• Last time: Stationary iterative solver for $Ax = b$

$A \in \mathbb{C}^{n \times n}$, $b \in \mathbb{C}^n$

$$x_{k+1} = B x_k + b \qquad (*)$$

<u>Thm</u> $(*)$ converges to a fixed point iff one of the following conditions hold:

1) $\rho(B) < 1$

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

It is possible for some norm we have $\|B\| \geq 1$ but $(*)$ converges

Let x^* be the fixed point of $(*)$, then $d_k = x_k - x^*$ satisfies $d_k = B^k d_0$

The convergence factor is defined as

$$\rho = \lim_{k \to +\infty} \left(\max_{x_0 \in \mathbb{R}^n} \frac{\|d_k\|}{\|d_0\|} \right)^{1/k}$$

$$= \lim_{k \to +\infty} \|B^k\|^{1/k}$$

$$= \rho(B)$$

That is, $\|d_k\| \approx [\rho(B)]^k \|d_0\|$

* if we set $\tau = -\log[\rho(B)] \longleftarrow$ convergence rate

then $\|d_k\| \approx e^{-\tau k} \|d_0\|$

* Larger spectral gap of B from $1 \Leftrightarrow$ Larger convergence rate

Set $\rho(B) = 1 - \varepsilon$, then $\tau = -\log(1-\varepsilon) \approx \varepsilon$ if $0 < \varepsilon \ll 1$

- Choices of B:

$$A = (A-C) + C$$

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

C invertible

$$x_{k+1} = C^{-1}(C-A)x_k + C^{-1}b$$

$$A = \underset{\text{lower }\Delta}{L} + \underset{\text{diagonal}}{D} + \underset{\text{upper }\Delta}{U}$$

- Jacobi iteration

Take $C = D$, $B_J = -D^{-1}(L+U) = I - D^{-1}A$

$$x_{k+1} = -D^{-1}(L+U)x_k + D^{-1}b \quad (J)$$

$$(J) \iff x_{k+1} = D^{-1}[b-(L+U)x_k]$$

$$\iff \boxed{x_i^{k+1} = \frac{1}{a_{ii}}\Big[b_i - \sum_{j\neq i} a_{ij}x_j^k\Big] \qquad i = 1, \cdots, n}$$

- Gauss-Seidel

Take $C = D+L$, $B_{GS} = -(D+L)^{-1}U$

$$x_{k+1} = -(D+L)^{-1}U x_k + (D+L)^{-1}b \quad (GS)$$

$$(GS) \iff (L+D)x_{k+1} = -Ux_k + b$$

$$\iff \boxed{x_i^{k+1} = \frac{1}{a_{ii}}\Big[b_i - \underbrace{\sum_{j<i} a_{ij}x_j^{k+1}}_{\substack{\text{Use } x_j \text{ computed}\\ \text{at the current step}}} - \sum_{j>i} a_{ij}x_j^k\Big], \quad i=1,\cdots,n}$$

(not good for parallel computing)

- Successive over-relaxation (SOR)

Take $c = L + D/\omega$, $\quad B_{SOR} = -(L + D/\omega)^{-1}[(1-\frac{1}{\omega})D + U]$

$$x_{k+1} = -\left(L+\frac{D}{\omega}\right)^{-1}\left[(1-\tfrac{1}{\omega})D+U\right]x_k + \left(L+\frac{D}{\omega}\right)^{-1}b \quad (SOR)$$

$$(SOR) \iff (\omega L + D)x_{k+1} = -[(\omega-1)D + \omega U]x_k + \omega b$$

$$\iff Dx_{k+1} = Dx_k + \omega[b - (D+U)x_k - Lx_{k+1}]$$

$$\iff \boxed{x_i^{k+1} = x_i^k + \frac{\omega}{a_{ii}}\Big[b_i - \sum_{j\geq i} a_{ij}x_j^k - \underbrace{\sum_{j<i} a_{ij}x_j^{k+1}}_{\text{similar as above}}\Big] \quad i=1,\dots,n}$$

other methods:

