Classical Numerical Analysis, Chapter 06

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

Chapter 06, Part 3 of 3
Linear Iterative Methods

A Special Matrix

A Special Matrix

In this section, we will consider the following matrix, which comes up again and again in numerical analysis: for $n \geq 2$,

$$A = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -1 & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \in \mathbb{R}^{(n-1)\times(n-1)}.$$

This matrix fits into a class of matrices known as TST matrices, that is, Toeplitz Symmetric Tridiagonal. Toeplitz means that the entries on diagonal and super- and sub-diagonals are all the same. The other terms are standard.

Theorem (convergence)

Let A *be the TST matrix defined on the previous slid. Denote by* $\mathrm{T_J}, \mathrm{T_{GS}} \in \mathbb{R}^{(n-1)\times(n-1)}$ *the error transfer matrices of the Jacobi, and Gauss–Seidel methods, respectively. Then* $\rho(\mathrm{T_J}) < 1$ *and* $\rho(\mathrm{T_{GS}}) < 1$.

Proof.

From the proof of an earlier theorem, we observe that the distinct eigenvalues of A are $\lambda_k = 2 - 2\cos(k\pi h)$, where $h = 1/n$, for $k = 1, \ldots, n-1$. This implies

$$0 < \lambda_1 < \cdots < \lambda_{n-1} < 4.$$

In addition, we recall that the k–th eigenvector of A, associated with λ_k, can be defined as

$$[\mathbf{w}_k]_i = \sin(k\pi i h).$$

Proof, Cont.

The error transfer matrix of the Jacobi method is

$$T_J = D^{-1}(D - A) = \frac{1}{2}(2I_{n-1} - A) = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 & 0 \\ \vdots & \ddots & 1 & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

Then

$$T_J \mathbf{w}_k = \left(1 - \frac{1}{2}\lambda_k\right)\mathbf{w}_k.$$

In other words, the eigenvalues of T_J are $\mu_k = \cos(k\pi h)$, $k = 1, \ldots, n-1$. Hence

$$\rho(T_J) = \mu_1 = -\mu_{n-1} = \cos(\pi h) < 1.$$

Proof, Cont.

By another theorem from this chapter, since A is SPD and tridiagonal,

$$\rho(\mathrm{T_{GS}}) = \rho^2(\mathrm{T_J}) = \cos^2(\pi h) < 1.$$

□

Theorem (tridiagonal matrices)

Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is HPD and tridiagonal. Then*

$$\rho(\mathsf{T}_{\mathrm{GS}}) = \rho^2(\mathsf{T}_{\mathrm{J}}) < 1,$$

and the optimal choice for ω *in the relaxation method is*

$$\omega_{\mathrm{opt}} = \frac{2}{1+\sqrt{1-\rho(\mathsf{T}_{\mathrm{GS}})}}.$$

With this choice,

$$\omega_{\mathrm{opt}} - 1 = \rho(\mathsf{T}_{\omega_{\mathrm{opt}}}) = \min_{\omega\in(0,2)} \rho(\mathsf{T}_\omega).$$

Optimized Convergence

Suppose, for $n \geq 2$,

$$A = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -1 & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \in \mathbb{R}^{(n-1)\times(n-1)}.$$

Then, we proved that

$$\rho(T_{GS}) = \rho^2(T_J) = \cos^2(\pi h).$$

Optimized Convergence, Cont.

Suppose that $n = 32$, so that $h = 1/32$. Then

$$\rho(\mathsf{T}_{\mathrm{GS}}) = \cos^2(\pi/32) \approx (0.99518472)^2 \approx 0.99039264 \ .$$

But

$$\omega_{\mathrm{opt}} = \frac{2}{1 + \sqrt{1 - \cos^2(\pi/32)}} \approx 1.82146519 \ .$$

Therefore,

$$\rho(\mathsf{T}_{\omega_{\mathrm{opt}}}) = \omega_{\mathrm{opt}} - 1 \approx 0.82146519 \ .$$

The convergence rate of the relaxation method with the optimal parameter ω_{opt} is much better that that for Gauss–Seidel.

