Finite difference method for elliptic problems: I

Praveen. C
praveen@math.tifrbng.res.in

Tata Institute of Fundamental Research
Center for Applicable Mathematics
Bangalore 560065
http://math.tifrbng.res.in/~praveen

January 13, 2013

Contents

General approach of numerical methods:

$$\text{Stability} + \text{Consistency} = \text{Convergence}$$

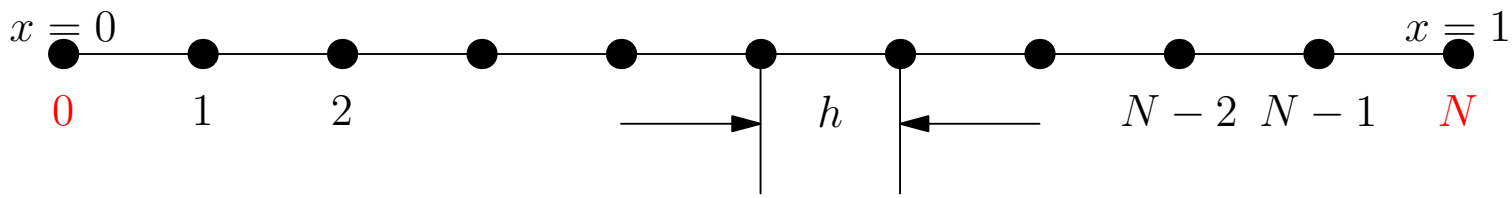
1-D boundary value problem

Differential equation:

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

Finite difference mesh: Let $N \geq 2$ be an integer and let

$$\text{mesh size:} \qquad h = \frac{1}{N}$$



$$\text{mesh points:} \qquad x_i = ih, \qquad i = 0, 1, \ldots, N$$

$$\Omega_h = \{x_i : i = 1, 2, \ldots, N-1\}, \qquad \Gamma_h = \{x_0, x_N\}, \qquad \bar{\Omega}_h = \Omega_h \cup \Gamma_h$$

$$U_i = \text{numerical approximation to } u(x_i)$$

$$\text{Need to find} \qquad U_1, U_2, \ldots, U_{N-1}$$

Finite difference approximation

Let $u : [0,1] \to \mathbb{R}$. By Taylor series

$$u(x_{i\pm1}) = u(x_i \pm h) = u(x_i) \pm hu'(x_i) + \frac{h^2}{2}u''(x_i) \pm \frac{h^3}{6}u'''(x_i) + \mathcal{O}\left(h^4\right)$$

Forward difference for u': $u \in \mathcal{C}^2[0,1]$

$$D_x^+ u(x_i) := \frac{u(x_{i+1}) - u(x_i)}{h} = u'(x_i) + \mathcal{O}\left(h\right)$$

Backward difference for u': $u \in \mathcal{C}^2[0,1]$

$$D_x^- u(x_i) := \frac{u(x_i) - u(x_{i-1})}{h} = u'(x_i) + \mathcal{O}\left(h\right)$$

Central difference for u'': $u \in \mathcal{C}^4[0,1]$

$$D_x^+ D_x^- u(x_i) = D_x^- D_x^+ u(x_i) = \frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1})}{h^2} = u''(x_i) + \mathcal{O}\left(h^2\right)$$

Note: If $u \in \mathcal{C}^3[0,1]$ then $D_x^+ D_x^- u(x_i) - u''(x_i) = \mathcal{O}\left(h\right)$.

Finite difference method: $u'' \to D_x^+ D_x^- u$

Find $(U_1, U_2, \ldots, U_{N-1})$ such that

$$\boxed{\begin{aligned}(AU)_i := -D_x^+ D_x^- U_i + c(x_i)U_i &= f(x_i) \qquad i = 1, 2, \ldots, N-1 \\ U_0 = 0, \qquad U_N &= 0\end{aligned}} \tag{2}$$

Then the hope is that $U_i \approx u(x_i)$ which we have to prove. In matrix notation

$$\begin{bmatrix} \frac{2}{h^2} + c(x_1) & -\frac{1}{h^2} & & & \bigcirc \\ -\frac{1}{h^2} & \frac{2}{h^2} + c(x_2) & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{2}{h^2} + c(x_{N-2}) & -\frac{1}{h^2} \\ \bigcirc & & & -\frac{1}{h^2} & \frac{2}{h^2} + c(x_{N-1}) \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{N-2} \\ U_{N-1} \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{N-2}) \\ f(x_{N-1}) \end{bmatrix}$$

or more compactly

$$AU = F, \qquad A \in \mathbb{R}^{(N-1)\times(N-1)}, \quad U \in \mathbb{R}^{N-1}, \quad F \in \mathbb{R}^{N-1}$$

Does solution exist, i.e., is the matrix A invertible ?

