Conditioning and Stability

→ Conditioning pertains to sensitivity of a mathematical problem to perturbations in inputs.

→ Stability pertains to perturbation behavior of an algorithm used to solve the mathematical problem on a computer.

(*) Conditioning of a problem :-

Solving a problem is like evaluating a function

$$\underset{\sim}{y} = \underset{\sim}{f}(\underset{\sim}{x})$$

Here $\underset{\sim}{x}$ represents input to the problem (the data), $\underset{\sim}{f}$ represents problem itself and $\underset{\sim}{y}$ represents its solution

* What happens to $\underset{\sim}{y}$ when given $\underset{\sim}{x}$ is perturbed slightly ?

If small changes in $\underset{\sim}{x}$ leads to large changes in $\underset{\sim}{y}$, we say the problem is ill-conditioned and usually we are interested in solving well-conditioned problem.

* Absolute condition number:-

If a small perturbation of $\underset{\sim}{x}$ is denoted by $\delta \underset{\sim}{x}$, then let the resulting perturbation in the solution be represented as δf

i.e $\delta f = f(\underset{\sim}{x} + \underset{\sim}{\delta x}) - f(\underset{\sim}{x})$

then the absolute condition number $\hat{K} = K(x)$ of the problem f at x is given by

$$K(x) = \max_{\delta x} \left(\frac{\|\delta f\|}{\|\delta x\|} \right) \quad \text{--- (1)}$$

for infinitesimally small $\underset{\sim}{\delta f}$ and $\underset{\sim}{\delta x}$

If $\underset{\sim}{f}$ has a derivative, we can evaluate the Jacobian matrix $\underset{\sim}{J}(x)$ as

$$J_{ij} = \frac{\partial f_i}{\partial x_j}$$

We have $\underset{\sim}{\delta f} \approx \underset{\sim}{J}(x)\, \underset{\sim}{\delta}(x)$ with equality as $\|\delta x\| \to 0$

$$K(x) = \max_{\delta x} \frac{\|J(x)\,\delta x\|}{\|\delta x\|}$$

$$\hat{K} = K(x) = \|J(x)\|$$

<u>Relative</u> <u>condition</u> <u>number</u>:-

Assume $\underset{\sim}{\delta f}$ and $\underset{\sim}{\delta x}$ are infinitesimal

$$\hat{K}^R = \max_{\underset{\sim}{\delta x}} \left[\frac{\dfrac{\|\delta f\|}{\|f(x)\|}}{\dfrac{\|\delta x\|}{\|x\|}} \right]$$

$$= \max_{\underset{\sim}{\delta x}} \left(\frac{\dfrac{\|\delta f\|}{\|\delta x\|}}{\dfrac{\|f(x)\|}{\|x\|}} \right)$$

$$\boxed{\hat{K}^R = \frac{\|J(x)\|}{\|f(x)\| / \|x\|}}$$

A problem is well-conditioned if $\underset{\sim}{K}$ is small (eg:- $1, 10, 10^2$). and ill-conditioned if K is large (eg: $10^6, 10^{16}$ ..)

Examples :-

(1) $f(x) = \frac{x}{2}$ $\quad x \in \mathbb{R}$

Input : x

output : $\frac{x}{2}$

$$J = \frac{df}{dx} = \frac{1}{2}$$

$$\hat{K} = \frac{\|J\|}{\frac{\|f(x)\|}{\|x\|}} = \frac{1/2}{\frac{[x/2]}{x}} = 1$$

well conditioned problem.

(2) $f(\underset{\sim}{x}) = x_1 - x_2$ where $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} = [1 \ -1]_{1\times 2}$$

$$\hat{K} = \frac{\|J\|_\infty}{\|f(\underset{\sim}{x})\|_\infty / \|\underset{\sim}{x}\|_\infty}$$

$\|J\|_\infty$ is max row sum $= 2$

$$\hat{K} = \frac{2}{\frac{|x_1 - x_2|}{\max\{|x_1|, |x_2|\}}}$$

$$= \frac{2 \max\{|x_1|, |x_2|\}}{|x_1 - x_2|}$$

if $|x_1 - x_2|$ is small ≈ 0, K is large and is not well conditioned when you are subtracting two numbers which are very close by.

$$\hat{K} = \frac{\|J\|_1}{\frac{\|f(x)\|_1}{\|x\|_1}} = \frac{1}{\frac{|x_1 - x_2|}{(|x_1|+|x_2|)}} = \frac{(|x_1|+|x_2|)}{|x_1 - x_2|}$$

<u>Root finding of quadratic equation:-</u>

$$x^2 - 2xp + 1 = 0 \qquad x_1 = p - \sqrt{p^2-1}$$
$$x_2 = p + \sqrt{p^2-1}$$

<u>Input</u>:- value of p

<u>output</u> :- x_1, x_2 which are the roots of quadratic equation.