- Relaxed Jacobi

  Take $c = D/\omega$, $\quad B_{JOR} = (1-\omega)I - \omega D^{-1}(L+U)$

  $$x_{k+1} = B_{JOR}x_k + \omega D^{-1}b$$

- Richardson's iteration

  Take $c = I/\omega$, $\quad B_R = I - \omega A$

  $$x_{k+1} = (I - \omega A)x_k + \omega b$$

- Symmetric SOR: (Better performance for symmetric A)

  $$(D + \omega L)x_{k+1/2} = [(1-\omega)D - \omega U]x_k + \omega b$$

  $$(D + \omega U)x_{k+1} = [(1-\omega)D - \omega L]x_{k+1/2} + \omega b$$

- When does Jacobi / GS / SOR converge ?

Def : A matrix A is strictly diagonally dominant if

$$\sum_{k\neq i} |a_{ik}| < |a_{ii}| \qquad \forall i = 1, \cdots, n$$

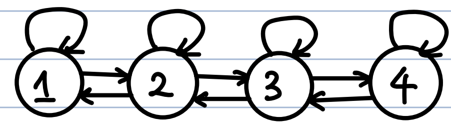
A matrix A is (weakly) diagonally dominant if

$$\sum_{k\neq i} |a_{ik}| \leq |a_{ii}|$$

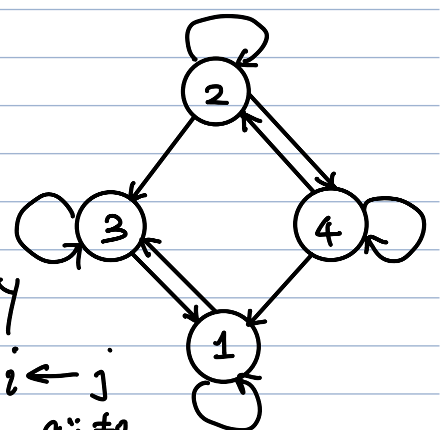
Def : A matrix A is reducible if there is a permutation matrix P such that $P^T A P = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$, where A_{11}, A_{22} are $r\times r$ and $(n-r)\times(n-r)$ square matrices. Otherwise we call A irreducible.

ex. If A is reducible, then we can permute the rows and columns of A such that $Ax=b$ becomes two small scale questions

$$A = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$ irreducible



$$B = \begin{bmatrix} 2 & 0 & -1 & 1 \\ 0 & 2 & 0 & 1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 0 & 2 \end{bmatrix}$$ reducible



Strong connectivity of Graph : $i \leftarrow j$ when $a_{ij} \neq 0$

Def A matrix A is irreducibly diagonally dominant if A is irreducible and

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}| , \quad i = 1, \cdots, n$$

with strict inequality for at least one i.

Lemma: If A is strictly diagonally dominant or irreducibly weakly diagonally dominant, then A is non-singular.

Thm The Jacobi iteration and G-S iteration converges for all initial guesses x_0 if A is strictly diagonally dominant or irreducibly diagonally dominant.

Pf: Aim to show $\rho(B_J)$, $\rho(B_{GS}) < 1$

- We first prove the results for strictly diagonally dominant matrices. Under diagonal dominance we have $a_{ii} \neq 0$, $\forall i = 1, \cdots, n$

For Jacobi. $B_J = I - D^{-1}A$

Clearly $\|B_J\|_\infty = \max_i \sum_j \left| \delta_{ij} - \frac{1}{a_{ii}} a_{ij} \right|$ $\quad \boxed{\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}}$

$$= \max_i \sum_{j \neq i} |a_{ij}| / |a_{ii}| < 1$$

So $\rho(B_J) \leq \|B\|_\infty < 1$

For G-S. $B_{GS} = -(D+L)^{-1} U$.

Let λ be the dominant eigenvalue of $-B_{GS}$. then $(|\lambda| = \rho(B_{GS}))$

$$0 = \det(\lambda I + B_{GS}) = \det(D+L)\det(\lambda(D+L)+U)$$

Since $\det(D+L) = \det(D) \neq 0$,

we know $\det(\lambda(D+L)+U) = 0$

when $|\lambda| \geq 1$, we know that $\lambda(D+L)+U$ is strictly diagonally dominat, Hence $\det(\lambda(D+L)+U) \neq 0$ which is contradiction.