Discrete inner product

Let V, W be two *grid functions* defined at the mesh points, vanishing at $i = 0, N$. Define the discrete inner product

$$\boxed{(V, W)_h = \sum_{i=1}^{N-1} h V_i W_i}$$

which resembles the L^2 inner product

$$(v, w) = \int_0^1 v(x) w(x) \mathrm{d}x$$

Lemma (Summation by parts)

Suppose V is a grid function defined at the mesh-points x_i, $i = 0, 1, \ldots, N$ and let $V_0 = V_N = 0$. Then

$$\left(-D_x^+ D_x^- V, V\right)_h = \sum_{i=1}^{N} h |D_x^- V_i|^2 \tag{3}$$

Discrete inner product

Proof: We do summation by parts,

$$
\begin{aligned}
\left(-D_x^+ D_x^- V, V\right)_h &= -\sum_{i=1}^{N-1} h(D_x^+ D_x^- V_i) V_i \\
&= -\sum_{i=1}^{N-1} \frac{V_{i+1}-V_i}{h} V_i + \sum_{i=1}^{N-1} \frac{V_i - V_{i-1}}{h} V_i \\
&= -\sum_{i=2}^{N} \frac{V_i - V_{i-1}}{h} V_{i-1} + \sum_{i=1}^{N-1} \frac{V_i - V_{i-1}}{h} V_i \quad \text{(shift indices)} \\
&= -\sum_{i=1}^{N} \frac{V_i - V_{i-1}}{h} V_{i-1} + \sum_{i=1}^{N} \frac{V_i - V_{i-1}}{h} V_i \quad (V_0 = V_N = 0) \\
&= \sum_{i=1}^{N} \frac{V_i - V_{i-1}}{h} (V_i - V_{i-1}) = \sum_{i=1}^{N} h |D_x^- V_i|^2
\end{aligned}
$$

Discrete analogue of $-\int_0^1 v'' v \mathrm{d}x = \int_0^1 (v')^2 \mathrm{d}x \qquad (v(0) = v(1) = 0)$

Existence of discrete solution

Let V be a grid function such that $V_0 = V_N = 0$ and let $c \geq 0$. Then

$$\begin{aligned}(AV, V)_h &= \left(-D_x^+ D_x^- V + cV, V\right)_h \\ &= \left(-D_x^+ D_x^- V, V\right)_h + (cV, V)_h \geq \sum_{i=1}^{N} h|D_x^- V_i|^2 \geq 0\end{aligned}$$

$$\text{If} \quad AV = 0 \quad \text{for some} \quad V \quad \text{then necessarily}$$

$$\sum_{i=1}^{N} h|D_x^- V_i|^2 = 0 \quad \Longrightarrow \quad D_x^- V_i = 0, \quad i = 1, \ldots, N$$

$$\Longrightarrow V_0 = V_1 = \ldots = V_N$$

But since $V_0 = V_N = 0$, we obtain that $V = 0$. Hence $AV = 0$ if and only if $V = 0$ from which we deduce that A is a non-singular matrix.

Theorem (Existence of FD solution)

Suppose c and f are continuous functions on $[0, 1]$ and $c(x) \geq 0$, $x \in [0, 1]$. Then the finite difference scheme (2) has a unique solution $U = A^{-1}F$.

Discrete norms

Discrete L^2 norm

$$\|U\|_h := \sqrt{(U,U)_h} = \left(\sum_{i=1}^{N-1} h|U_i|^2 \right)^{\frac{1}{2}}$$

Discrete Sobolev norm

$$\|U\|_{1,h} := \left(\|U\|_h^2 + \|D_x^- U]\|_h^2 \right)^{\frac{1}{2}}$$

where

$$\|V]\|_h^2 := \sum_{i=1}^{N} h|V_i|^2 \qquad \text{(includes last grid point } i = N)$$

Using this notation

$$(AV,V)_h \geq \|D_x^+ V]\|_h^2 \qquad \text{(equality if } c \equiv 0)$$

Using a discrete version of Poincare-Friedrichs inequality, we will show that

$$(AV,V)_h \geq c_0 \|V\|_{1,h}^2$$

where c_0 is a positive constant. This is a *discrete coercivity property.*

Lemma (Discrete Poincare-Friedrichs inequality)

Let V be a mesh function with $V_0 = V_N = 0$. Then $\exists\ c_* > 0$, independent of V and h, such that

$$\|V\|_h^2 \le c_* \|D_x^- V]\|_h^2 \tag{4}$$

for all such V.

