The four subspaces

Given an m by n matrix A; there are four subspaces related to this matrix (and one another):

- Column space $C(A)$: Space that consists of all Ax (all linear Combinations of its columns).
  $C(A)$ is a subspace of R^m. Dimension of $C(A)$ is rank(A) r.

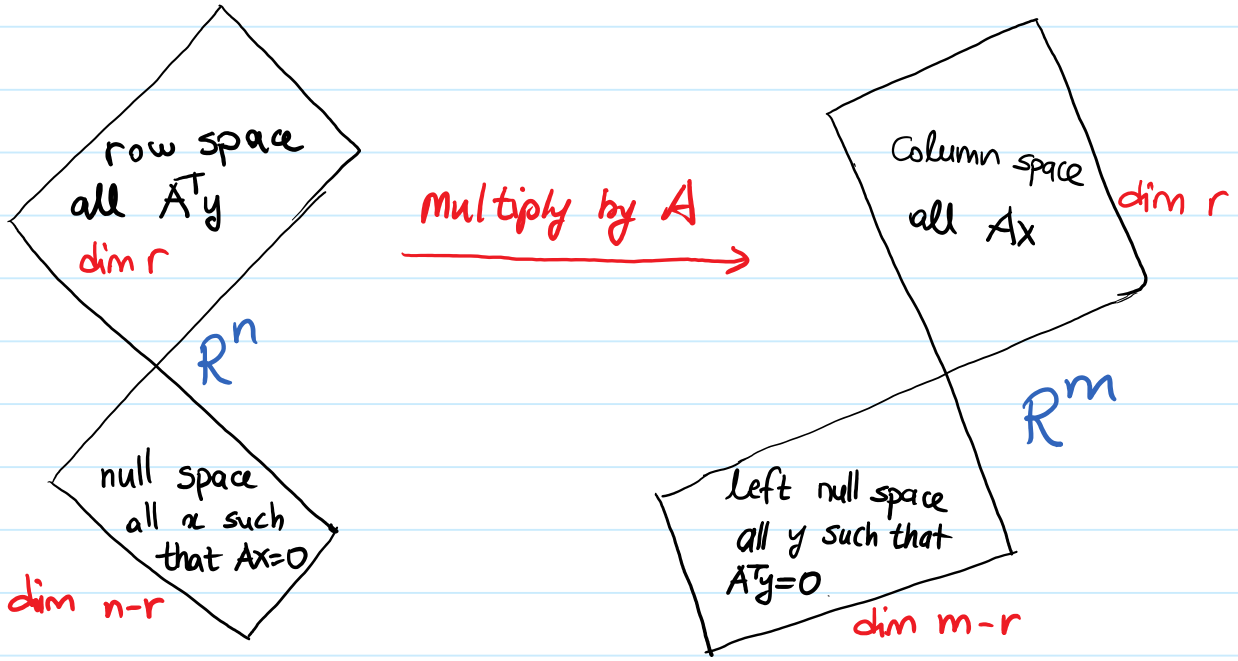
- Null space of A $N(A)$: Space of all vectors x such that $Ax=0$.
  (All solutions to $Ax=0$. Zero vector is obviously one of them).
  $N(A)$ is a subspace of R^n. Dimension of $N(A)$ is $n-r$.

- Row space of A $C(A^T)$: (Note: We write row space as column space of the transpose) Space that consists of all A^Ty
  (all linear combinations of its rows)
  $C(A^T)$ is a subspace of R^n. Dimension of $C(A^T)$ is rank(A) r.

- Left null space of A $N(A^T)$:
  Space of all vectors y such that $A^Ty=0$.
  (All solutions to $A^Ty=0$. zero vector is one of them).
  $N(A^T)$ is a subspace of R^m. Dimension of $N(A^T)$ is $m-r$.

! $\dim(C(A)) = r = \dim(C(A^T))$

Dimension of column space and row space of a matrix is equal to each other and is the rank of the matrix.



Fundamental Thm of Linear algebra (part 1):

The column space and row space both have dimension r.

