Def.: The dot product of $x, y \in \mathbb{R}^n$ is

$$x^T y = x_1 y_1 + \dots + x_n y_n$$
$$(x \cdot y)$$

this is an example of an inner product

- $x^T y = y^T x$

  $(x+y)^T z = x^T z + y^T z$

  $(c\vec{x})^T \vec{y} = c(\vec{x}^T \vec{y})$

- $\forall x \in \mathbb{R}^n$, $x^T x \geq 0$.

  $x^T x = 0 \iff x = \vec{0}$.

Def.: The <u>length</u> or <u>norm</u> of $\vec{x} \in \mathbb{R}^n$ is

$$\|\vec{x}\| = \sqrt{x^T x} = \sqrt{x_1^2 + \dots + x_n^2}$$

Def.: the distance between $\vec{x}, \vec{y} \in \mathbb{R}^n$ is

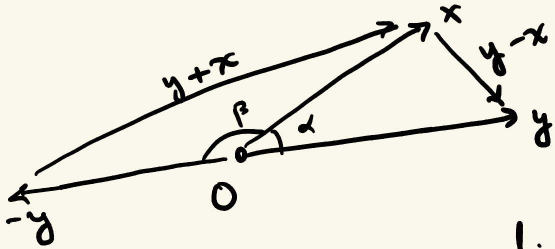
$$d(x, y) = \|\vec{x} - \vec{y}\|$$

Def.: A <u>unit vector</u> is a vector $\vec{x}$ s.t. $\|\vec{x}\| = 1$.

Note: $\frac{x}{\|x\|}$ is always a unit vector $(\vec{x} \neq \vec{0})$.

Def. $\vec{x}, \vec{y}$ are orthogonal or perpendicular if

$$x^T y = 0.$$



x and y are perpendicular iff

$$\|y+x\| = \|y-x\|$$

$$\iff (y+x)^T(y+x) = (y-x)^T(y-x)$$

$$\iff \cancel{y^Ty} + 2x^Ty + \cancel{x^Tx} = \cancel{y^Ty} - 2x^Ty + \cancel{x^Tx}$$

$$\iff 4x^Ty = 0 \iff x^Ty = 0.$$

Note: $\vec{0}^T x = 0 \ \forall \vec{x}$, so 0 is orthogonal to every vector.

Def.: Let $W \subseteq \mathbb{R}^n$ be a subspace. The <u>orthogonal complement</u> of W is

$$W^\perp = \{ v \in \mathbb{R}^n \mid v^T w = 0 \ \forall w \in W \}$$

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

Prop: A matrix. Let $W = \mathrm{Ran}\,A$ ($\mathrm{col}\,A$). Then

$$W^\perp = \ker(A^T).$$

Pf.: $A = \begin{pmatrix} v_1 & \cdots & v_m \end{pmatrix}$

Note that x is perpendicular to W iff its perpendicular to a spanning list of W.

$$A^T = \begin{pmatrix} v_1^T \\ \vdots \\ v_m^T \end{pmatrix}$$

$$A^T \vec{x} = \begin{pmatrix} v_1^T x \\ \vdots \\ v_m^T x \end{pmatrix} = \vec{0} \quad \text{iff} \quad x \perp v_i$$

for each v_i
$i = 1, \dots, m$.

$\square$

Prop: Let $W \subseteq \mathbb{R}^n$ be a subspace.
Then $\dim W + \dim W^\perp = n$.

Pf: Let $v_1, \dots, v_m$ be a basis for W.
Let $A = (v_1 \ \dots \ v_m)$.

So $\operatorname{Ran} A = W$.

Rank-nullity $\Rightarrow$ $\underbrace{\dim \operatorname{Ran} A^T}_{\substack{\text{rank thm} \\ = \dim \operatorname{Ran} A}} + \dim \ker A^T = n$

$\Rightarrow \dim W + \dim W^\perp = n$. $\square$

Prop: $(W^\perp)^\perp = W$.

If $x \in W$, and $y \in W^\perp$, then $x^T y = 0$, so
$x \in (W^\perp)^\perp$, so $W \subseteq (W^\perp)^\perp$.

$\dim W^\perp + \dim (W^\perp)^\perp = n$ $\qquad \dim W + \dim W^\perp = n$

$\dim (W^\perp)^\perp = n - \dim W^\perp$
$= \dim W$

So W is a full-dimension subspace of $(W^\perp)^\perp$.

