Numerical Analysis – Lecture 2[1]

Theorem 1.7 (The Householder–John theorem) *If the real symmetric matrices A and $A-B-B^\top$ are both positive definite and B is real then the spectral radius of $H=-(A-B)^{-1}B$ is strictly less than one.*

Proof Let λ be an eigenvalue of H, so $H\boldsymbol{v}=\lambda\boldsymbol{v}$ holds where $\boldsymbol{v}\neq\boldsymbol{0}$ is an eigenvector. (Note that both λ and $\boldsymbol{v}$ may have nonzero imaginary parts when H is not symmetric, e.g. in the Gauss–Seidel method.) The definition of H provides $-B\boldsymbol{v}=\lambda(A-B)\boldsymbol{v}$, and the value of λ is different from one because A is nonsingular. Thus we deduce

$$\bar{\boldsymbol{v}}^\top B\boldsymbol{v}=\frac{\lambda}{\lambda-1}\bar{\boldsymbol{v}}^\top A\boldsymbol{v}. \tag{1.2}$$

Moreover, writing $\boldsymbol{v}=\boldsymbol{v}_{\mathrm{R}}+\mathrm{i}\boldsymbol{v}_{\mathrm{I}}$, where $\boldsymbol{v}_{\mathrm{R}}$ and $\boldsymbol{v}_{\mathrm{I}}$ are real, we find the identity $\bar{\boldsymbol{v}}^\top A\boldsymbol{v}=\boldsymbol{v}_{\mathrm{R}}^\top A\boldsymbol{v}_{\mathrm{R}}+\boldsymbol{v}_{\mathrm{I}}^\top A\boldsymbol{v}_{\mathrm{I}}$, so positive definiteness implies $\bar{\boldsymbol{v}}^\top A\boldsymbol{v}>0$ and $\bar{\boldsymbol{v}}^\top(A-B-B^\top)\boldsymbol{v}>0$. It follows from equation (1.2), $\bar{\boldsymbol{v}}^\top B\boldsymbol{v}=\lambda/(\lambda-1)\bar{\boldsymbol{v}}^\top A\boldsymbol{v}$, $\bar{\boldsymbol{v}}^\top B^\top\boldsymbol{v}=\bar{\lambda}/(\bar{\lambda}-1)\bar{\boldsymbol{v}}^\top A\boldsymbol{v}$ and from the fact that B is real that the last inequality is the condition

$$0<\bar{\boldsymbol{v}}^\top A\boldsymbol{v}-\bar{\boldsymbol{v}}^\top B\boldsymbol{v}-\bar{\boldsymbol{v}}^\top B^\top\boldsymbol{v}=\left(1-\frac{\lambda}{\lambda-1}-\frac{\bar{\lambda}}{\bar{\lambda}-1}\right)\bar{\boldsymbol{v}}^\top A\boldsymbol{v}=\frac{(1-|\lambda|^2)\bar{\boldsymbol{v}}^\top A\boldsymbol{v}}{|\lambda-1|^2}.$$

Now $\lambda\neq1$ implies $|\lambda-1|^2>0$. Hence, recalling that $\bar{\boldsymbol{v}}^\top A\boldsymbol{v}>0$, we see that $1-|\lambda|^2$ is positive. Therefore $|\lambda|<1$ occurs for every eigenvalue of H as required. □

Corollary 1.8 (Application to Example 1.2) *Both Jacobi and Gauss–Seidel methods converge when A is the matrix of Example 1.2.*

Proof Positive definiteness of the symmetric matrix A has been already established in Lemma 1.3. For Jacobi's method, $A-B-B^\top$ is the same as A except that the signs of the off-diagonal elements are reversed. Therefore the proof of Lemma 1.3 shows too that $A-B-B^\top$ is positive definite: recall that the proof depended on the *modulus* of off-diagonal elements, not on their sign! Moreover, for the Gauss–Seidel method, $A-B-B^\top$ is just the diagonal part of A, all the off-diagonal elements being zero, so this matrix is also positive definite. Therefore Theorem 1.7 implies $\rho(H)<1$ in both cases. It follows from Revision 1.5 that the corollary is true. □

Technique 1.9 (Relaxation). It is often possible to improve the efficiency of Method 1.4 (simple iteration) by *relaxation.* Specifically, instead of letting $(A-B)\boldsymbol{x}^{(k+1)}=-B\boldsymbol{x}^{(k)}+\boldsymbol{b}$, $k=0,1,\ldots$, we let

$$(A-B)\tilde{\boldsymbol{x}}^{(k+1)}=-B\boldsymbol{x}^{(k)}+\boldsymbol{b}\qquad\text{and}\qquad\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\omega(\tilde{\boldsymbol{x}}^{(k+1)}-\boldsymbol{x}^{(k)}),\qquad k=0,1,\ldots,$$

where ω is a real constant called the *relaxation parameter.* Note that $\omega=1$ corresponds to the former, "unrelaxed" iteration.

Good choice of ω leads to small spectral radius of the iteration matrix: clearly, it should be less than one, but ideally it should be the least possible: the smaller the spectral radius, the faster the iteration converges. To choose ω, we need to determine the iteration matrix $\tilde{H}$. First, we relate $\boldsymbol{x}^{(k+1)}$ to $\boldsymbol{x}^{(k)}$ by eliminating $\tilde{\boldsymbol{x}}^{(k+1)}$ from the last displayed equation. Multiplying the equation for $\boldsymbol{x}^{(k+1)}$ by $A-B$ we obtain

$$\begin{aligned}(A-B)\boldsymbol{x}^{(k+1)}&=(A-B)[(1-\omega)\boldsymbol{x}^{(k)}+\omega\tilde{\boldsymbol{x}}^{(k+1)}]\\&=(1-\omega)(A-B)\boldsymbol{x}^{(k)}+\omega(-B\boldsymbol{x}^{(k)}+\boldsymbol{b})\\&=[(1-\omega)A-B]\boldsymbol{x}^{(k)}+\omega\boldsymbol{b}.\end{aligned}$$

[1] Please email all corrections and suggestions to these notes to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartII/Handouts.html`.

