Chapter 12 Eigenvalues & Iterative methods for Solving Linear Systems.

§ 12.2 Iterative Methods for $Ax=b$.

Let A be an $n \times n$ matrix. If n is very large, iterative methods are very useful.
(read example of the Poisson's eqn. on the unit cube in 3D on page 327.)
Iterative methods deal with big size & storage difficulties, they can do much faster than GE.
To avoid having many iterations during the computation process, lin. system $Ax=b$ can be modified: a preconditioner or matrix splitting can be done. For instance, one may replace the system by the preconditioned system $M^{-1}Ax = M^{-1}b$ (M is a preconditioner), such that $M^{-1}A$ is close to identity ($\Rightarrow M^{-1}Ax = M^{-1}b$ is easy to solve).

Note: there is no need to form $M^{-1}A$, we only need to compute the product of $M^{-1}A$ with a given vector v. How?
Compute $Av \Rightarrow$ solve $Mw = Av$ for w: $w = M^{-1}Av$.

Examples of M: - the diagonal of A
- lower (or upper) triangle of A

§ 12.2.2 Simple Iteration.

Solving $Ax = b$.

Let M be a preconditioner.

Given an initial guess $x^{(0)}$, compute

$\underbrace{r^{(0)} = b - Ax^{(0)}}_{\text{residual}}$, and solve $Mz^{(0)} = r^{(0)}$ for $z^{(0)}$.

residual $\longrightarrow$ can't use error!

(Note : $r^{(0)} = Ax - Ax^{(0)} = A(x - x^{(0)}) = A\underbrace{e^{(0)}}_{\text{error}}$)

For $k = 1, 2, \ldots$

Set $x^{(k)} = x^{(k-1)} + z^{(k-1)}$

Compute $r^{(k)} = b - Ax^{(k)}$

Solve $Mz^{(k)} = r^{(k)}$ for $z^{(k)}$

unknown

$e^{(0)} = \underbrace{x}_{\text{sol.}} - \underbrace{x^{(0)}}_{\text{guess}}$

$= A^{-1}b - x^{(0)}$

can't compute!

Idea: If M is such that $M^{-1}A \approx I$, with $Mz^{(0)} = r^{(0)}$ easy to solve, then we can approximate $e^{(0)}$ by $M^{-1}r^{(0)} = z^{(0)}$ $\Rightarrow$ $x^{(1)} = x^{(0)} + z^{(0)}$, and repeat.

Depending on M used, this algorithm goes by different names:

- Jacobi iteration if $M = \text{diag}(A)$
- The Gauss-Seidel method if $M = \text{tril}(A)$ (low triange of A)
- The Successive Overrelaxation Method (SOR) if $M = \omega^{-1}D - L$

≈1950

ω is a parameter to accelerate convergence

D: diag(A)

L: strict low triangle of A

Let us take a look at this simple iteration from another pt of view:

Write $A = M - N$ $\leftarrow$ matrix splitting.
(for some matrices M, N)

Then $Ax = (M-N)x = b \Rightarrow Mx = Nx + b$

or $x = M^{-1}Nx + M^{-1}b$.

If $x^{(0)}$ is an initial guess, then for $k \geq 1$, $\underbrace{x^{(k)} = M^{-1}Nx^{(k-1)} + M^{-1}b}$

Same as in the simple iteration,

Since $M^{-1}N = M^{-1}\underbrace{(M-A)}_{''N} = I - M^{-1}A$

$$\Rightarrow x^{(k)} = (I - M^{-1}A)x^{(k-1)} + M^{-1}b$$
$$= x^{(k-1)} - M^{-1}(\underbrace{b - Ax^{(k-1)}}_{r^{(k-1)}}) \rightsquigarrow \text{residual!}$$
$$= x^{(k-1)} + z^{(k-1)}$$

$\hookrightarrow$ sol. to $Mz^{(k-1)} = r^{(k-1)}$

Example: $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Approximate solution to $Ax = b$ using two steps of the Jacobi & Gauss-Seidel methods. Let $x^{(0)} = (0,0)^T$