$\square$

Thm (orthogonal decomposition)

Let W be a subspace of $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$.

Then x can be written uniquely as

$$x = x_W + x_{W^\perp}, \text{ where } x_W \in W, x_{W^\perp} \in W^\perp.$$

orthogonal decomp.

(i.e. $W, W^\perp$ is a basis of subspaces for $\mathbb{R}^n$)

Pf: lin indp:

Suppose $\vec{u} + \vec{v} = \vec{0}$, $\vec{u} \in W$, $\vec{v} \in W^\perp$.

Then $\vec{u} = -\vec{v} \in W^\perp$.

So $\vec{u} \in W^\perp$ and $\vec{u} \in W$.

$\implies \vec{u}^T\vec{u} = 0 \implies \vec{u} = 0 \implies \vec{v} = 0$.

spanning:

Let $v_1, \dots, v_m$ be a basis of W.

$v_{m+1}, \dots, v_n$ basis of $W^\perp$.

Then $v_1, v_2, \dots, v_n$ are lin indp, so they form a basis for $\mathbb{R}^n$. Any $\vec{x}$ can be written

$$\vec{x} = \underbrace{c_1v_1 + \dots + c_mv_m}_{\in W} + \underbrace{c_{m+1}v_{m+1} + \dots + c_nv_n}_{\in W^\perp}$$

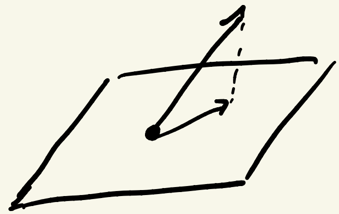
$\square$

$\operatorname{Ran}(A)^\perp = \ker(A^T)$

$\operatorname{Ran}(A^T)^\perp = \ker(A)$

$\operatorname{Ran}(A^T) = \ker(A)^\perp$

$\operatorname{Ran}(A) = \ker(A^T)^\perp$



<u>Thm:</u> $A = m \times n$. $W = \operatorname{Ran}(A)$, $x \in \mathbb{R}^m$.

Then
$$A^T A \vec{c} = A^T \vec{x} \quad (*)$$
is consistent in c, and $x_W = Ac$ for any sol'n c.

<u>Pf:</u> Let $x = x_W + x_{W^\perp}$ be the orthogonal decomp.

$x_W \in \operatorname{Ran}(A)$ by definition, so $A\vec{c} = x_W$ for some $\vec{c}$.

$x - x_W \in \operatorname{Ran}(A)^\perp = W^\perp = \ker(A^T)$, so

$$A^T(x - x_W) = \vec{0}$$

$$\iff A^T(\vec{x} - A\vec{c}) = \vec{0}$$

$$\iff A^T A \vec{c} = A^T \vec{x}.$$

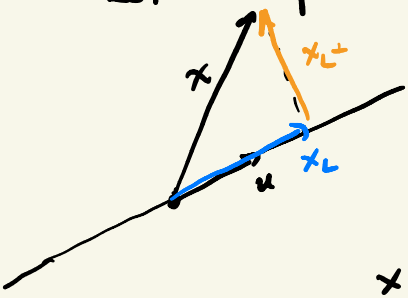
so $(*)$ is consistent.

And if $\vec{c}$ satisfies $(*)$ then $x - A\vec{c} \in W^\perp$, so

$A\vec{c} = x_W$ by uniqueness. $\square$

Ex: (projection onto a line)

Let $L = \text{span}(u)$. $u \in \mathbb{R}^n$.



c is 1×1 (a scalar)

$$u^T u c = u^T x$$

$$c = \frac{u^T x}{u^T u} = \frac{u \cdot x}{u \cdot u}$$

$$x_L = cu = \frac{u \cdot x}{u \cdot u} u = \frac{u \cdot x}{\|u\|^2} u$$

$$= (u \cdot x) \frac{u}{\|u\|^2}$$

Prop: Let W be a subspace of $\mathbb{R}^n$. Define $T: \mathbb{R}^n \to \mathbb{R}^n$ by $T(x) = x_W$. Then

1. T is linear
2. $T(x) = x \iff x \in W$
3. $T(x) = 0 \iff x \in W^\perp$
4. $T \circ T = T$
5. $\text{Ran}\, T = W$.

Pf: 1) Let $x, y \in \mathbb{R}^n$.

$$x = x_W + x_{W^\perp}, \quad y = y_W + y_{W^\perp}$$

$$x + y = \underbrace{(x_W + y_W)}_{\in W} + \underbrace{(x_{W^\perp} + y_{W^\perp})}_{\in W^\perp}$$

So $T(x+y) = x_W + y_W$

similar for scaling

2) If $x \in W$, then $x = \underset{\in W}{x} + \underset{\in W^\perp}{0}$

$x_W = x \Rightarrow T(x) = x$.

