Basics in Applied Mathematics – Part I: Numerics

Lecture 2: Conditioning, stability and norms – linear systems

JProf. Dr. Diyora Salimova

Department of Applied Mathematics, University of Freiburg

22 October 2026

Last time, and today

Last time: floating-point numbers $\mathbb{F}$; standard model (1.2): each basic operation exact up to a relative error $u \approx 1.1 \cdot 10^{-16}$; **cancellation** (1.3) amplifies errors that are already there (Examples 1.1, 1.5).

After this lecture you can

(i) distinguish a well-conditioned *problem* from a stable *algorithm*,

(ii) compute $\|A\|_1$, $\|A\|_\infty$, $\|A\|_2$ and bound data-error effects in $Ax = b$ by $\operatorname{cond}(A) = \|A\|\,\|A^{-1}\|$,

(iii) describe Gaussian elimination as $A = LU$, with its cost $\frac{2}{3}n^3$ and the need for pivoting,

(iv) use the Cholesky factorisation for SPD matrices and exploit band structure.

1.3 Conditioning and stability · 1.4 Norms · 2.1–2.2 Linear systems, $\operatorname{cond}(A)$ · 2.3–2.4 Substitution, LU, pivoting · 2.5 Cholesky · 2.6 Banded matrices

Is the *problem* sensitive to perturbations of its data? (independent of any algorithm)

Idea: relative change of the output $\approx \kappa_{\mathrm{rel}}(x) \cdot$ relative change of the input.

Definition 1.6 (condition numbers)

Problem $f\colon \mathbb{R}^n \to \mathbb{R}^m$, input x, output $f(x)$:

$$\kappa_{\mathrm{abs}}(x) = \lim_{\varepsilon \to 0} \sup_{0<\|\Delta x\| \le \varepsilon} \frac{\|f(x+\Delta x) - f(x)\|}{\|\Delta x\|},$$

$$\kappa_{\mathrm{rel}}(x) = \lim_{\varepsilon \to 0} \sup_{0<\|\Delta x\| \le \varepsilon} \frac{\|f(x+\Delta x) - f(x)\|/\|f(x)\|}{\|\Delta x\|/\|x\|} \qquad (x \neq 0,\ f(x) \neq 0).$$

well-conditioned: $\kappa_{\mathrm{rel}}(x)$ moderate **ill-conditioned**: $\kappa_{\mathrm{rel}}(x)$ large
(what counts: $\kappa_{\mathrm{rel}}(x)\times$ relative input error against the accuracy you need)

Differentiable functions

Lemma 1.7

$f\colon \mathbb{R} \to \mathbb{R}$ differentiable at x (with $x \neq 0$, $f(x) \neq 0$):

$$\kappa_{\mathrm{abs}}(x) = |f'(x)|, \qquad \kappa_{\mathrm{rel}}(x) = \frac{|f'(x)|\,|x|}{|f(x)|}.$$

Example 1.8. $\sqrt{x}$ $(x > 0)$: $\kappa_{\mathrm{rel}} = \frac{1}{2}$ · $x - 1$: $\kappa_{\mathrm{rel}} = \frac{|x|}{|x-1|} \to \infty$ as $x \to 1$ (cancellation)
e^x: $\kappa_{\mathrm{rel}} = |x|$ · $\sin x$: $\kappa_{\mathrm{rel}} = |x \cot x|$, large near $k\pi$, $k \neq 0$

Remark 1.9. $\kappa_{\mathrm{rel}}(x)$ is a *worst-case amplification factor*: a small relative input error ε can produce an output error of about $\kappa_{\mathrm{rel}}(x)\,\varepsilon$ — and no algorithm that only sees the perturbed input can repair this.

Well-conditioned $\Rightarrow$ accurate result? Example 1.1 says no: $\kappa_{\mathrm{rel}}(x) \le 2$ there, yet f loses k digits. The *algorithm* matters too.

Algorithm $\tilde{f}$ for f: finitely many operations, each subject to (1.2).

