Conditioning and stability

Zecheng Zhang

March 27, 2023

In the abstract, we can view a problem as $f : X \to Y$ where X, Y are two spaces. A well-conditioned problem is one with the property that all small perturbation of x lead to only small changes in $f(x)$.

1 Relative condition number

Denote $\delta f = f(x + \delta x) - f(x)$. The relative conditioning number is defined as

$$\kappa(x) = \lim_{\delta \to 0} \sup_{\|\delta x\| \le \delta} \left(\frac{\|\delta f\|}{\|f(x)\|} / \frac{\|\delta x\|}{\|x\|}\right). \tag{1}$$

One can assume δx and δf are infinitesimal, then

$$\kappa(x) = \sup_{\|\delta x\|} \left(\frac{\|\delta f\|}{\|f(x)\|} / \frac{\|\delta x\|}{\|x\|}\right). \tag{2}$$

When f is differentiable, we can express the quantity in terms of the Jacobian of f,

$$\kappa = \frac{\|J(x)\|}{\|f(x)\| / \|x\|}. \tag{3}$$

A problem is well-conditioned if κ is small (e.g., 1, 10, 100), and a problem is ill-conditioned if κ is large (e.g., 10^6 or bigger).

Example 1.1. Consider $x \to x/2$.

conditioning number:
relative change in f / relative change in x.

Example 1.2. Consider $x \to \sqrt{x}$, $x > 0$.

Example 1.3. Consider $f(x) = x_1 - x_2$.

2 Conditioning of matrix multiplication

Let $A \in \mathbb{R}^{m \times n}$, we consider the problem of computing Ax given a x. We want to know how Ax will change if there is a perturbation in x. The conditioning number of A is defined as,

$$\kappa = \sup_{\delta x} \left(\frac{\|A(x + \delta x) - Ax\|}{\|Ax\|} \Big/ \frac{\|\delta x\|}{\|x\|}\right) = \sup_{\delta x} \frac{\|A\delta x\|}{\|\delta x\|} \Big/ \frac{\|Ax\|}{\|x\|}. \tag{4}$$

Note that sup is over all δx and $\frac{\|Ax\|}{\|x\|}$ is independent with respect to sup, it follows that,

$$\kappa = \frac{\|x\|}{\|Ax\|} \sup_{\delta x} \frac{\|A\delta x\|}{\|\delta x\|} = \|A\| \frac{\|x\|}{\|Ax\|}, \tag{5}$$

where $\|A\|$ is the operator norm (it is L_2 norm if $\|\cdot\|$ is the L_2 vector norm). Note that, the condition number depends both on A and x.

2. Conditioning of matrix multiplication.

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

change in f

change in x

$$\kappa = \sup_{\delta x} \frac{\| A(x+\delta x) - Ax \|}{\| Ax \|} \Big/ \frac{\| \delta x \|}{\| x \|}$$

$$= \underbrace{\sup_{\delta x > 0} \frac{\| A \cdot \delta x \|}{\| \delta x \|}} \Big/ \frac{\| Ax \|}{\| x \|}$$

$$= \| A \| \cdot \frac{\| x \|}{\| Ax \|}$$

Remark: if $\| A \cdot \delta x \|_{L^2}$, $\| \delta x \|_{L^2}$, $\Rightarrow \| A \|_{L^2} = \sigma_1$

Remark: Suppose $A \in \mathbb{R}^{m \times m}$, rank $(A) = m$

$$\| x \| = \| A^{-1} A x \| \le \| A^{-1} \| \cdot \| Ax \|$$

$$\frac{\| x \|}{\| Ax \|} \le \| A^{-1} \|$$

$$\Rightarrow \underbrace{\kappa(x) = \| A \| \cdot \frac{\| x \|}{\| Ax \|}}_{(*)} \le \| A \| \cdot \| A^{-1} \|$$

Remark:

Suppose $\|\text{vector}\|_{L_2}$, $\|A\|_{L^2}$, $A \in \mathbb{R}^{mm}$

want to show $k = \|A\| \cdot \|A^{-1}\|$

$\|V_m\| = 1 = |\lambda| \cdot \|x\|$

Take $V_m = \lambda x$, $\lambda \in \mathbb{R}$, $\lambda \neq 0$, V_m is m-th right singular vector of A.