Examine sensitivity of x_2 w.r.t p

$$\hat{K} = K(p) = \frac{\|J\|}{\frac{\|f(p)\|}{\|p\|}} \quad ; \quad J = \frac{dx_2}{dp}$$

$$K(p) = \left|\frac{dx_2}{dp}\right| \times \frac{|p|}{|x_2|}$$

$$= \left|1 + \frac{p}{\sqrt{p^2-1}}\right| \times \frac{|p|}{|p + \sqrt{p^2-1}|}$$

$$= \left| \frac{p + \sqrt{p^2 - 1}}{\sqrt{p^2-1}} \right| \times \frac{|p|}{|p + \sqrt{p^2-1}|}$$

$$= \frac{|p|}{\sqrt{p^2-1}}$$

The root x_2 is sensitive for p close to 1 with $K(p) \to \infty$ and insensitive for large p with $K \simeq 1$

<u>Eigenvalues</u> of <u>matrix</u> :-

$$A x = \lambda x$$

$$\det(A - \lambda I) = 0$$

<u>Input</u> :- A

<u>Output</u> : eigenvalues λ

$$A = \begin{pmatrix} 1 & 1000 \\ 0 & 1 \end{pmatrix} \quad \text{eigenvalues of } A = \{1, 1\}$$

$$\hat{A} = \begin{pmatrix} 1 & 1000 \\ 0.001 & 1 \end{pmatrix} \quad \text{eigenvalues of } A \text{ after perturbation } \{0, 2\}$$

Consider a symmetric matrix $A = A^T$

λ and $\lambda + \delta\lambda$ are corresponding eigenvalues A and $A + \delta A$, then $\underbrace{|\delta\lambda| \le \|\delta A\|_2}$ — (1)

Relative condition number $= \max_{\|\delta A\|} \dfrac{\dfrac{|\delta\lambda|}{|\lambda|}}{\dfrac{\|\delta A\|_2}{\|A\|_2}}$

use ① $\rightarrow$ $\boxed{\text{Relative condition number } \hat{k} = \dfrac{\|A\|_2}{|\lambda|}}$

Conditioning of matrix-vector multiplication:-

$y = Ax$

Fixed A

Input:- x

Output:- Ax

Consider the problem of computing Ax for fixed A and input x

$$\boxed{k = \max_{\delta x} \frac{\dfrac{\|\delta f\|}{\|f\|}}{\dfrac{\|\delta x\|}{\|x\|}}}$$

$$k = \max_{\delta x} \frac{\dfrac{\|A(x+\delta x) - Ax\|}{\|Ax\|}}{\dfrac{\|\delta x\|}{\|x\|}}$$

$$= \max_{\delta x} \frac{\dfrac{\|A\delta x\|}{\|Ax\|}}{\dfrac{\|\delta x\|}{\|x\|}} = \underbrace{\left(\max_{\delta x} \frac{\|A\delta x\|}{\|\delta x\|}\right) \times \frac{\|x\|}{\|Ax\|}}$$

$$\hat{k} = \frac{\|A\|\,\|x\|}{\|Ax\|} \quad \text{— ②}$$

$\hat{k}$ for a give A and at x

To loosen above equality to get a bound of $\underset{\sim}{x}$, let us assume $\underset{\sim}{A}$ is square and non-singular

$$\underset{\sim}{x} = \underset{\sim}{A}^{-1} \underset{\sim}{A} \underset{\sim}{x}$$

$$\|\underset{\sim}{x}\| = \|\underset{\sim}{A}^{-1} \underset{\sim}{A} \underset{\sim}{x}\| \leq \|\underset{\sim}{A}^{-1}\| \|\underset{\sim}{A} \underset{\sim}{x}\|$$

$$\frac{\|\underset{\sim}{x}\|}{\|\underset{\sim}{A} \underset{\sim}{x}\|} \leq \|\underset{\sim}{A}^{-1}\| \quad - \text{(3)}$$