Definition 1.10

C of moderate size (independent of x and u), admissible inputs $x \neq 0$, $f(x) \neq 0$: $\tilde{f}$ is

(i) **forward stable** if $\dfrac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} \le C\,\big(1 + \kappa_{\mathrm{rel}}(x)\big)\,u$ for all x;

(ii) **backward stable** if for every x there is $\tilde{x}$ with

$$\tilde{f}(x) = f(\tilde{x}) \qquad \text{and} \qquad \frac{\|\tilde{x} - x\|}{\|x\|} \le C\,u.$$

forward stable: as accurate as the conditioning permits
backward stable: *exact* result for slightly perturbed data

Backward stability is the stronger property

Remark 1.11. Fix x with $\kappa_{\mathrm{rel}}(x) < \infty$. If $\tilde{f}(x) = f(\tilde{x})$ with $\|\tilde{x} - x\|/\|x\| \le Cu$, then

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \frac{\|f(\tilde{x}) - f(x)\|}{\|f(x)\|} \le \big(\kappa_{\mathrm{rel}}(x) + o(1)\big)\, C\, u \qquad \text{as } u \to 0.$$

The rule to keep in mind

$$\text{error of the result} \lesssim \underbrace{\text{conditioning of the problem}}_{\text{given by the problem}} \times \underbrace{\text{backward error of the algorithm}}_{\text{choose a good algorithm}}$$

Definition 1.14. A **norm** on $\mathbb{R}^n$: $\|\cdot\|\colon \mathbb{R}^n \to [0,\infty)$ with

$$\|x\| = 0 \iff x = 0, \qquad \|\alpha x\| = |\alpha|\,\|x\|, \qquad \|x+y\| \le \|x\| + \|y\|.$$

$$\|x\|_1 = \sum_j |x_j|, \qquad \|x\|_2 = \sqrt{x^\top x}, \qquad \|x\|_\infty = \max_j |x_j| \qquad \text{(norms: Proposition 1.15)}$$

Theorem 1.16 (equivalence, with constants)

$$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\,\|x\|_\infty, \qquad \|x\|_\infty \le \|x\|_1 \le n\,\|x\|_\infty, \qquad \|x\|_2 \le \|x\|_1 \le \sqrt{n}\,\|x\|_2$$

(sharp: $x = e_1$ on the left, $x = (1,\dots,1)^\top$ on the right)

Theorem 1.17. Any two norms on $\mathbb{R}^n$ are equivalent: $c\,\|x\|_a \le \|x\|_b \le C\,\|x\|_a$ $(0 < c \le C)$.

Frobenius norm $\|A\|_F = \big(\sum_{i,j} a_{ij}^2\big)^{1/2}$. More useful: how much can A stretch a vector?

Definition 1.19 (induced norm)

Vector norms on $\mathbb{R}^n$ and on $\mathbb{R}^m$, both written $\|\cdot\|$; for $A \in \mathbb{R}^{m\times n}$

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|.$$

$$\|Ax\| \le \|A\|\,\|x\| \quad \text{for all } x, \tag{1.4}$$

and $\|A\|$ is the smallest such constant.

Proposition 1.20. The induced norm is a norm, $\|I\| = 1$, $\|AB\| \le \|A\|\,\|B\|$.

Explicit formulas

Theorem 1.22

For $A \in \mathbb{R}^{m\times n}$: $\|A\|_1 = \max_k \sum_i |a_{ik}|$ (maximal absolute column sum),
$\|A\|_\infty = \max_i \sum_k |a_{ik}|$ (maximal absolute row sum),
$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)}$ (square root of the largest eigenvalue of $A^\top A$).

By hand (one minute): $A = \begin{pmatrix} 1 & 2 \\ -3 & 1 \end{pmatrix}$: $\|A\|_1 = ?$ $\|A\|_\infty = ?$

$\|A\|_1 = 4$ $(x = e_1)$, $\|A\|_\infty = 4$ $(x = (-1, 1)^\top)$, $\|A\|_2 = \sqrt{(15 + \sqrt{29})/2} \approx 3.19$
(Example 1.25)

The 2-norm and the spectral radius

$B \in \mathbb{R}^{n\times n}$: $\rho(B) = \max\{|\lambda| : \lambda \in \mathbb{C} \text{ eigenvalue of } B\}$ · Q orthogonal: $Q^\top Q = I$

Proposition 1.23

(i) $\|A^\top\|_2 = \|A\|_2$; (ii) $\|QA\|_2 = \|A\|_2 = \|AP\|_2$ for orthogonal Q, P;
(iii) A symmetric $\Rightarrow \|A\|_2 = \rho(A)$; in general $\|A\|_2 = \sqrt{\rho(A^\top A)}$

Proposition 1.24

$\rho(A) \le \|A\|$ for every $A \in \mathbb{R}^{n\times n}$ and every induced norm.