$$A V_m = \sigma_m u_m$$

$$A \lambda x = \sigma_m u_m$$

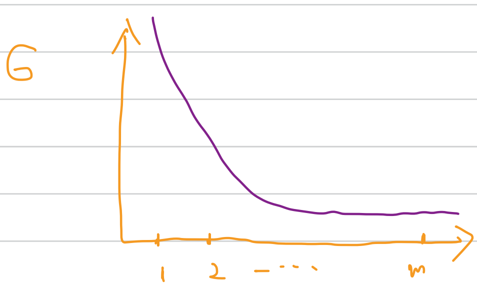
$$A x = \frac{\sigma_m}{\lambda} u_m$$

$$\|Ax\| = \frac{\sigma_m}{|\lambda|} \|u_m\| = \frac{\sigma_m}{|\lambda|}$$

substitute back into:

$$k = \|A\| \cdot \frac{\|x\|}{\|Ax\|} = \|A\| \cdot \frac{\|x\| \cdot |\lambda|}{\sigma_m} = \|A\| \cdot \frac{1}{\sigma_m}$$

$\|A\|_{L^2}$ = the 1st singular value. $= \dfrac{\sigma_1}{\sigma_m}$



If we have a fast decay in singular values, by POD, we only need to use a few u_i to represent the col(A).

$\longrightarrow$ the problem is not well-conditioned.

The only thing left to prove is $\|A^{-1}\| = \frac{1}{\sigma_m}$.

$$A v_i = \sigma_i u_i ,$$ $v_i\ \sigma_i\ u_i$ are standard singular xxx

$$v_i = \sigma_i A^{-1} u_i$$

$$A^{-1} u_i = \frac{1}{\sigma_i} v_i ,$$ $\Rightarrow$ $\frac{1}{\sigma_i}$ is a singular value of A^{-1}

we know $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m > 0$

$$\frac{1}{\sigma_1} \leq \dots \leq \frac{1}{\sigma_m}$$

$$\Rightarrow \quad \|A^{-1}\|_{L^2} = \frac{1}{\sigma_m}.$$

no assumption on A

$$\Rightarrow \quad k = \|A\| \cdot \|A^{-1}\|$$ $\rightsquigarrow$ definition: conditioning number of A (relative to $\|\cdot\|$).

Thm 2.1.

$A \in \mathbb{R}^{m \times m}$ invertible.

Last remark.

$A \in \mathbb{R}^{m\times n}$, $\operatorname{rank}(A) = n$. $m \geq n$

$$\kappa \leq \|A\| \cdot \|A^+\|, \text{ where}$$

$$A^+ = (A^* A)^{-1} A^* \longrightarrow \text{pseudo-inverse of } A.$$

motivation: $(A^* A) x = A^* b$

$$x = (A^* A)^{-1} A^* \cdot b$$

From $(*)$, $\kappa = \|A\| \cdot \dfrac{\|x\|}{\|Ax\|}$

$$\|x\| = \|\underbrace{(A^* A)^{-1} A^* \cdot A}_{I} x\| \leq \|A^+\| \cdot \|Ax\|$$

substitute back into $(*)$,

$$\kappa \leq \|A\| \cdot \|A^+\|.$$

Remark 1. Suppose A is nonsingular square matrix. We have $\|x\| = \|A^{-1}Ax\| \le \|A^{-1}\|\|Ax\|$, this further implies that,

$$\kappa \le \|A\|\|A^{-1}\|, \tag{6}$$

or

$$\kappa = c\|A\|\|A^{-1}\|, \tag{7}$$

for some positive constant $c = \frac{\|x\|}{\|Ax\|}/\|A^{-1}\|$.

Theorem 2.1. Let $A \in \mathbb{R}^{m\times n}$ be invertiable and let us consider $Ax = b$. The problem of computing b given x has conditioning number,

Ax (matrix multiplication)

$$\kappa = \|A\|\frac{\|x\|}{\|b\|} \le \|A\|\|A^{-1}\|, \tag{8}$$

$x = A^{-1}b$

with the perturbation in x. The problem of computing x given b has the conditioning number,

$$\kappa = \|A^{-1}\|\frac{\|b\|}{\|x\|} \le \|A^{-1}\|\|A\|, \tag{9}$$

remark

with the perturbation in b. If we use the L_2 norm, the first equality holds if x is a multiple of a right singular vector of A corresponding to the minimal singular value. The second equality holds if b is a multiple of a left singular vector of A corresponding to the largest singular value.

Definition 2.2. We will call $\kappa(A) = \|A\|\|A^{-1}\|$ the condition of A relative to norm $\|\cdot\|$ and denote it as $\kappa(A) = \|A\|\|A^{-1}\|$. The conditioning number is attached to matrix A not to the problem and x. If $\kappa(A)$ is small, A is called well-conditioned, otherwise, it is called ill-conditioned. If A is singular, we write $\kappa(A) = \infty$.

