Sensitivity of a linear system

Condition number of a matrix

Solving linear systems in floating point

Given

$$\mathbf{A}\mathbf{x} = \mathbf{b}\,,$$

the formal solution is

$$\widehat{x} = \mathbf{A}^{-1}\mathbf{b}$$

However,

- $\mathbf{A}$ is only known approximately
- $\mathbf{b}$ is only known approximately
- A numerical algorithm introduces some errors (at least roundoff).

Can only compute an *approximate* solution, $\mathbf{x}^*$.

Solving linear systems in floating point

Let $\mathbf{x}^*$ is an *approximate* solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$.
We want to have $\mathbf{x}^*$ “close” to the true solution $\widehat{\mathbf{x}}$.

- Deviation $\mathbf{e} = \widehat{\mathbf{x}} - \mathbf{x}^*$
- *Residual* $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}^* \equiv \mathbf{A}\mathbf{e}$

Require that either the deviation or the residual is “small”. But what is small.

Need a distance measure for vectors.

Vector norms

A *norm* in the vector space $\mathbb{R}^m$, $\|\cdot\|$, is a mapping $\mathbb{R}^m \to \mathbb{R}$, such that
$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m, \forall \alpha \in \mathbb{R}$

1. $\|\mathbf{x}\| \geqslant 0$ and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$
3. $\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$ triangle inequality

Vector norms

A *norm* in the vector space $\mathbb{R}^m$, $\|\cdot\|$, is a mapping $\mathbb{R}^m \to \mathbb{R}$, such that
$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m, \forall \alpha \in \mathbb{R}$

1. $\|\mathbf{x}\| \geqslant 0$ and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$
2. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$ triangle inequality

Can define *absolute error*, $\epsilon_{\mathrm{abs}} = \|\widehat{\mathbf{x}} - \mathbf{x}^*\|$, *relative error* $\epsilon_{\mathrm{rel}} = \|\widehat{\mathbf{x}} - \mathbf{x}^*\| / \|\mathbf{x}^*\|$.

Convergence

A sequence $\mathbf{x}^{(n)}$ *converges* to $\widehat{\mathbf{x}}$ if

$$\|\mathbf{x}^{(n)} - \widehat{\mathbf{x}}\| \to 0\,, \qquad n \to 0$$

Vector norms

p-norms

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_m|^p)^{1/p}$$

for $p \geqslant 1$.

1-norm $\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_m|$

2-norm $\|\mathbf{x}\|_2 = \left(|x_1|^2 + \cdots + |x_m|^2\right)^{1/2}$

∞-norm $\|\mathbf{x}\|_\infty = \max_{1 \leqslant j \leqslant m} |x_j|$

Vector norms

p-norms

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_m|^p)^{1/p}$$

for $p \geqslant 1$.

$$\begin{aligned}
\text{1-norm} \quad \|\mathbf{x}\|_1 &= |x_1| + \cdots + |x_m| \\
\text{2-norm} \quad \|\mathbf{x}\|_2 &= \left(|x_1|^2 + \cdots + |x_m|^2\right)^{1/2} \\
\infty\text{-norm} \quad \|\mathbf{x}\|_\infty &= \max_{1 \leqslant j \leqslant m} |x_j|
\end{aligned}$$

All p-norms are equivalent:

$$\|\mathbf{x}\|_\infty \leqslant \|\mathbf{x}\|_2 \leqslant \|\mathbf{x}\|_1 \leqslant m\|\mathbf{x}\|_\infty$$

Matrix norms

Let $\mathbf{A}$ and $\mathbf{B}$ are an m-by-n matrices.

