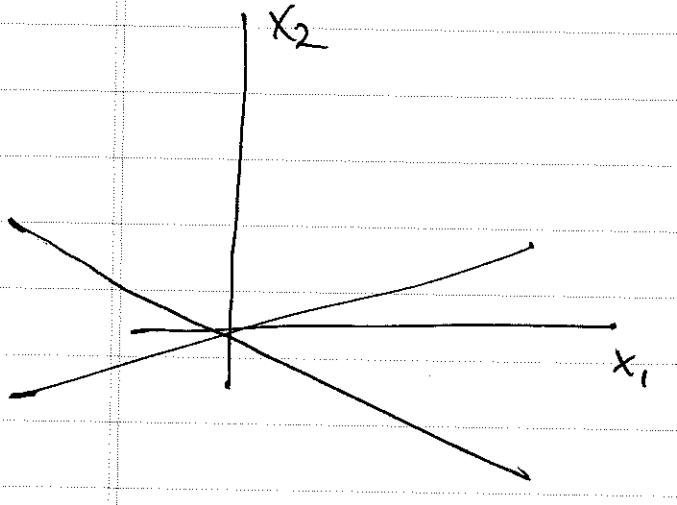
Example: Consider the set of equations

$$\left.\begin{aligned} x_2 &= \varepsilon x_1 \\ x_2 &= -\varepsilon x_1 \end{aligned}\right\} \cdot \quad (\varepsilon << 1)$$




$$\underbrace{\begin{bmatrix} -\varepsilon & 1 \\ \varepsilon & 1 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We compute $A^T A = \begin{bmatrix} -\varepsilon & \varepsilon \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -\varepsilon & 1 \\ \varepsilon & 1 \end{bmatrix} = \begin{bmatrix} 2\varepsilon^2 & 0 \\ 0 & 2 \end{bmatrix}$

If $\varepsilon << 1$, then clearly the max eigenvalue of $A^T A$ is $1$, so $\|A\|_2 = 1$.

$$A^{-1} = -\frac{1}{2\varepsilon} \begin{bmatrix} 1 & -1 \\ -\varepsilon & -\varepsilon \end{bmatrix} = \frac{1}{2\varepsilon} \begin{bmatrix} -1 & 1 \\ \varepsilon & \varepsilon \end{bmatrix}.$$

$$(A^{-1})^T A^{-1} = \frac{1}{4\varepsilon^2} \begin{bmatrix} -1 & \varepsilon \\ 1 & \varepsilon \end{bmatrix} \begin{bmatrix} -1 & 1 \\ \varepsilon & \varepsilon \end{bmatrix}$$

$$= \frac{1}{4\varepsilon^2} \begin{bmatrix} 1+\varepsilon^2 & -1+\varepsilon^2 \\ -1+\varepsilon^2 & 1+\varepsilon^2 \end{bmatrix}$$

To compute the eigenvalues of $(A^{-1})^T A^{-1}$,

Notice that

$$(A^{-1})^T A^{-1} = \frac{1}{4\varepsilon^2}\left( \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} + \begin{bmatrix} \varepsilon^2 & \varepsilon^2 \\ \varepsilon^2 & \varepsilon^2 \end{bmatrix} \right).$$

We know that the eigenvectors for

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ are $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$

with eigenvalues 2 and 0 respectively.

We can use this info to rewrite $(A^{-1})^T A^{-1}$, i.e. "diagonalize" it, as:

$$(A^{-1})^T A^{-1} = \frac{1}{4\varepsilon^2} Q \left( \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} + \varepsilon^2 \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \right) Q^T$$

with $Q = Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$.

Then, we can read off the max eigenvalue of $(A^{-1})^T A^{-1}$ as $\frac{1}{2\varepsilon^2}$, for $\varepsilon \ll 1$.

So $\|A^{-1}\|_2 = \left(\frac{1}{2\varepsilon^2}\right)^{1/2} = \frac{1}{\sqrt{2}\,\varepsilon}$.

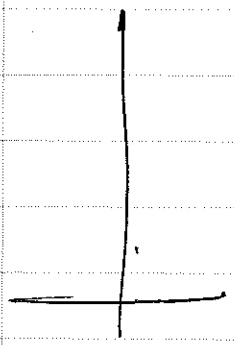
This implies:

$$K(A) = \|A\|_2 \, \|A^{-1}\|_2$$

$$= 1 \cdot \frac{1}{\sqrt{2}\,\varepsilon} .$$

Note that $\varepsilon \to 0$, i.e. the lines get flatter, $A$ becomes <u>ill-conditioned</u>.

