A discrete maximum principle

Recall that we constructed $\{U_i\}_{i=0}^{N+1}$ by setting

$$\begin{cases} -p\,\Delta_h^{(3)} U_i + q\,\delta_h U_i + r U_i = f(x_i) & \text{for } i = 1, \dots, N \\ U_0 = g_a \text{ and } U_{N+1} = g_b \end{cases} \qquad (*)$$

We now want to prove that this system is well posed, meaning that it admits a unique solution $\{U_i\}_{i=0}^{N+1}$ and that this solution is "controlled" by $f(x)$ and g_a and g_b. (we will see what this means precisely in a minute).

Theorem: (Discrete Maximum principle)

Assume that $p > 0$, $r \geq 0$ and h is chosen small enough such that $p - \frac{1}{2}|q|h > 0$. Let $\{V_i\}_{i=0}^{N+1}$ be an arbitrary vector satisfying $-p\,\Delta_h^{(3)} V_i + q\,\delta_h V_i + r V_i \leq 0$ for $i = 1, \dots, N$.

Then, for $r = 0$: $\max\limits_{0 \leq i \leq N+1} V_i \leq \max(V_0, V_{N+1})$.

and equality holds if and only if V is constant.

And for $r > 0$: $\max\limits_{1 \leq i \leq N} V_i \leq 0$

Proof: Suppose the maximal value $\max_{0\le i\le N+1} V_i$ is obtained for an interior index i_* (meaning $1\le i^* \le N$). Then, by assumption

$$-p\Delta_h^{(3)} V_{i_*} + q\delta_h V_{i_*} + r V_{i_*} \le 0 \quad \text{, which written out:}$$

$$\left(-p+\tfrac{1}{2}qh\right)V_{i_*+1} + \left(2p+rh^2\right)V_{i_*} + \left(-p-\tfrac{1}{2}qh\right)V_{i_*-1} \le 0.$$

Rearranging: $\left(2p+rh^2\right)V_{i_*} \le \underbrace{\left(p-\tfrac{1}{2}qh\right)}_{>0}V_{i_*+1} + \underbrace{\left(p+\tfrac{1}{2}qh\right)}_{>0}V_{i_*-1}$

by assumption

$$\le \left(p-\tfrac{1}{2}qh\right)V_{i_*} + \left(p+\tfrac{1}{2}qh\right)V_{i_*}$$

because $V_{i_*} \ge V_{i_*-1}$ and $V_{i_*} \ge V_{i_*+1}$

$$\le 2pV_{i_*}$$

Now, if $r>0$ then $(2p+rh^2)V_{i_*} \le 2pV_{i_*}$ implies $rh^2 V_{i_*} \le 0$ which shows $\max_{1\le i\le N} V_i \le V_{i_*} \le 0$.

Similarly, for $r=0$ we end up with

$$2pV_{i_*} \le \left(p-\tfrac{1}{2}qh\right)V_{i_*-1} + \left(p+\tfrac{1}{2}qh\right)V_{i_*+1} \le 2pV_{i_*}.$$

and we have $V_{i_*-1} \le V_{i_*}$ and $V_{i_*+1} \le V_{i_*}$. The only possible solution is $V_{i_*-1} = V_{i_*} = V_{i_*+1}$ (not necessarily 0).

Now we can repeat the argument for index $i_\gamma - 1$ and $i_\gamma + 1$ until we hit $i_\gamma - 1 = 0$ and $i_\gamma + 1 = N+1$. This shows that the $\{V_i\}$ has to be constant, which proves the statement. $\square$

We can now prove the following theorem which assets that the finite difference approximation is "well posed":

Theorem: Given the task to find $\{U_i\}_{i=0}^{N+1}$ such that

$$\begin{cases} -p\,\Delta_h^{(3)} U_i + q\,\delta_h U_i + r\,U_i = f(x_i) & \text{for } i = 1, \dots, N \\ U_0 = g_a \text{ and } U_{N+1} = g_b & (*) \end{cases}$$

and assume that $p > 0$, $r \geq 0$ and h is chosen small enough such that $p - \frac{1}{2}|q|h > 0$. Then $(*)$ is uniquely solvable and the solution satisfies

$$\max_{0 \leq i \leq N+1} |U_i| \leq \frac{(b-a)^2}{8p} \max_i |f(x_i)| + \max(|g_a|, |g_b|)$$

— a priori estimate —

<u>Proof</u>: • Let us prove <u>uniqueness</u> first. Suppose there are two solutions $\{U_i\}$ and $\{V_i\}$ to (*). Then, the difference $\{W_i\}$ defined by $W_i = U_i - V_i$ solves

$$\begin{cases} -p\,\Delta_h^{(3)} W_i + q\,\delta_h W_i + r W_i = 0 & \text{for } i=1,\dots,N \\ W_0 = 0 \text{ and } W_{N+1} = 0 \end{cases}$$

and we can now apply the discrete maximum principle which gives $\max\limits_{0\le i\le N+1} W_i \le 0$. But the system is also solved by $\{-W_i\}$:

$$\begin{cases} -p\,\Delta_h^{(3)}(-W_i) + q\,\delta_h(-W_i) + r(-W_i) = 0 & \text{for } i=1,\dots,N \\ -W_0 = 0 \text{ and } -W_{N+1} = 0 \end{cases}$$

and the discrete maximum principle implies $\max\limits_{0\le i\le N+1}(-W_i) \le 0$, which is equivalent to $\min\limits_{0\le i\le N+1} W_i \ge 0$.