$N(A)$ has dimension $n-r$.

$N(A^T)$ has dimension $m-r$.

Example: $A=\begin{bmatrix}1&2&3\\2&4&6\end{bmatrix}$ Rank(A) = 1.

$C(A)$: line through $\begin{bmatrix}1\\2\end{bmatrix}$ $\in R^2$ Dimension: 1

$N(A)$: plane through $\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}-2\\1\\0\end{bmatrix}, \begin{bmatrix}-3\\0\\1\end{bmatrix}$ $\left(\text{or all } c\begin{bmatrix}-2\\1\\0\end{bmatrix}+d\begin{bmatrix}-3\\0\\1\end{bmatrix}\right) \in R^3$ Dimension: 2.

$C(A^T)$: line through $\begin{bmatrix}1\\2\\3\end{bmatrix}$ $\in R^3$ Dimension: 1

$N(A^T)$: line through $\begin{bmatrix}-2\\1\end{bmatrix}$ $\in R^2$ Dimension: 1.

Repeat: A set of vectors $u_1, u_2, \cdots, u_n$

1. are LI if no non-zero combination of them gives zero vector.
2. span a subspace V if any vector in V can be written as a linear combination of those vectors.
3. are a basis for a subspace V if they are LI and span V.

ORTHOGONALITY

Definition:

Vectors u and v are orthogonal if $u \cdot v = u^T v = 0$.
(Their dot product produces zero.)

Example:

$$u = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad u^T v = 1 \cdot 0 + 0 \cdot 1 = 0.$$

Example:

$$u = \begin{bmatrix} a \\ b \end{bmatrix} \quad v = \begin{bmatrix} -b \\ a \end{bmatrix} \qquad u^T v = a(-b) + ba = 0.$$

Orthogonal spaces:

Two subspaces V and W are orthogonal if
$v^T w = 0$ for $\underline{\underline{\text{all}}}$ v in V and all w in W.

Example: xy-plane and z-axis line

xy-plane: all $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$, z-axis line: all $\begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}$

$$\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} = 0 \quad \text{for all } x, y, z \in R$$

Example (four subspaces):

- Null space and row space $(N(A)$ and $C(A^T))$ are orthogonal subspaces of R^n.

$$Ax = 0 \quad \text{implies} \quad \begin{aligned} \text{row } 1 \cdot x &= 0 \\ \text{row } 2 \cdot x &= 0 \\ &\vdots \\ \text{row } m \cdot x &= 0 \end{aligned}$$

$$\underbrace{x^T}_{\text{any vector from null space}}(\underbrace{A^Ty}_{\text{any vector from row space}}) = \text{(when } x \in N(A) \text{ so } Ax = 0\text{)}$$

$$= (x^TA^T)y = (Ax)^Ty = 0^Ty = 0.$$

Or similarly:

$$(A^Ty)^Tx = y^TAx = y^T0 = 0$$

Thus A^Ty in $C(A^T)$ and x in $N(A)$ are orthogonal.

Example:

Column space and left null space are orthogonal:

$$\underbrace{(Ax)}_{\text{a vector from } C(A)}{}^T \underbrace{y}_{\text{a vector from left nullspace.}} = x^TA^Ty = x^T0 = 0.$$

Example:

The only vector orthogonal to itself is zero vector:

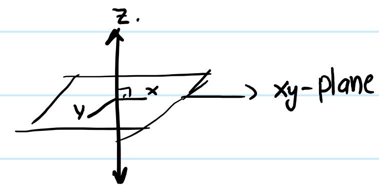
$$x^Tx = 0 \Rightarrow x \text{ is zero-vector.}$$

Orthogonal complements

notation for orthogonal complement of V.

Orthogonal complement $V^{\perp}$ of a subspace V contains every vector that is perpendicular to V.

Example: orthogonal complement of $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$ is $\begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}$.