A *norm* in the space $\mathbb{R}^{m,n}$, $\|\cdot\|$, is a mapping $\mathbb{R}^{m,n} \to \mathbb{R}$, such that $\forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m,n}$, $\forall \alpha \in \mathbb{R}$

1. $\|\mathbf{A}\| \geqslant 0$ and $\|\mathbf{A}\| = 0$ iff $\mathbf{A} = \mathbf{0}$
2. $\|\alpha \mathbf{A}\| = |\alpha| \|\mathbf{A}\|$
3. $\|\mathbf{A} + \mathbf{B}\| \leqslant \|\mathbf{A}\| + \|\mathbf{B}\|$ triangle inequality

Matrix norms

Frobenius norm

$$\|\mathbf{A}\|_F = \sqrt{\sum_{j=1}^{m} \sum_{k=1}^{n} |a_{jk}|^2}$$

p-norms

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

A matrix p-norm is *subordinate* to (or, is *induced* by) a vector p-norm.

Matrix norms

For p-norms, have

4. $\|\mathbf{Ax}\| \leqslant \|\mathbf{A}\| \cdot \|\mathbf{x}\|$
5. $\|\mathbf{AB}\| \leqslant \|\mathbf{A}\| \cdot \|\mathbf{B}\|$

   (p-norms are *mutually consistent*)

Matrix norms

For p-norms, have

4. $\|\mathbf{Ax}\| \leqslant \|\mathbf{A}\| \cdot \|\mathbf{x}\|$
5. $\|\mathbf{AB}\| \leqslant \|\mathbf{A}\| \cdot \|\mathbf{B}\|$
(p-norms are *mutually consistent*)

NB: Consistency is not guaranteed by properties 1-4. Consider,

$$\|\mathbf{A}\|_{\triangle} = \max_{k,l} |a_{kl}|$$

and

$$\mathbf{A} = \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Common matrix norms: 1-norm

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & & a_{mn} \end{pmatrix}$$

$p = 1$ (a.k.a. the *column norm*)

$$\|\mathbf{A}\|_1 = \max_{1<k<n} \sum_{j=1}^{m} |a_{jk}|$$

is subordinate to the vector 1-norm.

Common matrix norms: ∞-norm

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & & a_{mn} \end{pmatrix}$$

$p = \infty$ (a.k.a. the *row norm*)

$$\|\mathbf{A}\|_\infty = \max_{1<j<m} \sum_{k=1}^{n} |a_{jk}|$$

is subordinate to the vector ∞-norm.

Common matrix norms: 2-norm

- 1- and ∞-norms are easy to compute, $O(n^2)$ operations
- 2-norm, subordinate to the vector 2-norm is

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$$

- In terms of SVD, $\|\mathbf{A}\|_2 = \sigma_{\max}$
- Note the naming: *Euclidean norm* typically means the Frobenius norm, not $\|\cdot\|_2$.

Sensitivity of a linear system

Suppose $\mathbf{A}$ is known exactly, and $\mathbf{b}$ is not; we only know an approximation, $\mathbf{b}^*$. How different are $\mathbf{x}$ and $\mathbf{x}^*$?

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad\qquad \mathbf{A}\mathbf{x}^* = \mathbf{b}^*$$

$\lVert\mathbf{b} - \mathbf{b}^*\rVert$	abs error of $\mathbf{b}$ (input)
$\lVert\mathbf{x} - \mathbf{x}^*\rVert$	abs error of $\mathbf{x}$ (result)

Sensitivity of a linear system

Suppose $\mathbf{A}$ is known exactly, and $\mathbf{b}$ is not; we only know an approximation, $\mathbf{b}^*$. How different are $\mathbf{x}$ and $\mathbf{x}^*$?