Proof: Using Cauchy-Schwarz inequality, we have

$$|V_i|^2 = \left| \sum_{j=1}^{i} (D_x^- V_j) h \right|^2 \le \left(\sum_{j=1}^{i} h \right) \sum_{j=1}^{i} h |D_x^- V_j|^2 = ih \sum_{j=1}^{i} h |D_x^- V_j|^2$$

$$\begin{aligned} \|V\|_h^2 &= \sum_{i=1}^{N-1} h|V_i|^2 \le \sum_{i=1}^{N-1} ih^2 \sum_{j=1}^{i} h|D_x^- V_j|^2 \le \left(\sum_{i=1}^{N-1} i \right) h^2 \sum_{j=1}^{N} h|D_x^- V_j|^2 \\ &\le \frac{(N-1)N}{2} h^2 \sum_{j=1}^{N} h|D_x^- V_j|^2 \le \frac{1}{2} \|D_x^- V]\|_h^2, \quad \text{since} \quad (N-1)N < \frac{1}{h^2} \end{aligned}$$

which proves (4) with $c_* = \frac{1}{2}$.

Discrete coercivity property:

$$(AV, V)_h \geq \|D_x^- V]\|_h^2 \geq \frac{1}{c_*} \|V\|_h^2$$

Combining

$$c_* (AV, V)_h \geq \|V\|_h^2 \qquad \text{and} \qquad (AV, V)_h \geq \|D_x^- V]\|_h^2$$

we get

$$(AV, V)_h \geq (1 + c_*)^{-1} (\|V\|_h^2 + \|D_x^- V]\|_h^2)$$

With $c_0 = (1 + c_*)^{-1} = \frac{2}{3}$, we have the coercivity property

$$(AV, V)_h \geq c_0 \|V\|_{1,h}^2 \tag{5}$$

Theorem (Stability of FD solution)

The scheme (2) is stable in the sense that

$$\|U\|_{1,h} \leq \frac{1}{c_0} \|f\|_h \tag{6}$$

Proof: Use coercivity (5) and Cauchy-Schwarz

$$c_0 \|U\|_{1,h}^2 \leq (AU, U)_h = (f, U)_h \leq \|f\|_h \|U\|_h \leq \|f\|_h \|U\|_{1,h} \qquad \square$$

Global error and truncation error

The global error between the true solution u and the numerical solution U is

$$e_i := u(x_i) - U_i, \qquad i = 0, 1, \ldots, N$$

Due to boundary conditions

$$e_0 = e_N = 0$$

Then

$$\begin{aligned} Ae_i &= Au(x_i) - AU_i = Au(x_i) - f(x_i) \\ &= -D_x^+ D_x^- u(x_i) + c(x_i)u(x_i) - [-u''(x_i) + c(x_i)u(x_i)] \\ &= u''(x_i) - D_x^+ D_x^- u(x_i), \qquad i = 1, 2, \ldots, N-1 \end{aligned}$$

Local truncation error: error in central difference approximation

$$\tau_i := u''(x_i) - D_x^+ D_x^- u(x_i)$$

Thus the error satisfies the equation

$$\boxed{\begin{aligned} (Ae)_i &= \tau_i, \qquad i = 1, 2, \ldots, N-1 \\ e_0 &= e_N = 0 \end{aligned}} \tag{7}$$

Theorem (Error estimate)

Let $f \in \mathcal{C}[0,1]$, $c \in \mathcal{C}[0,1]$ with $c(x) \geq 0$ and suppose that the solution of (1) belongs to $\mathcal{C}^4[0,1]$. Then

$$\|u - U\|_{1,h} \leq \frac{h^2}{8} \|u^{(4)}\|_\infty \tag{8}$$

Proof: Using Taylor series with remainder term, show that

$$\tau_i = u''(x_i) - D_x^+ D_x^- u(x_i) = -\frac{h^2}{12} u^{(4)}(\xi_i), \qquad \xi_i \in [x_{i-1}, x_{i+1}]$$

so that

$$|\tau_i| \leq \frac{h^2}{12} \sup_{x_{i-1} \leq x \leq x_{i+1}} |u^{(4)}(x)| \leq \frac{h^2}{12} \sup_{0 \leq x \leq 1} |u^{(4)}(x)| \tag{9}$$

Applying stability result (6) to (7) we obtain

$$\|e\|_{1,h} \leq \frac{1}{c_0} \|\tau\|_h = \frac{1}{c_0} \left(\sum_{i=1}^{N-1} h|\tau_i|^2 \right)^{\frac{1}{2}} \leq \frac{h^2}{12 c_0} \|u^{(4)}\|_\infty$$

which yields the error estimate (8) since $c_0 = \frac{2}{3}$.

