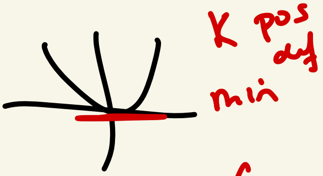
Yesterday...

minimize

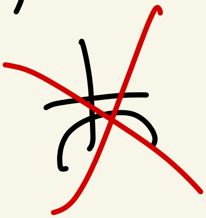
$$p(x) = x^T K x - 2x^T f + c$$

K pos def.



$$x^* = K^{-1} f$$

$$p(x^*) = \underline{c} - f^T x^*$$



Minimize

$\|w-b\|^2$, where $w \in W$ a subspace $\subseteq \mathbb{R}^n$.

and $b \in \mathbb{R}^n$

Two ways to figure out

$$\min \{ \|w-b\|^2 \mid w \in W \}.$$

- $w = x_1 w_1 + \dots + x_k w_k$ $\quad w_1 \dots w_k$ basis of W.

  Solve for $x_1 \dots x_k$

let $W \subseteq \mathbb{R}^n$ be a subspace, $b \in \mathbb{R}^n$. minimize $\|w-b\|^2$.

Suppose $w_1 \dots w_k$ is a basis of W.

$A = (w_1 \dots w_k)$. Then we know that $\mathrm{img}(A) = \mathrm{span}(w_1 \dots w_k)$

$= W.$

On the other hand

$$\mathrm{img}(A) = \{A\underline{x} \mid x \in \mathbb{R}^k\}.$$

$$\min\{\|w-b\|^2 \mid w \in W\}$$

$$= \min\{\|Ax - b\|^2 \mid x \in \mathbb{R}^k\}$$

$$\|Ax-b\|^2 = (Ax-b)\cdot(Ax-b)$$

$$= (Ax-b)^T(Ax-b)$$

$$(Ax-b)^T(Ax-b)$$

$$= \left((Ax)^T - b^T\right)(Ax-b)$$

$$= \left(x^TA^T - b^T\right)(Ax-b)$$

$$= x^TA^TAx - b^TAx - x^TA^Tb + b^Tb$$

$$= x^T\underbrace{A^TA}x - \underbrace{b^TAx \overset{?}{-} x^TA^Tb} + \|b\|^2$$

These two terms are equal!

$$\text{"}=\text{"}\ x^TKx - 2x^Tf + \underbrace{\|b\|^2}_{\text{constant}}$$

$K = A^TA$ gram matrix for $w_1 \ldots w_k$.

Side calculation $\longrightarrow$

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

Pf $\quad b^T A x \in \mathbb{R}^1$

So $(b^T A x) = (b^T A x)^T$ trivially

$$= x^T A^T b^{T^T} = x^T A^T x$$

$$\|Ax - b\|^2$$

$$= x^T A^T A x - \underbrace{b^T A x - x^T A^T b}_{\text{equal}} + \|b\|^2$$

$$= x^T \underbrace{A^T A}_{\substack{\big\downarrow \\ K}} x - 2x^T \underbrace{A^T b}_{\substack{\big\downarrow \\ f}} + \|b\|^2$$

Therefore $\quad x^* = K^{-1} f \quad$ (yesterday)

$$x^* = (A^T A)^{-1} A^T b \quad \rightsquigarrow$$

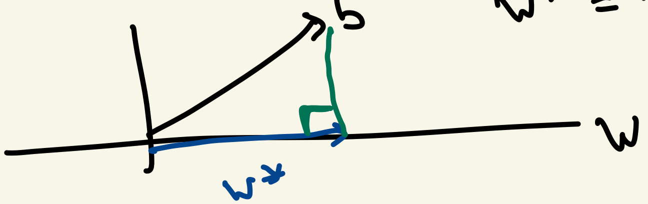
$x^* = (A^TA)^{-1}A^Tb$ minimizes $\|Ax - b\|^2$.

