, then if we define

$$\underline{v}_1 = \underline{x}_1$$

$$\underline{v}_j = \underline{x}_j - \sum_{i=1}^{j-1} \left(\frac{\underline{x}_j \cdot \underline{v}_i}{\underline{v}_i \cdot \underline{v}_i} \right) \underline{v}_i$$

we have that $\{\underline{v}_1, \ldots, \underline{v}_k\}$ is an orthogonal basis for H.

Moreover for each j we have
$\mathrm{Span}(\underline{v}_1, \ldots, \underline{v}_j) = \mathrm{Span}(\underline{x}_1, \ldots, \underline{x}_j)$.

Proof Orthogonal decomposition theorem

$\Rightarrow$ $\underline{v}_j \cdot \underline{v}_i = 0$ for all $i < j$

(by induction on j).

Write $H_k = \mathrm{Span}(\underline{x}_1, \dots, \underline{x}_k)$.

- Then $\underline{v}_1 = \underline{x}_1$ so

  $$\mathrm{Span}(\underline{v}_1) = \mathrm{Span}(\underline{x}_1) = H_1.$$

- For $j > 1$ suppose we have proved

  $$\mathrm{Span}(\underline{v}_1, \dots, \underline{v}_{j-1}) = H_{j-1}.$$

  Then $\underline{v}_j = \underline{x}_j + (\text{lin. comb. of } \underline{v}_1, \dots, \underline{v}_{j-1})$

  & $\underline{x}_j = \underline{v}_j + (\;\text{''} \quad \text{''} \quad \text{''}\;)$

  So $\mathrm{Span}(\underline{v}_1, \dots, \underline{v}_{j-1}, \underline{v}_j) =$

  $\mathrm{Span}(\underline{v}_1, \dots, \underline{v}_{j-1}, \underline{x}_j) =$

  $\mathrm{Span}(\underline{x}_1, \dots, \underline{x}_{j-1}, \underline{x}_j)$ [by induction]

  $= H_j$.

- So we have proved (by induction on j) that for all j,

  $$\mathrm{Span}(\underline{x}_1, \dots, \underline{x}_j) = \mathrm{Span}(\underline{v}_1, \dots, \underline{v}_j).$$

Now we prove (again by induction on j)
that $\underline{v}_1, \ldots, \underline{v}_j$ are orthogonal and non-zero.

- $\underline{v}_1 = \underline{x}_1 \neq \underline{0}$ because $\underline{x}_1$ lies in a basis.
- If $\underline{v}_1, \ldots, \underline{v}_{j-1}$ are orthogonal and non-zero,
then the orthogonal decomposition theorem
implies $\underline{v}_j \in H_{j-1}^{\perp}$. ($\Rightarrow \underline{v}_j \cdot \underline{v}_i = 0$ for all $i < j$)
Moreover $\underline{v}_j \neq \underline{0}$ because if $\underline{v}_j = \underline{0}$
then $\underline{x}_j \in H_{j-1}$, contradicting the
fact that $\underline{x}_1, \ldots, \underline{x}_j$ are linearly independent.
- Hence $\underline{v}_1, \ldots, \underline{v}_k$ are orthogonal and
non-zero. $\Rightarrow$ linearly independent.
$\Rightarrow$ they form a basis for H.

Example 6.31 (Gram-Schmidt in action)

Let $\underline{x}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$, $\underline{x}_2 = \begin{pmatrix}0\\1\\1\\2\end{pmatrix}$, $\underline{x}_3 = \begin{pmatrix}0\\0\\2\\6\end{pmatrix}$

& let $H = \text{Span}(\underline{x}_1, \underline{x}_2, \underline{x}_3)$.

Apply Gram-Schmidt to find an <u>orthogonal</u> basis for H.

(<u>N.B.</u> $\underline{x}_1, \underline{x}_2, \underline{x}_3$ is linearly independent so forms a <u>basis</u> for H.)

<u>Step 1</u> Set $\underline{v}_1 = \underline{x}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$.

<u>Step 2</u> Set $\underline{v}_2 = \underline{x}_2 - \left(\frac{\underline{x}_2 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1}\right)\underline{v}_1$

$$= \begin{pmatrix}0\\1\\1\\2\end{pmatrix} - \frac{4}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \begin{pmatrix}-1\\0\\0\\1\end{pmatrix}.$$

<u>Step 3</u> Set $\underline{v}_3 = \underline{x}_3 - \left(\frac{\underline{x}_3 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1}\right)\underline{v}_1 - \left(\frac{\underline{x}_3 \cdot \underline{v}_2}{\underline{v}_2 \cdot \underline{v}_2}\right)\underline{v}_2$

$$= \begin{pmatrix}0\\0\\2\\6\end{pmatrix} - \frac{8}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{6}{2}\begin{pmatrix}-1\\0\\0\\1\end{pmatrix}$$

$$= \begin{pmatrix}0\\0\\2\\6\end{pmatrix} - \begin{pmatrix}2\\2\\2\\2\end{pmatrix} + \begin{pmatrix}3\\0\\0\\-3\end{pmatrix} = \begin{pmatrix}1\\-2\\0\\1\end{pmatrix}.$$

So $\{\underline{v}_1, \underline{v}_2, \underline{v}_3\} = \left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}-1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}1\\-2\\0\\1\end{pmatrix}\right\}$

is an orthogonal basis for H.