---



# Intro to least squares

S'pose we have a matrix which is not necessarily square, $A \in \mathbb{R}^{m \times n}$.

(1) $m < n$: "underconstrained", i.e. more unknowns ($n$) than equations ($m$). Generically, this problem has infinitely many solutions.

(2) $m > n$: "over constrained", i.e. more equations ($m$) than unknowns ($n$). Generically, the problem has no solutions.

By "problem" above we mean the linear set of equations $A\underline{x} = \underline{b}$.

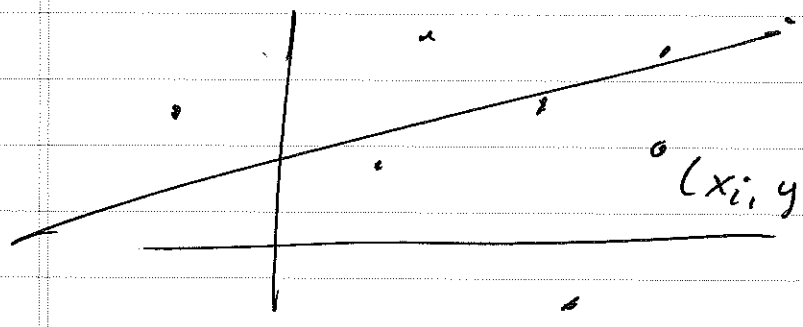
For case (2), we can try to find $\underline{x}$ so that $A\underline{x} \simeq \underline{b}$, i.e. they are "close" in some norm.

Example: (very common)

We are given a set of data points in the plane:

$$(x_i, y_i) \quad i = 1, \ldots, m.$$

We want to find a "line of best fit"




So, we seek an equation $y = \gamma x + \beta$

$\gamma$ = slope, $\beta$ = y-intercept

are the parameters we need to determine

we want $y_i \simeq \gamma x_i + \beta \quad i = 1, \ldots, m$,

or in matrix form

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{= \underline{b}} \simeq \underbrace{\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}}_{= A \in \mathbb{R}^{m \times 2}} \begin{bmatrix} \gamma \\ \beta \end{bmatrix}_{= \underline{x}}$$

One approach is to minimize a functional over possible values of $(\gamma, \beta) \in \mathbb{R}^2$.

$$\min_{\underline{x} \in \mathbb{R}^2} \| A\underline{x} - \underline{b} \|_2$$

$$= \min_{\underline{x} \in \mathbb{R}^2} f(\underline{x})^{1/2}.$$

We might as well minimize $f(\underline{x})$ directly to avoid the square root.

$$\min_{\underline{x} \in \mathbb{R}^2} f(\underline{x}) = \min_{\underline{x} \in \mathbb{R}^2} \| A\underline{x} - \underline{b} \|_2^2.$$

More generally stated, least squares is the problem of solving,

For $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$, $\underline{b} \in \mathbb{R}^m$:

$$\min_{\underline{x} \in \mathbb{R}^n} \| A\underline{x} - \underline{b} \|_2^2.$$

We can do some manipulations to solve this minimization problem:

$$\|A\underline{x} - \underline{b}\|_2^2 = (A\underline{x} - \underline{b})^T (A\underline{x} - \underline{b})$$

$$= (\underline{x}^T A^T - \underline{b}^T)(A\underline{x} - \underline{b})$$

$$= \underline{x}^T A^T A \underline{x} - \underline{x}^T A^T \underline{b} - \underline{b}^T A \underline{x} + \underline{b}^T \underline{b}.$$

$$= \underline{x}^T A^T A \underline{x} - 2\,\underline{x}^T A^T \underline{b} + \underline{b}^T \underline{b}.$$

So, we see that $f(\underline{x})$ is a quadratic function:

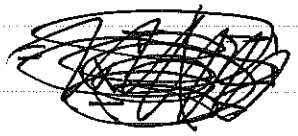
$$f(\underline{x}) = \underline{x}^T A^T A \underline{x} - 2\,\underline{x}^T A^T \underline{b} + \underline{b}^T \underline{b}.$$

At a minimize $x^*$ of $f$, it is necessary that $\nabla f(x^*) = 0$.