If $T(x) = x$, then $x_W = x$, so $x_{W^\perp} = 0$

$\Rightarrow x \in W$.

3) If $T(x) = 0$, then $x = \underset{\in W^\perp}{0} + x_{W^\perp}$

so $x \in W^\perp$.

If $x \in W^\perp$, then $x = \underset{\in W}{0} + \underset{\in W^\perp}{x} \Rightarrow T(x) = 0$.

4) $x = x_W + x_{W^\perp}$. $T(x) = x_W$

$x_W = \underset{\in W}{x_W} + \underset{\in W^\perp}{0} \Rightarrow T^2(x) = T(x_W) = x_W$.

$\Rightarrow T \circ T = T$.

5) $T(x) \in W \quad \forall x \in \mathbb{R}^n$, so $\operatorname{Ran} T \subseteq W$.

$T(x) = x$ for any $x \in W$, so $\operatorname{Ran} T = W$.

Thm (Cauchy-Schwarz) (skip?)

$$|x^T y| \le \|x\| \|y\| \quad \forall x, y \in \mathbb{R}^n$$

Pf: If $y = \vec{0}$, done. Suppose $y \neq \vec{0}$.

$t \in \mathbb{R}$: $0 \le \|\vec{x} - t\vec{y}\|^2 = (x - ty)^T (x - ty)$

$$= \|x\|^2 - 2t(x^T y) + t^2 \|y\|^2$$

This is a quadratic in t minimized at

$$t = \frac{x^T y}{\|y\|^2}$$

$$\Rightarrow 0 \le \|x\|^2 - 2\frac{(x^T y)^2}{\|y\|^2} + \frac{(x^T y)^2}{\|y\|^2}$$

$$\Rightarrow 0 \le \|x\|^2 \|y\|^2 - (x^T y)^2$$

$$\Rightarrow (x^T y)^2 \le \|x\|^2 \|y\|^2$$

$$\Rightarrow |x^T y| \le \|x\| \|y\|. \quad \square$$

$\implies \|x+y\| \leq \|x\| + \|y\|.$ $\square$

Prop: Pythagorean thm

If $x^T y = 0$ (x, y orthogonal), then

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Pf: $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2x^T y.$

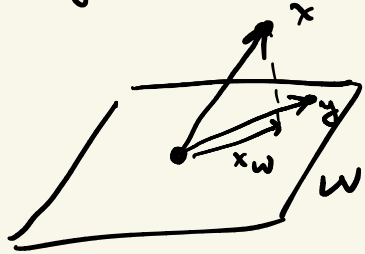
$x^T y = 0.$ $\square$

Thm Let W be a subspace of $\mathbb{R}^n$, let $x \in \mathbb{R}^n$.

The pt in W closest to W is x_W.

i.e. $\|x-y\| \geq \|x-x_W\| \ \forall y \in W$.

And if $\|x-y\| = \|x-x_W\|$, then $y = x_W$.



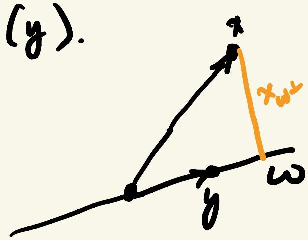
Pf.: $x - y = \underbrace{x - x_W}_{\in W^\perp} + \underbrace{x_W - y}_{\in W}$

So $\|x-y\|^2 = \|x-x_W\|^2 + \|x_W - y\|^2$
$\geq \|x-x_W\|^2$.