$W = Ax$ we W. what is the actual vector $w^* \in W$ s.t. w^* is closest to b?

$$w^* = Ax^* = \underbrace{A(A^TA)^{-1}A^T}_{\text{from HW}} b.$$

$\text{proj}_W b = A(A^TA)^{-1}A^Tb$ from HW

the projection should be the closest vector w^* to b s.t. $w^* \in W$.



$$w^* = \text{proj}_W b$$
$$= A(A^TA)^{-1}A^Tb.$$

Ex $W = \text{Span}\left(\begin{pmatrix}1\\2\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\2\end{pmatrix}\right)$. $b = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}$

Closest distance from b to W?

which vector $w^* \in W$ is closest to b?

$$w^* = A(A^TA)^{-1}A^Tb$$

where $A = \begin{pmatrix}1 & 0\\2 & 0\\-1 & 1\\0 & 2\end{pmatrix}$

$A^TA = \begin{pmatrix}6 & -1\\-1 & 5\end{pmatrix}$ $(A^TA)^{-1} = \frac{1}{29}\begin{pmatrix}5 & 1\\1 & 6\end{pmatrix}$

$$w^* = \begin{pmatrix}1 & 0\\2 & 0\\-1 & 1\\0 & 2\end{pmatrix}\frac{1}{29}\begin{pmatrix}5 & 1\\1 & 6\end{pmatrix}\begin{pmatrix}1 & 2 & -1 & 0\\0 & 0 & 1 & 2\end{pmatrix}\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$$

$= \frac{1}{29}\begin{pmatrix}5\\10\\-4\\2\end{pmatrix}$ is closest to $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$ from W.

On the other hand

$$w^* = \mathrm{proj}_W b$$

① G-S basis of W

② $w^* = \langle u_1, b \rangle u_1 + \langle u_2, b \rangle u_2$

These methods are the same.

$$d = \| w^* - b \|$$

$$= \sqrt{\|b\|^2 - f^T x^*} = \frac{1}{29}\left(2\sqrt{274}\right)$$

is the minimal distance from $b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ to $W = \mathrm{Span}\left(\begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix}\right)$.

$(A^TA)^{-1}A^T$ matrix helps computation quite nicely.

5.4 Least squares

Def Let $Ax = b$ be a system of equations. b may or not be in $\mathrm{img}(A)$.

Then the vector x^* which minimizes the distance $\|Ax - b\|^2$ is called the least squares solution to $Ax = b$.

Letting $W = \mathrm{img}(A)$. Find x^* by

$$x^* = (A^TA)^{-1}A^Tx$$

any matrix A may not have independent columns

$\underline{(A^TA)}^{-1}$ might not exist.

let A be an $m \times n$ matrix.

When is A^TA invertible?

If the columns of A are independent

$K = A^TA$ = Gram matrix of independent vectors

$\implies$ always positive definite

$\implies$ $K = (A^TA)^{-1}$ invertible.

When are columns of $A_{m\times n}$ independent?

columns are independent $\iff$ $\text{rank}(A) = n$. (every column has a pivot)

If $n < m$.

$$\iff \underline{\ker(A) = 0.}$$

A $m \times n$ (which we want to do least squares to)

$x^* = (A^T A)^{-1} A^T b$. $(A^T A)^{-1}$ needs to exist.

Claim: If A is $m \times n$, $\underline{n < m}$, and $\ker(A) = 0$, then $(A^T A)^{-1}$ exists.

$$A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$\dim \mathbb{R}^n < \dim \mathbb{R}^m$$

Linear transformation from smaller vector space to a bigger one.

$$A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$\dim \mathbb{R}^n \quad < \quad \dim \mathbb{R}^n$$

Linear transformation from smaller vector space to a bigger one.

If $\ker(A) = 0.$ Claim:

$$\boxed{\dim(\operatorname{img}(A)) = n}$$

$$\|$$

$$\operatorname{rank}(A)$$

Since $\dim(\operatorname{img}(A)) = n$

$$\operatorname{img}(A) \subseteq \mathbb{R}^m \qquad m > n.$$

We can think of $\operatorname{img}(A)$ just as a copy of $\mathbb{R}^n$ but tilted and stuck inside of $\mathbb{R}^m$ somehow.