From eqn (2) $\hat{k} = \dfrac{\|\underset{\sim}{A}\| \|\underset{\sim}{x}\|}{\|\underset{\sim}{A} \underset{\sim}{x}\|}$ ✓

using (3) $\hat{k} \leq \|\underset{\sim}{A}\| \|\underset{\sim}{A}^{-1}\| \quad - \text{(4)}$

What about for a give $\underset{\sim}{A}$, if you want to compute $\underset{\sim}{A}^{-1} \underset{\sim}{b}$ from a given input $\underset{\sim}{b}$?

Input :- $\underset{\sim}{b}$ Fixed $\underset{\sim}{A}$, $\underset{\sim}{A}^{-1}$

Output :- $\underset{\sim}{A}^{-1} \underset{\sim}{b} = \underset{\sim}{x}$

$$\hat{k} = \frac{\|\underset{\sim}{A}^{-1}\| \|\underset{\sim}{b}\|}{\|\underset{\sim}{A}^{-1} \underset{\sim}{b}\|}$$

$$\hat{k} = \frac{\|A^{-1}\| \|b\|}{\|\underset{\sim}{x}\|}$$

using (4), even for the problem of $\underset{\sim}{A}^{-1} \underset{\sim}{b}$

$$\hat{k} \leq \|\underset{\sim}{A}^{-1}\| \|\underset{\sim}{A}\|$$

Result:-

Let $\underset{\sim}{A} \in \mathbb{R}^{m \times m}$ be non-singular and consider the equation $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$, the problem of computing $\underset{\sim}{b}$ given $\underset{\sim}{x}$ has condition number

$$\hat{K} = \frac{\|\underset{\sim}{A}\| \|\underset{\sim}{x}\|}{\|\underset{\sim}{b}\|} \leq \|\underset{\sim}{A}\| \|\underset{\sim}{A}^{-1}\|$$

with respect to perturbations of $\underset{\sim}{x}$.

The problem of computing $\underset{\sim}{x}$ given $\underset{\sim}{b}$ has condition number

$$\hat{K} = \frac{\|\underset{\sim}{A}^{-1}\| \|\underset{\sim}{b}\|}{\|\underset{\sim}{x}\|} \leq \|\underset{\sim}{A}\| \|\underset{\sim}{A}^{-1}\|$$

Condition number of a matrix:-

The condition number of $\underset{\sim}{A}$ (relative to norm $\|\cdot\|$) denoted by $\hat{K}(\underset{\sim}{A}) = \|\underset{\sim}{A}\| \|\underset{\sim}{A}^{-1}\|$

If $\hat{K}(\underset{\sim}{A})$ is small, $\underset{\sim}{A}$ is said to be well-conditioned

if $\hat{K}(A)$ is large, $\underset{\sim}{A}$ is said to be ill conditioned.

If $\underset{\sim}{A}$ is singular, or close to singular $K(A) \to \infty$

In the 2-norm

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

$\sigma_1 \rightarrow$ max singular value of $\underset{\sim}{A}$

$(1/\sigma_m) \rightarrow$ max singular value of $\underset{\sim}{A}^{-1}$

$\sigma_m \rightarrow$ min singular of $\underset{\sim}{A}$

$$K(\underset{\sim}{A}) = \frac{\sigma_1}{\sigma_m}$$

For a square matrix, non zero singular values are square roots of non-zero eigenvalues of A^TA or AA^T.

$$K(A) = \frac{\sqrt{\lambda_{max}(A^TA)}}{\sqrt{\lambda_{min}(A^TA)}}$$

$$K(A) = \frac{|\lambda_{max}(\underset{\sim}{A})|}{|\lambda_{min}(\underset{\sim}{A})|} \quad (\text{If } \underset{\sim}{A} \text{ is symmetric})$$

If $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ $(m > n)$, $K(\underset{\sim}{A})$ is defined in terms of pseudoinverse A^+.

$$K(A) = \|A\| \|A^+\|$$

$$\boxed{A^+ = A^T(\underset{\sim}{A}\underset{\sim}{A}^T)^{-1}}$$

<u>Conditioning of system of equations</u>:-

So far we fixed $\underset{\sim}{A}$ and perturbed $\underset{\sim}{x}$ or $\underset{\sim}{b}$

→ What about if we perturb A?

