Numerical Solution of Partial Differential Equations

Endre Süli

Mathematical Institute
University of Oxford
2022

Lecture 7

Iterative solution of linear systems

We shall develop a simple iterative method for the approximate solution of systems of linear algebraic equations of the form

$$AU = F,$$

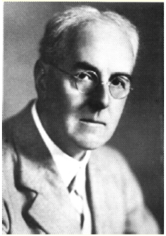
where $A \in \mathbb{R}^{M \times M}$ is a symmetric matrix with positive eigenvalues, which are contained in a nonempty closed interval $[\alpha, \beta]$, with $0 < \alpha < \beta$, $U \in \mathbb{R}^{M}$ is the vector of unknowns and $F \in \mathbb{R}^{M}$ is a given vector.

To this end, we consider the following iteration for the approximate solution of the linear system $AU = F$:

$$U^{(j+1)} := U^{(j)} - \tau(AU^{(j)} - F), \qquad j = 0, 1, \ldots, \tag{1}$$

where $U^{(0)} \in \mathbb{R}^M$ is a given initial guess, and $\tau > 0$ is a parameter to be chosen so as to ensure that the sequence of iterates $\{U^{(j)}\}_{j=0}^{\infty} \subset \mathbb{R}^M$ converges to $U \in \mathbb{R}^M$ as $j \to \infty$.

We shall explore the speed of convergence of this 'linear stationary iterative method', called the Richardson iteration[1].



[1]Lewis Fry Richardson, FRS (11 October 1881 – 30 September 1953).

We begin by observing that $U = U - \tau(AU - F)$. Therefore, upon subtraction of (1) from this equality we find that, for $j = 0, 1, \ldots,$

$$U - U^{(j+1)} = U - U^{(j)} - \tau A(U - U^{(j)}) = (I - \tau A)(U - U^{(j)}), \qquad (2)$$

where $I \in \mathbb{R}^{M\times M}$ is the identity matrix. Consequently,

$$U - U^{(j)} = (I - \tau A)^j (U - U^{(0)}), \qquad j = 1, 2, \ldots.$$

Recall that if $\|\cdot\|$ is a(ny) norm on $\mathbb{R}^M$, then the *induced matrix norm* is defined, for a matrix $B \in \mathbb{R}^{M\times M}$, by

$$\|B\| := \sup_{V\in\mathbb{R}^M\setminus\{0\}} \frac{\|BV\|}{\|V\|}.$$

Thanks to this definition, $\|BV\| \leq \|B\|\|V\|$ for all $V \in \mathbb{R}^M$, and hence, by induction $\|B^j V\| \leq \|B\|^j\|V\|$ for all $j = 1, 2 \ldots$, and all $V \in \mathbb{R}^M$.

Therefore, with $B := I - \tau A$ and $V := U - U^{(0)}$, we have that

$$\|U - U^{(j)}\| = \|(I - \tau A)^j(U - U^{(0)})\| \leq \|I - \tau A\|^j\|U - U^{(0)}\|. \quad (3)$$

To bound $\|I - \tau A\|$, we need a few tools from linear algebra.

(1) First, note that $\mathbb{R}^M$ is a finite-dimensional linear space, and in a finite-dimensional linear spaces all norms are equivalent.[2] Thus, if the sequence $\{U^{(j)}\}_{j=0}^{\infty}$ converges to U in one particular norm on $\mathbb{R}^M$, it will also converge to U in any other norm on $\mathbb{R}^M$. For simplicity, we shall therefore assume that the norm $\|\cdot\|$ on $\mathbb{R}^M$ is the Euclidean norm:

$$\|V\| := \left(\sum_{i=1}^{M} V_i^2\right)^{1/2}, \qquad V = (V_1, \ldots, V_M)^{\mathrm{T}} \in \mathbb{R}^M.$$

[2] Suppose that $\mathcal{V}$ is a linear space and $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on $\mathcal{V}$; then $\|\cdot\|_1$ and $\|\cdot\|_2$ are said to be *equivalent* if there exist positive constants C_1 and C_2 such that $C_1\|V\|_1 \leq \|V\|_2 \leq C_2\|V\|_1$ for all $V \in \mathcal{V}$.

