

Why matrix norm is important?

Note Title

LECTURE 06

you know that a computer cannot represent real numbers exactly unless they are dyadic numbers.

1. $d_1 d_2 \dots d_t \times 2^e$

So, suppose you want to solve

$$A x = b, \quad A \in \mathbb{R}^{m \times m}$$

But in reality, you have to encode A, x, b on the computer.

Let $A' = fl(A)$, $x' = fl(x)$, $b' = fl(b)$.

i.e., you end up solving

$$A' x' = b'$$

Suppose for the moment, let's assume $b' = b$ for simplicity.

Now, you want to know the relative error of the solution:

$$\begin{aligned} \frac{\|x' - x\|}{\|x'\|} &= \frac{\|x' - A^{-1}b\|}{\|x'\|} \\ \text{rel. error in solution} &= \frac{\|x' - A^{-1}A'x'\|}{\|x'\|} \\ &= \frac{\|A^{-1}(A - A')x'\|}{\|x'\|} \\ &\leq \frac{\|A^{-1}\| \cdot \|A - A'\| \cdot \|x'\|}{\|x'\|} \end{aligned}$$

LECTURE 06

$$= \underbrace{\|A\| \|A^{-1}\|}_{\text{amplification factor}} \underbrace{\frac{\|A - A'\|}{\|A\|}}_{\text{relative error in matrix}}$$

Now define the **condition number** of A by

$$\kappa(A) = \text{cond}(A) := \|A\| \cdot \|A^{-1}\|$$

If $\kappa(A)$ is large, then A is pretty bad, i.e., $\exists$ large error in solution $x = A^{-1}b$.

- Roughly speaking, to compute A^{-1} or the solution of $Ax = b$, we lose $\approx \log_{10} \kappa(A)$ digits.
- In particular, if A is singular, $\kappa(A) = +\infty$.

Least Squares Problem

Since the error analysis of Gaussian elimination & LU decomposition are subtle and difficult, we'll first talk about the least squares problem, then talk about the projections, QR decomposition, etc.

The Least Squares Problem was conceived by Gauss and Legendre around 1800 in the fields of astronomy & geodesy, in particular, model fitting to measured data.

Want to solve

$$(*) \quad A x = b$$

where $A \in \mathbb{R}^{m \times n}$, $m > n$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ overdetermined
more equations than unknowns

In general, (*) has no exact solution unless $b \in \text{range}(A)$.

this usually does not happen!

$\Rightarrow$ Check the size of the residual

$$r := b - Ax \in \mathbb{R}^m$$

and want r as small as possible.

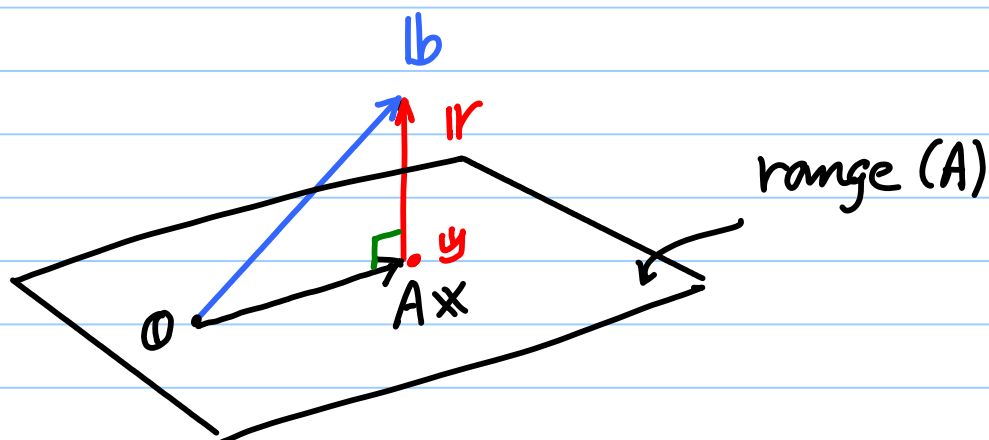
Given $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $b \in \mathbb{R}^m$
Find $x \in \mathbb{R}^n$ s.t. $\|b - Ax\|_2 \rightarrow \min$

LECTURE 06

This is called a general linear least squares problem.

Why 2-norm is used?

⇒ its geometric interpretation!



Thm. Let $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $b \in \mathbb{R}^m$.

