When solving \( Ax = b \), the \textit{absolute} condition number:

\[
\| x - x' \| \leq \| A \| \cdot \| x - x' \| \leq \| A^{-1} \| \cdot \| b - b' \|
\]

Relative condition number:

\[
\frac{\| x - x' \|}{\| x \|} \leq \frac{\| A \| \cdot \| A^{-1} \| \cdot \| b - b' \|}{\| b \|}
\]

The most important case: \( \| \cdot \| = \| \cdot \|_2 \).

\[
\kappa_2(A) = \sqrt{\frac{\lambda_1}{\lambda_n}} \quad \text{where} \quad \lambda_1 \text{ is largest eigenvalue of } A^T A
\]

\( \lambda_n \) is the smallest eigenvalue.

Write \( A \) in SVD form:

\[
A = U \Sigma V^T
\]

\( U \) and \( V \) are orthogonal matrices.

Then \( A^T A = V \Sigma^T U^T U \Sigma V^T = V \Sigma^2 V^T \)

Eigenvalues of \( A^T A \) \( \Rightarrow \) \( \sigma_1^2 > \sigma_2^2 > \ldots > \sigma_n^2 \)

So therefore:

\[
\kappa_2(A) = \sqrt{\frac{\sigma_1}{\sigma_n}} = \sqrt{\frac{\sigma_1^2}{\sigma_n^2}} = \frac{\sigma_1}{\sigma_n}
\]
Consequences of $K_2(A)$:

$$\frac{\|\bar{x} - \bar{x}'\|}{\|\bar{x}\|} \leq K_2(A) \cdot \frac{\|\tilde{b} - \tilde{b}'\|}{\|\tilde{b}\|}$$

True problem is $A\bar{x} = \tilde{b}$.

Imagine that $\tilde{b}'$ is the floating point representation of $\tilde{b}$, meaning $\tilde{b}' = \text{round}(\tilde{b})$.

If machine precision is $\varepsilon$, then $\|\tilde{b} - \tilde{b}'\| \sim \varepsilon \|\tilde{b}\| \sim \varepsilon \sim 10^{-16}$.

Hence, $\frac{\|\bar{x} - \bar{x}'\|}{\|\bar{x}\|} \leq K_2(A) \cdot \varepsilon$.

The number of significant digits lost in solving $A\bar{x} = \tilde{b}$ is $\sim \log_{10}(K_2(A) \cdot \varepsilon)$.

Ex: Double precision floating point $\Rightarrow \varepsilon \sim 10^{-16}$

Compute $K_2 = 10^{10}$

$$\Rightarrow \frac{\|\bar{x} - \bar{x}'\|}{\|\bar{x}\|} \leq \varepsilon \cdot K_2 = 10^{-6}$$

This doesn't mean that $\frac{\|\bar{x} - \bar{x}'\|}{\|\bar{x}\|}$ cannot be smaller, but it puts a bound on how bad it can be.
**Least Squares**

Two canonical problems in linear algebra:

1. Solve $A\hat{x} = b$ where $A$ is a square $m \times m$ matrix.

2. Find "the best" solution to a system $A\hat{x} = b$ where $A$ is an $m \times n$ matrix.

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

In general, not solvable because $m > n$.

One version of "the best" solution is the least squares solution:

**Least squares:** For $A \in \mathbb{R}^{m \times m}$, with $m > n$, find $\hat{x}$ such that $\|A\hat{x} - b\|_2$ is as small as possible.

The fact that this is the 2-norm is important.

**Example:** $A \in \mathbb{R}^{3 \times 2}$

Geometrically, minimize the distance between $b$ and $A\hat{x}$.

We also have that

\[
\|A\hat{x} - b\|_2^2 = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2 = f(\hat{x})
\]

If all you knew was calculus, then you could solve this by finding the solution to

\[
\begin{align*}
\frac{df}{dx_i} &= 0 \\
\frac{df}{dx_j} &= 0
\end{align*}
\]

Finding a zero of $f'$ is equivalent to minimizing or maximizing.
Let \( \hat{r} = b - Ax \) = residual vector.

\[ \Rightarrow \| \hat{r} \| = \text{distance from } \text{col}(A) \text{ to } b. \]

If \( \hat{x} \) is the least square solution, then \( \hat{r} \perp \text{col}A \), in particular, \( A^T \hat{r} = 0 \).

\[ \Rightarrow A^T \hat{r} = 0 = A^T (b - Ax) \]
\[ = A^T b - A^T Ax \]
\[ \Rightarrow \text{The least square solution } \hat{x} \text{ solves } A^T A \hat{x} = A^T b. \]

From a numerical point of view, how best to find \( \hat{x} \) to \( \min \| A\hat{x} - b \|_2^2 \)?

