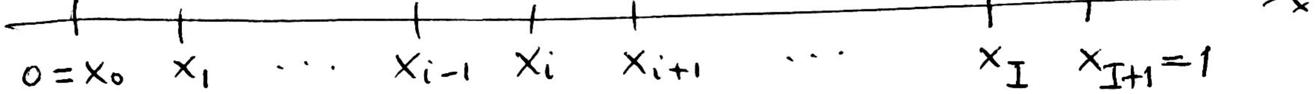
Lecture 4 (9/9/21)

Approximation of 2-point boundary value problems

Consider $\Omega = (0,1)$ (dimension $d=1$) and model elliptic BVP

$$\begin{cases} L[u] := -a(x)\, u'' + b(x)\, u' + c(x)\, u = f(x) & x \in \Omega \\ u(0) = \alpha, \quad u(1) = \beta \end{cases}$$

Consider a uniform partition $\mathcal{T} = \{x_i\}_{i=0}^{I+1}$ of Ω



Use centered finite differences at $x = x_i$

$$-a(x_i)\, \frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1})}{h^2}$$
$$+\, b(x_i)\, \frac{u(x_{i+1}) - u(x_{i-1})}{2h}$$
$$+\, c(x_i)\, u(x_i) \quad = \quad f(x_i) + \underbrace{\tau(x_i)}_{\text{Truncation error}}$$

$$|\tau(x_i)| \leq C h^2 \|u\|_{W^4_\infty(x_{i-1}, x_{i+1})}$$

The FDM method reads as follows: find $\underset{\sim}{U} = (U_i)_{i=0}^{I+1}$ such that

$$L_h[\underset{\sim}{U}]_i := -a_i\, \frac{U_{i-1} - 2U_i + U_{i+1}}{h^2}$$
$$+\, b_i\, \frac{U_{i+1} - U_{i-1}}{2h}$$
$$+\, c_i U_i = f_i \qquad \forall\, 1 \leq i \leq I$$

where $a_i = a(x_i)$, $b_i = b(x_i)$, $c_i = c(x_i)$, $f_i = f(x_i)$, and

$$U_0 = u(0) = \alpha, \qquad U_{I+1} = u(\underbrace{x_{I+1}}_{=1}) = \beta$$

In compact form this reads

$$L_h[\underset{\sim}{U}]_i := \left(-\frac{a_i}{h^2}-\frac{b_i}{2h}\right)U_{i-1}+\left(2\frac{a_i}{h^2}+c_i\right)U_i+\left(-\frac{a_i}{h^2}+\frac{b_i}{2h}\right)U_{i+1}=f_i$$

This is a linear tridiagonal algebraic system.

<u>Properties</u>

1. Matrix form : $K = \begin{bmatrix} \ddots & \ddots & \ddots \end{bmatrix} \in \mathbb{R}^{I\times I}$

$$-\frac{a_i}{h^2}-\frac{b_i}{2h} \qquad 2\frac{a_i}{h^2}+c_i \qquad -\frac{a_i}{h^2}+\frac{b_i}{2h} \qquad (i\text{th row})$$

$$\Longrightarrow \quad K\underset{\sim}{U} = \underset{\sim}{F} = \begin{bmatrix} f_1+\left(\frac{a_1}{h^2}+\frac{b_1}{2h}\right)\alpha \\ \vdots \\ f_i \\ \vdots \\ f_I+\left(\frac{a_I}{h^2}-\frac{b_I}{2h}\right)\beta \end{bmatrix} \in \mathbb{R}^I$$

$\longleftarrow$ ith row

2. Diagonal coeff

$$k_{ii} = 2\frac{a_i}{h^2}+c_i > 0$$

because of the following <u>assumptions</u>

(i) $0 < a_- \leq a(x) \leq a_+ \quad \forall x \in \bar{\Omega} \quad (a_-, a_+ \in \mathbb{R})$

(ii) $c(x) \geq 0$

(iii) $a, b, c, f \in C^0[0,1]$

$$\Longrightarrow \boxed{k_{ii} \geq \frac{2a_-}{h^2}} \qquad (1)$$

3. Off-diagonal coeffs

$$k_{i,i+1} = -\frac{a_i}{h^2}+\frac{b_i}{2h} < 0$$

We need to impose a condition on h for $k_{i,i+1} < 0$.