Remark 2. If $\|\cdot\| = \|\cdot\|_2$, $\|A\| = \sigma_1$ and $\|A^{-1}\| = 1/\sigma_m$, it follows that $\kappa(A) = \frac{\sigma_1}{\sigma_m}$

Remark 3. When $A \in \mathbb{C}^{m\times n}$ of full rank and $m \ge n$. The conditioning number is defined in terms of the pseudo-inverse, i.e.,

$$\kappa(A) = \|A\|\|A^{+}\|, \tag{10}$$

where $A^{+} = (A^{*}A)^{-1}A^{*}$ is called the pseudo-inverse of A.

3 Conditioning of a system of eqautions

We considered the case when A is fixed and perturbed x or b. What if we perturb A? Specifically, b is fixed and let us consider solving x from $Ax = b$ given a small change in A, We have,

$$(A + \delta A)(x + \delta x) = b \tag{11}$$
$$Ax + A\delta x + \delta Ax + \delta A\delta x = b. \tag{12}$$

Using $Ax = b$ and dropping the high order infinitesimal $\delta A\delta x$, it follows that $A\delta x + \delta Ax = 0$, or $\delta x = -A^{-1}\delta Ax$. Taking a norm, $\|\delta x\| \le \|A^{-1}\|\|\delta A\|\|x\|$, or,

$$\frac{\|\delta x\|}{\|x\|}/\frac{\|\delta A\|}{\|A\|} \le \|A^{-1}\|\|A\| = \kappa(A). \tag{13}$$

Equality holds when $\|\delta x\| = \|A^{-1}\|\|\delta A\|\|x\|$. It can be shown that for any A and b such δA exists. This leads us to the following result.

Theorem 3.1. Let b be fixed and consider the problem $x = A^{-1}b$, where A is nonsingular. The conditioning number associated with this problem with respect to perturbation in A is:

$$\kappa = \|A\|\|A^{-1}\| = \kappa(A). \tag{14}$$

Section 3,

$Ax = b$, we have perturbation in A, but b is fixed.

$$(A + \underbrace{\delta A}_{\text{original err}})(x + \underbrace{\delta x}_{\text{due to the err in } A\ (\delta A)}) = \underbrace{b}_{\text{fixed.}}$$

$$Ax + A\delta x + \delta A \cdot x + \underbrace{\cancel{\delta A \delta x}}_{\text{high order}} = b \qquad \doteq 0$$

① $Ax = b$

② want to consider $\delta x \to 0$, $\delta A \to 0$,

$$A\delta x + \delta A \cdot x = 0$$

def: $\dfrac{\text{relative change in } f}{\text{relative change in } x}$

$$\delta x = -A^{-1}\delta A \cdot x$$

$$\frac{\|\delta x\|}{\|x\|} \Big/ \frac{\|\delta A\|}{\|A\|} \le \|A^{-1}\| \cdot \|A\| = k(A)$$

4 Floating point

Computers use a finite number of bits to represent real numbers, they can only represent only a finite subset of real numbers. This has two limitations. Firstly, the represented number cannot be arbitrarily large or small. Secondly, there must be gaps between them.

In IEEE double precision arithmetic (one way to store numbers/digital representation of number in the computer), the interval $[1, 2]$ is represented by the discrete subset:

$$1, 1 + 1 \times 2^{-52}, 1 + 2 \times 2^{-52}, ..., 2 + 2^{52} \times 2^{-52}. \tag{15}$$

In general, the interval $[2^j, 2^{j+1}]$ is represented by 15 times 2^j. The gap between the two adjacent numbers is never larger than $2^{-52} \approx 2.22 \times 10^{-16}$ in relative sense.

IEEE double precision is an example of an arithmetic system based on a floating-point **F** representation of real numbers. Here **F** is a discrete subset of real numbers (example, Equation 15) which is used to digitally represent real numbers. Let us now define the machine epsilon ϵ_m. This number is half the distance between 1 and the next larger floating point number. It has the following property.

Property 4.0.1. For all $x \in \mathbb{R}$, ther exists $x' \in F$ such that $|x - x'| \leq \epsilon_m |x|$.

This is in a relative sense since if $x > 0$, $|1 - x'/x| \leq \epsilon_m$.

Let $fl : \mathbb{R} \to F$ be a function giving the closet floating-point approximation to a real number (rounded to one floating number). Then the above property can be stated in terms of ft: for all $x \in \mathbb{R}$, there exists ϵ with $\epsilon < \epsilon_m$, there exists ϵ with $|\epsilon| < \epsilon_m$ such that $fl(x) = x(1 + \epsilon)$.

The difference between a real number and its closest floating-point approximation is always smaller than the machine ϵ_m in relative sense.