Last Time:

Matrix conditioning: \( \text{For } Ax = b \)\n
\[ \| x - x' \| = \| A b - A' b' \| \quad \text{any norm} \]

\[ \leq \| A^T \| \| b - b' \| \]

Absolute condition number.

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

\( K(A) \)

In the 2-norm, recall that

\[ \| A \|_2 = \max \sqrt{\lambda(A^T A)} \leq \text{largest eigenvalue of } A^T A. \]

\( \Rightarrow \) If \( A = U S V^T \) by SVD, then

\[ A^T A = V S^2 V^T, \quad \Rightarrow \| A \|_2 = \sigma_1 \leq \text{largest singular value of } A. \]

\[ A^{-1} = V S^{-1} U^T, \quad \Rightarrow \| A^{-1} \|_2 = \frac{1}{\sigma_n} \leq \sigma_n \text{ is smallest singular value of } A. \]

\( \Rightarrow \) \( K_2(A) = \frac{\sigma_1}{\sigma_n} \)
Consequences of $K_2$: Recall:

\[
\frac{\|x - x'\|}{\|x\|} \leq K_2 \frac{\|b - b'\|}{\|b\|}
\]

Imagine $b'$ is the rounded version of $b$, so then:

\[
\frac{\|b - b'\|}{\|b\|} \sim O(e)
\]

which implies

\[
x' \text{ is the solution to } Ax' = b'
\]

This means that the "number of correct digits in } x'\" is bounded by $-\log K_2 e$.

\[
\text{Ex: } e \approx 10^{-16}
\]

\[
K_2 \approx 10^{10}
\]

\[
\Rightarrow \frac{\|x - x'\|}{\|x\|} \leq 10^4 \Rightarrow \text{ we may expect lose log } K_2 \text{ digits in solving } Ax = b \text{, this does not mean that } \text{the } \|r = b' - Ax'\| \text{ is large.}
\]

Depends on \underline{backward stability}.\]
The SVD

The factorization \( A = U S V^T \) exists for any matrix \( A \), with \( U, V \) orthogonal and \( S = \text{diag} (\sigma_1, \ldots, \sigma_n) \), \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0 \) (obviously real).

**Proof by construction:**

Let \( \nu_j \) be the eigenvectors of \( A^T A \) (assume \( n \) distinct \( \lambda_j > 0 \) for now).

Then \( (AV_i, AV_j) = (V_i, A^T A V_j) = (V_i, \lambda_j V_j) = \lambda_j (V_i, V_j) \)

\[ = 0 \quad \text{since} \]

\( A^T A \) is real symmetric (HW question).

This means that \( (AV_i, AV_j) = 0 \), i.e.

\( AV_i, AV_j \) are orthogonal, Each is in the column space of \( A \) (obviously), and therefore form a basis for \( \text{col}(A) \).

\[ \text{col}(A) = \text{span} \left\{ AV_1, \ldots, AV_n \right\}. \]

Also, \( \| AV_i \|^2 = (AV_i, AV_i) = (V_i, \lambda_i V_i) = \lambda_i. \)

Set \( U_j = \frac{1}{\lambda_j} AV_j. \)
Set the matrix:

\[ U = \begin{pmatrix} u_1 & u_2 & \ldots & u_n \end{pmatrix} \]

\[ V = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} \]

\[ S = \begin{pmatrix} \sigma_1 & \sigma_2 & \ldots & \sigma_m \\ \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_2 & \ldots & \sigma_m \end{pmatrix} \]

Then trivially,

\[ AV = A \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} \]

\[ = (Av_1, \ldots, Av_n) = S \begin{pmatrix} \sigma_1 & \sigma_2 & \ldots & \sigma_m \\ \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_2 & \ldots & \sigma_m \end{pmatrix} \]

\[ = (Av_1, \ldots, Av_n) \]

so \( A = USV^T \).

What if \( A \) is not full rank? \( \dim \text{col}(A) < n \).

**Truncated SVD**

Then \( A = U_{nnk} \begin{pmatrix} \Sigma_{nnk} \end{pmatrix} V_{nnk}^T \)
We will discuss methods for computing the SVD after eigenvals, least square. (Computing $ATA$ is not recommended.)

**Next: Least squares:**

Often, $A$ is not square, but we would still like to find the "best" solution to $A\mathbf{x} = \mathbf{b}$.

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

**Geometrically:** Minimize distance between $\mathbf{b}$ and $A\mathbf{x}$.

We have that $\| A\mathbf{x} - \mathbf{b} \|_2^2$ (2-norm notation suppressed):

$$
= \sum_{i=1}^{n} \left( b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2
$$

Many other interpretations/applications:

- Calculus
- Statistics
You probably learned that least squares problems can be solved by instead solving:

\[ A^T A x = A^T b \]  (the normal equation)

but this squares the condition number!

(Aside: the normal equation can be obtained by computing the minimum of \( \| A x - b \|_2^2 \) by taking partial derivatives, etc.)

Also:

\[ r = b - A x \]

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

\[ A^T r = 0 = A^T b - A^T A x \]

\[ A^T A x = A^T b \] has at least one solution.

How else can least squares be solved? Another matrix factorization... QR.