Fundamental Theorem of Linear Algebra Part 2:

- $N(A)$ is the orthogonal complement of row space $C(A^T)$ in $\mathbb{R}^n$.
- $N(A^T)$ is the orthogonal complement of column space $C(A)$ in $\mathbb{R}^m$.

Idea: Every x can be split into a row space component x_r and null space component x_n:

$$Ax = A(x_r + x_n) = Ax_r + \underbrace{Ax_n}_{=0} = Ax_r.$$

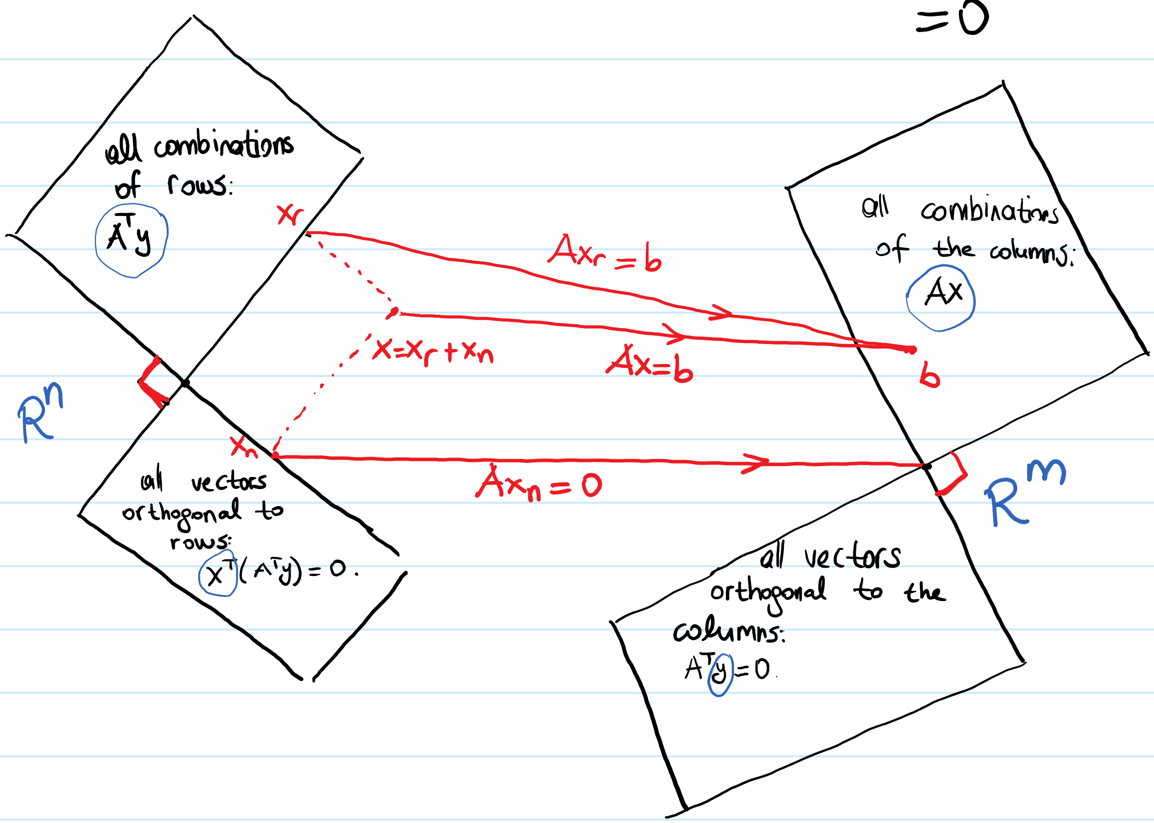
Example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

row space: $c\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + d\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow$ row space component: $x_r = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$

null space: $c\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow$ null space component: $x_n = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}}_{=0} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$



Combining bases:

Any n LI vectors in R^n span R^n. So they are a basis for R^n.
Any n vectors that span R^n are LI. So they are a basis for R^n.

PROJECTIONS

Motivation:

Assume we want to solve $Ax = b$. But what if there is no solution?