For the 2×2 matrix above: $\rho(A) = \sqrt{7} \approx 2.65 \le 3.19$.

Given $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$: find x with $Ax = b$. A **regular** (nonsingular: $\det A \neq 0 \iff \ker A = \{0\}$): unique solution for every b.

Data $b \to b + \Delta b$ (measurement, rounding); solution $x \to x + \Delta x$ with $A\,\Delta x = \Delta b$. How large is $\|\Delta x\|/\|x\|$?

Throughout: $\|\cdot\|$ a vector norm on $\mathbb{R}^n$ and the matrix norm it induces.

Definition 2.1 (condition number of a matrix)

For regular $A \in \mathbb{R}^{n\times n}$ and an induced matrix norm,

$$\operatorname{cond}(A) = \|A\|\,\|A^{-1}\|; \qquad \operatorname{cond}_p(A) \text{ for the norm induced by } \|\cdot\|_p.$$

Perturbation of the right-hand side

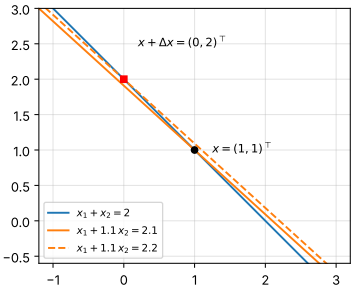
Theorem 2.2

A regular, $Ax = b \neq 0$, $A(x + \Delta x) = b + \Delta b$. Then

$$\frac{\|\Delta x\|}{\|x\|} \;\leq\; \operatorname{cond}(A)\,\frac{\|\Delta b\|}{\|b\|}.$$

The bound is **sharp**: for every regular A there are $b \neq 0$, $\Delta b \neq 0$ with equality.

In the picture (∞-norm): $\operatorname{cond}_\infty(A) = 44.1$; $\Delta b = (0, 0.1)^\top$ changes b by $4.8\,\%$ and x by $100\,\%$ — amplification 21.



$x_1 + x_2 = 2,\;\; x_1 + 1.1\,x_2 = 2.1$

Remark 2.3. $\operatorname{cond}(A) \geq \|AA^{-1}\| = 1$, $\operatorname{cond}(\alpha A) = \operatorname{cond}(A)$ $(\alpha \neq 0)$: the size of $\det A$ says nothing about conditioning ($A = \alpha I$: $\det A = \alpha^n$, $\operatorname{cond}(A) = 1$). $\operatorname{cond}(A)$ is the *worst case* over all b of κ_{rel} of $b \mapsto A^{-1}b$.

Theorem 2.5 (perturbation of matrix and right-hand side)

A regular, $Ax = b \neq 0$, $\|A^{-1}\|\,\|\Delta A\| < 1$. Then $A + \Delta A$ is regular and $(A + \Delta A)(x + \Delta x) = b + \Delta b$ gives

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\operatorname{cond}(A)}{1 - \operatorname{cond}(A)\,\dfrac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|} \right).$$

Proof: Lemma 2.4 ($\|B\| < 1 \Rightarrow I + B$ regular, $\|(I + B)^{-1}\| \leq 1/(1 - \|B\|)$) applied to $(A + \Delta A)\,\Delta x = \Delta b - \Delta A\, x$.

Relative data errors $\leq \varepsilon$ with $\operatorname{cond}(A)\,\varepsilon \ll 1$ $\Rightarrow$ relative error of $x \lesssim 2\operatorname{cond}(A)\,\varepsilon$.
Stored data carry errors $\approx u$ $\Rightarrow$ **errors of order $\operatorname{cond}(A)\,u$ are possible whatever the algorithm**.

Residual and error

Question: a computed $\tilde{x}$ has a tiny residual $r = b - A\tilde{x}$. Is $\tilde{x}$ accurate?