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad\qquad \mathbf{A}\mathbf{x}^* = \mathbf{b}^*$$

$\|\mathbf{b} - \mathbf{b}^*\|$	abs error of $\mathbf{b}$ (input)
$\|\mathbf{x} - \mathbf{x}^*\|$	abs error of $\mathbf{x}$ (result)

Condition number relates the error of the result to the error of inputs

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|}{\|\mathbf{x}\|} \propto \frac{\|\mathbf{b} - \mathbf{b}^*\|}{\|\mathbf{b}\|}$$

Condition number

- Assuming $\mathbf{A}\mathbf{x} = \mathbf{b}$, define the *residual*

$$\begin{aligned} \mathbf{r} &= \mathbf{b} - \mathbf{A}\mathbf{x}^* \\ &= \mathbf{b} - \mathbf{b}^* \\ &= \mathbf{A}(\mathbf{x} - \mathbf{x}^*) \qquad \Rightarrow \qquad \mathbf{x} - \mathbf{x}^* = \mathbf{A}^{-1}\mathbf{r} \end{aligned}$$

Condition number

- Assuming $\mathrm{A}\mathrm{x} = \mathrm{b}$, define the *residual*

$$\begin{aligned} \mathbf{r} &= \mathbf{b} - \mathbf{A}\mathbf{x}^* \\ &= \mathbf{b} - \mathbf{b}^* \\ &= \mathbf{A}(\mathbf{x} - \mathbf{x}^*) \qquad \Rightarrow \quad \mathbf{x} - \mathbf{x}^* = \mathbf{A}^{-1}\mathbf{r} \end{aligned}$$

- $$\frac{\|\mathbf{b}\|}{\|\mathbf{x}\|} = \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \leqslant \|\mathbf{A}\|$$

Condition number

- Assuming $\mathbf{A}\mathbf{x} = \mathbf{b}$, define the *residual*

$$\begin{aligned} \mathbf{r} &= \mathbf{b} - \mathbf{A}\mathbf{x}^* \\ &= \mathbf{b} - \mathbf{b}^* \\ &= \mathbf{A}(\mathbf{x} - \mathbf{x}^*) \qquad \Rightarrow \quad \mathbf{x} - \mathbf{x}^* = \mathbf{A}^{-1}\mathbf{r} \end{aligned}$$

- $$\frac{\|\mathbf{b}\|}{\|\mathbf{x}\|} = \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \leqslant \|\mathbf{A}\|$$

- Define $\delta\mathbf{x} = \dfrac{\|\mathbf{x} - \mathbf{x}^*\|}{\|\mathbf{x}\|}$ and $\delta\mathbf{b} = \dfrac{\|\mathbf{b} - \mathbf{b}^*\|}{\|\mathbf{b}\|} \equiv \dfrac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$

Condition number

$$
\begin{aligned}
\delta \mathbf{x} = \frac{\|\mathbf{x} - \mathbf{x}^*\|}{\|\mathbf{x}\|} &= \frac{\|\mathbf{A}^{-1}\mathbf{r}\|}{\|\mathbf{x}\|} \\
&\leqslant \frac{\|\mathbf{A}^{-1}\| \cdot \|\mathbf{r}\|}{\|\mathbf{x}\|} \\
&= \frac{\|\mathbf{A}^{-1}\| \cdot \|\mathbf{r}\|}{\|\mathbf{x}\|} \frac{\|\mathbf{b}\|}{\|\mathbf{b}\|} \\
&\leqslant \|\mathbf{A}^{-1}\| \, \|\mathbf{A}\| \, \delta \mathbf{b}
\end{aligned}
$$

$$
\operatorname{cond} \mathbf{A} = \|\mathbf{A}^{-1}\| \, \|\mathbf{A}\|
$$

Condition number

$$\operatorname{cond} \mathbf{A} = \|\mathbf{A}^{-1}\| \, \|\mathbf{A}\|$$

- Condition number is scale invariant

$$\operatorname{cond}(\alpha \mathbf{A}) = \operatorname{cond} \mathbf{A} \qquad \forall \alpha \in \mathbb{R}$$

- Condition number is norm-dependent
- If $\operatorname{cond} \mathbf{A} \gg 1$, the matrix is *ill-conditioned*

- For errors in both $\mathbf{A}$ and $\mathbf{b}$,

$$\delta \mathbf{x} = \operatorname{cond} \mathbf{A} (\delta \mathbf{A} + \delta \mathbf{b})$$