Fix $\underset{\sim}{b}$ and consider $f : A \to x = \underset{\sim}{A}^{-1}\underset{\sim}{b}$

inputs :- $\underset{\sim}{A}$

output :- $\underset{\sim}{x}$

$\underset{\sim}{A}$ is perturbed by $\delta\underset{\sim}{A}$ and let $\underset{\sim}{x}$, the output be perturbed by $\delta\underset{\sim}{x}$

$$(\underset{\sim}{A} + \delta\underset{\sim}{A})(\underset{\sim}{x} + \delta\underset{\sim}{x}) = b \quad -(1)$$

$$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b} \quad -(2)$$

Using (1) and (2) above, we have

$$(\delta\underset{\sim}{A})\underset{\sim}{x} + \underset{\sim}{A}(\delta\underset{\sim}{x}) = 0$$

$$\delta\underset{\sim}{x} = -\underset{\sim}{A}^{-1}(\delta\underset{\sim}{A})\underset{\sim}{x}$$

$$\|\delta\underset{\sim}{x}\| = \|-\underset{\sim}{A}^{-1}(\delta\underset{\sim}{A})\underset{\sim}{x}\|$$

$$\leq \|\underset{\sim}{A}^{-1}\| \, \|(\delta\underset{\sim}{A})\underset{\sim}{x}\|$$

$$\leq \|\underset{\sim}{A}^{-1}\| \, \|\delta\underset{\sim}{A}\| \, \|\underset{\sim}{x}\| \quad -(3)$$

$$\hat{K} = \max_{\delta\underset{\sim}{A}} \frac{\dfrac{\|\delta\underset{\sim}{x}\|}{\|\underset{\sim}{x}\|}}{\dfrac{\|\delta\underset{\sim}{A}\|}{\|\underset{\sim}{A}\|}}$$

Using eq 3

$$\underbrace{\left(\frac{\dfrac{\|\delta x\|}{\|x\|}}{\dfrac{\|\delta A\|}{\|A\|}}\right) \leq \|\underset{\sim}{A}^{-1}\| \, \|A\|}$$

If perturbation $\delta \underset{\sim}{A}$ exists which makes the above inequality an equality

then

$\hat{K}$ of solving system of equations

$$\hat{K} = \max_{\delta \underset{\sim}{A}} \left(\frac{\frac{\|\delta \underset{\sim}{x}\|}{\|\underset{\sim}{x}\|}}{\frac{\|\delta \underset{\sim}{A}\|}{\|\underset{\sim}{A}\|}} \right) = K(\underset{\sim}{A})$$

Result:-

For a fixed $\underset{\sim}{b}$, Input $\rightarrow \underset{\sim}{A}$, output $\underset{\sim}{x} = \underset{\sim}{A}^{-1}\underset{\sim}{b}$

condition number of this problem with respect to perturbations in $\underset{\sim}{A}$ is

$$\hat{K} = \|\underset{\sim}{A}\|\|\underset{\sim}{A}^{-1}\| = K(\underset{\sim}{A})$$

eg:-

$$K(\underset{\sim}{A}) = 10^6$$

$$\frac{\|\delta \underset{\sim}{x}\|}{\|\underset{\sim}{x}\|} \leq 10^6 \times \frac{\|\delta \underset{\sim}{A}\|}{\|\underset{\sim}{A}\|} \qquad \varepsilon_M \simeq 10^{-8}$$

$$\leq 10^6 \times 10^{-8} \qquad \frac{\|\delta \underset{\sim}{A}\|}{\|\underset{\sim}{A}\|} \leq O(\varepsilon_M)$$

$$\approx O(10^{-2})$$

Stability of Algorithms:-

(1) **Desirable** :- Exact solutions to numerical problems

(2) **Reality** :- Problems are continuous while computer arithmetic is discrete

(3) Stability tells us what it means to get the best answer even if this is not the exact answer!

Algorithm:-

An abstract way to think about solving a problem is evaluating function

$f : X \rightarrow Y$ $\quad$ X : vector space of data

$y = f(x)$ $\quad$ Y: vector space of solutions

where $x \in X$, $y \in Y$

An algorithm can be viewed as a different function $\tilde{f}$ that usually takes the same data $x \in X$ and maps it to a result which is a collection of floating point numbers that belongs to Y.