To compute $\nabla f(\underline{x})$, let

$$C = A^T A = (c_{ij})$$
$$\underline{d} = A^T \underline{b}.$$

Then $f(\underline{x}) = \underline{x}^T C \underline{x} - 2\,\underline{x}^T \underline{d} + \underline{b}^T \underline{b}.$



let's write out $\underline{x}^T C \underline{x}$ in component form:

First, $(C\underline{x})_k = \sum_{i=1}^{m} C_{ki}\, x_i$.

Then $\underline{x}^T C \underline{x} = \sum_{k=1}^{m} x_k (C\underline{x})_k$

$$= \sum_{k=1}^{m} \sum_{i=1}^{m} x_k C_{ki} x_i .$$

Now, we compute the contribution of this term to the $j$th component of $\nabla f$.

$$\frac{\partial}{\partial x_j}\left(\underline{x}^T C \underline{x}\right) = \frac{\partial}{\partial x_j}\left(\sum_{k=1}^{m}\sum_{i=1}^{m} x_k C_{ki} x_i\right).$$

$$= \sum_{k=1}^{m}\sum_{i=1}^{m}\left(\frac{\partial x_k}{\partial x_j} C_{ki} x_i + x_k C_{ki} \frac{\partial x_i}{\partial x_j}\right)$$

$$= \sum_{k=1}^{m}\sum_{i=1}^{m}\left(\delta_{kj} C_{ki} x_i + \delta_{ij} C_{ki} x_k\right).$$

$$= \sum_{i=1}^{m}\sum_{k=1}^{m} \delta_{kj} C_{ki} x_i + \sum_{k=1}^{m}\sum_{i=1}^{m} \delta_{ij} C_{ki} x_k$$

$$= \sum_{i=1}^{m} C_{ji} x_i + \sum_{k=1}^{m} C_{kj} x_k$$

Note that $C = A^T A$, so $C$ is symmetric $\Rightarrow C_{ij} = C_{ji}$. Then we can write:

$$\frac{\partial}{\partial x_j}\left(\underline{x}^T C \underline{x}\right) = 2 \sum_{i=1}^{m} C_{ji} x_i$$

$$= 2\left(A^T A \underline{x}\right)_j$$

Similarly, you can show:

$$\frac{\partial}{\partial x_j}\left(2\underline{x}^T \underline{d}\right) = 2 d_j$$

$$= 2\left(A^T \underline{b}\right)_j$$

and trivially $\frac{\partial}{\partial x_j}\left(b^T b\right) = 0$.

So: $$\frac{\partial f}{\partial x_j} = 2\left(A^T A \underline{x}\right)_j - 2\left(A^T \underline{b}\right)_j$$

implying that $\nabla f = A^T A \underline{x} - A^T \underline{b}$.

So then $\nabla f(\underline{x}^*) = 0$ is equivalent to

$$A^T A \underline{x}^* = A^T \underline{b}. \qquad (3).$$

(3) is called the <u>normal equations</u>.

The solution to the normal equations $\underline{x}^*$ will minimize the function $f(\underline{x})$:

$$f(\underline{x}) = \| A\underline{x} - \underline{b} \|_2^2.$$

(!!) Nice geometric picture, a la. Fundamental Theorem of Linear Algebra. (very important).

Note that by (3), we have

$$A^T (A\underline{x}^* - \underline{b}) = 0.$$

Take any $\underline{y} \in \mathbb{R}^n$, then

$$\underline{y}^T A^T (A\underline{x}^* - \underline{b}) = 0.$$

This statement says that

$A\underline{x}^* - \underline{b}$ is orthogonal to the

Range of $A$, i.e. $\text{Ran}(A)$.