General framework

Linear differential equation

$$\begin{aligned} Lu &= f \qquad \text{in} \qquad \Omega \\ lu &= g \qquad \text{on} \qquad \Gamma \end{aligned} \tag{10}$$

Finite difference approximation

$$\begin{aligned} L_h U &= f_h \qquad \text{in} \qquad \Omega_h \\ l_h U &= g_h \qquad \text{on} \qquad \Gamma_h \end{aligned} \tag{11}$$

Two key steps:

(1) Show that the scheme is stable:

$$|||U|||_{\Omega_h} \leq C_s \left(\|f_h\|_{\Omega_h} + \|g_h\|_{\Gamma_h} \right)$$

where $C_s > 0$ is independent of f, g, h.

General framework

(2) Show that the scheme is consistent: Local truncation error

$$\begin{aligned} \tau_{\Omega_h} &= L_h u - f_h \qquad \text{in} \quad \Omega_h \\ \tau_{\Gamma_h} &= l_h u - g_h \qquad \text{in} \quad \Gamma_h \end{aligned}$$

For Dirichlet problem, $\tau_{\Gamma_h} = 0$. Assuming sufficiently smooth solution u show that

$$\|\tau_{\Omega_h}\|_{\Omega_h} + \|\tau_{\Gamma_h}\|_{\Gamma_h} \leq C_\tau h^p \qquad \text{as} \quad h \to 0$$

where $C_\tau > 0$ independent of h but might depend on u and $p > 0$.

Lax equivalence theorem

Suppose the finite difference scheme (11) is stable and consistent. Then it is a convergent approximation of (10).

Proof: Define the global error $e = u - U$. Then

$$L_h e = L_h u - L_h U = L_h u - f_h = \tau_{\Omega_h}$$

General framework

and similarly

$$l_h e = \tau_{\Gamma_h}$$

Error is governed by the equation

$$\begin{aligned} L_h e &= \tau_{\Omega_h} \quad \text{in} \quad \Omega_h \\ l_h e &= \tau_{\Gamma_h} \quad \text{in} \quad \Gamma_h \end{aligned}$$

By stability and consistency of the scheme

$$|||u - U|||_{\Omega_h} = |||e|||_{\Omega_h} \leq C_s(\|\tau_{\Omega_h}\|_{\Omega_h} + \|\tau_{\Gamma_h}\|_{\Gamma_h}) \leq C_s C_\tau h^p$$

Convergence of U now follows since

$$|||u - U|||_{\Omega_h} \to 0 \quad \text{as} \quad h \to 0$$

The quantity p is called the order of accuracy of the scheme. It is desirable to have a large value of p since we can get more accurate solution with smaller number of grid points.

We will next show some results in the maximum norm. We begin with some definitions.

Definitions

- *Non-negative matrix*: A matrix A is said to be non-negative if all its entries are non-negative. We will indicate this property by writing $A \geq 0$.
- *Non-negative vector*: A vector V is said to be non-negative if all its entries are non-negative. We will indicate this property by writing $V \geq 0$.
- *Monotone matrix*: A real, square matrix A is said to be monotone if it is invertible and the matrix A^{-1} is non-negative.
- *M-matrix*: A real, square matrix $A = (a_{ij})$ is called an M-matrix if
  - $a_{ii} > 0$ and $a_{ij} \leq 0$ for $i \neq j$
  - A^{-1} is non-negative

  Thus an M-matrix is also a monotone matrix.

Theorem (Characterization of monotone matrices)

A real matrix A of order n is monotone if and only if the inclusion

$$\{v \in \mathbb{R}^n : Av \geq 0\} \subset \{v \in \mathbb{R}^n : v \geq 0\}$$

is satisfied.