<u>Accuracy</u>:- A good algo should have an $\tilde{f}$ that closely approximates the underlying problem f.

→ <u>Absolute</u> <u>error</u> of <u>computation</u>:-

$$\|\tilde{f}(x) - f(x)\|$$

→ <u>Relative</u> <u>error</u> of <u>computation</u>:-

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|}$$

We say that $\tilde{f}$ is an accurate algorithm for f if for all relevant input data x,

Forward relative error

$$\boxed{\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = O(\varepsilon_M)}$$

"On the order of machine epsilon"

If f is ill-conditioned

i.e $\max_{\delta x} \dfrac{\frac{\|\delta f\|}{\|f\|}}{\frac{\|\delta x\|}{\|x\|}} = \hat{K}$ is very large

$\dfrac{\|\delta x\|}{\|x\|} \approx O(\varepsilon_M)$

$\dfrac{\|\delta f\|}{\|-\|} \lessapprox K\, O(\varepsilon_M)$

Very ambitious to esign an algo $\tilde{f}$ such that relative error in computation is $O(\varepsilon_M)$

Instead of aiming for accuracy of an algo, the most we can aim is for stability!

We can say that an algorithm $\tilde{f}$ for solving a problem f is stable if for all (relevant) input data x

$$\frac{\|\tilde{f}(x) - f(\tilde{x})\|}{\|f(\tilde{x})\|} = O(\varepsilon_M)$$

for some $\tilde{x}$ satisfying $\dfrac{\|\tilde{x} - x\|}{\|x\|} = O(\varepsilon_M)$
$\leq C\,\varepsilon_M$

i.e A stable algorithm gives nearly right answer to nearly right question.

<u>Backward stability</u> :-

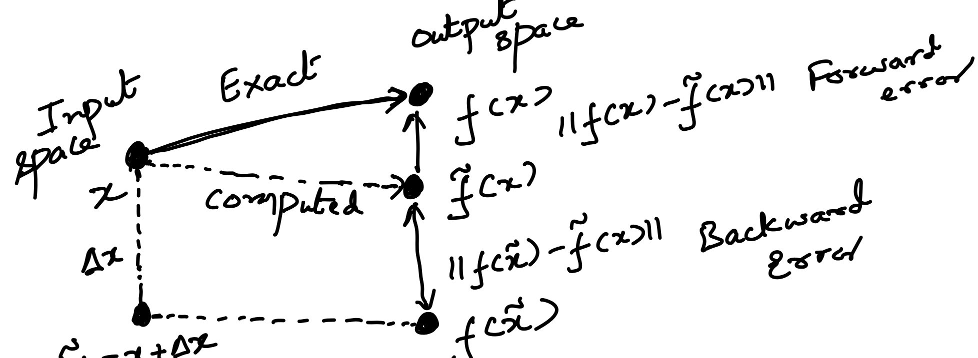
<u>Backward error</u> :- $\|f(\tilde{x}) - \tilde{f}(x)\|$

What is the input $\tilde{x}$ for which my algo with input x has exactly computed a solution for

$$f(\tilde{x}) = \tilde{f}(x)$$

Then such an algo $\tilde{f}$ for problem f is backward stable. (i.e Backward error is zero)

i.e Exactly right answer to a nearly right question!



Stability of Floating point arithmetic operations!

$+, -, \times, /$ (Classical arithmetic operations)

Floating point analogues $\oplus, \ominus, \otimes, \oslash$

Recall Floating point axioms :-

$$fl(x) = x(1+\varepsilon) \text{ where } |\varepsilon| < \varepsilon_M$$

$$x \circledast y = x * y(1+\varepsilon) \text{ where } |\varepsilon| < \varepsilon_M$$

Example 1 :-

Floating point arithmetic for $\ominus$

$\underset{\sim}{x} \in \mathbb{R}^2$ $\quad \underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$f(\underset{\sim}{x}) = x_1 - x_2$

$f(x_1, x_2) = x_1 - x_2$

$x_1 \longrightarrow fl(x_1)$; $x_2 \longrightarrow fl(x_2)$

$fl(x_1) = x_1(1+\varepsilon_1)$; $fl(x_2) = x_2(1+\varepsilon_2)$

$\varepsilon_1, \varepsilon_2$ satisfy $|\varepsilon_1|, |\varepsilon_2| = O(\varepsilon_M)$