1. That means b is not in column space $C(A)$.
2. What is the vector $\hat{x}$ that produces $A\hat{x}$ which is "closest" to b?

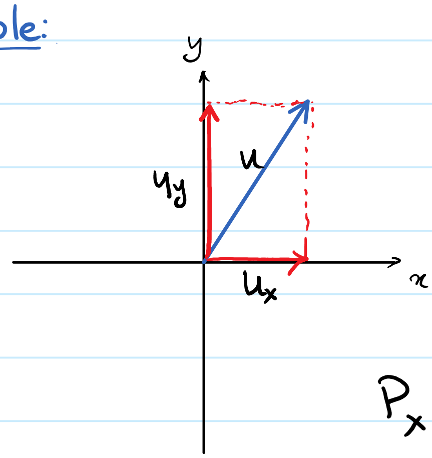
With projections, we will find the answer to this question.

motivation example:

You are trying to get home (b) via public transportation (A) and x is the strategy you choose.

However you know none of the strategies take you to your doorstep, so you choose the one $(\hat{x})$ that takes you to the closest point $(A\hat{x})$ to home (b).

Example:



$u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $u_x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $u_y = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$

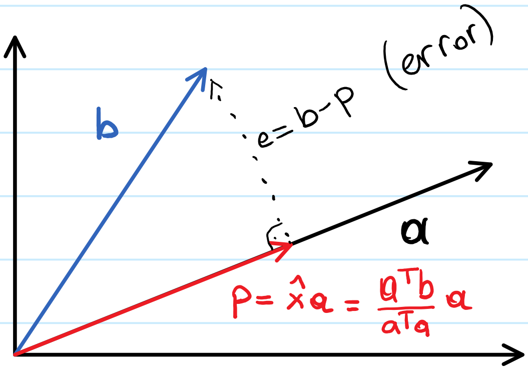
u_x, and u_y are projections of u onto x and y axis.

$P_x u = u_x$, $P_x = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ $P_x^2 = P_x$

$P_y u = u_y$, $P_y = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ $P_y^2 = P_y$.

P_x, and P_y are called projection matrices.

Projection onto a line



Fact: p is a scalar times a: $p = \hat{x}a$
Fact: e is perpendicular to a

$$a^T e = 0$$
$$a^T(b - \hat{x}a) = 0$$
$$a^T b = \hat{x}\, a^T a \Rightarrow \hat{x} = \frac{a^Tb}{a^Ta}$$

$$Pb = p = \hat{x}a = a\hat{x} = a\frac{a^Tb}{a^Ta}$$
$$\Rightarrow P = \frac{aa^T}{a^Ta}.$$

- The projection p is a scalar times a: $\hat{x}a$
- Error $e = b - p = b - \hat{x}a$ is perpendicular to a: $a^Te = a^T(b - \hat{x}a) = 0$
- Projection matrix: $Pb = p$

We first find $\hat{x}$, then $p = \hat{x}a$, the projection matrix P.

> **Projection onto a line**
> Projection of b through a line a is $p = \hat{x}a = \frac{a^Tb}{a^Ta}a$
> Corresponding projection matrix is $P = \frac{aa^T}{a^Ta}$

Example:

Project $b = \begin{bmatrix}1\\2\\3\end{bmatrix}$ onto $a = \begin{bmatrix}1\\1\\0\end{bmatrix}$

$$\hat{x} = \frac{a^Tb}{a^Ta} = \frac{1+2+0}{1+1+0} = \frac{3}{2} \Rightarrow p = \hat{x}a = \frac{3}{2}\begin{bmatrix}1\\1\\0\end{bmatrix}$$

$$P = \frac{aa^T}{a^Ta} = \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix}\begin{bmatrix}1 & 1 & 0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 1 & 0\\1 & 1 & 0\\0 & 0 & 0\end{bmatrix}$$

$P^2 = P$. (Exercise). and $e = b - p = \begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}3/2\\3/2\\0\end{bmatrix} = \begin{bmatrix}-1/2\\1/2\\3\end{bmatrix}$ $e^Ta = 0$.