Proof: (a) If A is monotone and the vector Av is non-negative then

$$v = A^{-1}(Av) \geq 0$$

(b) Conversely, suppose the inclusion is satisfied. Then

$$\left.\begin{array}{rcl} Av = 0 & \Longrightarrow & v \geq 0 \\ A(-v) = 0 & \Longrightarrow & -v \geq 0 \end{array}\right\} \quad \Longrightarrow \quad v = 0$$

Hence A is non-singular and A^{-1} exists. The j'th column vector of A^{-1} is

$$b_j = A^{-1}e_j, \qquad e_j = [0, \ldots, 0, 1, 0, \ldots, 0]^\top$$
$$\uparrow \quad j\text{'th position}$$

so that

$$Ab_j = e_j \geq 0 \quad \Longrightarrow \quad b_j \geq 0 \quad \Longrightarrow \quad A^{-1} \geq 0 \quad \square$$

Theorem

Suppose that c is non-negative. Then the matrix A in (2) is monotone.

Proof: Let $A \in \mathbb{R}^{(N-1)\times(N-1)}$ be the matrix of the finite difference scheme. Due to above characterization, it is enough to show that

$$Av \geq 0 \quad \Longrightarrow \quad v \geq 0$$

Given any vector $v \in \mathbb{R}^{N-1}$ such that $Av \geq 0$, let $p \in \{1, \ldots, N-1\}$ be an integer satisfying

$$v_p \leq v_i \quad \text{for} \quad i = 1, 2, \ldots, N-1 \qquad (\text{ i.e., } v_p = \min_{1 \leq i \leq N-1} v_i)$$

We have to show that $v_p \geq 0$. (a) If $p = 1$

$$0 \leq (2 + c_1 h^2)v_1 - v_2 = (1 + c_1 h^2)v_1 + (v_1 - v_2) \leq (1 + c_1 h^2)v_1$$

(b) If $\underline{2 \leq p \leq N-2}$ then

$$0 \leq -v_{p-1} + (2 + c_p h^2) v_p - v_{p+1} \leq c_p h^2 v_p$$

(c) If $\underline{p = N-1}$ then

$$0 \leq -v_{N-2} + (2 + c_{N-1} h^2) v_{N-1} \leq (1 + c_{N-1} h^2) v_{N-1}$$

Hence we have

$$\min_{1 \leq i \leq N-1} v_i \geq 0 \quad \text{if} \quad c_i > 0, \quad 2 \leq i \leq N-2$$

It remains to look at the case where atleast one of the c_i, $2 \leq i \leq N-2$ is zero. We already know that A is invertible (even if $c \equiv 0$). Now the matrix $A + \alpha I$ is monotone for every $\alpha > 0$. This implies that

$$(A + \alpha I)^{-1} \geq 0$$

The elements of $(A + \alpha I)^{-1}$ are continuous functions of $\alpha \geq 0$ and hence it follows that $A^{-1} \geq 0$.

Matrix norm

For any square matrix $M = (m_{ij}) \in \mathbb{R}^{n\times n}$ and vector norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^+$

$$\|M\| := \max_{V\neq 0} \frac{\|MV\|}{\|V\|} = \max_{\|V\|=1} \|MV\|$$

In particular

$$\|M\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |m_{ij}|$$

Theorem (Error in max norm)

Suppose that c is non-negative. If the solution u of the BVP (1) satisfies $u \in \mathcal{C}^4[0,1]$ then we have the bound

$$\max_{1\leq i\leq N-1} |u_i - U_i| = \|u - U\|_\infty \leq \frac{h^2}{96} \sup_{0\leq x\leq 1} |u^{(4)}(x)|$$

Proof: (1) We first show the stability result

$$\left\|A^{-1}\right\|_\infty \leq \frac{1}{8} \qquad \left(\left\|A^{-1}\right\|_\infty \leq \left\|A_0^{-1}\right\|_\infty \leq \frac{1}{8}\right) \tag{12}$$

Let A_0 be the matrix A with $c = 0$. Since A, A_0 are monotone

$$A^{-1} \geq 0 \qquad \text{and} \qquad A_0^{-1} \geq 0$$

Since c is non-negative

$$A - A_0 = \text{diag}(c_i) \geq 0$$

so that

$$A_0^{-1} - A^{-1} = A_0^{-1}(A - A_0)A^{-1} \geq 0$$

Then using the expression of the matrix norm $\|\cdot\|_\infty$ we obtain

$$\left\|A^{-1}\right\|_\infty \leq \left\|A_0^{-1}\right\|_\infty \leq ?$$

Observe that

$$A_0^{-1} \geq 0 \quad \Longrightarrow \quad \left\|A_0^{-1}\right\|_\infty = \left\|A_0^{-1}E\right\|_\infty, \quad E = [1, 1, \ldots, 1]^\top \in \mathbb{R}^{N-1}$$