Algo :- $fl(x_1) \ominus fl(x_2)$

$\Rightarrow [x_1(1+\varepsilon_1)] \ominus [x_2(1+\varepsilon_2)]$

$$fl(x_1) \ominus fl(x_2) = [x_1(1+\varepsilon_1) - x_2(1+\varepsilon_2)](1+\varepsilon_3)$$

where $|\varepsilon_3| = O(\varepsilon_M)$

$$= x_1(1+\varepsilon_1)(1+\varepsilon_3) - x_2(1+\varepsilon_2)(1+\varepsilon_3)$$

$$= x_1(1+\underbrace{\varepsilon_1+\varepsilon_3+\varepsilon_1\varepsilon_3}) - x_2(1+\underbrace{\varepsilon_2+\varepsilon_3+\varepsilon_2\varepsilon_3})$$

$$= x_1(1+\varepsilon_4) - x_2(1+\varepsilon_5)$$

where $|\varepsilon_4|, |\varepsilon_5| \leq 2\varepsilon_M + O(\varepsilon_M^2) = O(\varepsilon_M)$

$$\tilde{f}(x_1, x_2) = fl(x_1) \ominus fl(x_2)$$

$$= \tilde{x}_1 - \tilde{x}_2 = f(\tilde{x}_1, \tilde{x}_2)$$

Algo is backward stable!

Example 2:- Outer product between two vectors :-

$\underline{x} \in \mathbb{R}^m$, $\underline{y} \in \mathbb{R}^n$

compute outer product $\underline{A} = \underline{x}\,\underline{y}^T$

$$\boxed{A_{ij} = x_i y_j} = m\times1 \; 1\times n$$

$\underline{A}$ $m\times n$

$\uparrow$ rank 1 matrix

$$\tilde{f}(\underline{x}, \underline{y}) = \tilde{A}_{ij}$$

$$= fl(x_i) \otimes fl(y_j)$$

$$\tilde{A}_{ij} = fl(x_i) \otimes fl(y_j)$$

$$= [fl(x_i) \times fl(y_j)](1+\varepsilon_3^{ij})$$

$$= [x_i(1+\varepsilon_1^i)\, y_j(1+\varepsilon_2^j)](1+\varepsilon_3^{ij})$$

$$= x_i y_j (1+\varepsilon_1^i)(1+\varepsilon_2^j)(1+\varepsilon_3^{ij})$$

$$= x_i y_j (1 + \underbrace{\varepsilon_1^i + \varepsilon_2^j + \varepsilon_1^i \varepsilon_2^j}) (1 + \varepsilon_3^{ij})$$

$$= x_i y_j (\underline{1 + \varepsilon_4^{ij}})(1 + \underline{\varepsilon_3^{ij}}) \leftarrow (i)$$

I want to verify

$$\boxed{\begin{aligned} \tilde{f}(\underset{\sim}{x}, \underset{\sim}{y}) &= \tilde{\underset{\sim}{x}} \tilde{\underset{\sim}{y}}^T \leftarrow (*) \\ &= (\underset{\sim}{x} + \delta \underset{\sim}{x})(\underset{\sim}{y} + \delta \underset{\sim}{y})^T \end{aligned}}$$

This algo is not backward stable because from eqn (i), it is not possible for my algo always to give a perturbed rank 1 matrix of the form $\tilde{\underset{\sim}{x}} \tilde{\underset{\sim}{y}}^T$

Is this algo stable? Exercise!

Example 3:-

Adding 1 to floating point number!

Let $x \in \mathbb{R}$ and $f(x) = x + 1$

$$\tilde{f}(x) = fl(x) \oplus 1$$

$$\tilde{f}(x) = (fl(x) + 1)(1 + \varepsilon_1)$$

$$= (x(1 + \varepsilon_2) + 1)(1 + \varepsilon_1)$$

$$= (1 + x + x\varepsilon_2)(1 + \varepsilon_1)$$

$= 1 + x + x\varepsilon_2 + \varepsilon_1 + \varepsilon_1 x + x\varepsilon_2\varepsilon_1$

$= 1 + x\left[1 + \varepsilon_1 + \varepsilon_2 + \frac{\varepsilon_1}{x}\right] \quad — \quad (2)$

For backward stability is $\tilde{f}(x) = f(\tilde{x})$

$= \tilde{x} + 1$

$= x(1+\varepsilon) + 1$?