We impose

$$-\frac{a_-}{h^2} + \frac{\|b\|_{L^\infty(\Omega)}}{2h} < 0$$

$$\Longrightarrow \quad \frac{a_-}{h^2} > \frac{\|b\|_{L^\infty(\Omega)}}{2h} \Longrightarrow \boxed{h < \frac{2a_-}{\|b\|_{L^\infty(\Omega)}}} \qquad (2)$$

Peclet condition

The ratio diffusion / advection dictates the size of h. If $b=0$, then there is no restriction on h. Moreover (2) implies

$$k_{i,i-1} < 0.$$

4. $\sum_j k_{ij} = \left(-\cancel{\frac{a_i}{h^2}} - \cancel{\frac{b_i}{2h}}\right) + \left(\cancel{2\frac{a_i}{h^2}} + c_i\right) + \left(-\cancel{\frac{a_i}{h^2}} + \cancel{\frac{b_i}{2h}}\right) = c_i \geq 0$

for all $1 \leq i \leq I$ and for $i=1$

$$\sum_j k_{1j} = \Big(\underbrace{2\frac{a_i}{h^2} + c_i\Big) + \Big(-\frac{a_i}{h^2}}_{\frac{a_i}{h^2}} + \frac{b_i}{2h}\Big) \underset{(2)}{\geq} 0$$

and the same for $i = I$.

5. Upwinding: Can we get around (2)?

Since b represents a transport term ($b>0$ means information comes from the left) we exploit this feature in discretizing b. Suppose $b_i \geq 0$

$$b(x_i)\, u'(x_i) \Longrightarrow b(x_i)\, \frac{u(x_i) - u(x_{i-1})}{h} + \underbrace{z_i}_{O(h)}$$

The discrete operator reads

$$L_h[\underline{U}]_i = \underbrace{\left(-\frac{a_i}{h^2} - \frac{b_i}{h}\right)}_{k_{i,i-1}} U_{i-1} + \underbrace{\left(2\frac{a_i}{h^2} + c_i + \frac{b_i}{h}\right)}_{k_{ii}} U_i + \underbrace{\left(-\frac{a_i}{h^2}\right)}_{k_{i,i+1}} U_{i+1} = f_i$$

We have

(3) $k_{ii} > 0, \ k_{i,i-1} < 0, \ k_{i,i+1} < 0, \ \sum_j k_{ij} \geq 0.$

without Peclet condition

<u>Theorem</u> (discrete maximum principle). Assume $K = (k_{ij})_{ij=1}^{I}$ satisfies (3) for all $1 \leq i \leq I$. Let $\underset{\sim}{U} \in \mathbb{R}^I$ satisfy

$$L_h[\underset{\sim}{U}]_i = f_i \leq 0 \quad \forall \ 1 \leq i \leq I$$

Then

(i) If $c = 0$, then

$$\max_{1 \leq i \leq I} \{U_i\} \leq \max \{\alpha, \beta\}$$

(ii) If $c \geq 0$, then

$$\max_{1 \leq i \leq I} \{U_i\} \leq \max \{0, \alpha, \beta\}.$$

<u>Proof</u> We prove (1) and leave (2) as an exercise. We argue by contradiction: suppose $1 \leq i \leq I$ is an absolute maximum

(4) $\quad U_{i-1} \leq U_i \quad , \quad U_{i+1} < U_i \quad$ (strict ineq)

Note that $c = 0$ implies

$$\sum_j k_{ij} = 0$$

which reads

$$k_{ii} = -(k_{i,i-1} + k_{i,i+1})$$

Consider

$$L_h[\underset{\sim}{U}]_i = k_{ii-1} U_{i-1} + k_{ii} U_i + k_{ii+1} U_{i+1} = f_i$$

whence

$$U_i = \frac{1}{k_{ii}} \Big(-k_{ii-1} U_{i-1} - k_{ii+1} U_{i+1} + \underbrace{f_i}_{\leq 0} \Big)$$

$$\le \frac{-1}{k_{ii}} \Big(k_{ii-1} \underbrace{U_{i-1}}_{(A) \Rightarrow\ \le U_i} + k_{ii+1} \underbrace{U_{i+1}}_{< U_i} \Big)$$

$$< \underbrace{\frac{-1}{k_{ii}} \left(k_{ii-1} + k_{ii+1} \right)}_{=1} U_i = U_i \qquad \text{contradiction.}$$

This proves the Theorem. ∎

<u>Remark</u> DMP applies to both centered differences w/ Peclet condition and upwinding.

