- Last time: Direct solvers for $Ax=b$. $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$
  - △ Pros:
    - * Compute x exactly if arithmetic is exact
    - * Rounding error analysis is well-studied
  - △ Cons:
    - * High complexity:

      GE/Cholesky: $O(n^3)$ FLOPs (0.1~1ns per FLOP)

      - When $n=10^3$, take 0.1s
      - When $n=10^6$, take $10^8 s \approx 3$ yrs

      — Impractical in complexity and memory
    - * Low complexity variants are tailored to specific matrix structures.
    - * Sparse A? — could lead to dense LU even A is sparse

- Today: Iterative solvers for $Ax=b$
  - * Instead of solving $Ax=b$ directly, generate an approximate sequence $x_1, x_2, x_3, \cdots$ such that $x_k \xrightarrow{k \to +\infty} x = A^{-1}b$
  - * Each iteration is cheap to evaluate: matrix-vector multiplication only, Bx.

    where $B=A$ or constructed from A

    (share "similar" structure with A).

ex: A sparse (mostly zero)

say, only m nonzero entries per row

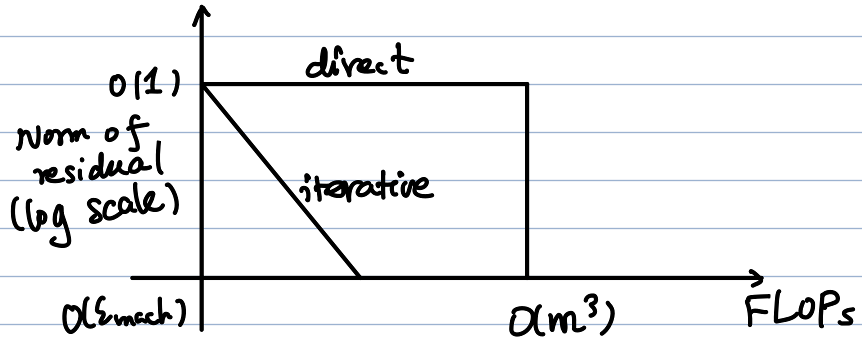
$\Rightarrow Ax$ cost $2mn$ FLOPs

ex: A is cyclic

$\Rightarrow Ax$ is $O(n\log n)$ FLOPs with FFTs

$$\begin{pmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{pmatrix}$$

* Solving the problem approximately is not a problem because even direct solvers suffer from rounding error



* Rounding error analysis of iterative solvers is not well-developed — only analyze approximation error in this class.

- Iterative solvers

Let $x_1, \dots, x_K \in \mathbb{C}^n$ be known vectors

Compute the next vector through

$$x_{K+1} = F_K(x_K, x_{K-1}, \dots, x_{K-m})$$

$\uparrow$ $(m+1)$-step methods

Easiest case: $m=0$, and F_K linear function

$$x_{k+1} = B_k x_k + f_k$$

<u>Stationary iterative method</u>: $B_k \equiv B \in \mathbb{C}^{n\times n}$, $f_k \equiv f \in \mathbb{C}^n$

↑
topic this week

More efficient iterative methods: <u>Krylov subspace methods</u>

↑
topic Later

- Convergence of vectors and matrices

Def: • Let $\{x_k\}_{k=0}^{\infty} \subseteq \mathbb{C}^n$, if $\exists x^* \in \mathbb{C}^n$ s.t.

$$\|x_k - x^*\| \longrightarrow 0 \quad \text{as} \quad k \to +\infty$$

Then we say $\{x_k\}_{k=0}^{\infty}$ converges to x^*.

• Let $\{A_k\}_{k=0}^{\infty} \subseteq \mathbb{C}^{m\times n}$, if $\exists A^* \in \mathbb{C}^{m\times n}$ s.t.

$$\|A_k - A^*\| \to 0 \quad \text{as} \quad k \to +\infty$$

Then we say $\{A_k\}_{k=0}^{\infty}$ converges to A^*.

Rk: – By the equivalence of norms on finite dimensional space, both definitions are independent of the choice of norm.

– Furthermore, by choosing $\|\cdot\|_\infty$ and $|\cdot|_\infty$ we know both convergence is equivalent to the convergence of entries.