Proposition 2.6

(i) A symmetric and regular with eigenvalues λ_i: $\operatorname{cond}_2(A) = \dfrac{\max_i |\lambda_i|}{\min_i |\lambda_i|}$

(ii) Q orthogonal: $\operatorname{cond}_2(Q) = 1$

(iii) A regular, $Ax = b \neq 0$, $\tilde{x}$ arbitrary, $r := b - A\tilde{x}$:

$$\frac{1}{\operatorname{cond}(A)} \frac{\|r\|}{\|b\|} \;\leq\; \frac{\|x - \tilde{x}\|}{\|x\|} \;\leq\; \operatorname{cond}(A) \frac{\|r\|}{\|b\|}$$

Answer: a small error is *guaranteed* if $\operatorname{cond}(A)\,\|r\|/\|b\|$ is small; otherwise the bound is inconclusive. A small residual alone is a small *backward* error, $A\tilde{x} = b - r$; the residual is computable, the error is not.
(ii) is why Chapter 3 works with orthogonal transformations.

Example 2.7: the Hilbert matrix

$H_n = \left(\frac{1}{i+j-1}\right)_{i,j=1}^{n}$, the Gram matrix of $1, t, \ldots, t^{n-1}$ in $L^2(0,1)$ (Chapter 3):
SPD, but $\operatorname{cond}_2(H_n) \approx c\,34^n/\sqrt{n}$ — exponentially ill-conditioned.

n	cond_2(H_n)	rel. error of numpy's solve for H_n x = H_n * ones
4	1.55e+04	3.37e-14
6	1.50e+07	4.08e-11
8	1.53e+10	3.71e-08
10	1.60e+13	2.27e-05
12	1.64e+16	1.29e-01
14	3.22e+17	3.02e+00

- the error grows roughly like $\operatorname{cond}_2(H_n)$ (Theorem 2.5, $\varepsilon \approx u$: rounded entries); $\operatorname{cond}_2(H_n)\,u$ is a worst-case scale (here 10–$100\times$ the error), not a prediction
- $n \geq 12$: $\operatorname{cond}_2(H_n)\,u \gtrsim 1$ — Theorem 2.5 no longer guarantees any accuracy; observed: $13\,\%$ ($n = 12$), $300\,\%$ ($n = 14$)
- cond_2 column computed in double precision: unreliable for $n = 14$ (true $1.9 \cdot 10^{19}$)

Lemma 2.8. L, M lower triangular: $\det L = \prod_i l_{ii}$; LM lower triangular (diagonal $l_{ii}m_{ii}$); L regular $\Rightarrow L^{-1}$ lower triangular (diagonal $1/l_{ii}$). Unit lower triangular matrices (diagonal 1) are closed under products and inverses. Same for upper triangular.

Algorithm 2.9 (backward and forward substitution; U, L regular)

$$Ux = b: \qquad x_i = \frac{1}{u_{ii}}\Big(b_i - \sum_{j=i+1}^{n} u_{ij}x_j\Big), \qquad i = n, n-1, \ldots, 1$$

$$Ly = b: \qquad y_i = \frac{1}{l_{ii}}\Big(b_i - \sum_{j=1}^{i-1} l_{ij}y_j\Big), \qquad i = 1, 2, \ldots, n$$

Cost: $\sum_{i=1}^{n} \big(2(n-i)+1\big) = n^2$ flops per substitution.

Idea: subtract multiples of one equation from the others until the system is triangular.

Algorithm 2.10 (Gaussian elimination without pivoting)

$A^{(1)} := A$, $b^{(1)} := b$. For $k = 1, \dots, n-1$:
pivot $a_{kk}^{(k)}$ (stop if 0), **multipliers** $l_{ik} := a_{ik}^{(k)} / a_{kk}^{(k)}$ $(i > k)$, update

$$a_{ij}^{(k+1)} := a_{ij}^{(k)} - l_{ik}\, a_{kj}^{(k)}, \qquad b_i^{(k+1)} := b_i^{(k)} - l_{ik}\, b_k^{(k)} \qquad (i > k). \tag{2.1}$$

After $n-1$ steps: $U := A^{(n)}$ upper triangular; solve $Ux = b^{(n)}$ by backward substitution ($u_{nn} = 0 \Rightarrow A$ singular).