Example:

P for $a = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ is $\frac{aa^T}{a^Ta} = \frac{1}{6}\begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ 2 & 2 & 4 \end{bmatrix}$

$$P\underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{b} = \frac{1}{6}\begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix} = \underbrace{\frac{2}{3}}_{\hat{x}}\underbrace{\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}}_{a} = p.$$

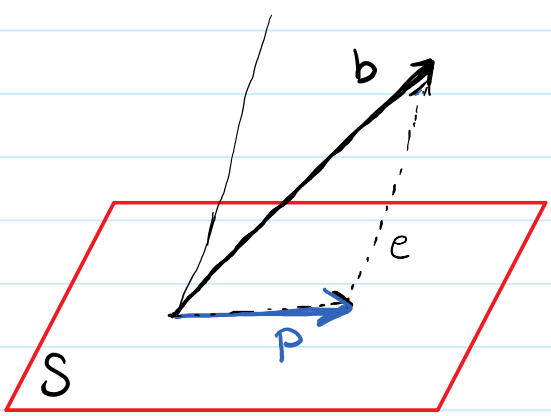
Projection onto a subspace:

Assume n vectors $a_1, a_2, \dots, a_n$ are linearly independent

Find $p = a_1\hat{x}_1 + a_2\hat{x}_2 + \dots + a_n\hat{x}_n = A\hat{x}$ closest to b.

Three steps: Find $\hat{x}$, Find $p = A\hat{x}$, find the projection matrix P such that $Pb = p$.

Idea: Use the fact that error ($e = b - p$) is perpendicular to the subspace:



$$\begin{aligned} p &= A\hat{x} \\ &= A(A^TA)^{-1}A^Tb \\ &= Pb \end{aligned}$$

Error perpendicular to subspace means: error is perpendicular to all the vectors in the subspace, some of which are $a_1, a_2, \dots, a_n$:

$$\begin{aligned} a_1^T(b - A\hat{x}) &= 0 \\ a_2^T(b - A\hat{x}) &= 0 \\ &\vdots \\ a_n^T(b - A\hat{x}) &= 0 \end{aligned} \Rightarrow \begin{bmatrix} — a_1^T — \\ — a_2^T — \\ \vdots \\ — a_n^T — \end{bmatrix}\begin{bmatrix} b - A\hat{x} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \Rightarrow A^T(b - A\hat{x}) = 0$$

$$A^T(b - A\hat{x}) = 0 \Rightarrow (A^TA)\hat{x} = A^Tb$$

(If A is full column rank then A^TA is invertible.)

$$\Rightarrow \hat{x} = (A^TA)^{-1}A^Tb \quad \text{and} \quad p = A\hat{x} = \underbrace{A(A^TA)^{-1}A^T}_{}b = Pb$$

Note: $P^2 = A\underbrace{(A^TA)^{-1}A^TA}_{I}(A^TA)^{-1}A^T = A(A^TA)^{-1}A^T = P.$

⋆ A^TA is invertible if and only if A has linearly independent columns.

Summary:

Assume $A_{m\times n}$ is full column rank. Then,

The combination $p = A\hat{x}$ closest to b comes from $\hat{x}$:

$$(A^TA)\hat{x} = A^Tb.$$

The symmetric A^TA is invertible: $\hat{x} = (A^TA)^{-1}A^Tb$. Then

$$p = A\hat{x} = A(A^TA)^{-1}A^Tb = Pb$$

And,

$$P = A(A^TA)^{-1}A^T$$

Example:

Find projection of $b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto column space of $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix}$.