① Jacobi: $M = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \Rightarrow$

$$x^{(1)} = x^{(0)} + \underbrace{z^{(0)}}_{M^{-1}r^{(0)}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} =$$

$r^{(0)} = b - Ax^{(0)}$

$$= \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

$$r^{(1)} = b - Ax^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

$$\Rightarrow z^{(1)} = M^{-1} r^{(1)} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 1/4 \end{pmatrix}$$

$$\Rightarrow x^{(2)} = x^{(1)} + z^{(1)} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} + \begin{pmatrix} 1/4 \\ 1/4 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 3/4 \end{pmatrix}$$

If we continue, we'll see that the sequence $x^{(0)}, x^{(1)}, x^{(2)}, \ldots,$ moves closer to the true solution $x = (1,1)^T$.

② The Gauss-Seidel: $\quad M = \underbrace{\begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix}}_{tril(A)}$

$$\Rightarrow x^{(1)} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{x^{(0)}} + \underbrace{\begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix}^{-1}}_{M^{-1}}\underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{r^{(0)}}$$

$$= \begin{pmatrix} 1/2 & 0 \\ 1/4 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 3/4 \end{pmatrix}$$

$$r^{(1)} = \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{b} - \underbrace{\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}}_{A}\underbrace{\begin{pmatrix} 1/2 \\ 3/4 \end{pmatrix}}_{x^{(1)}} = \begin{pmatrix} 3/4 \\ 0 \end{pmatrix} \Rightarrow$$

$$z^{(1)} = M^{-1} r^{(1)} = \begin{pmatrix} 1/2 & 0 \\ 1/4 & 1/2 \end{pmatrix}\begin{pmatrix} 3/4 \\ 0 \end{pmatrix} = \begin{pmatrix} 3/8 \\ 3/16 \end{pmatrix}$$

$$\Rightarrow x^{(2)} = x^{(1)} + z^{(1)} = \begin{pmatrix} 7/8 \\ 15/16 \end{pmatrix}$$

$\rightsquigarrow$ closer to $x = (1,1)^T$ than $x^{(1)}$ in Jacobi method.

(See Ex 12.2.2 p. 331)

Convergence (§12.2.3)

Questions to ask:

① Does the simple iteration converge to the true solution x^*?

② If so, how fast does it converge?

Look at the kth step:

$$e^{(k)} = x^* - x^{(k)} = A^{-1}b - x^{(k)}$$
$$= A^{-1}b - (I - M^{-1}A)x^{(k-1)} - M^{-1}b$$
$$= \underbrace{A^{-1}b - x^{(k-1)}}_{e^{(k-1)}} + \left(M^{-1}A\,x^{(k-1)} - M^{-1}b\right)$$
$$= e^{(k-1)} + M^{-1}A\underbrace{\left(x^{(k-1)} - \underbrace{A^{-1}b}_{x^*}\right)}_{e^{(k-1)}}$$
$$= (I - M^{-1}A)e^{(k-1)} \quad \Rightarrow$$

$$e^{(k)} = (I - M^{-1}A)^2 e^{(k-2)} = \dots = (I - M^{-1}A)^k e^{(0)}$$

and

$$\|e^{(k)}\| \leq \|(I - M^{-1}A)^k\| \, \|e^{(0)}\|$$

Theorem (12.2.1) The norm of the error $\|e^{(k)}\| \to 0$ & $x^{(k)} \to A^{-1}b$ as $k \to \infty$, for every initial error $e^{(0)}$, $\boxed{\text{iff}}$

$$\lim_{k\to\infty} \|(I - M^{-1}A)^k\| = 0$$

(Proof: p. 333)

So, for stopping criteria, one can choose:

1) max. number of iterations achieved;
2) iterate while $\| r^{(k)} \|_2 > \varepsilon_{tolerance}$. (e.g.; 10^{-6})

Ex: See the MATLAB code for Jacobi iteration.

To explain the result in Thm (12.2.1) further, we'll use the spectral radius:

Def: The spectral radius of $n\times n$ matrix G is $\rho(G) = \max\{|\lambda|, \lambda = \mathrm{eig}(G)\}$.