- row operations do not change the solution set: $Ux = b^{(n)}$ and $Ax = b$ have the same solution
- step k leaves rows $1, \dots, k$ unchanged, creates the zeros below the pivot and updates only the trailing block $i, j > k$

Gaussian elimination is the factorisation $A = LU$

Lemma 2.11

Pivots nonzero, $\ell_k := (0, \dots, 0, l_{k+1,k}, \dots, l_{nk})^\top$, $E_k := I - \ell_k e_k^\top$. Then $A^{(k+1)} = E_k A^{(k)}$, $E_k^{-1} = I + \ell_k e_k^\top$, and

$$A = LU, \qquad L = E_1^{-1} \cdots E_{n-1}^{-1} = I + \sum_{k=1}^{n-1} \ell_k e_k^\top \quad \text{(multipliers below the unit diagonal)}.$$

Definition 2.12. $A = LU$, L unit lower triangular, U upper triangular: an **LU decomposition**.

Theorem 2.13

A regular A has an LU decomposition $\iff$ $A_k := (a_{ij})_{i,j \le k}$ is regular for all $k < n$. Then it is unique, U is regular, L, U come from Algorithm 2.10 without breakdown, and $a_{kk}^{(k)} = \det A_k / \det A_{k-1}$ with $\det A_0 := 1$ (2.3).

$A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ is regular, but $A_1 = (0)$: no LU decomposition — the elimination stops at once.

$L(Ux) = b$: (1) $Ly = b$ forward substitution, (2) $Ux = y$ backward substitution.

Cost

decomposition: $\sum_{k=1}^{n-1} \big(2(n-k)^2 + (n-k)\big) = \frac{2}{3}n^3 + O(n^2)$ flops

the two substitutions: $2n^2 - n \approx 2n^2$ flops (no divisions for the unit diagonal of L)

Question: why not compute A^{-1} and form $A^{-1}b$?

- A^{-1} costs $\approx 2n^3$ flops (three times as much), and $A^{-1}b$ need not have a small residual — no accuracy advantage
- factorise *once*; every further right-hand side costs $2n^2$ (time steps in Chapter 10)
- $\det A = \pm \prod_k u_{kk}$ (sign from the row exchanges); against over-/underflow compute the sign and $\log|\det A| = \sum_k \log|u_{kk}|$ (`numpy.linalg.slogdet`)

Without pivoting the elimination **fails** for $\left(\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right)$ — and it is **unstable** when a pivot is small relative to the entries below it:

Example 2.15. $A = \left(\begin{smallmatrix}\varepsilon&1\\1&1\end{smallmatrix}\right)$, $\varepsilon = 10^{-17}$: $l_{21} = 10^{17}$, $u_{22} = \mathrm{fl}(1 - 10^{17}) = -10^{17}$, $\tilde{L}\tilde{U} = \left(\begin{smallmatrix}\varepsilon&1\\1&0\end{smallmatrix}\right)$: $a_{22} = 1$ is lost, although $\mathrm{cond}_2(A) \approx 2.6$.

Partial pivoting: in step k exchange rows so that the pivot is the entry of largest modulus among $a_{ik}^{(k)}$, $i = k, \ldots, n$ $\Rightarrow$ all $|l_{ik}| \le 1$.

Theorem 2.16

For every regular A there are a permutation matrix P, a unit lower triangular L with $|l_{ij}| \le 1$ and a regular upper triangular U with $PA = LU$. Solve $Ax = b$: $Ly = Pb$, then $Ux = y$.

With the rows exchanged, Example 2.15 is harmless: $l_{21} = \varepsilon$, $u_{22} = \mathrm{fl}(1 - \varepsilon)$.

Definition 2.18

$A \in \mathbb{R}^{n\times n}$ is **symmetric positive definite (SPD)** if $A = A^\top$ and $x^\top A x > 0$ for all $x \neq 0$. (semidefinite: $A = A^\top$, $x^\top A x \geq 0$)

Ubiquitous: $B^\top B$ for B of full column rank (Chapter 3), Hessians at non-degenerate minima, discretised elliptic operators (Example 2.25), nonsingular covariance matrices.