The convergence of matrix can be characterized by testing on vectors

<u>Prop</u> Let $\{A_k\}_{k=1}^{\infty}$ then the following statements are equivalent

1) $\lim_{k\to+\infty} A_k = 0$

2) $\lim_{k\to+\infty} A_k x = 0$, $\forall x \in \mathbb{C}^n$

Pf: 1) ⇒ 2), $\|A_k x\| \le \|A_k\| \, \|x\| \to 0$ as $k \to +\infty$

2) ⇒ 1), $A_k e_j \to 0$ $\forall j = 1, \cdots, n$

thus each column of A_k converges to zero and A_k converges to zero ▨

For stationary iteration, we have

$$x_{k+1} = Bx_k + f = B^2 x_{k-1} + Bf + f$$

$$= \cdots = \underline{B^{k+1} x_0} + \sum_{i=0}^{k} B^i f$$

We thus need to understand the convergence of $\{B^k\}_{k=1}^{\infty}$

The convergence of $\{B^k\}_{k=1}^{\infty}$ is closely related to the spectrum of B. We call the maximal magnitude of the eigenvalue of a matrix the spectral radius, denoted by

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$$

The spectrum radius is closely related to the matrix norm.

Lemma 1) Let $\|\cdot\|$ be a submultiplicative matrix norm over $\mathbb{C}^{n\times n}$ then $\forall A \in \mathbb{C}^{n\times n}$, we have

$$\rho(A) \le \|A\|$$

$\|\cdot\|$ is constructed from A and ε →

2) $\forall A \in \mathbb{C}^{n\times n}$ and $\varepsilon > 0$, $\exists$ a subordinate norm $\|\cdot\|$ over $\mathbb{C}^{n\times n}$ such that $\|A\| \le \rho(A) + \varepsilon$

Pf: 1) Let λ be an eigenvalue of A s.t. $\rho(A) = |\lambda|$ with eigenvector $x \in \mathbb{C}^n$. There exists $y \in \mathbb{C}^n$ s.t.

the matrix xy^T is nonzero, thus $\|xy^T\| \neq 0$

Now we have

$$\rho(A)\,\|xy^T\| \overset{\text{homogeneity of norm}}{=} \|Axy^T\| \overset{\text{submultiplicativity}}{\leq} \|A\|\,\|xy^T\|$$

$$\Rightarrow \quad \rho(A) \leq \|A\|$$

2) We shall use the Jordan decomposition of A:

$$J = P^{-1}AP$$

$$= \begin{bmatrix} \lambda_1 & \delta_1 & & & \\ & \lambda_2 & \delta_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & \delta_{n-1} \\ & & & & \lambda_n \end{bmatrix}$$

$\delta_i = 0$ or 1, $i = 1, \cdots, n-1$. P nonsingular

Define $D_\varepsilon = \operatorname{diag}(1, \varepsilon, \varepsilon^2, \cdots, \varepsilon^{n-1})$

Then

$$J_\varepsilon = D_\varepsilon^{-1} J D_\varepsilon = \begin{bmatrix} \lambda_1 & \varepsilon\delta_1 & & & \\ & \lambda_2 & \varepsilon\delta_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & \varepsilon\delta_{n-1} \\ & & & & \lambda_n \end{bmatrix}$$

Clearly, $\|J_\varepsilon\|_\infty = \max\limits_{1\leq i\leq n} \left(|\lambda_i| + \varepsilon|\delta_i|\right) \leq \rho(A) + \varepsilon$

Now consider the vector norm

$$\|x\| := \|D_\varepsilon^{-1}P^{-1}x\|_\infty$$

Then the subordinate norm $\|A\| \overset{\text{exercise}}{=} \|D_\varepsilon^{-1}P^{-1}APD_\varepsilon\|_\infty$

$$= \|J_\varepsilon\|_\infty \leq \rho(A) + \varepsilon$$

∎

The following theorem relates the convergence of $\{B^k\}_{k=0}^{\infty}$. the spectral radius of B and norm of B :

<u>Thm</u> Let $B \in \mathbb{C}^{n\times n}$, then the following three statements are equivalent:

1) $\lim\limits_{k\to+\infty} B^k = 0$

2) $\rho(B) < 1$

Not for all norms

It is possible that $\|B\| \geq 1$ from some norms

3) there exists a subordinate norm such that $\|B\| < 1$.

Pf: 1) $\Rightarrow$ 2) Let (λ, x) be an eigenpair of B.

then $\|B^k x\| = |\lambda|^k \|x\| \to 0$ as $k \to \infty$

Implies that $|\lambda| < 1$

2) $\Rightarrow$ 3) Proved in the lemma

3) $\Rightarrow$ 1) $\|B^k\| \leq \|B\|^k \longrightarrow 0$ as $k \to +\infty$ ▨

Remark: Requiring $\|B\| < 1$ for some specific $\|\cdot\|$ is a sufficient but not necessary condition for the convergence. The significance of 3) is that it gives you a hope to check convergence with commonly used matrix norms.