And if $\|x-y\| = \|x-x_W\|$, we see $\|x_W - y\| = 0$,

so $x_W = y$. $\square$

Let $x, y \in \mathbb{R}^n$, $y \neq \vec{0}$. Let $W = \text{span}(y)$.



$$x_W = \frac{x^T y}{\|y\|^2} y$$

$$\|x_{W^\perp}\|^2 = \|x - x_W\|^2$$

$$= \left\|x - \frac{x^T y}{\|y\|^2} y\right\|^2$$

$$= \left(x - \frac{x^T y}{\|y\|^2} y\right)^T \left(x - \frac{x^T y}{\|y\|^2} y\right)^T$$

$$= \|x\|^2 - \frac{2x^Ty}{\|y\|^2} x^Ty + \frac{(x^Ty)^2}{\|y\|^4} \|y\|^2$$

$$= \|x\|^2 - \frac{(x^Ty)^2}{\|y\|^2} \geq 0$$

$$\Longrightarrow (x^Ty)^2 \leq \|x\|^2 \|y\|^2$$

$$\Longrightarrow \boxed{|x^Ty| \leq \|x\| \, \|y\|}$$

Cauchy Schwarz inequality

<u>Prop (triangle inequality)</u>

$$\|x+y\| \leq \|x\| + \|y\|$$

<u>Pf.</u> $\|x+y\|^2 = (x+y)^T(x+y)$

$$= \|x\|^2 + 2x^Ty + \|y\|^2$$

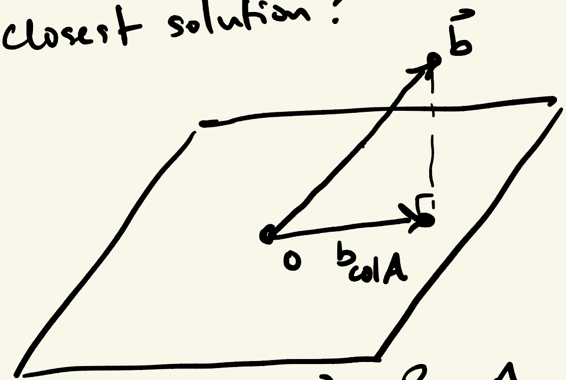
$$\leq \|x\|^2 + 2|x^Ty| + \|y\|^2$$

$$\overset{CS}{\leq} \|x\|^2 + 2\|x\|\|y\| + \|y\|^2$$

$$= (\|x\| + \|y\|)^2. \quad \square$$

Least Squares

$A\vec{x} = \vec{b}$ inconsistent. Closest solution?



$W = \text{Ran } A$

Want to minimize

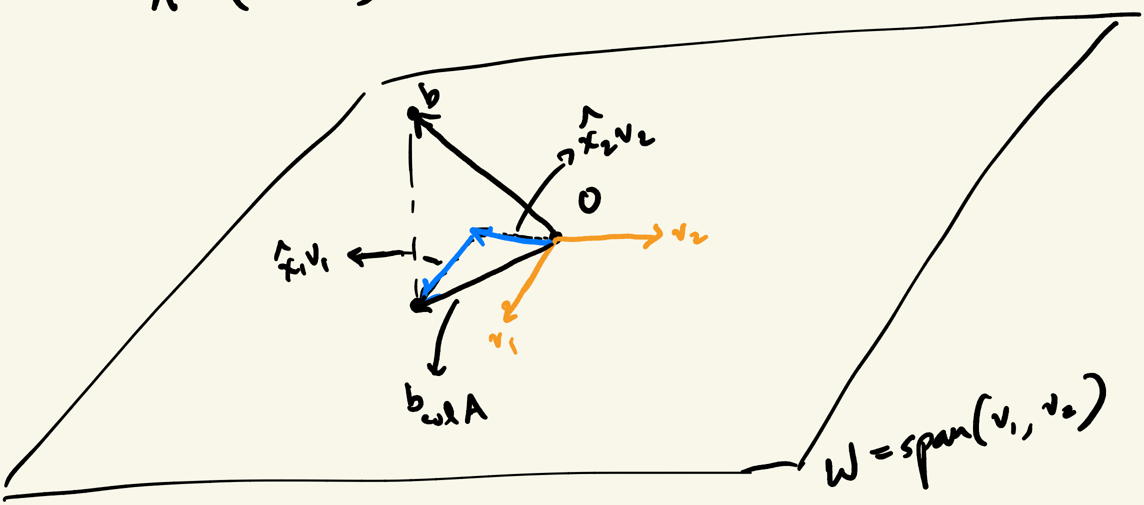
$\|b - Ax\|$, i.e.