Non–Stationary Two–Layer Methods

Chebyshev's Method

Let us consider the explicit method:

$$\frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\alpha_{k+1}} + \mathrm{A}\mathbf{x}_k = \mathbf{f},$$

where $\alpha_k > 0$. This is a non–stationary method defined by setting

$$\mathrm{B}_k = \mathrm{B}_{\mathrm{C},k} = \frac{1}{\alpha_k}\mathrm{I}_n$$

and is known as *Chebyshev's method*. It is like Richardson's method, except that we will choose the parameter α_k adaptively, in such a way that, after m steps, the quantity $\|\mathbf{e}_m\|_2 = \|\mathbf{x} - \mathbf{x}_m\|_2$ is as small as possible.

Theorem (convergence of Chebyshev's method)

Let $\mathrm{A} \in \mathbb{C}^{n\times n}$ *be HPD with spectrum* $\sigma(\mathrm{A}) = \{\lambda_i\}_{i=1}^n$, $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. *For a given* $m \in \mathbb{N}$, *the quantity* $\|\mathbf{e}_m\|_2$ *is minimized if*

$$\alpha_k = \frac{\alpha_0}{1+\rho_0 t_k}, \quad k = 1, \ldots, m,$$

where

$$\alpha_0 = \frac{2}{\lambda_1+\lambda_n}, \quad \rho_0 = \frac{\kappa_2(\mathrm{A})-1}{\kappa_2(\mathrm{A})+1}, \quad t_k = \cos\left[\frac{(2k-1)\pi}{2m}\right].$$

In this case we have

$$\|\mathbf{e}_m\|_2 \leq \frac{2\rho_1^m}{1+\rho_1^{2m}}\|\mathbf{e}_0\|_2, \quad \rho_1 = \frac{\sqrt{\kappa_2(\mathrm{A})}-1}{\sqrt{\kappa_2(\mathrm{A})}+1}.$$

Proof.

Let us merely sketch the proof. The error is governed by

$$\mathbf{e}_{k+1} - \mathbf{e}_k + \alpha_{k+1}\mathrm{A}\mathbf{e}_k = \mathbf{0}.$$

Proof, Cont.

Thus,

$$\mathbf{e}_{k+1} = (\mathrm{I}_n - \alpha_{k+1}\mathrm{A})\mathbf{e}_k = \cdots = (\mathrm{I}_n - \alpha_{k+1}\mathrm{A})\cdots(\mathrm{I}_n - \alpha_1\mathrm{A})\mathbf{e}_0.$$

This implies that

$$\mathbf{e}_m = \mathrm{T}_m\mathbf{e}_0, \qquad \mathrm{T}_m = (\mathrm{I}_n - \alpha_m\mathrm{A})(\mathrm{I}_n - \alpha_{m-1}\mathrm{A})\cdots(\mathrm{I}_n - \alpha_1\mathrm{A}).$$

Since A is Hermitian, so is T_m and, therefore, $\|\mathrm{T}_m\|_2 = \rho(\mathrm{T}_m)$. It follows that

$$\|\mathbf{e}_m\|_2 \leq \rho(\mathrm{T}_m)\|\mathbf{e}_0\|_2,$$

and it suffices to minimize $\rho(\mathrm{T}_m)$.

Proof, Cont.

Notice that $\nu \in \sigma(\mathsf{T}_m)$ iff $\nu = \prod_{i=1}^{m}(1-\alpha_i\lambda)$, for some $\lambda \in \sigma(\mathsf{A})$. Now, since A is HPD

$$\rho(\mathsf{T}_m) = \max_{i=1}^{n} \left| \prod_{j=1}^{m}(1-\alpha_j\lambda_i) \right| \leq \max_{\zeta\in[\lambda_1,\lambda_n]} |p(\zeta)|,$$

where

$$p(\zeta) = \prod_{j=1}^{m}(1-\alpha_j\zeta).$$

Solving this problem is beyond the scope of our present discussion. For now, let us just say this is possible to do. The solution is given by shifted and rescaled roots of the Chebyshev polynomials T_m, which will imply all the formulas given above. See Chapter 10. □

Method of Minimal Residuals

Let $A \in \mathbb{C}^{n\times n}$ be HPD. Given an arbitrary guess $\mathbf{x}_k$, the residual is defined by $\mathbf{r}_k = \mathbf{f} - A\mathbf{x}_k$. Notice that

$$A\mathbf{e}_k = A(\mathbf{x} - \mathbf{x}_k) = \mathbf{f} - A\mathbf{x}_k = \mathbf{r}_k.$$