Section 6.7: Least squares problems

An important application of linear algebra to statistics is trying to fit a straight line to some data. Suppose we have a set of data points (x_i, y_i), and we want to find the best possible m and c such that

$$y = mx + c$$

is a good approximation to

$$y_i = mx_i + c$$

We can think of this as an (probably inconsistent) overdetermined linear system of equations we're trying to solve for two unknowns m and c.

More generally suppose we have an overdetermined system

$$Ax = b$$

where $A \in \mathbb{R}^{m \times n}$ (with $m > n$) and $b \in \mathbb{R}^m$. We want to find $x \in \mathbb{R}^n$ such that

$$\| Ax - b \|$$

is as small as possible.

__Terminology__ Such an $\underline{x}$ is called a __least squares solution__ of $A\underline{x} = \underline{b}$.
(N.B. it is __not__ a __solution__ of $A\underline{x} = \underline{b}$!)

__Method__ We know $A\underline{x} \in \mathrm{col}(A)$
so we want to find the __closest__ point in $\mathrm{col}(A)$ to the vector $\underline{b}$.

- This is just the vector $\underline{c} = \mathrm{proj}_{\mathrm{col}A}(\underline{b})$.
- Then we solve $A\underline{x} = \underline{c}$ instead.

~~__But__ this is typically a vastly overdetermined system so it would be very inefficient to solve this directly.~~

The __trick__ is to multiply by A^T.

__Why?__ By the orthogonal decomposition theorem

$\underline{b} = \underline{c} + (\underline{b} - \underline{c})$ where $\underline{c} = \mathrm{col}(A)$ & $\underline{b} - \underline{c} \in \mathrm{col}(A)^{\perp}$

Also $\mathrm{col}(A)^{\perp} = N(A^T)$ (Theorem 6.12)

$$\Rightarrow A^T(\underline{b} - \underline{c}) = \underline{0}.$$

So now we have

$$Ax = c$$

$$\Rightarrow A^T A x = A^T c$$

$$= A^T b$$

$$\boxed{\begin{aligned} A^T &\in \mathbb{R}^{n \times m} \\ A^T b &\in \mathbb{R}^{n} \\ A^T A &\in \mathbb{R}^{n \times n} \end{aligned}}$$

This new system of equations is called the system of <u>normal equations</u> for $Ax = b$.

In fact:

<u>Theorem 6.33</u> If $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ (where $m > n$), then the set of <u>least squares solutions</u> of $Ax = b$ is the same as the set of <u>solutions</u> of $A^T A x = A^T b$.

<u>Proof</u> We've proved that every L.S.S. of $Ax = b$ also satisfies $A^T A x = A^T b$.

<u>Conversely</u>, if $A^T A x = A^T b$

then $A^T (Ax - b) = 0$

so $Ax - b \in N(A^T) = \mathrm{col}(A)^{\perp}$.

$$\Rightarrow \quad b = \underset{\substack{\uparrow \\ \text{in col}A}}{Ax} + \underset{\substack{\uparrow \\ \text{in }(\text{col}A)^\perp}}{(b - Ax)}$$

<u>is the</u> orthogonal decomposition of b

$$\Rightarrow \quad A\underline{x} = \text{proj}_{\text{col}A}(b) = \underline{c}, \text{ as required.}$$

<u>Example 6.34</u> Find the least squares solution(s) of the inconsistent system

$$\underbrace{\begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} \leftarrow \underline{b}.$$

<u>Solution</u> Compute $A^TA =$

$$\begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}.$$

$$\& \ A^Tb = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}.$$

So normal equations are ($A^TAx = A^Tb$, i.e.)

$$\begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

Gaussian elimination: $\left(\begin{array}{cc|c} 17 & 1 & 19 \\ 1 & 5 & 11 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 5 & 11 \\ 17 & 1 & 19 \end{array}\right)$

$$\rightarrow \left(\begin{array}{cc|c} 1 & 5 & 11 \\ 0 & -84 & \end{array}\right)$$

$\ldots\ldots$

$\Rightarrow \quad x_1 = 1, \quad x_2 = 2.$

If A^TA is invertible we can use the inverse matrix to solve the normal equations:

$$A^TA x = A^T b$$

$$\Rightarrow \quad x = (A^TA)^{-1} A^T b.$$

(Theorem 6.35)

[Example ctd.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

$$= \frac{1}{84} \begin{pmatrix} 5 & -1 \\ -1 & 17 \end{pmatrix} \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

$$= \frac{1}{84} \begin{pmatrix} 84 \\ 168 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

]