Or, stated another way,

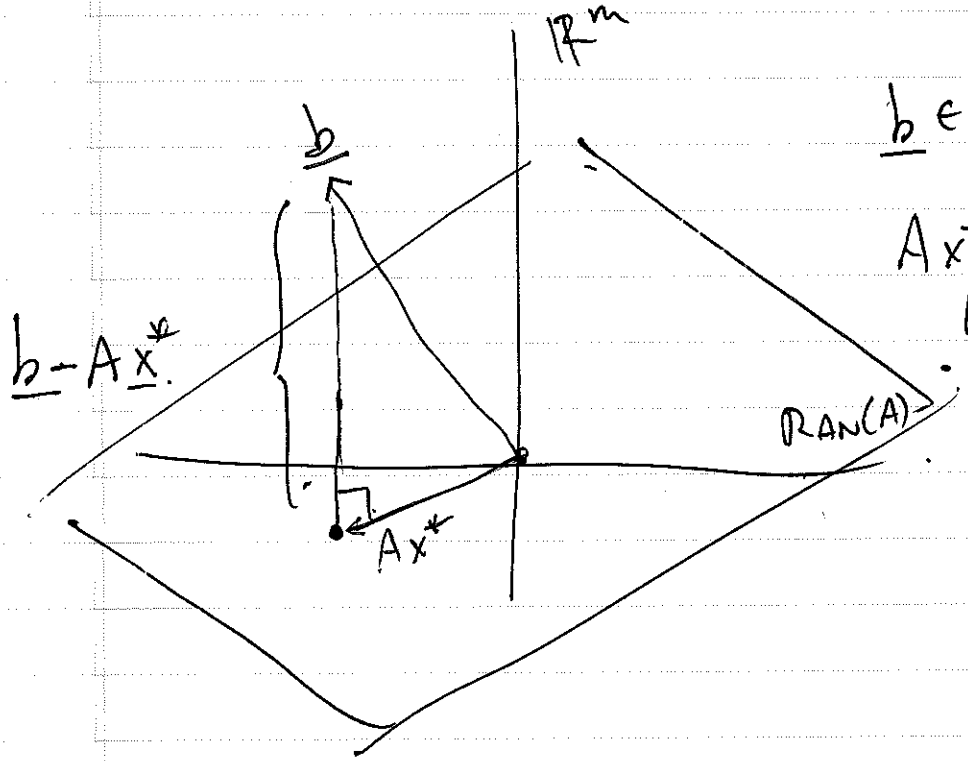
$$Ax^* - b \in \ker(A^T),$$

but $\ker(A^T) \oplus \text{RAN}(A) = \mathbb{R}^m$

by the Fundamental Theorem of Linear Algebra, so then

$$A\underline{x}^* - \underline{b} \perp \underline{z}$$

For any vector $\underline{z} \in \text{Ran}(A)$.




$\underline{b} \in \mathbb{R}^m$.

$Ax^* \in \mathbb{R}^m$, and by def'n $Ax^* \in \text{Ran}(A)$.

Example: Setting up normal equations for least squares problem.

let's say we are given a set of data

$$(1, 1) = (x_1, y_1)$$
$$(4, 1) = (x_2, y_2)$$
$$(9, 2) = (x_3, y_3)$$

And we want to fit a function to this data of the form

$$y = \alpha x^{1/2} + \beta.$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \approx \begin{bmatrix} x_1^{1/2} & 1 \\ x_2^{1/2} & 1 \\ x_3^{1/2} & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \underline{x}.$$

$$= \underline{b} \qquad = \underline{\underline{A}}.$$

$$\underline{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 2\sqrt{2} & 1 \end{bmatrix}.$$

Normal equations are

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

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 2\sqrt{2} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 + 2\sqrt{2} & 3 \\ 13 & 3 + 2\sqrt{2} \end{bmatrix}.$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 + 2\sqrt{2} \end{bmatrix}$$

Then $\underline{x}^*$ satisfying $\nabla f(x^*) = 0$

solves the system:

$$\begin{bmatrix} 3+2\sqrt{2} & 3 \\ 13 & 3+2\sqrt{2} \end{bmatrix} \underline{x}^* = \begin{bmatrix} 4 \\ 3+2\sqrt{2} \end{bmatrix}$$

Thm 2.12: S'pose $A \in \mathbb{R}^{m \times n}$, $m \geq n$.

Then $A$ can be written in the form

$$A = \hat{Q}\,\hat{R},$$

where $\hat{R}$ is upper triangular, $n \times n$ matrix, and $\hat{Q}$ is an $m \times n$ which satisfies

$$\hat{Q}^T \hat{Q} = I_n = n \times n \text{ identity.}$$

If $\text{rank}(A) = n$, then $\hat{R}$ is nonsingular.

$\square$

We can use the QR factorization to stably compute a solution to the least squares problem.