(2) A symmetric matrix $B \in \mathbb{R}^{M \times M}$ has real eigenvalues, and the associated set of orthonormal eigenvectors spans the whole of $\mathbb{R}^M$.

Denoting by $\{e_i\}_{i=1}^M$ the (orthonormal) eigenvectors of B and by λ_i, $i = 1, \dots, M$, the corresponding eigenvalues, for any vector

$$V = \alpha_1 e_1 + \cdots + \alpha_M e_M,$$

expanded in terms of the eigenvectors of B, thanks to orthonormality, the Euclidean norms of V and BV can be expressed, respectively, as follows:

$$\|V\| = \left(\sum_{i=1}^{M} \alpha_i^2 \right)^{1/2} \quad \text{and} \quad \|BV\| = \left(\sum_{i=1}^{M} \alpha_i^2 \lambda_i^2 \right)^{1/2}.$$

Clearly, $\|BV\| \leq \max_{i=1,\dots,M} |\lambda_i| \, \|V\|$ for all $V \in \mathbb{R}^M$, and the inequality becomes an equality if V is the eigenvector of B associated with the largest in absolute value eigenvalue of B. Therefore,

$$\|B\| = \max_{i=1,\dots,M} |\lambda_i|,$$

where now $\| \cdot \|$ is the matrix norm induced by the Euclidean norm.

We now return to (3) to find that $\|I - \tau A\|$ on the r.h.s. of (3), where $\|\cdot\|$ denotes the matrix norm induced by the Euclidean norm, is equal to the largest in absolute value eigenvalue of the symmetric matrix $I - \tau A$.

As the eigenvalues of A are assumed to belong to the interval $[\alpha, \beta]$, where $0 < \alpha < \beta$, and the parameter τ is by assumption positive, the eigenvalues of $I - \tau A$ are contained in the interval $[1 - \tau\beta, 1 - \tau\alpha]$, whereby

$$\|I - \tau A\| \leq \max\{|1 - \tau\beta|, |1 - \tau\alpha|\}.$$

As $\tau > 0$ is a free parameter, we need to choose it so that the iterative method (1) converges as fast as possible. We see from (3) that it is therefore desirable to choose τ so that $\|I - \tau A\|$ is as small as possible, and less than 1.

We shall therefore seek $\tau > 0$ s.t.

$$\min_{\tau > 0} \max\{|1 - \tau\beta|, |1 - \tau\alpha|\} < 1. \qquad \text{Thus: } \tau = \frac{2}{\alpha + \beta}.$$

In summary then, the iterative method proposed for the approximate solution of the linear system $AU = F$ is the one stated in (1), with $\tau := \frac{2}{\alpha+\beta}$, and $[\alpha, \beta]$ being a closed subinterval of $(0, \infty)$ that contains all eigenvalues of the symmetric matrix $A \in \mathbb{R}^{M\times M}$.

Example 1

Consider the boundary-value problem

$$\begin{aligned} -u''(x) + c\,u(x) = f(x), \qquad & x \in (0,1), \\ u(0) = 0, \quad & u(1) = 0, \end{aligned}$$

where $c \geq 0$ and $f \in C([0,1])$. The finite difference approximation of this boundary-value problem on the mesh $\{x_i : i = 0, \ldots, N\}$ of uniform spacing $h = 1/N$, with $N \geq 2$, and $x_i = ih$, $i = 0, \ldots, N$, is given by

$$\begin{aligned} -\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + c\,U_i = f(x_i), \quad & i = 1, \ldots, N-1, \\ U_0 = 0, \quad & U_N = 0. \end{aligned} \tag{4}$$

In terms of matrix notation this can be rewritten as the linear system:

$$AU = F \tag{5}$$

where A is an $(N-1) \times (N-1)$ symmetric tridiagonal matrix, $U = (U_1, \ldots, U_{N-1})^{\mathrm{T}}$, and $F = (f(x_1), \ldots, f(x_{N-1}))^{\mathrm{T}}$.