<u>Properties of K</u>

1. K is <u>nonsingular</u>: it suffices to show
$$K \underset{\sim}{U} = \underset{\sim}{0} \implies \underset{\sim}{U} = \underset{\sim}{0}$$
Choose $\alpha = \beta = 0$ and $f = 0$. Then
$$K \underset{\sim}{U} = L_h[\underset{\sim}{U}] \le \underset{\sim}{0} \implies \underset{\sim}{U} = (U_i)_{i=1}^{I} \le \underset{\sim}{0}$$
Likewise
$$K(-\underset{\sim}{U}) = \underset{\sim}{0} \le \underset{\sim}{0} \implies -\underset{\sim}{U} \le \underset{\sim}{0}$$
So $\underset{\sim}{U} = \underset{\sim}{0}$.

2. $K^{-1} \ge 0$ (componentwise)
Write
$$K^{-1} = [\underset{\sim}{v_1}, \underset{\sim}{v_2}, \dots, \underset{\sim}{v_I}] \in \mathbb{R}^{I \times I}$$
and notice
$$K K^{-1} = I \implies K \underset{\sim}{v_i} = \underset{\sim}{e_i} \ge \underset{\sim}{0} \quad \forall i$$
$$\implies \underset{\sim}{v_i} \ge \underset{\sim}{0} \quad \forall i$$
(take $\alpha = \beta = 0$, $\underset{\sim}{f} = \underset{\sim}{e_i} \ge \underset{\sim}{0}$ and apply DMP).

Def A matrix K is an <u>M-matrix</u> iff property (3) holds and $\sum_j k_{ij} > 0$ for some $1 \le i \le I$.

Construction of Discrete Subsolutions

Let $w \in C^4(\bar{\Omega})$ satisfy

$$\begin{cases} L[w] \le -2 & \Omega \\ w(0),\ w(1) \ge 1 \end{cases}$$

Recall $w(x) = c\, e^{\lambda x}$ is such a function for λ large.
Let $\underset{\sim}{W} = (W_i)_{i=1}^{I}$ with $W_i = w(x_i)$. We have shown

$$L[w](x_i) = L_h[\underset{\sim}{W}]_i + \tau_i$$

truncation error $= \begin{cases} ch^2 & \text{centered} \\ ch & \text{upwinding} \end{cases}$

Consequently

$$L_h[\underset{\sim}{W}]_i = L[w](x_i) - \tau_i$$
$$\le -2 - \tau_i \le -1$$

provided h is sufficiently small $(|\tau_i| \le 1)$. In addition $W_0 = w(x_0) \ge 1, \quad W_{I+1} = w(x_{I+1}) \ge 1.$

<u>Theorem</u> (stability) The discrete solution $\underset{\sim}{U} \in \mathbb{R}^I$ of $L_h[\underset{\sim}{U}] = \underset{\sim}{F}$ satisfies

$$\|\underset{\sim}{U}\|_\infty \le \max\{|\alpha|, |\beta|\} + \Lambda \|\underset{\sim}{F}\|_\infty$$

where Λ is a constant that depends on Ω, a, b, c.

Proof 1. Consider auxiliary vector

$$\underset{\sim}{V} = \underset{\sim}{U} + c\underset{\sim}{W}$$

where $c>0$ is to be determined. Compute

$$L_h[\underset{\sim}{V}]_i = \underbrace{L_h[\underset{\sim}{U}]_i}_{F_i} + c\underbrace{L_h[\underset{\sim}{W}]_i}_{\leq -1} \leq F_i - c \leq 0$$

provided $c = \|\underset{\sim}{F}\|_\infty$. Apply DMP to $\underset{\sim}{V}$

$$V_i \leq \max\{V_0, V_{I+1}, 0\}.$$

2.

$$U_i + cW_i = V_i \geq U_i$$

and

$$V_0 = U_0 + cW_0 = \alpha + \|\underset{\sim}{F}\|_\infty W(0)$$

$$V_{I+1} = \beta + \|\underset{\sim}{F}\|_\infty W(1)$$

This implies the assertion. ∎