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$$

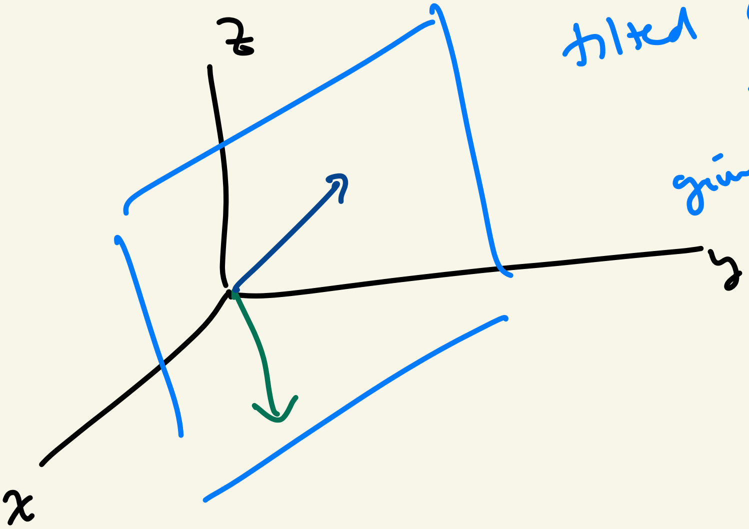
2 ind columns

$\Longrightarrow$ rank 2

$\Longrightarrow$ $\ker(A) = 0$.

$\Longrightarrow$ $\dim(\operatorname{img}(A)) = 2$

Img(A) is a 2D subspace of $\mathbb{R}^3$.

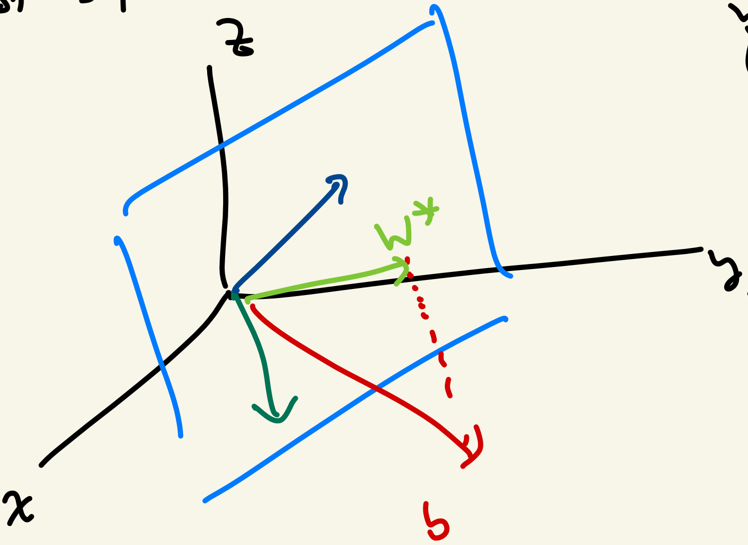


tilted copy of $\mathbb{R}^2$ given by img(A).

$$\begin{tikzcd} A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} : \mathbb{R}^2 \arrow[r] & \mathbb{R}^3 \arrow[l, dashed, bend left, "(A^TA)^{-1}A^T"] \end{tikzcd}$$

Square matrices don't have inverses but $(A^TA)^{-1}A^T$ is an almost inverse if $\ker A = 0.$

least squares



$x^* = (A^TA)^{-1}A^Tb$ is coordinates of w^* in terms the columns of A

x^* is coefficients of $w_1 \ldots w_k$ which make w^*.

$$\begin{tikzcd} (A^TA)^{-1}A^T : \mathbb{R}^3 \arrow[r] \arrow[d] & \mathbb{R}^2 \\ b & x^* \end{tikzcd}$$

Another way to think about x^* is it's closest variable to being a solution to $Ax = b$.

If $Ax = b$ has an actual solution z, then $z = x^*$.

Since $d_{min} = \|Ax^* - b\| = 0$

so $x^* = z$.

Note: If $\ker(A) \neq 0$, then there is a not a unique least squares solution!