We need to explore the associated eigenvalue problem $AU = \Lambda U$:

$$\begin{bmatrix} \frac{2}{h^2}+c & -\frac{1}{h^2} & & & \mathbf{0} \\ -\frac{1}{h^2} & \frac{2}{h^2}+c & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{2}{h^2}+c & -\frac{1}{h^2} \\ \mathbf{0} & & & -\frac{1}{h^2} & \frac{2}{h^2}+c \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{N-2} \\ U_{N-1} \end{bmatrix} = \Lambda \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{N-2} \\ U_{N-1} \end{bmatrix}.$$

We will show that the eigenvalues of A are

$$\Lambda_k = c + \frac{4}{h^2}\sin^2\frac{k\pi h}{2}, \qquad k = 1, 2, \ldots, N-1$$

and the corresponding eigenvectors are, respectively,

$$(U^k(x_1), \ldots, U^k(x_{N-1}))^{\mathrm{T}}, \qquad k = 1, \ldots, N-1,$$

where

$$U^k(x) := \sin(k\pi x), \quad x \in \{x_0, x_1, \ldots, x_N\}, \qquad k = 1, 2, \ldots, N-1.$$

The algebraic eigenvalue problem $AU = \Lambda U$ is simply a restatement, on the mesh $\{x_i \,:\, i = 0, \ldots, N\}$ of uniform spacing $h = 1/N$, with $N \geq 2$, and $x_i = ih$, $i = 0, \ldots, N$, of the finite difference eigenvalue problem:

$$-\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + c\, U_i = \Lambda U_i, \quad i = 1, \ldots, N-1,$$

$$U_0 = 0, \quad U_N = 0.$$

A simple calculation yields the nontrivial solution: $U_i := U^k(x_i)$, where

$$U^k(x) := \sin(k\pi x), \quad x \in \{x_0, x_1, \ldots, x_N\} \quad \text{and} \quad \Lambda_k := c + \frac{4}{h^2}\sin^2\frac{k\pi h}{2}$$

for $k = 1, 2, \ldots, N-1$.

This can be verified by inserting

$$U_i = U^k(x_i) = \sin(k\pi x_i) \quad \text{and} \quad U_{i\pm1} = U^k(x_{i\pm1}) = \sin(k\pi x_{i\pm1})$$

into the finite difference scheme and noting that

$$\sin(k\pi x_{i\pm1}) = \sin(k\pi(x_i \pm h)) = \sin(k\pi x_i)\cos(k\pi h) \pm \cos(k\pi x_i)\sin(k\pi h)$$

and

$$1 - \cos(k\pi h) = 2\sin^2 \frac{k\pi h}{2}$$

for $k = 1, 2, \ldots, N-1$ and $i = 1, 2, \ldots, N-1$.

Clearly,

$$c + 8 \leq \Lambda_k \leq c + \frac{4}{h^2} \qquad \text{for all } k = 1, 2, \ldots, N-1.$$

The first of these inequalities follows by noting that

$$\Lambda_k \geq \Lambda_1 = c + \frac{4}{h^2} \sin^2 \frac{\pi h}{2} \qquad \text{for } k = 1, \ldots, N-1$$

and $\sin x \geq \frac{2\sqrt{2}}{\pi} x$ for $x \in [0, \frac{\pi}{4}]$ (recall that $h \in [0, \frac{1}{2}]$ because $N \geq 2$, whereby $0 < \frac{\pi h}{2} \leq \frac{\pi}{4}$).

The second inequality is the consequence of $0 \leq \sin^2 x \leq 1$ for all $x \in \mathbb{R}$.

Example 2

Now consider the elliptic boundary-value problem

$$\begin{aligned} -\Delta u + cu &= f(x,y) && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma := \partial\Omega, \end{aligned}$$

where $c \geq 0$ is a real number and $f \in C(\overline{\Omega})$, whose finite difference approximation posed on a uniform mesh $\{(x_i, y_j) \,:\, i,j = 0,\dots,N\}$ of spacing $h = 1/N$, $N \geq 2$, in the x and y directions, is

$$\begin{aligned} -\frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{h^2} - \frac{U_{i,j+1} - 2U_{i,j} + U_{i,j-1}}{h^2} + c\,U_{i,j} &= f(x_i, y_j), && i,j = 1,\dots,N-1, \\ U_{i,j} &= 0 && \text{for } (x_i, y_j) \in \Gamma_h, \end{aligned} \tag{6}$$

where, Γ_h is the set of all mesh-points on Γ. This, too, can be rewritten as a system of linear algebraic equations of the form $AU = F$, where now A is a symmetric $(N-1)^2 \times (N-1)^2$ matrix with positive eigenvalues, $\Lambda_{k,m}$, $k, m = 1,\dots,N-1$.