Hence, $\rho(B_{GS}) = |\lambda| < 1$.

- In the case when A is irreducibly diagonally dominant,

For GS, the proof of strictly diagonolly dominat can be directly extended.

For Jacobi, let λ be the dominant eigenvalue of B_J.

Following the same proof in the strictly diagonally domdnant case, we know that $|\lambda| \leq \|B\|_\infty \leq 1$. We then show $|\lambda| < 1$:

Since λ is an eigenvalue of B_J, we know that

$B_J - \lambda I = D^{-1}(D-A) - \lambda I$ is singular.

Hence, $D(1-\lambda) - A$ is singular. If $|\lambda| = 1$, then

$D(1-\lambda) - A$ would be irreducibly diagonally dominant:

$$|a_{ii}(1-\lambda) - a_{ii}| = |\lambda a_{ii}| = |a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$$

But then we know $D(1-\lambda) - A$ is non-singular. achieving a contradiction. Hence $|\lambda| < 1$ and $\rho(B_J) < 1$. $\blacksquare$

Another set of matrices that we are interested in is Hermitian positive definite matrices.

Thm Let A be positive definite Hermitian, then

$$\rho(B_J) < 1 \iff 2D - A \text{ positive definite}$$

Pf: $B_J = I - D^{-1}A = D^{-1/2}(I - D^{-1/2}AD^{-1/2})D^{1/2}$

Thus $\rho(B_J) = \rho(I - D^{-1/2}AD^{-1/2})$

Let μ be the eigenvalue of $D^{-1/2}AD^{-1/2}$, then

$$|1-\mu| < 1$$

$$\iff 0 < \mu < 2$$

$$\iff 2I - D^{-1/2}AD^{-1/2} \text{ positive definite}$$

$$\iff D^{1/2}(2I - D^{-1/2}AD^{-1/2})D^{1/2} = 2D - A \text{ positive definite}$$

∎

The following is a necessary condition for the convergence of SOR

Thm If $\rho(B_{SOR}) < 1$ and $a_{ii} \neq 0$, $i = 1, \dots, n$, then $\omega \in (0,2)$

Pf: Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of

$$B_{SOR} = (\omega L + D)^{-1}[(1-\omega)D - \omega U]$$

Then $\lambda_1 \cdot \dots \cdot \lambda_n = \det(B_{SOR})$

$$= \det[(\omega L + D)^{-1}] \det[(1-\omega)D - \omega U]$$

$$= (1-\omega)^n$$

Hence, $1 > \rho(B_{SOR}) \geq \sqrt[n]{\prod_{i=1}^{n} |\lambda_i|} = |\omega - 1| \Rightarrow \omega \in (0,2)$ ∎

In particular, when A is Hermitian positive definite, $\omega \in (0,2)$ is sufficient

Thm If A is Hermitian positive definite, and $\omega \in (0,2)$ then SOR converges.

Pf: Let λ be an eigenvalue of B_{SOR}

Then $[(1-\omega)D - \omega U]x = \lambda(D+\omega L)x \quad (\Delta)$

Note that $2[(1-\omega)D - \omega U]$

$$= (2-\omega)D - \omega D - 2\omega U$$

$A = L+D+U$

$$= (2-\omega)D - \omega A - \omega(U-L)$$

and $2(D+\omega L)$

$$= (2-\omega)D + \omega A - \omega(U-L)$$

b.c. $A^* = A$, we know $L^* = U$

thus $x^*(U-L)x$

$= x^*(U-U^*)x$

$= 2i\, x^* \operatorname{Im}(U) x$

$=: iu$

Multiplying (Δ) by x^* we have

$$\lambda = \frac{(2-\omega)d - \omega a - i\omega u}{(2-\omega)d + \omega a - i\omega u}$$

where $d = x^* D x$, $a = x^* A x$, $iu = x^*(U-L)x$

Since A is Hermitian positive definite,

$$d > 0, \quad a > 0, \quad u \in \mathbb{R}$$

Hence, $$|\lambda|^2 = \frac{[(2-\omega)d - \omega a]^2 + \omega^2 u^2}{[(2-\omega)d + \omega a]^2 + \omega^2 u^2} < 1$$