From eq (2)

$$1 + x\left[\underbrace{1 + \varepsilon_1 + \varepsilon_2 + \frac{\varepsilon_1}{x}}_{\text{As } x \to 0}\right] \quad \varepsilon \neq O(\varepsilon_M)$$

This is not backward stable

Unstable Algos :-

→ Calculation of eigenvalues of a matrix by finding roots of characteristic polynomial

→ Since λ is eigenvalue of a matrix $\underset{\sim}{A}$, then the characteristic polynomial is $p(\lambda) = \det(\underset{\sim}{A} - \lambda \underset{\sim}{I}) = 0$

roots $p(\lambda) = 0$ are eigenvalues of A

Algo :-

① Find the coefficients of $p(\lambda) = \det(\underset{\sim}{A} - \lambda I)$

② Find its roots

The problem of finding roots of a polynomial given its coefficients is an ill-conditioned problem!

i.e roots are sensitive to small errors in polynomial coefficients!

Recall $\underset{\sim}{K} = \dfrac{\|A\|_2}{\lambda}$

Eg :- $\underset{\sim}{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Eigen values of $\underset{\sim}{A}$ are not sensitive to perturbations of the entries. If we design a stable algo, we should be able to compute eigenvalues with relative errors of $O(\epsilon_M)$

Let us see the order of error in roots if we compute the roots of characteristic polynomial for the above problem:

$\Rightarrow \quad x^2 - 2x + 1$

Let us analyse the error in roots for perturbation in the coefficient p for $x^2 - px + 1$

Roots :- $\dfrac{p \pm \sqrt{p^2 - 4}}{2}$

Let us say my error in estimating the coefficient p be $|\varepsilon| < \varepsilon_m$

$$\tilde{p} = p(1+\varepsilon)$$

New Roots :- $\dfrac{p(1+\varepsilon) \pm \sqrt{[p(1+\varepsilon)]^2 - 4}}{2}$

$$= \frac{p(1+\varepsilon) \pm \sqrt{p^2(1+\varepsilon)^2 - 4}}{2}$$

For our problem at hand $p = 2$

New roots for $p=2$:- $\dfrac{2(1+\varepsilon) \pm \sqrt{4(1+\varepsilon)^2 - 4}}{2}$

$$= (1+\varepsilon) \pm \sqrt{\cancel{1} + \varepsilon^2 + 2\varepsilon - \cancel{1}}$$

$$\Rightarrow (1+\varepsilon) \pm \sqrt{\cancel{\varepsilon^2} + 2\varepsilon}$$

Error in root :- $\dfrac{|(1+\varepsilon) \pm \sqrt{2\varepsilon} - 1|}{|1|}$

Dominant Error in my root $\approx O(\sqrt{\varepsilon_M})$

$\varepsilon_M \approx 10^{-16}$

$\sqrt{\varepsilon_M} \approx 10^{-8}$

Accuracy of Backward stable Algo :-

Thm :- If a backward stable algo is applied to solve a problem f with condition number K, the relative forward errors satisfy

$$\frac{\|f(\tilde{x}) - f(x)\|}{\|f(x)\|} = O(K(x)\varepsilon_M)$$

Pf :- By definition of backward stability we have $\tilde{f}(x) = f(\tilde{x})$

for $\dfrac{\|\tilde{x} - x\|}{\|x\|} = O(\varepsilon_M)$

we also know

$$K(x) = \max_{\delta x} \frac{\frac{\|\delta f\|}{\|f\|}}{\frac{\|\delta x\|}{\|x\|}}$$

$$\frac{\left(\frac{\|f(\tilde{x}) - f(x)\|}{\|f(x)\|}\right)}{\frac{\|\tilde{x} - x\|}{\|x\|}} \leq K(x)$$

But B.S, $f(\tilde{x}) = \tilde{f}(x)$

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} \leq K(x) \underbrace{\frac{\|\tilde{x} - x\|}{\|x\|}}_{O(\varepsilon_M)}$$

$O(K(x)\, \varepsilon_M)$