Options: (1) Solve the normal equations:
\[ A^T A \hat{x} = A^T b \]

\[ \text{Drawback: Solving } A^T A \hat{x} = A^T b \text{ has a condition number that is the square of } \frac{\| \hat{r} \|}{\| \hat{x} \|}. \]

Example: If \( A \) is \( mn \times n \) with rank \( n \), then
\[ A = U_{mn \times mn} S_{mn \times n} V_{n \times n}^T \Rightarrow \text{cond}(A) = \frac{\sigma_1}{\sigma_n} \]
\[ A^T A = V S^2 V^T \Rightarrow \text{cond}(A^T A) = \frac{\sigma_2^2}{\sigma_n^2} \]

So if \( \text{cond}(A) = 10^4 \), then \( \text{cond}(A^T A) = 10^8 \).

Option (2) Use calculus to \( \min \| A\hat{x} - b \| \)
Goal: \( \min \| Ax - b \| \) without solving the normal equations.

\[ \nabla^2 \mathbf{x} = \mathbf{b} - \mathbf{A}^2 \mathbf{x} \]

closest point

col A
to \( \mathbf{b} \)

Idea: Construct a linear system that is consistent (i.e., has a solution), without computing \( \mathbf{A}^2 \mathbf{A} \).

Instead of solving \( \mathbf{A} \mathbf{x} = \mathbf{b} \) (which has no solution), solve \( \mathbf{A} \mathbf{\hat{x}} = \mathbf{\tilde{b}} \leftarrow \mathbf{\tilde{b}} = \text{orthogonal projection of } \mathbf{b} \) onto the column space of \( \mathbf{A} \).

\[ \mathbf{A} \mathbf{\hat{x}} = \text{proj}_{\text{col} \mathbf{A}} \mathbf{\tilde{b}} \leftarrow \mathbf{\tilde{b}} \]

How do we compute \( \text{proj}_{\text{col} \mathbf{A}} \mathbf{\tilde{b}} \)?

How do we compute a projection of one vector onto another?

In 2D,

The orthogonal projection of \( \mathbf{\tilde{y}} \) onto \( \mathbf{\hat{x}} \) is the component of \( \mathbf{\tilde{y}} \) pointing in the \( \mathbf{\hat{x}} \) direction.

\[
\text{proj}_{\mathbf{\hat{x}}} \mathbf{\tilde{y}} = \begin{pmatrix} \mathbf{\tilde{y}} \cdot \mathbf{\hat{x}} \end{pmatrix} \begin{pmatrix} \mathbf{\hat{x}} \end{pmatrix} / \| \mathbf{\hat{x}} \|^2
\]

unit vector pointing in the direction of \( \mathbf{\hat{x}} \).

In 3D:

\[
\text{proj}_{\mathbf{\hat{x}}} \mathbf{\tilde{y}} = \text{the component of } \mathbf{\tilde{y}} \text{ in the } \mathbf{\hat{x}} \text{-plane.}
\]

\[
\text{proj}_{\mathbf{\hat{x}}} \mathbf{\tilde{y}} = (\mathbf{\tilde{y}}, \mathbf{\hat{x}}) \mathbf{\hat{x}} + (\mathbf{\tilde{y}} - (\mathbf{\tilde{y}}, \mathbf{\hat{x}}) \mathbf{\hat{x}}) \mathbf{\hat{x}}
\]

\[
\text{proj}_{\mathbf{\hat{x}}} \mathbf{\tilde{y}} = \mathbf{\hat{x}} + \mathbf{\tilde{y}} - (\mathbf{\tilde{y}}, \mathbf{\hat{x}}) \mathbf{\hat{x}}
\]

the remaining part of \( \mathbf{\tilde{y}} \) after projecting onto \( \mathbf{\hat{x}} \).

If \( (\mathbf{\hat{x}}, \mathbf{\hat{z}}) = 0 \) (orthogonal), then:

\[
\text{proj}_{\mathbf{\hat{x}}} \mathbf{\tilde{y}} = (\mathbf{\tilde{y}}, \mathbf{\hat{x}}) \mathbf{\hat{x}} + (\mathbf{\tilde{y}}, \mathbf{\hat{z}}) \mathbf{\hat{z}}
\]

Gram Schmidt process.
So this suggests that we want to find an orthogonal basis for \( \text{col}(A) \), \( U = \text{span}\{\hat{u}_1, \ldots, \hat{u}_n\} \) and then project \( \tilde{b}' \) onto \( U \), forming \( \hat{b}' \).

\[
\tilde{b}' = (\hat{b}, \hat{u}_1)\hat{u}_1 + (\hat{b}, \hat{u}_2)\hat{u}_2 + \ldots + (\hat{b}, \hat{u}_n)\hat{u}_n
\]

\[
= \begin{pmatrix} \hat{u}_1 & \hat{u}_2 & \ldots & \hat{u}_n \end{pmatrix} \begin{pmatrix} (\hat{b}, \hat{u}_1) \\ (\hat{b}, \hat{u}_2) \\ \vdots \\ (\hat{b}, \hat{u}_n) \end{pmatrix} \tilde{b}
\]

\[
= U \hat{U}^T \tilde{b}
\]

orthogonal projector of \( \tilde{b} \) onto the column space of \( A \).

\( \Rightarrow \) The linear system \( AX = U \hat{U}^T \tilde{b} \) is consistent.