Let us again consider the non–stationary method

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_{k+1}(\mathbf{f} - A\mathbf{x}_k) = \mathbf{x}_k + \alpha_{k+1}\mathbf{r}_k,$$

where we will choose the iteration parameter to minimize $\|\mathbf{r}_{k+1}\|_2$. Let us apply A to the method to get

$$A\mathbf{x}_{k+1} = A\mathbf{x}^k + \alpha_{k+1}A\mathbf{r}_k,$$

which is equivalent to

$$\mathbf{r}_{k+1} = \mathbf{f} - A\mathbf{x}_{k+1} = \mathbf{f} - A\mathbf{x}_k - \alpha_{k+1}A\mathbf{r}_k.$$

To sum up,

$$\mathbf{r}_{k+1} = T_{k+1}\mathbf{r}_k, \qquad T_{k+1} = I_n - \alpha_{k+1}A.$$

Method of Minimal Residuals, Cont.

Computing the 2–norm of the residual, we find

$$\begin{aligned}\|\mathbf{r}_{k+1}\|_2^2 &= (\mathbf{r}_k - \alpha_{k+1}\mathrm{A}\mathbf{r}_k, \mathbf{r}_k - \alpha_{k+1}\mathrm{A}\mathbf{r}_k)_2 \\ &= \|\mathbf{r}_k\|_2^2 + \alpha_{k+1}^2\|\mathrm{A}\mathbf{r}_k\|_2^2 - 2\alpha_{k+1}(\mathrm{A}\mathbf{r}_k, \mathbf{r}_k)_2,\end{aligned}$$

which, being a positive quadratic in α_{k+1}, is clearly minimized by setting

$$\alpha_{k+1} = \frac{(\mathrm{A}\mathbf{r}_k, \mathbf{r}_k)_2}{\|\mathrm{A}\mathbf{r}_k\|_2^2}.$$

Thus we arrive at the so–called *method of minimal residuals*. This method is presented in the listing at the end of the chapter. It will converge faster than Richardson's method, as we show in the following result.

Theorem (convergence)

Let A *be HPD with spectrum* $\sigma(\mathsf{A}) = \{\lambda_i\}_{i=1}^n$, $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, *then*

$$\|\mathsf{A}\mathbf{e}_k\|_2 \leq \rho_\star^k \|\mathsf{A}\mathbf{e}_0\|_2, \quad \rho_\star = \frac{\kappa_2(\mathsf{A}) - 1}{\kappa_2(\mathsf{A}) + 1}.$$

Proof.

Since $\mathbf{r}_{k+1} = \mathsf{T}_{k+1}\mathbf{r}_k$ in the 2–norm is minimized by the given choice of α_{k+1}, we must have that

$$\|\mathbf{r}_{k+1}\|_2 = \|\mathsf{T}_{k+1}\mathbf{r}_k\|_2 \leq \|\mathsf{T}\mathbf{r}_k\|_2,$$

where $\mathsf{T} = \mathsf{I}_n - \alpha\mathsf{A}$ and any choice of $\alpha \in \mathbb{C}$. In particular, we can set $\alpha = \alpha_\star = \frac{2}{\lambda_1 + \lambda_n}$ to get the error transfer matrix of Richardson's method with its optimal choice of parameter. In this case, $\|\mathsf{T}\|_2 = \rho_\star < 1$. Since $\mathsf{A}\mathbf{e}_k = \mathbf{r}_k$, the result follows. □

The Method of Minimal Corrections

Let us consider now an non–stationary, two–layer method of the form

$$\frac{1}{\alpha_{k+1}}\mathsf{S}\left(\mathbf{x}_{k+1}-\mathbf{x}_k\right)+\mathsf{A}\mathbf{x}_k=\mathbf{f},$$

where $\alpha_{k+1}>0$ and $\mathsf{S}\in\mathbb{C}^{n\times n}$ is invertible. This conforms to our standard non–stationary iteration framework by setting $\mathsf{B}_{k+1}=\frac{1}{\alpha_{k+1}}\mathsf{S}$. We define the *correction* to be $\mathbf{w}_k=\mathsf{S}^{-1}\mathbf{r}_k$ and notice that

$$\mathbf{x}_{k+1}=\mathbf{x}_k+\alpha_{k+1}\mathsf{S}^{-1}(\mathbf{f}-\mathsf{A}\mathbf{x}_k)=\mathbf{x}_k+\alpha_{k+1}\mathsf{S}^{-1}\mathbf{r}_k=\mathbf{x}_k+\alpha_{k+1}\mathbf{w}_k.$$

Let us now assume that S is HPD so that it has an associated energy norm. The method of minimal corrections then chooses α_{k+1} so as to minimize $\|\mathbf{w}_{k+1}\|_{\mathsf{S}}^2$.