But $A_0^{-1}E$ is the finite difference approximation to the solution of

$$\begin{aligned} -v''(x) &= 1 \qquad x \in \Omega = (0,1) \\ v(0) = 0, \qquad v(1) &= 0 \end{aligned}$$

The solution is

$$v(x) = \frac{1}{2}x(1-x) \quad \Longrightarrow \quad v^{(3)}(x) = v^{(4)}(x) = 0$$

Hence the finite difference solution $A_0^{-1}E$ is exact at the nodes, i.e.,

$$(A_0^{-1}E)_i = v(x_i), \qquad 1 \le i \le N-1$$

so that

$$\left\|A_0^{-1}E\right\|_\infty = \max_{1\le i\le N-1} |v(x_i)| \le \max_{0\le x\le 1} |v(x)| = \frac{1}{8}$$

(2) We have already seen the error equation (7) for $e_i = u(x_i) - U_i$

$$\begin{aligned} (Ae)_i &= \tau_i, \qquad i = 1, 2, \ldots, N-1 \\ e_0 &= e_N = 0 \end{aligned}$$

Hence using (12) and (9)

$$\|e\|_\infty = \left\|A^{-1}\tau\right\|_\infty \le \left\|A^{-1}\right\|_\infty \|\tau\|_\infty \le \frac{1}{8}\frac{h^2}{12} \max_{0\le x\le 1} |u^{(4)}(x)|$$

which proves the desired result. □

Maximum principle (Differential equation)

Suppose that $c(x) \ge 0$ and

$$\left.\begin{aligned} -u''(x) + c(x)u(x) \ge 0, \quad 0 \le x \le 1 \\ u(0) \ge 0, \quad u(1) \ge 0 \end{aligned}\right\} \quad \Longrightarrow \quad u(x) \ge 0$$

Maximum principle (FDM)

Suppose A is monotone. Then

$$\left.\begin{aligned} AU \ge 0 \\ U_0 \ge 0, \quad U_N \ge 0 \end{aligned}\right\} \quad \Longrightarrow \quad U \ge 0$$

Steady diffusion-convection-reaction

$$Au := -au'' + bu' + cu = f, \quad \text{in} \quad \Omega = (0,1) \tag{13}$$
$$u(0) = u_0, \quad u(1) = u_1$$

$a(x), b(x), c(x)$ are smooth functions and $a > 0$, $c \geq 0$ in $\bar{\Omega}$

Finite difference approximation of PDE

$$(AU)_j := -a_j \frac{U_{j-1} - 2U_j + U_{j+1}}{h^2} + b_j \frac{U_{j+1} - U_{j-1}}{2h} + c_j U_j = f_j \tag{14}$$
$$U_0 = u_0, \quad U_N = u_1$$

or, for $j = 1, 2, \ldots, N-1$

$$-(a_j + \frac{1}{2} h b_j) U_{j-1} + (2a_j + h^2 c_j) U_j - (a_j - \frac{1}{2} h b_j) U_{j+1} = h^2 f_j$$

Steady diffusion-convection-reaction

Discrete maximum principle

Assume that h is so small that $a_j \pm \frac{1}{2}hb_j \geq 0$ and that U satisfies $AU_j \leq 0$ $(AU_j \geq 0)$.

1. If $c = 0$, then

$$\max_{0\leq j\leq N} U_j = \max\{U_0, U_N\} \qquad \left(\min_{0\leq j\leq N} U_j = \min\{U_0, U_N\}\right)$$

2. If $c \geq 0$, then

$$\max_{0\leq j\leq N} U_j \leq \max\{U_0, U_N, 0\} \qquad \left(\min_{0\leq j\leq N} U_j \geq \min\{U_0, U_N, 0\}\right)$$

Steady diffusion-convection-reaction

1. Since $c = 0$ and $AU_j \leq 0$

$$\begin{aligned} U_j &= \frac{(a_j + \frac{1}{2}hb_j)}{2a_j} U_{j-1} + \frac{(a_j - \frac{1}{2}hb_j)}{2a_j} U_{j+1} + \frac{h^2}{2a_j} AU_j \\ &\leq \frac{(a_j + \frac{1}{2}hb_j)}{2a_j} U_{j-1} + \frac{(a_j - \frac{1}{2}hb_j)}{2a_j} U_{j+1} \\ &\leq \max(U_{j-1}, U_{j+1}) \qquad \text{for} \quad 1 \leq j \leq N-1 \end{aligned}$$

Assume that U has an interior maximum at U_j, i.e.,

$$U_j = \max_{0 \leq k \leq N} U_k$$

But this contradicts the above inequality unless $U_j = U_{j-1} = U_{j+1}$ which means that $U_j = \text{constant} = U_0 = U_N$. Hence the maximum of $\{U_j\}_0^N$ must occur on the boundary.