$x^* = (A^T A)^{-1} A^T b$ makes no sense

If $z \in \ker(A)$. x^* is a least squares solution

then $w = x^* + z$ is also a least squares solution.

x^* minimizes $\|Ax - b\|^2$, $w = x^* + z$, $\underline{z \in \ker(A)}$

$$\underline{\|Aw - b\|^2}$$

$$= \|A(x^* + z) - b\|^2$$

$$= \|Ax^* + \overset{0}{\cancel{Az}} - b\| = \underline{\|Ax^* - b\|^2}$$

which is minimal.

So $\|Aw - b\|^2$ is also minimal.

$w = x^* + z$ is also a least squares solution.

Need $\ker(A) = 0$ to get a unique solution

$x^* = (A^TA)^{-1}A^Tb$ works.

Thm let A be $m \times n$, st either

$\text{rank}(A) = n \iff \ker(A) = 0$.

Then $Ax = b$ has a unique least squares solution $x^* = \underline{(A^TA)^{-1}A^T b}$.

Ex :

$$\underset{\substack{\| \\ A}}{\begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & -1 & -2 \\ 2 & 1 & -1 \end{pmatrix}} \underset{\substack{\| \\ x}}{\begin{pmatrix} x \\ y \\ z \end{pmatrix}} = \underset{\substack{\| \\ b}}{\begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \\ 2 \end{pmatrix}}$$

5×3 matrix.

Tell by row reduction

① . $\text{rank}(A) = 3$

: all columns are independent

: $\ker(A) = 0$

② $Ax = b$ has no solution!

$$\underset{\underset{A}{\parallel}}{\begin{pmatrix}1&2&0\\3&-1&1\\-1&2&1\\1&-1&-2\\2&1&-1\end{pmatrix}}\underset{\underset{x}{\parallel}}{\begin{pmatrix}x\\y\\z\end{pmatrix}}=\underset{\underset{b}{\parallel}}{\begin{pmatrix}1\\0\\-1\\2\\2\end{pmatrix}}$$

Since $\ker(A) = 0$ There is a unique least squares solution!

$$x^* = \underbrace{(A^TA)}^{-1}\underbrace{A^Tb}.$$

$$K = A^TA = \begin{pmatrix}16&-2&-2\\-2&11&2\\-2&2&7\end{pmatrix}^{*}$$

$$f = A^Tb = \begin{pmatrix}8\\0\\-7\end{pmatrix}$$

closest $\begin{pmatrix}x\\y\\z\end{pmatrix}$ ↓ to being a solution!

$$x^* = K^{-1}f = (A^TA)^{-1}A^Tb$$

$$= \cancel{\frac{1}{556}\begin{pmatrix}355\\-58\\674\end{pmatrix}} = \begin{pmatrix}0.4119\\.2482\\-.9532\end{pmatrix}$$

Calculator . . .

Suppose we have a bunch of data

$$(t_1, y_1) \dots (t_m, y_n)$$

Suppose $y = \alpha + \beta t$ is supposed to relate t, y.

Which α, β fit best?

actual data

supposed y-value

$$e_1 = \varepsilon_1 = y_1 - (\alpha + \beta t_1)$$

$$\varepsilon_2 = y_2 - (\alpha + \beta t_2)$$

$$\vdots$$

$$\varepsilon_m = y_m - (\alpha + \beta t_m)$$

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - \underbrace{\begin{pmatrix} \alpha + \beta t_1 \\ \vdots \\ \alpha + \beta t_m \end{pmatrix}}_{\text{matrix product}}$$

$$\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_m\end{pmatrix} = \begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix} - \underbrace{\begin{pmatrix}\alpha+\beta t_1\\ \vdots\\ \alpha+\beta t_m\end{pmatrix}}_{}$$

$$\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_m\end{pmatrix} = \begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix} - \begin{pmatrix}1 & t_1\\ \vdots & \vdots\\ 1 & t_m\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

$$\updownarrow$$

$$\varepsilon = y - \underset{\substack{\shortparallel\\ A}}{\begin{pmatrix}1 & t_1\\ 1 & t_2\\ \vdots & \vdots\\ 1 & t_m\end{pmatrix}} x$$

$$\varepsilon = y - Ax.$$

α, β fit best when

$\sum \varepsilon_i^2$ is minimal.

(least squares)

$$\varepsilon = y - Ax.$$

α, β fit best when $\sum \varepsilon_i^2$ is minimal.