Step 1: $A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$ $\quad A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$

$A^TA\hat{x} = A^Tb \Rightarrow \hat{x} = \begin{bmatrix} 5/3 \\ 0 \end{bmatrix}$ $\quad \hat{p} = A\hat{x} = \frac{5}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\quad P = A(A^TA)^{-1}A^T.$

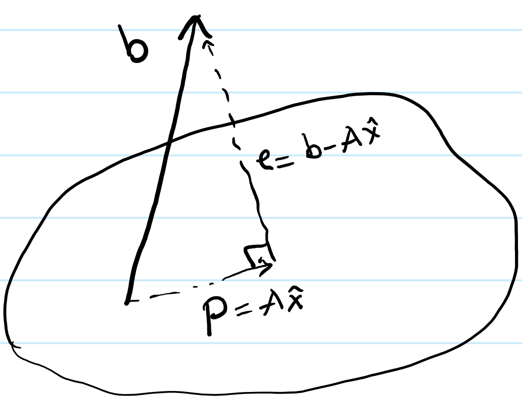
Least Squares Approximation

Often $Ax=b$ has no solution. Reason: too many equations.

Meaning b is not in column space of A.

Least squares solution: Pick a vector $\hat{x}$ such that square of length of error $e = b - Ax$ is minimum.

The projection p of b onto column space of A is connected to the least squares solution $\hat{x}$:



The closest point in $C(A)$ to b is $p = A\hat{x}$, the projection of b onto $C(A)$. In other words, $\hat{x}$ in the projection gives the smallest $\|b - A\hat{x}\|^2$ and thus is the least squares solution.

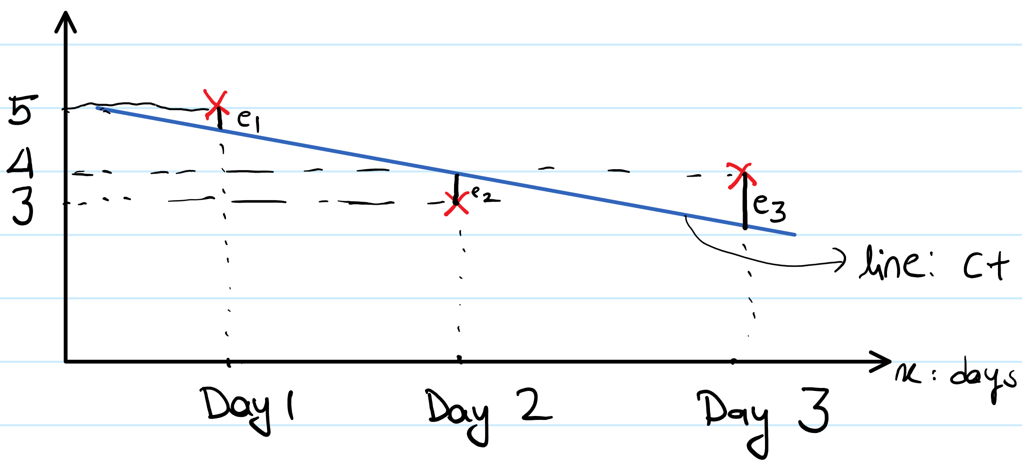
Algebra check:

$$0 = \frac{\partial \|b - A\hat{x}\|^2}{\partial \hat{x}} = \frac{\partial (b - A\hat{x})^T (b - A\hat{x})}{\partial \hat{x}} = \frac{\partial \left(b^T b - 2x^T A^T b + \hat{x}^T A^T A \hat{x}\right)}{\partial \hat{x}}$$

$$= -2A^T b + 2A^T A \hat{x} \Rightarrow A^T A \hat{x} = A^T b$$

$$\boxed{\text{The partial derivatives of } \|b - Ax\|^2 \text{ are zero when } A^T A \hat{x} = A^T b}$$

Example: (Line fit)



Fit a line into 3 points: $(1,5)$, $(2,4)$, $(3,5)$

Answer: Impossible! Instead we will find a line that minimizes least squares error: $\|b - Ax\|^2$.

$$\begin{aligned} c + d\,1 &= 5 \\ c + d\,2 &= 3 \\ c + d\,3 &= 4 \end{aligned} \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}$$

No solution!

