553.385 Numerical Linear Algebra, Spring 2022
Section 9

Hongyu Cheng

hongyucheng@jhu.edu

April 7, 2022

Theorem 1 (Banach fixed point)**.** *Let (X, d) be a complete metric space and $F : X \mapsto X$ a contractive mapping ($\exists \lambda < 1$ such that $d(F(x), F(y)) \leq \lambda d(x, y)$ for all $x, y \in X$). Then F admits a unique fixed point. Furthermore, it can be obtained as the limit of any sequence defined by $x_{n+1} = F(x_n)$ with $x_0 \in X$.*

Proof. Let $x_0 \in X$ and consider the sequence defined iteratively by $x_{n+1} = F(x_n)$ for all $n \in \mathbb{N}$, we can show that $\forall n \in \mathbb{N}$ and $\forall p \in \mathbb{N}$ we have

$$d(x_n, x_{n+p}) \leq \frac{\lambda^n}{1-\lambda} d(x_0, x_1)$$

$\lambda < 1$ results that $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ such that $\forall n \geq N$ and $p \in \mathbb{N}$, $d(x_n, x_{n+p}) \leq \varepsilon$. This shows that (x_n) is a Cauchy sequence and as X is complete, it means that (x_n) is converging in X, i.e. $\exists x_* \in X$ such that $x_n \to x_*$. Now, taking the limit in $x_{n+1} = F(x_n)$ and using the fact that F is continuous (since it is contractive), we have $x_* = F(x_*)$ which shows that x_* is a fixed point of F.

As for the uniqueness, if $\tilde{x}$ is another fixed point of F, we have

$$d(x_*, \tilde{x}) = d(F(x_*), F(\tilde{x})) \leq \lambda d(x_*, \tilde{x}) \Longrightarrow d(x_*, \tilde{x}) = 0, \text{ i.e. } \tilde{x} = x_*$$

□

Theorem 2 (Convergence theorems in lecture notes)**.** *$\forall \mathbf{x}^0 \in \mathbb{R}^n$, the iteration $\mathbf{x}^{k+1} = \mathbf{T}\mathbf{x}^k + \mathbf{c}$ converges to the unique solution $\mathbf{x}$ of $\mathbf{x} = \mathbf{T}\mathbf{x} + \mathbf{c}$ iff $\rho(\mathbf{T}) < 1$.*

Proof. See lecture notes. □

Iterative techniques for solving linear systems

Given $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, consider solving $Ax = b$. We write A as $A = D - L - U$, that is

$$\underbrace{\begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & \cdots & a_{n,n} \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{n,n} \end{pmatrix}}_{D} - \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ -a_{21} & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ -a_{n_1} & \cdots & -a_{n,n-1} & 0 \end{pmatrix}}_{L} - \underbrace{\begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & -a_{n-1,n} \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{U}$$

Jacobi method

$$\begin{aligned} A\mathbf{x} = \mathbf{b} &\Longrightarrow (D - L - U)\mathbf{x} = \mathbf{b} \\ &\Longrightarrow D\mathbf{x} = (L+U)\mathbf{x} + \mathbf{b} \\ &\Longrightarrow \mathbf{x} = \underbrace{D^{-1}(L+U)}_{T_J}\mathbf{x} + \underbrace{D^{-1}\mathbf{b}}_{\mathbf{c}} \end{aligned}$$

By theorem 2, our iteration is

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

that is,

$$x_i^{(k+1)} = \frac{-\sum\limits_{j\neq i} a_{ij}x_j^{(k)} + b_i}{a_{ii}}, \quad i = 1, 2, \ldots, n$$

or

$$x_i^{(k+1)} - x_i^{(k)} = \frac{-\sum\limits_{j=1}^{n} a_{ij}x_j^{(k)} + b_i}{a_{ii}}, \quad i = 1, 2, \ldots, n$$

The second iteration structure is what we used in `demo_alg071.m`, and the stop criterion is given by

$$\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\| \leq \text{TOL}$$

Gauss-Seidel method

$$\begin{aligned} A\mathbf{x} = \mathbf{b} &\Longrightarrow (D - L - U)\mathbf{x} = \mathbf{b} \\ &\Longrightarrow (D-L)\mathbf{x} = U\mathbf{x} + \mathbf{b} \\ &\Longrightarrow \mathbf{x} = \underbrace{(D-L)^{-1}U}_{T_G}\mathbf{x} + \underbrace{(D-L)^{-1}\mathbf{b}}_{\mathbf{c}} \end{aligned}$$