Then, $x \in \mathbb{R}^n$ is the minimizer of

$$\|r\|_2 = \|b - Ax\|_2$$

$$\Leftrightarrow r \perp \text{range}(A)$$

$$\Leftrightarrow A^T r = 0$$

$$\Leftrightarrow A^T A x = A^T b \quad (\text{The normal equation})$$

$$\Leftrightarrow A(A^T A)^{-1} A^T b = Ax$$

Furthermore,

$A^T A$ is nonsingular $\Leftrightarrow A$ is full rank

Consequently, $\text{rank}(A) = n$ if $m \geq n$

The solution x is unique $\Leftrightarrow A$: full rank

LECTURE 06

(Proof) These statements are essentially obvious from the figure.

also $\mathbf{r} \perp \text{range}(A)$

$$\Leftrightarrow \mathbf{r} \perp \mathbf{a}_j \quad 1 \leq j \leq n$$

$$\Leftrightarrow \mathbf{r}^T [\mathbf{a}_1 \cdots \mathbf{a}_n] = \mathbf{0}$$

$$\Leftrightarrow \mathbf{r}^T A = \mathbf{0} \quad \Leftrightarrow A^T \mathbf{r} = \mathbf{0}$$

$$\begin{aligned} \text{Then } A^T \mathbf{r} &= A^T (\mathbf{b} - A\mathbf{x}) \\ &= A^T \mathbf{b} - A^T A \mathbf{x} \\ &= \mathbf{0} \end{aligned}$$

$$\Leftrightarrow A^T A \mathbf{x} = A^T \mathbf{b} \quad \checkmark$$

Now we can also show the uniqueness of the **orthogonal projection** of $\mathbf{b}$ onto $\text{range}(A)$ as follows:

$$\text{Let } \mathbf{y} = A (A^T A)^{-1} A^T \mathbf{b} \in \text{range}(A)$$

which obviously minimize $\underbrace{\|\mathbf{b} - \mathbf{y}\|_2}_{= \mathbf{r}}$

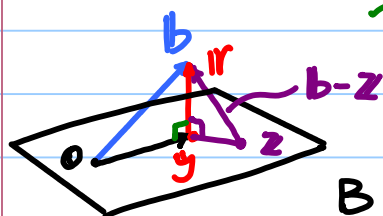
Suppose $\exists \mathbf{z} \neq \mathbf{y}$ s.t.

$$\|\mathbf{b} - \mathbf{z}\|_2 = \|\mathbf{b} - \mathbf{y}\|_2, \quad \mathbf{z} \in \text{range}(A)$$

Then, $\mathbf{y} - \mathbf{z} \in \text{range}(A)$.

So, $\mathbf{y} - \mathbf{z} \perp \mathbf{b} - \mathbf{y}$

$$\Leftrightarrow \underbrace{\|\mathbf{b} - \mathbf{z}\|_2^2}_{\uparrow} = \underbrace{\|\mathbf{b} - \mathbf{y}\|_2^2}_{\text{Pythagoras!}} + \|\mathbf{y} - \mathbf{z}\|_2^2$$



But these two are equal by the assumption.

$$\text{so } \|\mathbf{y} - \mathbf{z}\|_2 = 0 \Leftrightarrow \mathbf{y} = \mathbf{z} \quad \#$$

LECTURE 06

- $(A^T A)^{-1} A^T$ is often called pseudo inverse of A , and denoted by A^+ infinitely many sol's.
- What happens if $m < n$? ↓ sol's.
This case is called underdetermined.
Need extra constraints to solve such LS problem, e.g., $\min \|x\|_2$
That is,
Find $x \in \mathbb{R}^n$, s.t.
 $\min \|x\|_2^2$ subject to $Ax = b$.

This is done by the Lagrange multipliers:

$$\text{Let } J(x) := x^T x + \lambda^T (b - Ax)$$

$$\lambda \in \mathbb{R}^m$$

Then want $\min J(x)$.

$$\frac{\partial J}{\partial x} = 2x - A^T \lambda = 0$$

$$\Leftrightarrow \hat{x} = \frac{1}{2} A^T \lambda, \text{ this } \hat{x} \text{ minimizes } J(x)$$

$$\text{Now, } b = Ax = \frac{1}{2} A A^T \lambda$$

$$\Leftrightarrow \lambda = 2 (A A^T)^{-1} b$$

$$\Leftrightarrow \hat{x} = \frac{1}{2} A^T \lambda = \underline{A^T (A A^T)^{-1} b}$$

compare this with $(A^T A)^{-1} A^T b$
in the case of $m \geq n$.

LECTURE 06

Example The LS polynomial fit

Given m distinct points $x_1, \dots, x_m \in \mathbb{R}$
and data $y_1, \dots, y_m \in \mathbb{R}$
Want to fit a polynomial of deg. $n-1$
$$p(x) = c_0 + c_1 x + \dots + c_{n-1} x^{n-1}$$

for some $n < m$.

Such a polynomial is a LS fit to the data if it minimizes the residual

$$(*) \quad \sum_{i=1}^m |p(x_i) - y_i|^2.$$

So,

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_{n-1} \end{bmatrix} \approx \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

A x b

vandermonde! $(*) = \|r\|_2^2 = \|b - Ax\|_2^2$

Julia Demo here!