(least squares)

in terms of data

$$\underset{\text{min.}}{\underline{\sum \varepsilon_i^2}} = \|\varepsilon\|^2 = \underbrace{\|y - Ax\|^2}_{\text{we know how to solve this}}$$

$$x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \underbrace{(A^TA)^{-1}}\,\underline{A^T y}$$

$$A^TA = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}$$

$$= \begin{pmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}$$

$$\underline{(A^TA)^{-1}} = \frac{1}{m\sum t_i^2 - (\sum t)^2} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{pmatrix}$$

$$\underline{A^T y} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

$$= \begin{pmatrix} \sum y_i \\ \sum t_i y_i \end{pmatrix}$$

$$y = \alpha + \beta t$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (A^T A)^{-1} A^T y$$

$$= \frac{1}{m(\sum t_i^2) - (\sum t_i)^2} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum t_i y_i \end{pmatrix}$$

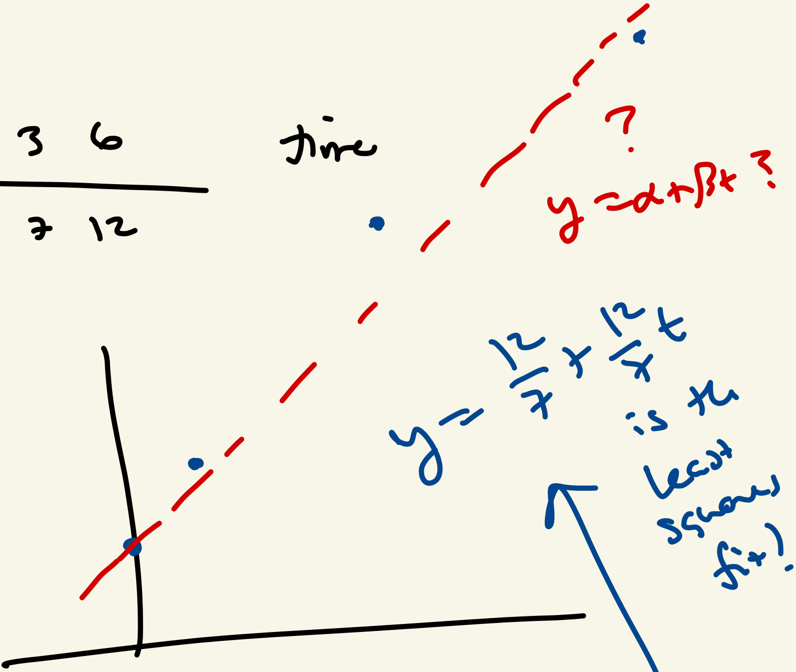
$$= \frac{1}{m(\sum t_i^2) - (\sum t_i)^2} \begin{pmatrix} (\sum t_i^2)(\sum y_i) - (\sum t_i)(\sum t_i y_i) \\ -(\sum t_i)(\sum y_i) + m(\sum t_i y_i) \end{pmatrix}$$

So we solved for α, β in terms of $(t_1, y_1) \cdots (t_m, y_m)$.

Ex:

t	0	1	3	6
y	2	3	7	12

time



$y = \frac{12}{7} + \frac{12}{7} t$ is the least squares fit!

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix} \quad y = \begin{pmatrix} 2 \\ 3 \\ 7 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (A^T A)^{-1} A^T y$$

$$A^T A = \begin{pmatrix} 4 & 10 \\ 10 & 46 \end{pmatrix}$$

$$A^T y = \begin{pmatrix} 24 \\ 96 \end{pmatrix}$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 10 & 46 \end{pmatrix}^{-1} \begin{pmatrix} 24 \\ 96 \end{pmatrix} = \underline{\begin{pmatrix} 12/7 \\ 12/7 \end{pmatrix}}$$

Same method for any polynomial fit!