Steady diffusion-convection-reaction

2. Case $c \geq 0$ and $AU_j \leq 0$:
   1. If $U_j \leq 0$, then we are done.
   2. Otherwise assume that $\max_j U_j = U_k > 0$ for some $1 \leq k \leq N-1$. Let (l, r) be the largest subinterval containing k such that $U_j > 0,\ j \in (l, r)$.
   3. We now have $\tilde{A}U_j := AU_j - c_j U_j \leq 0$ in (x_l, x_r). Applying result of Part 1, we have $U_k = \max\{U_l, U_r\}$.
   4. But then x_l and x_r cannot both be interior points of Ω, for then either U_l or U_r would be positive, and the interval (x_l, x_r) would not be the largest subinterval with $U_j > 0$. This implies that $U_k = \max\{U_0, U_N\}$.

Remark: The conditions in the theorem ensure that A is an M-matrix.

Remark: Key concept used in proof was convexity. An M-matrix gives convexity property to the scheme.

Remark: For proof of maximum principle in continuous case, see e.g., Larsson and Thomee.

Mesh Peclet number

- Note that

$$a_j = \text{viscosity coefficient} \qquad b_j = \text{convection speed}$$

The condition $a_j \pm \frac{1}{2} h b_j \geq 0$ requires that

$$P_j = \frac{h|b_j|}{a_j} \leq 2$$

Here, P_j is called *mesh Peclet number*. If $b_j \equiv 0$, i.e., there is no convection, then the condition is trivially satisfied for all h.

- When convection is large, we have to choose a small mesh h, which increases computational cost and hence is not desirable.
- For non-linear problems, the speed b will depend on the solution, which is itself unknown.
- These problems arise because we chose a central difference approximation for the term bu' which has a hyperbolic character.

Numerical solution

- Consider the boundary value problem

$$-u''(x) = f(x), \quad x \in (a, b)$$

with boundary condition

$$u(a) = u_a, \qquad u(b) = u_b$$

- At $i = 1$

$$\frac{2}{h^2}U_1 - \frac{1}{h^2}U_2 = f_1 + \frac{1}{h^2}u_a$$

- For $i = 2, \ldots, N-2$

$$-\frac{1}{h^2}U_{i-1} + \frac{2}{h^2}U_i - \frac{1}{h^2}U_{i+1} = f_i$$

- At $i = N-1$

$$-\frac{1}{h^2}U_{N-2} + \frac{2}{h^2}U_{N-1} = f_{N-1} + \frac{1}{h^2}u_b$$

FDM for $-u'' = f$

For $N = 11$, putting all equations together

$$\frac{1}{h^2}\begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \\ U_5 \\ U_6 \\ U_7 \\ U_8 \\ U_9 \\ U_{10} \end{bmatrix} = \begin{bmatrix} f_1 + \frac{u_a}{h^2} \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \\ f_{10} + \frac{u_b}{h^2} \end{bmatrix}$$