Thus, the iteration matrix is

$$\tilde{H} = (A-B)^{-1}[(1-\omega)A - B] = I - \omega(A-B)^{-1}A.$$

Recall the 'unrelaxed' iteration matrix $H = -(A-B)^{-1}B = (A-B)^{-1}(A-B-A) = I-(A-B)^{-1}A$. Substituting $(A-B)^{-1}A = I - H$, we deduce that

$$\tilde{H} = I - \omega(I-H) = (1-\omega)I + \omega H. \tag{1.3}$$

Suppose that $\omega \neq 0$. Then (1.3) proves that

$$\lambda \in \sigma(\tilde{H}) \qquad \Leftrightarrow \qquad \lambda = 1 - \omega + \omega\mu, \quad \mu \in \sigma(H),$$

where $\sigma(C)$ is the set of the eigenvalues (the *spectrum*) of the square matrix C. Therefore one may try to choose $\omega \in \mathbb{R} \setminus \{0\}$ to minimize

$$\rho(\tilde{H}) = \max\{|1-\omega+\omega\mu| \,:\, \mu \in \sigma(H)\}.$$

In general, $\sigma(H)$ is unknown, but often we have some information about it which can be utilized to find a 'good' (rather than 'best') value of ω. For example, suppose that it is known that $\sigma(H)$ is real and resides in the interval $[\alpha, \beta]$, where $-1 < \alpha < \beta < 1$. In that case we seek ω to minimize

$$\max\{|1-\omega+\omega\mu| \,:\, \mu \in [\alpha,\beta]\}.$$

Since maxima of the function above occur at endpoints, for optimal ω we have $|1-\omega+\omega\alpha| = |1-\omega+\omega\beta|$, and this is satisfied by $\omega_{\mathrm{opt}} = 2/(2-\alpha-\beta)$. (You can easily prove that $\omega_{\mathrm{opt}} \in (0,2)$.)

Approach 1.10 (An optimization calculation). We continue to assume that A is symmetric and positive definite. Therefore the quadratic function

$$F(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^\top A\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \tag{1.4}$$

is bounded below, and its least value occurs when $\boldsymbol{x}$ satisfies $\nabla F(\boldsymbol{x}) = \boldsymbol{0}$, which is equivalent to $\boldsymbol{x}$ being a solution of the system $A\boldsymbol{x} = \boldsymbol{b}$ of Problem 1.1. Therefore, when an iterative method generates the sequence $\boldsymbol{x}^{(k+1)}$, $k = 0,1,2,\ldots$, it may be helpful to force the condition $F(\boldsymbol{x}^{(k+1)}) < F(\boldsymbol{x}^{(k)})$ for every $k \in \mathbb{Z}_+$. This remark can provide an alternative useful way of choosing ω in Technique 1.9, especially if ω is allowed to depend on k.

We now turn to algorithms of the following form. We pick any starting vector $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$. For $k = 0,1,2,\ldots$, the calculation stops if $\|\nabla F(\boldsymbol{x}^{(k)})\| = \|A\boldsymbol{x}^{(k)} - \boldsymbol{b}\|$ is acceptably small. Otherwise, a *search direction* $\boldsymbol{d}^{(k)}$ is generated that satisfies the *descent condition* $[\mathrm{d}F(\boldsymbol{x}^{(k)} + \omega\boldsymbol{d}^{(k)})/\mathrm{d}\omega]_{\omega=0} < 0$. Then the value of ω that minimizes $F(\boldsymbol{x}^{(k)} + \omega\boldsymbol{d}^{(k)})$, $\omega > 0$, is calculated, and we call it $\omega^{(k)}$. Finally, the kth iteration sets $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \omega^{(k)}\boldsymbol{d}^{(k)}$. Thus the strict inequalities $F(\boldsymbol{x}^{(k+1)}) < F(\boldsymbol{x}^{(k)})$ and $\omega^{(k)} > 0$ are achieved.

There is a convenient form of the descent condition that has been mentioned. Specifically, because the definition (1.4) implies the identity

$$F(\boldsymbol{x}^{(k)} + \omega\boldsymbol{d}^{(k)}) = F(\boldsymbol{x}^{(k)}) + \omega{\boldsymbol{d}^{(k)}}^\top \boldsymbol{g}^{(k)} + \tfrac{1}{2}\omega^2 {\boldsymbol{d}^{(k)}}^\top A\boldsymbol{d}^{(k)}, \qquad \omega \in \mathbb{R}, \tag{1.5}$$

where $\boldsymbol{g}^{(k)} = \nabla F(\boldsymbol{x}^{(k)})$, the search direction has to satisfy ${\boldsymbol{d}^{(k)}}^\top \boldsymbol{g}^{(k)} < 0$, which is possible, because termination occurs when $\boldsymbol{g}^{(k)}$ is zero. Further, $\omega^{(k)}$ is the ω that minimizes the quadratic equation (1.5), so it has the value

$$\omega^{(k)} = -\frac{{\boldsymbol{d}^{(k)}}^\top \boldsymbol{g}^{(k)}}{{\boldsymbol{d}^{(k)}}^\top A\boldsymbol{d}^{(k)}}.$$