Proposition 2.19

A SPD $\Rightarrow$ (i) A regular; (ii) all leading principal submatrices A_k are SPD; (iii) all eigenvalues positive, $\det A > 0$; (iv) $a_{ii} > 0$ and $\max_{i,j} |a_{ij}| = \max_i a_{ii}$.

(ii) + Theorem 2.13 $\Rightarrow$ LU without pivoting; for SPD matrices the symmetry halves the work:

The Cholesky decomposition

Theorem 2.20

A is SPD $\iff$ $A = LL^\top$ with L lower triangular with positive diagonal entries; then L is unique.

Algorithm 2.21 (column by column)

Compare the entries of $A = LL^\top$, $a_{ij} = \sum_{k \le \min(i,j)} l_{ik} l_{jk}$, column by column: for $j = 1, \dots, n$

$$l_{jj} = \Big(a_{jj} - \sum_{k<j} l_{jk}^2 \Big)^{1/2}, \qquad l_{ij} = \frac{1}{l_{jj}} \Big(a_{ij} - \sum_{k<j} l_{ik} l_{jk} \Big) \quad (i > j).$$

Remark 2.22. cost $\frac{1}{3}n^3 + O(n^2)$ flops (half of LU); no pivoting ($|l_{ik}| \le \sqrt{a_{ii}}$), backward stable;
test for positive definiteness of a symmetric A: in exact arithmetic a radicand ≤ 0 appears $\iff$ A is not positive definite (in $\mathbb{F}$ a nearly singular SPD matrix may fail too).

Example 2.23: Cholesky by hand

$$A = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 10 & 2 \\ -2 & 2 & 5 \end{pmatrix}: \quad \text{column 1: } \; l_{11} = \sqrt{4} = 2, \; l_{21} = 2/2 = 1, \; l_{31} = -2/2 = -1$$

column 2: $l_{22} = \sqrt{10 - 1^2} = 3$, $l_{32} = \big(2 - (-1)\cdot 1\big)/3 = 1$

column 3: $l_{33} = \sqrt{5 - (-1)^2 - 1^2} = \sqrt{3}$

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 3 & 0 \\ -1 & 1 & \sqrt{3} \end{pmatrix}, \qquad \text{all radicands positive } \Rightarrow \; A \text{ is SPD.}$$

$\left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right)$: $l_{11} = 1$, $l_{21} = 2$, radicand $1 - 2^2 = -3 < 0$ $\Rightarrow$ not positive definite (eigenvalues 3, -1).

Sheet 1, Exercise 4(b): the same for two other matrices.

(Half-)bandwidth $w \geq 1$: $a_{ij} = 0$ for $|i-j| > w$ ($\leq 2w+1$ nonzero diagonals, $O(nw)$ storage).

Proposition 2.24

A of bandwidth w, elimination without pivoting does not break down. Then every $A^{(k)}$ has bandwidth w, $l_{ik} = 0$ for $i-k > w$, $u_{kj} = 0$ for $j-k > w$: no fill outside the band. Cost: $O(nw^2)$ (decomposition), $O(nw)$ (substitution).

Tridiagonal ($w = 1$): Thomas algorithm, $O(n)$. Partial pivoting: controlled fill-in, still $O(nw^2)$; SPD or strictly diagonally dominant: no pivoting.

Summary

- **conditioning** is a property of the problem (κ_{rel}, Definition 1.6), **stability** a property of the algorithm (Definition 1.10); error $\lesssim$ conditioning $\times$ backward error
- **norms**: $\|Ax\| \le \|A\|\,\|x\|$, $\|A\|_1$ max. absolute column sum, $\|A\|_\infty$ max. absolute row sum, $\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)}$
- $\mathrm{cond}(A) = \|A\|\,\|A^{-1}\| \ge 1$ bounds the amplification of relative data errors in $Ax = b$; a small residual guarantees a small error if $\mathrm{cond}(A)\,\|r\|/\|b\|$ is small
- Gaussian elimination with partial pivoting: $PA = LU$ in $\frac{2}{3}n^3$ flops, each right-hand side $2n^2$; SPD: Cholesky $A = LL^\top$ in $\frac{1}{3}n^3$, no pivoting, a test for positive definiteness (exact arithmetic); bandwidth w: $O(nw^2)$