Find $\hat{x}$ s.t. $\|b - A\hat{x}\|$

is small as possible.

Orthogonal decomp does this!

$$A = (v_1 \;\; v_2) \qquad \hat{x} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}$$



Def: A <u>least squares sol'n</u> to $Ax=b$ is one that minimizes $\|b-Ax\|$

Thm: The least-squares solutions to $Ax=b$ are the solutions of

$$A^T A x = A^T b.$$

Pf: least squares sol'n comes from solving $Ax = b_{\text{col}A}$. (i.e. setting $W = \text{col}A$, $b_{\text{col}A} = b_W$). If $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$, then $A\hat{x} = b_W$. $\square$

Ex: $Ax=b$, $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$.

This is inconsistent.

$$A^T A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$$

$$A^T b = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$$

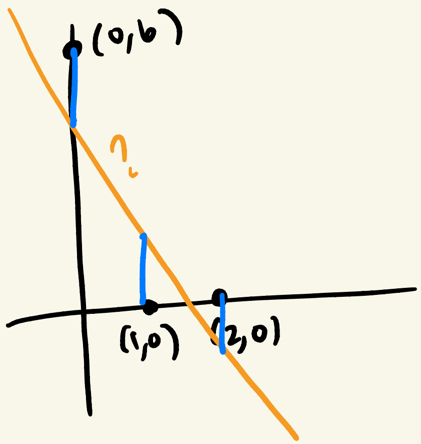
$$\left(\begin{array}{cc|c} 5 & 3 & 0 \\ 3 & 3 & 6 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 5 \end{array}\right)$$

$$\hat{x} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$$

$$\|b - A\hat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$

Closest Ax can get to b is within $\sqrt{6}$ units.

Ex:



$$y = Mx + B$$

$$6 = M \cdot 0 + B$$

$$0 = M \cdot 1 + B$$

$$0 = M \cdot 2 + B$$

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} M \\ B \end{pmatrix}$$

$$b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$$

We solved this! $\hat{x} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$.

$y = -3x + 5$ is the line minimizing sum of squared errors.

specifically, $\hat{x}$ minimizes $\|b - Ax\|$

And $A\hat{x}$ is a vector of $-3x+5$ evaluated at the pts $0, 1, 2$. $(6,0,0)^T$ is the vector of y-values we wanted for $x = 0, 1, 2$ resp.

<u>Rmk:</u> Gauss invented this method to find a best-fit ellipse for an asteroid trajectory. (1801)

Orthogonal sets

Def: A set of nonzero vectors $\{u_1, \dots, u_m\}$ is orthogonal if $u_i^T u_j = 0$ when $i \neq j$. It is orthonormal if, additionally, $u_i^T u_i = 1 \ \forall i$.

Ex: $\{e_1, \dots, e_n\} \subseteq \mathbb{R}^n$ is orthonormal.

Prop: An orthogonal set is lin indp.

Pf: $c_1 u_1 + \dots + c_m u_m = 0$

take dot prod of both sides w/ u_i gives

$$c_i \|u_i\|^2 = 0 \Rightarrow c_i = 0. \quad \square$$

Prop: Suppose $W \subseteq \mathbb{R}^n$ has an orthogonal basis $u_1, \dots, u_m$. Let $x \in \mathbb{R}^n$. Then

$$x_W = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \dots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m$$

$$\Big(\overbrace{x - \Big(\frac{x\cdot u_1}{u_1\cdot u_1}u_1 + \cdots + \frac{x\cdot u_m}{u_m\cdot u_m}u_m\Big)}^{x^*}\Big)\cdot u_j$$

$$= x\cdot u_j - \frac{x\cdot u_j}{u_j\cdot u_j}\,u_j\cdot u_j = x\cdot u_j - x\cdot u_j = 0.$$

This holds for each j. So x^* is orthogonal to $\operatorname{span}(u_1, \dots, u_m) = W$. So $x^* \in W^\perp$.

$$x = \underset{\substack{\Uparrow \\ W}}{\sum \frac{x\cdot u_j}{\|u_j\|^2}u_j} + \underset{\substack{\Uparrow \\ W^\perp}}{x^*}$$

this term must be x_W. $\square$

If $u_1, \dots, u_m$, this simplifies to

$$x_W = \sum_{j=1}^{m}(x\cdot u_j)u_j$$