Instead;

Find $\hat{x} = \begin{bmatrix} c \\ d \end{bmatrix}$ that minimizes $\|b - A\hat{x}\|^2$:

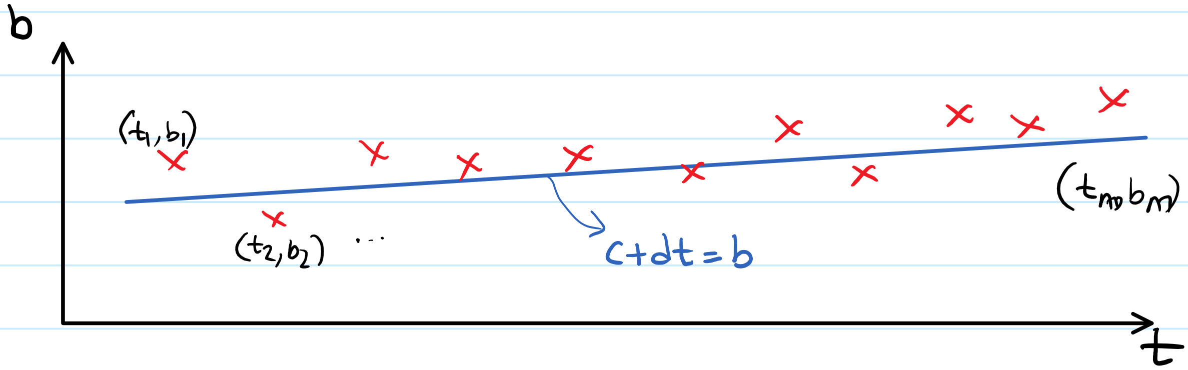
$$A^T A \hat{x} = A^T b \Rightarrow \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \hat{x} = \begin{bmatrix} 12 \\ 23 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 5 \\ -1/2 \end{bmatrix}$$

And $$p = A\hat{x} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ -1/2 \end{bmatrix} = \begin{bmatrix} 9/2 \\ 4 \\ 7/2 \end{bmatrix} \in C(A)$$

$$e = b - Ax = \begin{bmatrix} 1/2 \\ -1 \\ 1/2 \end{bmatrix} \qquad E = \|e\|^2 = 1/2^2 + 1^2 + 1/2^2 = \frac{3}{2}.$$

Fitting a straight line:

Suppose we have points $(t_1,b_1), (t_2,b_2), \dots, (t_m, b_m)$



$Ax=b$:
$$\begin{aligned} c + dt_1 &= b_1 \\ c + dt_2 &= b_2 \\ &\vdots \\ c + dt_m &= b_m \end{aligned}$$

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \quad x = \begin{bmatrix} c \\ d \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$\Rightarrow$ Least squares solution:
$$A^TA\hat{x} = A^Tb$$
$$\begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$$

Generalization: We are fitting m points (number of measurements) by n parameters.

Reason why least squares: Derivative of square is linear.