$\omega \in (0,2)$

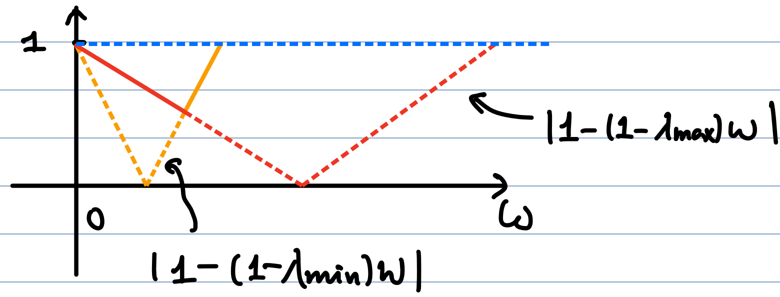
- How to find the optimal ω in relaxation ?

Goal: minimize $\rho(B)$

<u>For relaxed Jacobi,</u> $B_{JOR} = (1-\omega)I - \omega \underbrace{D^{-1}(L+U)}_{B_J}$

eigenvalues of B_{JOR} are $1-\omega+\omega\lambda_i$. λ_i eig values of B_J

suppose λ_i's are all real, $|\lambda_i|<1$ (so Jacobi converges)



ω^* is optimal when $1-(1-\lambda_{max})\omega^* = (1-\lambda_{min})\omega^* - 1$

$$\Rightarrow \quad \omega^* = \frac{1}{1-\bar{\lambda}} \qquad \bar{\lambda} := \frac{\lambda_{max}+\lambda_{min}}{2}$$

the optimal $\rho(B_{JOR}) = 1 - \dfrac{1-\lambda_{max}}{1-\bar{\lambda}}$

$$= \frac{\lambda_{max}-\lambda_{min}}{2(1-\bar{\lambda})} < \lambda_{max}$$

when $\lambda_{max} \neq -\lambda_{min}$

<u>For SOR,</u> the analysis is more involved.

<u>Thm</u> For matrices consistently ordered, i.e. if eigenvalues of $-D^{-1}(\alpha L+\alpha^{-1}U)$ are independent of α, then

(1) $[\rho(B_J)]^2 = \rho(B_{GS})$ ← GS converges twice as fast as Jacobi

(2) optimal ω^* for SOR is $\omega^* = \dfrac{2}{1+\sqrt{1-\rho(B_J)^2}}$

and optimal $\rho(B_{SOR}) = \dfrac{1-\sqrt{1-\rho(B_J)^2}}{1+\sqrt{1-\rho(B_J)^2}}$

Remark: Any tridiagonal matrix with nonzero diagonal entries are consistently ordered.

- Stationary iterative solver are less commonly used in practice.
  1) It is hard to guarantee convergence
  2) Even if converges, the convergence rate is slow
  3) G-S / SOR are hard to parallel

In special cases such as multigrid method. Jacobi is optimal

- Stationary iterative solvers can be used as preconditioner:

  The iterative scheme

$$x_{k+1} = C^{-1}(C-A)x_k + C^{-1}b$$

can be viewed as solving the system

$$[I - C^{-1}(C-A)]\,x = C^{-1}b$$

$$\Leftrightarrow \quad C^{-1}Ax = C^{-1}b \longleftarrow \text{preconditioned syste}$$

$$C_J = D$$

$$C_{GS} = D - L$$

$$C_{SOR} = \frac{1}{\omega}(D - \omega L)$$

* There is a hope that $\kappa(C^{-1}A) \ll \kappa(A)$

* In iterative solvers that used only matrix-vector product, to compute $C^{-1}Ax = [I - C^{-1}(C-A)]x$, we can do

Step 1: $r = (C-A)x$

Step 2: $w = C^{-1}r$

Step 3: $w \leftarrow x - w$