By theorem 2, our iteration is

$$\mathbf{x}^{(k+1)} = T_G\mathbf{x}^{(k)} + \mathbf{c} = (D-L)^{-1}U\mathbf{x}^{(k)} + (D-L)^{-1}\mathbf{b}$$

that is,

$$x_i^{(k+1)} = \frac{-\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + b_i}{a_{ii}}, \; i = 1, \ldots, n$$

Comments: For stop criterion, given the tolerance TOL$=\varepsilon$, when k large enough,

$$\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\| \approx \left\|\mathbf{x} - \mathbf{x}^{(k)}\right\| \approx [\rho(T)]^k \left\|\mathbf{x} - \mathbf{x}^{(0)}\right\| < \varepsilon$$

which is equivalent to

$$\begin{aligned} &k\underbrace{\log(\rho(T))}_{<0} + \log\left(\left\|\mathbf{x} - \mathbf{x}^{(0)}\right\|\right) < \log(\varepsilon) \\ \Longleftrightarrow\ &k > \frac{\log(\varepsilon)}{\log(\rho(T))} - \underbrace{\frac{\log\left(\left\|\mathbf{x} - \mathbf{x}^{(0)}\right\|\right)}{\log(\rho(T))}}_{\text{constant}} \end{aligned}$$

Theorem 3. *If A strictly diagonally dominant, then for any choice of $\mathbf{x}^{(0)}$ both Jacobi and Gauss-Seidel methods converge and $\|T_G\|_\infty \leq \|T_J\|_\infty < 1$*

Definition 1 (Irreducible matrix)**.** *A matrix A is said to be irreducible if $\nexists$ permutation matrix P such that*

$$P^\top A P = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ O & \tilde{A}_{22} \end{bmatrix},\ \tilde{A}_{11} \in \mathbb{E}(p,p),\ p > 0$$

Definition 2 (Irreducibly diagonally dominant)**.** *A matrix is said to be Irreducibly diagonally dominant if it is irreducible and*

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}| \quad \textit{and} \quad |a_{ss}| > \sum_{j \neq s} |a_{sj}| \textit{ for at least one } s$$

Theorem 4. *A matrix A is irreducible iff $G(A)$ is connected.*

Theorem 5. *If A is irreducibly diagonally dominant, then the Jacobi method converges.*

Relaxation Methods

SOR (Successive Over-Relaxation)

$$\begin{aligned} A\mathbf{b} = \mathbf{b} &\Longrightarrow (D - L - U)\mathbf{x} = \mathbf{b} \\ &\Longrightarrow (\omega D - \omega L - \omega U)\mathbf{x} = \omega \mathbf{b} \\ &\Longrightarrow (D - (1-\omega)D - \omega L - \omega U)\mathbf{x} = \omega \mathbf{b} \\ &\Longrightarrow (D - \omega L)\mathbf{x} = [(1-\omega)D + \omega U]\mathbf{x} + \omega \mathbf{b} \\ &\Longrightarrow \mathbf{x} = \underbrace{(D - \omega L)^{-1}[(1-\omega)D + \omega U]}_{T(\omega)}\mathbf{x} + \underbrace{(D - \omega L)^{-1}\omega \mathbf{b}}_{\mathbf{c}} \end{aligned}$$

In general, the iteration is

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

then,

$$\mathbf{x}^{(k+1)} = (1-\omega)\mathbf{x}^{(k)} + \omega D^{-1}\left[L\mathbf{x}^{(k+1)} + U\mathbf{x}^{(k)} + \mathbf{b}\right]$$

Recall our Gauss-Seidel iteration, then

$$\begin{aligned} \mathbf{x}^{(k+1)} &= (1-\omega) \underbrace{\mathbf{x}^{(k)}}_{\text{last iteration}} + \omega \underbrace{D^{-1}\left[L\mathbf{x}^{(k+1)} + U\mathbf{x}^{(k)} + \mathbf{b}\right]}_{\text{Gauss-Seidel iteration}} \\ &= (1-\omega)\mathbf{x}^{(k)} + \omega \mathbf{x}_{GS}^{(k+1)} \end{aligned}$$

which is a "weighted average" (or more precisely, an affine combination) of the new Gauss-Seidel value with the one obtained during the previous iteration. Here ω is the so-called relaxation parameter. If $\omega = 1$ we simply have the Gauss-Seidel method. For $\omega > 1$ of overrelaxation (accelerated convergence of GS), and for $\omega < 1$ of underrelaxation (can converge if GS non-convergent).