We have thus shown that W_i is identically zero.

• <u>Existence</u>: (*) is a linear system with N equations for N unknowns. We know from Linear Algebra that uniqueness implies existence of a solution.

We prove the last part only for $g=0$:

Introduce a function $w(x) = \left\{ \frac{1}{2p}\left(x - \frac{1}{2}(a+b)\right)^2 - \frac{(b-a)^2}{8p} \right\} \|f\|_\infty$ $=$ quadratic polynomial $=$

where we have set $\|f\|_\infty = \max_{1\le i\le N} |f(x_i)|$.

Let us compute $-p\partial_x^2 w(x) + r w(x)$:

$$-p\partial_x^2 w(x) + r w(x) = -p\left(\frac{1}{2p}2\right)\|f\|_\infty + \underbrace{r w(x)}_{\substack{\le 0 \\ \text{by construction}}}$$

$$\le -\|f\|_\infty$$

Because $w(x)$ is a quadratic polynomial it has vanishing 3rd derivatives. This implies that

$$-p\Delta_h^{(3)} w(x) + r w(x) \overset{w \text{ is quadratic polynomial!}}{=} -p\partial_x^2 w(x) + r w(x) \le -\|f\|_\infty .$$

Let us now set $W_i := w(x_i)$. Then, $\pm U_i + W_i$ satisfies

$$-p\Delta_h^{(3)}(\pm U_i + W_i) + r(\pm U_i + W_i) \le \pm f(x_i) - \|f\|_\infty \le 0$$

and we get $\max_{1\le i\le N}(\pm U_i + W_i) \le \max(\pm U_0 + W_0, \pm U_{N+1} + W_{N+1})$

$$= \max(\pm g_a, \pm g_b)$$

$$\le \max(|g_a|, |g_b|)$$

As a last ingredient we need to take the max on the left side apart:

$$\max_{1\le i\le N}\left(\pm U_i + W_i\right) \geq \max_{1\le i\le N}\left(\pm U_i + \min_{1\le i\le N} W_i\right)$$

$$\geq \max_{1\le i\le N}\left(\pm U_i\right) + \min_{1\le i\le N} W_i$$

and we have

$$\min_{1\le i\le N} W_i = -\frac{(b-a)^2}{8p}\|f\|_\infty$$

This implies: $\max_{1\le i\le N}\left(\pm U_i\right) \leq \max\left(|g_a|, |g_b|\right) + \frac{(b-a)^2}{8p}\|f\|_\infty.$ $\square$

Homework: How would you change $w(x)$ to accommodate $q\neq 0$?

With the stability result at hand we are now in a position to prove convergence:

$$\boxed{\begin{array}{l}\text{Corollary: Let } u(x) \text{ be the solution to } -p\partial_x^2 u + q\partial_x u + ru = f \\ \text{with } u(a) = g_a \text{ and } u(b) = g_b, \text{ and let } \{U_i\} \text{ be given by} \\ \begin{cases} -p\Delta_h^{(3)} U_i + q\delta_h U_i + rU_i = f(x_i) \text{ for } i=1,\dots,N \\ U_0 = g_a \text{ and } U_{N+1} = g_b \end{cases} \quad \boxed{q=0} \\ \text{Then, } \max_{1\le i\le N} |U_i - u(x_i)| \leq \frac{(b-a)^2}{96} h^2 \max_{x\in[a,b]} |\partial_x^{(4)} u(x)| \end{array}}$$

Proof.: We again introduce a truncation error:

$$\tau_i := -p\,\Delta_h^{(3)} u(x_i) + r\,u(x_i) - f(x_i)$$

- On the one hand this implies

$$\tau_i = -p\Delta_h^{(3)} u(x_i) + r u(x_i) - \left\{ -p\partial_x^2 u(x_i) - r u(x_i) \right\}$$

$$= -p\left\{ \Delta_h^{(3)} u(x_i) - \partial_x^2 u(x_i) \right\} = -p\frac{1}{12} h^2 \partial_x^{(4)} u(\underset{\substack{\uparrow \\ \text{some suitable } \xi_i \in (x_{i-1}, x_{i+1})}}{\xi_i})$$

$$\Rightarrow \max_{1\le i\le N} |\tau_i| \le \frac{p}{12} h^2 \max_{\xi\in[a,b]} |\partial_x^{(4)} u(\xi)|$$

- On the other hand the error $e_i = u(x_i) - U_i$ satisfies

$$-p\Delta_h^{(3)} e_i + r e_i = \tau_i \quad \text{with } e_0 = e_{N+1} = 0$$

and the stability estimate reads:

$$\max_{0\le i\le N+1} |e_i| \le \frac{(b-a)^2}{8p} \max_i |\tau_i| \le \frac{(b-a)^2}{96} h^2 \max_{\xi\in[a,b]} |\partial_x^{(4)} u(\xi)|$$

$\square$

Homework: How do you need to modify the Corollary for the case $q \neq 0$?