The eigenvalue problem $AU = \Lambda U$ is simply a restatement of the finite difference eigenvalue problem:

$$-\frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{h^2} - \frac{U_{i,j+1} - 2U_{i,j} + U_{i,j-1}}{h^2} + c\, U_{i,j} = \Lambda U_{i,j}, \qquad i,j = 1,\dots,N-1,$$

$$U_{i,j} = 0 \qquad \text{for } (x_i, y_j) \in \Gamma_h,$$

where, Γ_h is the set of all mesh-points on $\Gamma = \partial\Omega$. Here, A is a symmetric $(N-1)^2 \times (N-1)^2$ matrix with positive eigenvalues

$$\Lambda_{k,m} = c + \frac{4}{h^2}\left(\sin^2\frac{k\pi h}{2} + \sin^2\frac{m\pi h}{2}\right),$$

with $c + 16 \le \Lambda_{k,m} \le c + \frac{8}{h^2}$, and eigenvectors/(discrete) eigenfunctions $U_{i,j} = U^{k,m}(x_i, y_j)$, for $i,j = 1,\dots,N-1$ and $k,m = 1,\dots,N-1$, where

$$U^{k,m}(x,y) = \sin(k\pi x)\sin(m\pi y).$$

Note

In the case of the finite difference scheme (4), $\alpha = c + 8$ and $\beta = c + \frac{4}{h^2}$, while in the case of (6), $\alpha = c + 16$ and $\beta = c + \frac{8}{h^2}$. In both cases

$$\frac{\beta - \alpha}{\beta + \alpha} = 1 - \mathrm{Const.}\, h^2 \in (0, 1);$$

thus, while the sequence of iterates $\{U^{(j)}\}_{j=0}^{\infty}$ defined by the iterative method (1) is guaranteed to converge to the solution U of the linear system $AU = F$ for each fixed $h > 0$, the right-hand side in the inequality

$$\|U - U^{(j)}\| \leq \left(\frac{\beta - \alpha}{\beta + \alpha}\right)^j \|U - U^{(0)}\| \tag{7}$$

signals that deterioration of the speed of convergence may occur as $h \to 0$.

An alternative, computable bound on the iteration error

By multiplying (2) by the matrix A and recalling that $AU = F$, one has

$$F - AU^{(j+1)} = (I - \tau A)(F - AU^{(j)}),$$

and therefore, by proceeding as above,

$$\|F - AU^{(j)}\| \leq \|I - \tau A\|^j \|F - AU^{(0)}\| \leq \left(\frac{\beta - \alpha}{\beta + \alpha}\right)^j \|F - AU^{(0)}\|. \quad (8)$$

As α and β are available (in the case of the simple boundary-value problems considered here, at least) as are F, A and the initial guess $U^{(0)}$, it is possible to quantify the number of iterations required to ensure that the Euclidean norm of the so-called *residual* $F - AU^{(j)}$ of the j-th iterate becomes smaller than a chosen tolerance $\texttt{TOL} > 0$.

A sufficient condition for this is that the right-hand side of (8) is smaller than TOL, which will hold as soon as

$$j > \log \frac{\|F - AU^{(0)}\|}{\texttt{TOL}} \left[\log\left(\frac{\beta+\alpha}{\beta-\alpha}\right)\right]^{-1}. \qquad (9)$$

In the case of the two boundary-value problems considered above,

$$\frac{\beta-\alpha}{\beta+\alpha} = 1 - \mathrm{Const.}h^2$$

and therefore (because $\log(1 - \mathrm{Const.}h^2) \sim -\mathrm{Const.}h^2$ as $h \to 0$) the right-hand side of the inequality (9) is $\sim \mathrm{Const.}\, h^{-2} \log(1/\texttt{TOL})$.

We see in particular that the smaller the value of the mesh-size h the larger the number of iterations j will need to be to ensure that

$$\|F - AU^{(j)}\| < \texttt{TOL}.$$