$$\underset{x}{\text{minimize}} \; \|b - Ax\|^2$$

Derivative with respect to x

$$A^TA\hat{x} = A^Tb \;\rightsquigarrow\; \hat{x} = (A^TA)^{-1}A^Tb$$

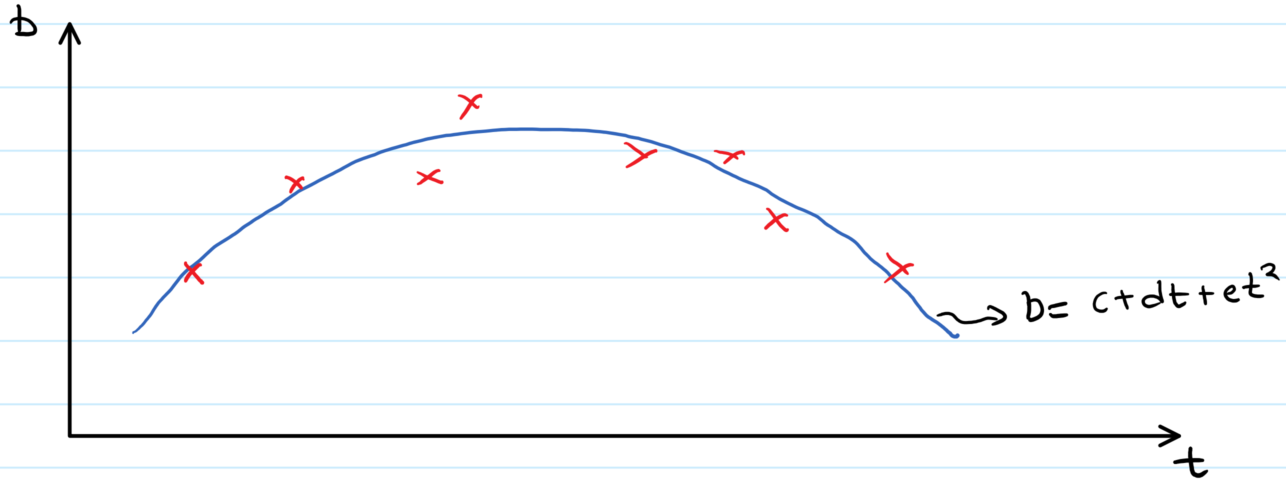
What if A has LD columns?

Many solutions. (will learn later how we could pick one.)

$A\hat{x} = p$ — $\hat{x}$ is a solution,

$A(\hat{x} + x_n) = p$ — so is $\hat{x} + x_n$, when x_n is taken from the null space.

Example: (Fit a parabola)



Points: $(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m)$

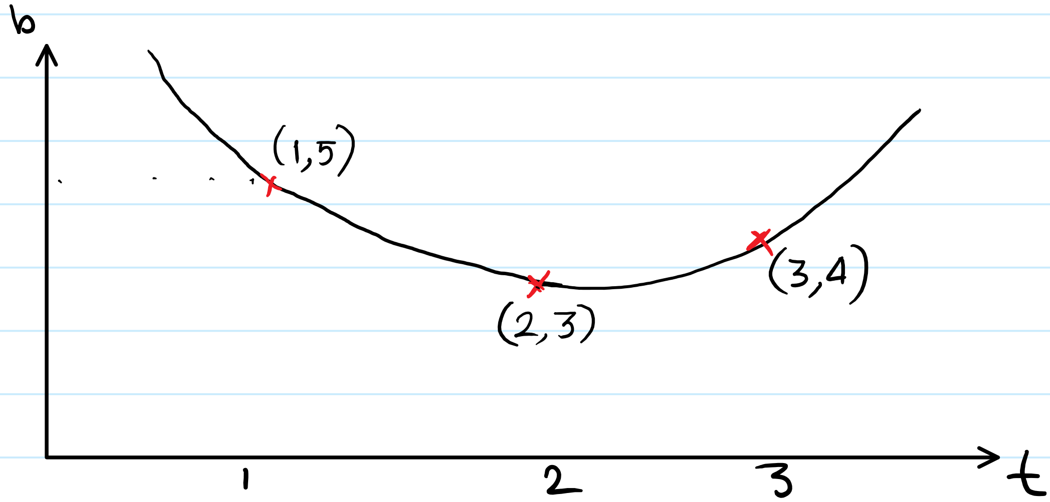
Parabola:

$$\begin{aligned} c + dt_1 + et_1^2 &= b_1 \\ c + dt_2 + et_2^2 &= b_2 \\ &\vdots \\ c + dt_m + et_m^2 &= b_m \end{aligned} \Rightarrow \underbrace{\begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} c \\ d \\ e \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{b}$$

No solution if $m > 3$ equations and no perfect measurements.

Least squares: $(A^TA)\hat{x} = A^Tb$

Fitting a parabola to previous example:



$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} c \\ d \\ e \end{bmatrix}}_{x} = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} \Rightarrow \begin{array}{l} c = 10 \\ d = -13/2 \\ e = 3/2 \end{array}$$

$E = \|b - A\hat{x}\|^2 = 0$ (no error term). b was in the column space of A.