Numerical Solution of Differential Equations I

Endre Süli

Mathematical Institute University of Oxford 2020

Lecture 16

The discrete maximum principle

Theorem (Discrete maximum principle for the θ -scheme)

The θ -scheme for the Dirichlet initial-boundary-value problem for the heat equation, with $0 \le \theta \le 1$ and $\mu(1-\theta) \le \frac{1}{2}$, yields a sequence of numerical approximations $\{U_j^m\}_{j=0,\dots,J;\ m=0,\dots,M}$ satisfying

$$U_{\min} \leq U_j^m \leq U_{\max}$$

where

$$U_{\min} = \min \left\{ \min \{ U_0^m \}_{m=0}^M, \min \{ U_j^0 \}_{j=0}^J, \min \{ U_J^m \}_{m=0}^M \right\}$$

and

$$U_{\max} = \max \left\{ \max\{U_0^m\}_{m=0}^M, \; \max\{U_j^0\}_{j=0}^J, \; \max\{U_J^m\}_{m=0}^M \right\}.$$

PROOF: We rewrite the θ -scheme as

$$(1 + 2\theta\mu) U_j^{m+1} = \theta\mu \left(U_{j+1}^{m+1} + U_{j-1}^{m+1} \right) + (1 - \theta)\mu \left(U_{j+1}^m + U_{j-1}^m \right) + [1 - 2(1 - \theta)\mu] U_j^m,$$

and recall that, by hypothesis,

$$\theta \mu \geq 0$$
 $(1-\theta)\mu \geq 0$, $1-2(1-\theta)\mu \geq 0$.

PROOF: We rewrite the θ -scheme as

$$(1 + 2\theta\mu) U_j^{m+1} = \theta\mu \left(U_{j+1}^{m+1} + U_{j-1}^{m+1} \right) + (1 - \theta)\mu \left(U_{j+1}^m + U_{j-1}^m \right) + [1 - 2(1 - \theta)\mu] U_j^m,$$

and recall that, by hypothesis,

$$\theta \mu \ge 0$$
 $(1 - \theta)\mu \ge 0$, $1 - 2(1 - \theta)\mu \ge 0$.

Suppose that U attains its maximum value at an internal mesh point U_j^{m+1} , $1 \leq j \leq J-1$, $0 \leq m \leq M-1$. If this is not the case, the proof is complete.

PROOF: We rewrite the θ -scheme as

$$(1 + 2\theta\mu) U_j^{m+1} = \theta\mu \left(U_{j+1}^{m+1} + U_{j-1}^{m+1} \right) + (1 - \theta)\mu \left(U_{j+1}^m + U_{j-1}^m \right) + [1 - 2(1 - \theta)\mu] U_j^m,$$

and recall that, by hypothesis,

$$\theta \mu \ge 0$$
 $(1 - \theta)\mu \ge 0$, $1 - 2(1 - \theta)\mu \ge 0$.

Suppose that U attains its maximum value at an internal mesh point U_j^{m+1} , $1 \leq j \leq J-1$, $0 \leq m \leq M-1$. If this is not the case, the proof is complete.

We define

$$U^{\star} = \max\{U_{j+1}^{m+1},\ U_{j-1}^{m+1},\ U_{j+1}^{m},\ U_{j-1}^{m},\ U_{j}^{m}\}.$$

Then,

$$(1 + 2\theta\mu) U_j^{m+1} \le 2\theta\mu U^* + 2(1 - \theta)\mu U^* + [1 - 2(1 - \theta)\mu]U^* = (1 + 2\theta\mu) U^*,$$

and therefore

$$U_j^{m+1} \leq U^*$$
.

Then,

$$(1 + 2\theta\mu) U_j^{m+1} \le 2\theta\mu U^* + 2(1 - \theta)\mu U^* + [1 - 2(1 - \theta)\mu]U^* = (1 + 2\theta\mu) U^*,$$

and therefore

$$U_j^{m+1} \leq U^*$$
.

However, also,

$$U^{\star} \leq U_j^{m+1},$$

as U_j^{m+1} is assumed to be the overall maximum value.

Then,

$$(1+2\theta\mu) U_j^{m+1} \le 2\theta\mu U^* + 2(1-\theta)\mu U^* + [1-2(1-\theta)\mu]U^* = (1+2\theta\mu) U^*,$$

and therefore

$$U_j^{m+1} \leq U^*$$
.