Theorem (12.2.2) ← relates $\rho(A)$ to norms of A.

For every matrix norm $\|\cdot\|$ & every $n\times n$ matrix G, $\rho(G) \le \|G\|$. Given $G_{n\times n}$ & any $\varepsilon > 0$, there is a matrix norm $|||\cdot|||$, induced by a certain vector norm, s.t. $|||G||| \le \rho(G) + \varepsilon$.

Proof: First part: $\rho(G) \le \|G\|$

Note that there exists at least one eigenvalue λ with $|\lambda| = \rho(G)$ (by def)
If $\vec{v}$ is an eigenvector associated with this λ, and if $V_{n\times n} = \underbrace{[\vec{v}|\vec{v}|\ldots|\vec{v}]}_{n}$,
then $GV = \lambda V$, and
for any matrix norm $\|\cdot\|$, $\underbrace{|\lambda|}_{\rho(G)} \cdot \|V\| = \|\lambda V\|$
$= \|GV\| \le \|G\| \cdot \|V\|$
$\Rightarrow \rho(G) \le \|G\|$.

Second part: p.p. 333-334, using Thm 12.1.4 (Schur form): $\exists Q, T$ s.t. $G = QTQ^*$
Q → unitary matrix $(Q^* = Q^{-1})$
T → upper-triangular

□

Theorem (12.2.3) $\longrightarrow$ Consequence of Thm (12.2.2)

Let G be an $n \times n$ matrix. Then $\lim_{k \to \infty} G^k = 0$ iff $\rho(G) < 1$.

Proof: ($\Rightarrow$) $\lim_{n \to \infty} G^k = 0$. If $\lambda v = Gv \Rightarrow$ (λ-eigvalue, v-eigenvector of G).

$G^k v = \lambda^k v \xrightarrow[k \to \infty]{} 0 \Rightarrow \lambda$ must be s.t. $|\lambda| < 1 \Rightarrow$ $\rho(G) < 1$.

($\Leftarrow$) Let $\rho(G) < 1 \Rightarrow \exists\, |||\cdot|||$ s.t. $\forall \varepsilon > 0$ (from Thm. 12.2.2)

$|||G||| \le \rho(G) + \varepsilon \Rightarrow |||G||| < 1$.

Then $|||G^k||| \le |||G|||^k \xrightarrow[k \to \infty]{} 0$. Since all the norms on the set of $n \times n$ matrices are equivalent, i.e. for any two norms, $\|\cdot\|$ & $|||\cdot|||$, $\exists$ constants C_1 & C_2 s.t.

$C_1 \|G\| \le |||G||| \le C_2 \|G\|$, then it follows that $\|G^k\| \xrightarrow[k \to \infty]{} 0$ for any $\|\cdot\|$.

If we take $\|\cdot\|_1$ (or $\|\cdot\|_\infty$), then since $\|\cdot\|_1$ is the maximum absolute col. sum, and $\|G^k\|_1 \xrightarrow[k \to \infty]{} 0 \Rightarrow$ all entries in G^k must approach 0 as $k \to \infty$, i.e., $\lim_{k \to \infty} G^k = 0$. $\square$

Theorem (12.2.4) ← Thm. (12.2.1) restated!

The norm of the error in the simple iteration converges to 0 & $x^{(k)} \to A^{-1}b$ as $k \to \infty$ for every $e^{(0)}$ (i.e., initial guess $x^{(0)}$) **iff** $\rho(I - M^{-1}A) < 1$.

→ Good! But it's not easy to check in practice. For instance, one does not know the spectral radius of $I - M^{-1}A$ or/and if this radius $\rho < 1$ or not. This condition is verified for some matrices. In particular, if A is strictly diagonally dominant (i.e. $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ → each diag. entry is greater in magnitude than the sum of the magnitudes of the off-diagonal entries in the row) Then the following theorem holds:

Theorem (12.2.5)

If A is strictly diagonally dominant, then the Jacobi iteration converges to the unique solution of the linear system $Ax = b$, for any initial guess $x^{(0)}$.