Theorem (convergence)

Let $\mathsf{A}, \mathsf{S} \in \mathbb{C}^{n\times n}$ *be HPD. Then*

$$\sigma(\mathsf{S}^{-1}\mathsf{A}) = \{\mu_i\}_{i=1}^{n} \subset (0,\infty).$$

Suppose

$$0 < \mu_1 \leq \cdots \leq \mu_n \quad \textit{and} \quad \kappa = \frac{\mu_n}{\mu_1}.$$

Then

$$\|\mathsf{A}\mathbf{e}_k\|_{\mathsf{S}^{-1}} \leq \rho_0^k \|\mathsf{A}\mathbf{e}_0\|_{\mathsf{S}^{-1}}, \qquad \rho_0 = \frac{\kappa - 1}{\kappa + 1}.$$

Proof.

Since S is HPD, the object $(\mathbf{x},\mathbf{y})_{\mathsf{S}} = (\mathsf{S}\mathbf{x},\mathbf{y})_2$ defines an inner product, and $\|\mathbf{x}\|_{\mathsf{S}} = \sqrt{(\mathbf{x},\mathbf{x})_{\mathsf{S}}}$ defines a norm. Since the matrix $\mathsf{S}^{-1}\mathsf{A}$ is self–adjoint and positive definite with respect to the inner product $(\,\cdot\,,\,\cdot\,)_{\mathsf{S}}$, we leave it as an exercise to the reader to show that the eigenvalues of $\mathsf{S}^{-1}\mathsf{A}$ are all real and positive.

Proof, Cont.

Furthermore, we can define the square root $S^{1/2}$. In particular, since S is HPD, there exist a unitary matrix $U \in \mathbb{C}^{n\times n}$ and a diagonal matrix $D = \mathbf{diag}\,(\nu_1, \ldots, \nu_n)$, with positive real diagonal entries, such that

$$S = UDU^H.$$

Consequently, we define $D^{1/2} = \mathbf{diag}\,(\sqrt{\nu_1}, \ldots, \sqrt{\nu_n})$, and set

$$S^{1/2} = UD^{1/2}U^H.$$

Clearly $S^{1/2}$ is HPD and $S^{1/2}S^{1/2} = S$. Let us introduce then the change of variables $\mathbf{v}_k = S^{1/2}\mathbf{w}_k$ and notice that, after a careful calculation,

$$\frac{1}{\alpha_{k+1}}\left(\mathbf{v}_{k+1} - \mathbf{v}_k\right) + C\mathbf{v}_k = \mathbf{0}, \tag{1}$$

with $C = S^{-1/2}AS^{-1/2}$.

Proof, Cont.

Finally, note that we must choose α_{k+1} in order to minimize

$$\|\mathbf{w}^{k+1}\|_S^2 = (S\mathbf{w}_{k+1}, w_{k+1})_2 = (S^{1/2}\mathbf{w}_{k+1}, S^{1/2}\mathbf{w}_{k+1})_2 = \|\mathbf{v}_{k+1}\|_2^2.$$

We can repeat some steps from the proof of a previous theorem to obtain

$$\alpha_{k+1} = \frac{(C\mathbf{v}_k, \mathbf{v}_k)_2}{\|C\mathbf{v}_k\|_2^2}.$$

Next, we observe the following identities

$$(C\mathbf{v}_k, \mathbf{v}_k)_2 = (S^{-1/2}AS^{-1/2}S^{1/2}\mathbf{w}_k, S^{1/2}\mathbf{w}_k)_2 = (A\mathbf{w}_k, \mathbf{w}_k)_2,$$
$$\|C\mathbf{v}_k\|_2^2 = (S^{-1/2}A\mathbf{w}_k, S^{-1/2}A\mathbf{w}_k)_2 = \|A\mathbf{w}_k\|_{S^{-1}}^2,$$
$$\|\mathbf{v}_{k+1}\|_2^2 = (S\mathbf{w}_{k+1}, \mathbf{w}_{k+1})_2 = \|A\mathbf{e}_{k+1}\|_{S^{-1}}^2.$$

The desired error estimate follows by using these. □