$$(t_1, y_1) \ldots (t_m, y_m)$$

$$y = \alpha_0 + \alpha_1 t^1 + \ldots + \alpha_n t^n$$

$$\varepsilon_i = y_i - (\underline{\alpha_0} + \alpha_1 t_{\underline{i}}^1 + \ldots + \alpha_n t_{\underline{i}}^n)$$

$$\varepsilon = y - A \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^n \\ 1 & t_2 & t_2^2 & \ldots & t_2^n \\ \vdots & & & & \\ 1 & t_m & t_m^2 & \ldots & t_m^n \end{pmatrix}$$

least squares is ...

$$\begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{pmatrix} = (A^T A)^{-1} A^T y$$

$$A = \begin{pmatrix} 1 & t_1 & \cdots & t_1^n \\ 1 & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \ddots \\ 1 & t_m & \cdots & t_m^n \end{pmatrix}$$ n+1 columns

is called a Vandermonde matrix.

A is a square matrix when

$\underline{m = n+1}.$

$$A = \begin{pmatrix} 1 & t_1 & \cdots & t_1^n \\ \vdots & \vdots & & \\ 1 & t_{n+1} & \cdots & t_{n+1}^n \end{pmatrix}_{n+1,\, n+1}$$

If $t_i \neq t_j$ then

A^{-1} exists in this case.

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

$$= A^{-1}\cancel{(A^T)^{-1}A^T}\, y$$

$$= A^{-1} y$$

determines which polynomial goes thru data points exactly

Ex: If $m=3$, then if I fit

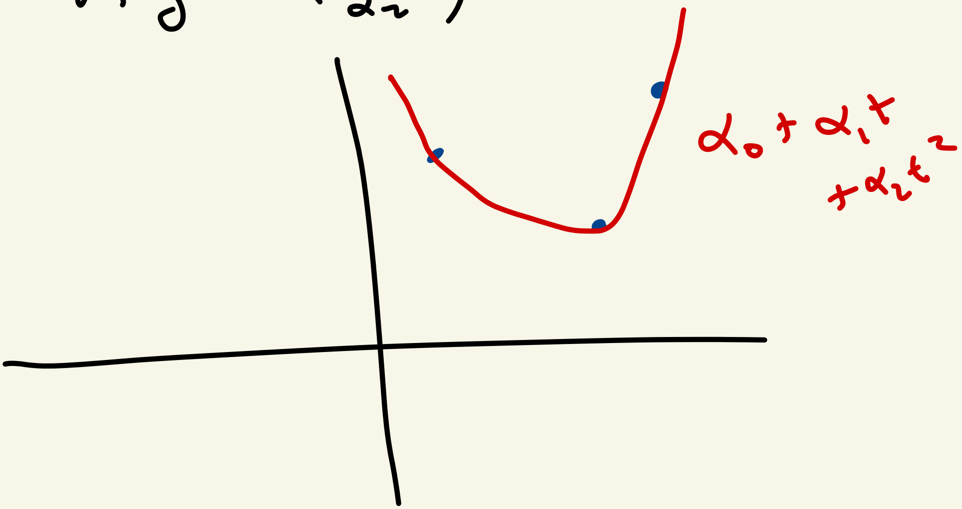
$(t_1, y_1) \dots (t_3, y_3)$

to a degree 2 polynomial

$$p(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$Ay = \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_2 \end{pmatrix}$$



Remember: A $m \times n$ $n < m$

$$A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

sticks $\mathbb{R}^n$ inside of $\mathbb{R}^m$ in some way

$$(A^T A)^{-1} A^T : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

best way to go backwards.

8.7 Singular Values

Take A, rank(n), $n \times m$.

$\ker(A) = 0$, n independent columns.

$K = A^T A$, pos def, invertible, symmetric

Eigenvalues of $A^T A$, $\lambda_i > 0$

Def: Given a matrix A,

the **singular values** of A are $\sigma_i = \sqrt{\lambda_i}$

λ_i is a eigenvalue of $K = A^T A$.

We don't A to be $\operatorname{rank}(A) = n$.

If $\operatorname{rank}(A) < n$.

$K = A^T A$ is positive **semi**-definite

$$\lambda_1 \dots \lambda_r > 0 \rightsquigarrow \sigma_i = \sqrt{\lambda_i}$$

$$\lambda_{r+1} \dots \lambda_n = 0.$$

Ex: $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ singular values?

$$K = A^T A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$K = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$\longrightarrow \quad \lambda = 0$$

$$\left\{ \begin{array}{ll} \lambda = 1 & v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \\ \lambda = 3 & v_3 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \end{array} \right.$$

The singular values of $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$

are $\sigma_1 = \sqrt{1} = 1$

$\sigma_2 = \sqrt{3} = \sqrt{3}$.