or

$$AU = b$$

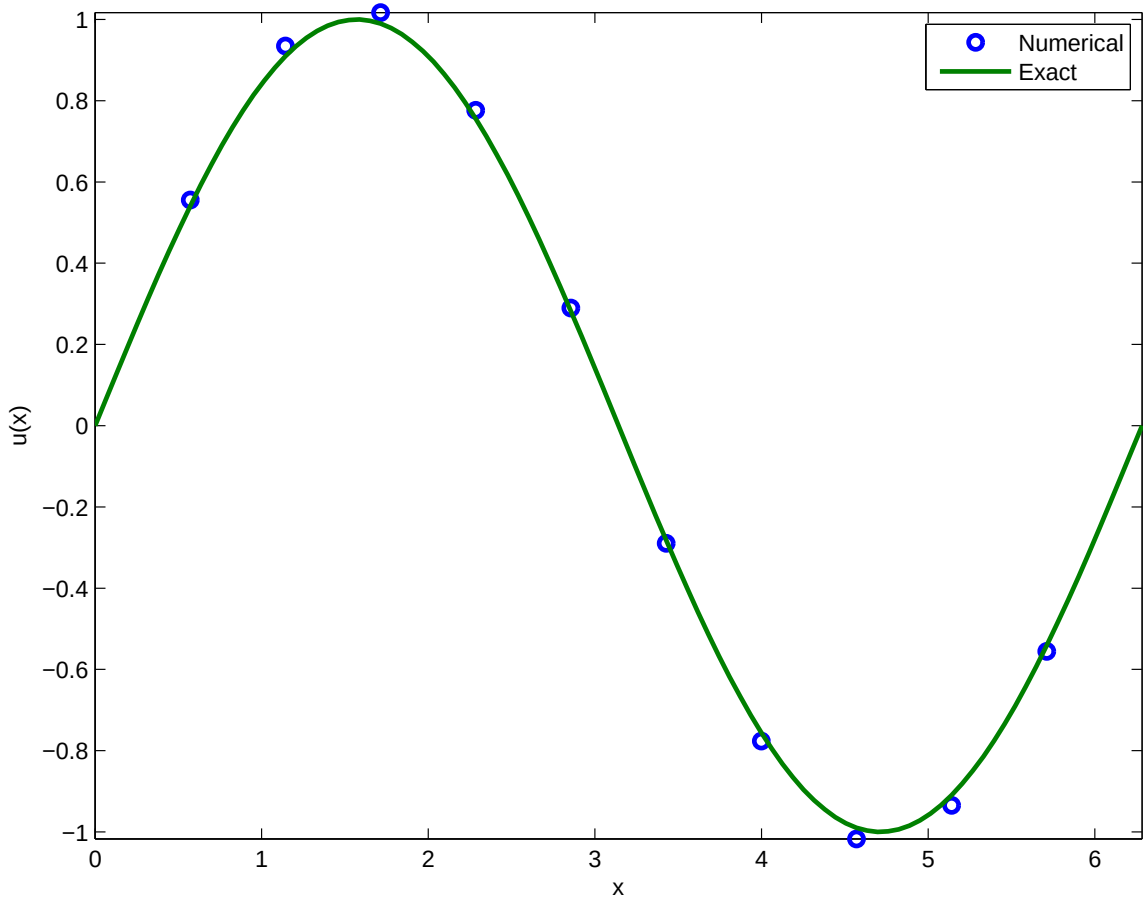
We have $N - 1$ equations for the $N - 1$ unknowns: $[U_1, U_2, \ldots, U_{N-1}]$

FDM for ODE

- We take

$$f(x) = \sin(x), \quad (a, b) = (0, 2\pi), \quad u(a) = u(b) = 0$$

- Efficient solution using Thomas Tri-diagonal algorithm



`bvp_1d.m`

Thomas tri-diagonal algorithm

General $n \times n$ tri-diagonal matrix A

$$\mathrm{A} = \begin{bmatrix} a_1 & c_1 & & 0 \\ b_2 & a_2 & \ddots & \\ & \ddots & & c_{n-1} \\ 0 & & b_n & a_n \end{bmatrix}$$

Make an LU decomposition

$$A = LU$$

with L lower triangular and U upper triangular matrix.

$$\mathrm{L} = \begin{bmatrix} 1 & & & 0 \\ \beta_2 & 1 & & \\ & \ddots & \ddots & \\ 0 & & \beta_n & 1 \end{bmatrix}, \quad \mathrm{U} = \begin{bmatrix} \alpha_1 & c_1 & & 0 \\ & \alpha_2 & \ddots & \\ & & \ddots & c_{n-1} \\ 0 & & & \alpha_n \end{bmatrix}$$

Thomas tri-diagonal algorithm

The α_i and β_i are obtained from

$$\alpha_1 = a_1, \qquad \beta_i = \frac{b_i}{\alpha_{i-1}}, \qquad \alpha_i = a_i - \beta_i c_{i-1}, \quad i = 2, 3, \ldots, n$$

We want to solve

$$Ax = b \qquad \Longrightarrow \qquad LUx = b$$

We do this in two steps: $\boxed{Ly = b}$ and $\boxed{Ux = y}$

1. Solve $Ly = b$ by forward substitution

$$\begin{aligned} y_1 &= b_1 \\ \beta_2 y_1 + y_2 &= b_2 \\ \beta_3 y_2 + y_3 &= b_3 \qquad \text{etc.} \end{aligned}$$

No need to store the full matrix A. Store only the three diagonals.

2. Solve $Ux = y$ by backward substitution.

$$\begin{aligned} \alpha_n x_n &= y_n \\ \alpha_{n-1} x_{n-1} + c_{n-1} x_n &= y_{n-1} \\ \alpha_{n-2} x_{n-2} + c_{n-2} x_{n-1} &= y_{n-2} \qquad \text{etc.} \end{aligned}$$

Solution obtained in $\mathcal{O}(N)$ operations.