<u>Sketch of Proof</u>: $M = \text{diag}(A) \Rightarrow$

$$M = \begin{pmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix} \Rightarrow$$

$G = I - M^{-1}A$ has 0's on the diagonal ($g_{ii} = 0$)

& $g_{ij} \underset{(i \neq j)}{=} -a_{ij}/a_{ii}$

$$\|G\|_\infty = \max_i \sum_{j \neq i} \left| \frac{a_{ij}}{a_{ii}} \right| = \max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}|$$

Since A is strictly diagonally dominant ($|a_{ii}| > \sum_{j \neq i} |a_{ij}|$) then $\|G\|_\infty < 1$

From Thm. 12.2.2, $\forall \|\cdot\|$, $\rho(G) \leq \|G\|$.
Thus, $\rho(G) = \rho(I - M^{-1}A) < 1 \Rightarrow$ by Thm. 12.2.4,
$x^{(k)} \underset{k \to \infty}{\longrightarrow} x^* = A^{-1}b$.

(Also: one can show that since A is strictly diagonally dominant, A is nonsingular $\Rightarrow$ solution to $Ax = b$ is unique) $\square$

<u>Note:</u> It can be shown that the Gauss-Seidel method converges to the true solution for strictly diagonally dominant matrices too.

- If A is real, SPD (symmetric & positive definite) $\Rightarrow$ the following Theorem holds:

Theorem 12.2.6: If A is a real, symmetric and positive definite matrix, and $A = I + L + L^T$, then the Gauss-Seidel method converges to the unique solution of $Ax = b$, for any initial vector $x^{(0)}$.

Sketch of Proof:

1) PD $\Rightarrow$ nonsingular $\Rightarrow$ sol. is unique.

2) $I - M^{-1}A = I - (I+L)^{-1}A = I - (I+L)^{-1}(I+L+L^T)$

$= I - I - (I+L)^{-1}L^T = -(I+L)^{-1}L^T$.

If $\rho(\underbrace{-(I+L)^{-1}L^T}_{G}) < 1 \Rightarrow$ we have convergence.

Consider $\lambda = eig(G)$ with $\vec{v}$ s.t. $\lambda\vec{v} = G\vec{v}$ and $\vec{v}$ is normalized, i.e. $\|\vec{v}\| = 1$.

$$\lambda\vec{v} = G\vec{v} = -(I+L)^{-1}L^T v$$

$$\Rightarrow \lambda(I+L)\vec{v} = -L^T\vec{v}$$

$$\lambda\vec{v}^*(I+L)\vec{v} = -\vec{v}^* L^T \vec{v}$$

($\vec{v}^*$ is a Hermitian transpose of $\vec{v}$, $\vec{v}^* = (\bar{v}_1, \dots \bar{v}_n)^T$, complex conjugates)

$$\lambda(\underbrace{\vec{v}^*\vec{v}}_{1} + \underbrace{\vec{v}^*L\vec{v}}_{\alpha+i\beta}) = \underbrace{-\vec{v}^*L^T\vec{v}}_{-(\alpha-i\beta)}$$

$$\Rightarrow \lambda(1+\alpha+i\beta) = -\alpha+i\beta$$

$$|\lambda(1+\alpha+i\beta)| = |-\alpha+i\beta|$$

$$|\lambda|\sqrt{(1+\alpha)^2+\beta^2} = \sqrt{\alpha^2+\beta^2}$$

$$\Rightarrow \quad |\lambda| = \frac{\sqrt{\alpha^2+\beta^2}}{\sqrt{\alpha^2+\beta^2+1+2\alpha}}$$

A is PD $\Rightarrow$ $\vec{v}^*A\vec{v} = \vec{v}^*(I+L+L^T)\vec{v}$

$$= 1+\underbrace{\vec{v}^*L\vec{v}}_{\alpha+i\beta} + \underbrace{\vec{v}^*L^T\vec{v}}_{\alpha-i\beta} = 1+2\alpha > 0$$

$\Rightarrow$ since $1+2\alpha<0 \Rightarrow |\lambda|<1 \Rightarrow \rho(G)<1$. $\square$