<u>Thm</u> Let $B \in \mathbb{C}^{n\times n}$. $\|\cdot\|$ a submultiplicative norm

then $$\lim_{k\to+\infty} \|B^k\|^{1/k} = \rho(B)$$

Pf: Since $\rho(B) \leq \|B\|$, we know that

Why? $\rho(B) = (\rho(B^k))^{1/k} \leq \|B^k\|^{1/k}$

On the other hand, consider the matrix

$$B_\varepsilon = (\rho(B)+\varepsilon)^{-1} B, \quad \text{where } \varepsilon \geq 0$$

Clearly. $\rho(B_\varepsilon) < 1$. and thus $B_\varepsilon^k \to 0$ as $k \to +\infty$

so we have

$$\|B^k\|^{1/k} = (\rho(B)+\varepsilon)\, \|B_\varepsilon^k\|^{1/k} < \rho(B)+\varepsilon$$

for sufficiently large k. Thus $\lim\limits_{k\to+\infty} \|B^k\|^{1/k} = \rho(B)$

▨

ex. Let $B = \begin{pmatrix} 1/2 & 0 \\ 1/4 & 1/2 \end{pmatrix}$, $\lambda_1 = \lambda_2 = 1/2$. so $\rho(B) = 1/2$

$$B^k = \begin{pmatrix} 1/2^k & 0 \\ k/2^k & 1/2^k \end{pmatrix}. \qquad \|B^k\|_\infty = \frac{1}{2^k}\left(1+\frac{k}{2}\right)$$

$$\|B^k\|_\infty^{1/k} = \frac{1}{2}\left(1+\frac{k}{2}\right)^{1/k} \to \frac{1}{2} = \rho(B)$$

▨

Consider stationary iteration

$$x_{k+1} = Bx_k + f$$

Suppose $x_k \longrightarrow x^*$ as $k \to +\infty$

Taking limit on both sides,

$$x^* = Bx^* + f$$

Let error $e_k = x_k - x^*$,

then $\{e_k\}_{k=0}^{\infty}$ satisfies

$$e_{k+1} = Be_k = B^k e_0$$

<u>Corollary</u> The stationary iteration converges if and only if one of the following conditions holds true

1) $\rho(B) < 1$

2) $\exists$ a subordinate norm s.t. $\|B\| < 1$.

the limit x^* is unique under either condition.

Pf: Let $x^* = Bx^* + f$, $y^* = By^* + f$, then

$$\|x^* - y^*\| = \|B(x^* - y^*)\| \le \|B\| \, \|x^* - y^*\| \Rightarrow \|x^* - y^*\| = 0$$ ▨

We can obtain different error bounds for the iteration

Thm Let $\|\cdot\|$ be a subordinate norm, then when $\|B\| = q < 1$, we have

$$\|x_k - x^*\| \le \frac{q}{1-q} \|x_k - x_{k-1}\| \qquad \text{(a posteriori)}$$

$$\|x_k - x^*\| \le \frac{q^k}{1-q} \|x_1 - x_0\| \qquad \text{(a priori)}$$

Pf: From $x_{k+1} = Bx_k + f$

$$x_k - x^* = B(x_{k-1} - x^*)$$

$$= B(x_{k-1} - x_k) + B(x_k - x^*)$$

$$\Rightarrow \|x_k - x^*\| \le q \|x_{k-1} - x_k\| + q \|x_k - x^*\|$$

$$\Rightarrow \|x_k - x^*\| \le \frac{q}{1-q} \|x_{k-1} - x_k\| \qquad \text{(a posteriori)}$$

$$= \frac{q}{1-q} \|B(x_{k-2} - x_{k-1})\|$$

$$\le \frac{q^2}{1-q} \|x_{k-2} - x_{k-1}\|$$

$$\le \cdots \le \frac{q^k}{1-q} \|x_1 - x_0\| \qquad \text{(a priori)}$$

▨

- From $e_k = B^k e_0$ we know $\|e_k\| \le \|B^k\| \, \|e_0\|$

on average, the contraction rate in each step is $\|B^k\|^{1/k} \to \rho(B)$

Usually use $\rho(B)$ to compare the convergence rate of the iteration.

- How to construct B ?

matrix splittings: $A = (A-C) + C$ (C invertible)

$$b = Ax = (A-C)x + Cx$$

$$\Rightarrow \quad x = \underbrace{C^{-1}(C-A)}_{B}\, x + \underbrace{C^{-1}b}_{f}$$

common choices:

$$A = \underset{\substack{\uparrow \\ \text{strictly} \\ \text{lower } \triangle \\ \text{part}}}{L} + \underset{\substack{\uparrow \\ \text{diagonal} \\ \text{part}}}{D} + \underset{\substack{\nwarrow \\ \text{strictly} \\ \text{upper } \triangle \\ \text{part}}}{U}$$

Want C easily invertible

- $C = D$ Jacobi iteration
- $C = L+D$ Gauss-Seidel iteration.
- $C = L + D/\omega$ Successive over-relaxation (SOR)
- $C = D/\omega$ Relaxed Jacobi iteration.