Recall: K symmetric,

then $K = Q \Lambda Q^T$

Λ diagonal Q orthogonal matrix

Spectral decomp.

Generalize this to non-square matrices!

Singular Value Decomposition :

Thm let A be $m \times n$, w/ singular values $\sigma_1 \ldots \sigma_r$ $\operatorname{rank}(A) = r$.

Then

$$A = P \Sigma Q^T$$

P, $m \times r$ $Q = r \times n$

Σ $r \times r$

where the columns of Q are orthonormal eigenvectors of $A^{+}A$ for $\lambda_i = \sigma_i^2$.

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

P has orthonormal columns given by

$$p_i = \frac{A q_i}{\sigma_i}.$$

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \boxed{P \cdot}\boxed{\Sigma}\boxed{Q^T}$$

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

$$\sigma_1 = 1 \qquad \sigma_2 = \sqrt{3}$$

$$\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\downarrow \qquad \downarrow$$

$$u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \qquad u_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\boxed{Q} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ q_1 & q_2 \end{pmatrix}$$

$$\boxed{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{pmatrix}$$

$$\boxed{P} = \begin{pmatrix} \frac{Aq_1}{\underset{1}{\sigma_1}} & \frac{Aq_2}{\underset{\sqrt{3}}{\sigma_2}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\begin{pmatrix}1&1&0\\0&1&1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}-1&1\\1&1\end{pmatrix}\begin{pmatrix}1&0\\0&\sqrt{3}\end{pmatrix}\begin{pmatrix}\frac{-1}{\sqrt{2}}&0&\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{6}}&\frac{2}{\sqrt{6}}&\frac{1}{\sqrt{6}}\end{pmatrix}$$

P Q^T

have orthonormal columns

orthonormal rows

let $A = P\Sigma Q^T$

almost $A^{-1} = (P\Sigma Q^T)^{-1}$

" $=$" $(Q^T)^{-1}\Sigma^{-1}P^{-1}$

" $=$" $Q\Sigma^{-1}P^T$

A^{-1} " $=$" $Q\begin{pmatrix}\frac{1}{\sigma_1}&&\\&\ddots&\\&&\frac{1}{\sigma_r}\end{pmatrix}P^T$

Def Given any matrix A, $m\times n$, $A = P\Sigma Q^T$, then the pseudoinverse is $A^+ = Q\Sigma^{-1}P^T$.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{pmatrix}\begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

$$A \quad = \quad P \quad \Sigma \quad Q^T$$

$$A^+ = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

$$(A^TA)^{-1}A^T : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

$$A^+ : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

Thm If $\mathrm{rank}(A) = n$ the

$$\underline{A^+} = \underline{(A^TA)^{-1}A^T}.$$

λ_i of $K = A^TA$

compute naturally

Pf : Let $A = P\Sigma Q^T$

$A^+ = Q\Sigma^{-1}P^T$.

$$A^TA = (P\Sigma Q^T)^T(P\Sigma Q^T)$$

$$= Q\Sigma^T \cancel{P^TP}\,\Sigma Q^T$$

orthonormal columns

$$= Q\Sigma^T\Sigma Q^T \quad , \Sigma^T = \Sigma$$

square diagonal

$$= Q\Sigma^2Q^T$$

$$(A^TA)^{-1} = Q\begin{pmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r^2} \end{pmatrix}Q^T \quad \text{(spectral decomp)}$$

$$= Q\Sigma^{-2}Q^T$$

$$\boxed{(A^TA)^{-1}A^T} = (Q^T\Sigma^{-2}Q)(P\Sigma Q^T)^T$$

orthonormal columns

$$= Q\Sigma^{-2}\cancel{Q^TQ}\,\Sigma^T P^T$$

$$= Q\Sigma^{-2}\Sigma P^T = Q\Sigma^{-1}P^T = \boxed{A^+}$$