Note: If $Ax=b$ and $A=D-L-U$ with $D=\text{diag}(A)$, $L=\begin{pmatrix} 0 & 0 & \cdots & 0 \\ -a_{21} & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ -a_{n1} & \cdots & -a_{nn-1} & 0 \end{pmatrix}$

and $U=\begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & -a_{n-1n} \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$

then the simple iteration $x^{(k)} = (I-M^{-1}A)x^{(k-1)}+M^{-1}b$ takes form:

① Jacobi: $M=D$

$$x^{(k)} = \underbrace{D^{-1}(L+U)}_{T_J}x^{(k-1)} + D^{-1}b$$

(2) Gauss-Seidel: $M = D - L \Rightarrow$

$$\underbrace{(D-L-U)}_{A} x = b \Rightarrow (D-L)x = Ux + b \Rightarrow$$

$$x = (D-L)^{-1} U x + (D-L)^{-1} b$$

$$\Rightarrow x^{(k)} = \underbrace{(D-L)^{-1} U}_{T_G} x^{(k-1)} + (D-L)^{-1} b$$

(3) Relaxation Techniques: $M = \omega^{-1} D - L$

($\omega = 1 \Rightarrow$ Gauss-Seidel)

$$A = (D-L-U) + (\omega^{-1}D - \omega^{-1}D) = (\omega^{-1}D - L) - (\omega^{-1}D - D) - U$$

$$\Rightarrow \underbrace{(\omega^{-1}D - L)}_{A} x = ((\omega^{-1} - 1)D + U)x + b$$

$$x = (\omega^{-1}D - L)^{-1}((\omega^{-1}-1)D + U)x + (\omega^{-1}D - L)^{-1} b$$

$$\Rightarrow x^{(k)} = \underbrace{(\omega^{-1}D-L)^{-1}((\omega^{-1}-1)D+U)}_{T_\omega} x^{(k-1)} + \underbrace{(\omega^{-1}D-L)^{-1} b}_{c_\omega}$$

$0 < \omega < 1 \rightsquigarrow$ under-relaxation techniques

$1 < \omega < 2 \rightsquigarrow$ over-relaxation techniques (SOR)

If A - SPD & $\omega \in (0,2) \Rightarrow$ convergence $\forall x^{(0)}$.

<u>Example:</u> $A = \begin{pmatrix} 2 & -1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{pmatrix}$, $b = \begin{pmatrix} -1 \\ 4 \\ 5 \end{pmatrix}$,

solution $x = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$

(a) Show that $\rho(T_J) > 1$

(b) Show that $\rho(T_G) < 1$

(a) $D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, $L = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$, $U = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{pmatrix}$

$D^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \Rightarrow T_J = D^{-1}(L+U) =$

$= \begin{pmatrix} 0 & 1/2 & -1/2 \\ -1 & 0 & -1 \\ 1/2 & 1/2 & 0 \end{pmatrix} \Rightarrow$ eigenvalues of T_J are $\lambda_{1,2} = \pm \frac{\sqrt{5}}{2} i$, $\lambda_3 = 0$

$\Rightarrow \rho(T_J) = \left| \pm \frac{\sqrt{5} i}{2} \right| = \frac{\sqrt{5}}{2} > 1$

$\Rightarrow$ by Thm. 12.2.4, Jacobi iteration diverges.

(b) $D - L = \begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{pmatrix}$, $(D-L)^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 1/4 & 1/2 \end{pmatrix}$

$\Rightarrow T_G = (D-L)^{-1} U = \begin{pmatrix} 0 & 1/2 & -1/2 \\ 0 & -1/2 & -1/2 \\ 0 & 0 & -1/2 \end{pmatrix}$

$\Rightarrow$ eigenvalues of T_G are $\lambda_{1,2} = -1/2$, $\lambda_3 = 0$

Thus, since $\rho(T_G) = 1/2 < 1$ then, by Thm 12.2.4, the Gauss-Seidel method converges to the solution of $Ax = b$ as $k \to \infty$.