However, also,

$$U^* \leq U_j^{m+1}$$
,

as U_j^{m+1} is assumed to be the overall maximum value. Hence,

$$U_j^{m+1}=U^*.$$

The same argument applies to these neighbouring points, and we can then repeat this process until the boundary at x = a or x = b or at t = 0 is reached, in a finite number of steps.

The same argument applies to these neighbouring points, and we can then repeat this process until the boundary at x = a or x = b or at t = 0 is reached, in a finite number of steps.

The maximum is therefore attained at a boundary point.

The same argument applies to these neighbouring points, and we can then repeat this process until the boundary at x = a or x = b or at t = 0 is reached, in a finite number of steps.

The maximum is therefore attained at a boundary point.

Similarly, the minimum is attained at a boundary point.

In summary then, for

$$\mu(1-\theta) \leq \frac{1}{2}$$

the $\theta\text{-scheme}$ satisfies the discrete maximum principle.

In summary then, for

$$\mu(1-\theta) \leq \frac{1}{2}$$

the θ -scheme satisfies the discrete maximum principle.

This is clearly more demanding than the ℓ_2 -stability condition:

$$\mu(1-2\theta) \le \frac{1}{2}$$
 for $0 \le \theta \le \frac{1}{2}$.

In summary then, for

$$\mu(1-\theta) \leq \frac{1}{2}$$

the θ -scheme satisfies the discrete maximum principle.

This is clearly more demanding than the ℓ_2 -stability condition:

$$\mu(1-2\theta) \le \frac{1}{2}$$
 for $0 \le \theta \le \frac{1}{2}$.

For example, the Crank-Nicolson scheme is unconditionally stable in the ℓ_2 norm, yet it only satisfies the discrete maximum principle when $\mu:=\frac{\Delta t}{(\Delta x)^2}\leq 1$.

Convergence of the θ -scheme in the maximum norm

We close our discussion of finite difference schemes for the heat equation in one space-dimension with the convergence analysis of the θ -scheme for the Dirichlet initial-boundary-value problem.

Convergence of the θ -scheme in the maximum norm

We close our discussion of finite difference schemes for the heat equation in one space-dimension with the convergence analysis of the θ -scheme for the Dirichlet initial-boundary-value problem.

We begin by rewriting the scheme as follows:

$$(1 + 2\theta\mu) U_j^{m+1} = \theta\mu \left(U_{j+1}^{m+1} + U_{j-1}^{m+1} \right) + (1 - \theta)\mu \left(U_{j+1}^{m} + U_{j-1}^{m} \right) + [1 - 2(1 - \theta)\mu] U_j^{m}.$$

Convergence of the θ -scheme in the maximum norm

We close our discussion of finite difference schemes for the heat equation in one space-dimension with the convergence analysis of the θ -scheme for the Dirichlet initial-boundary-value problem.

We begin by rewriting the scheme as follows:

$$(1 + 2\theta\mu) U_j^{m+1} = \theta\mu \left(U_{j+1}^{m+1} + U_{j-1}^{m+1} \right) + (1 - \theta)\mu \left(U_{j+1}^m + U_{j-1}^m \right) + [1 - 2(1 - \theta)\mu] U_j^m.$$

The scheme is considered subject to the initial condition

$$U_j^0 = u_0(x_j), \qquad j = 1, \dots, J-1,$$

and the boundary conditions

$$U_0^{m+1} = A(t_{m+1}), \quad U_1^{m+1} = B(t_{m+1}), \quad m = 0, \dots, M-1.$$

The **consistency error** for the θ -scheme is defined by

$$T_{j}^{m} = \frac{u_{j}^{m+1} - u_{j}^{m}}{\Delta t} - (1 - \theta) \frac{u_{j+1}^{m} - 2u_{j}^{m} + u_{j-1}^{m}}{(\Delta x)^{2}} - \theta \frac{u_{j+1}^{m+1} - 2u_{j}^{m+1} + u_{j-1}^{m+1}}{(\Delta x)^{2}},$$

where $u_j^m \equiv u(x_j, t_m)$,

The **consistency error** for the θ -scheme is defined by

$$T_{j}^{m} = \frac{u_{j}^{m+1} - u_{j}^{m}}{\Delta t} - (1 - \theta) \frac{u_{j+1}^{m} - 2u_{j}^{m} + u_{j-1}^{m}}{(\Delta x)^{2}} - \theta \frac{u_{j+1}^{m+1} - 2u_{j}^{m+1} + u_{j-1}^{m+1}}{(\Delta x)^{2}},$$

where $u_i^m \equiv u(x_j, t_m)$, and therefore

$$(1+2\theta\mu) u_j^{m+1} = \theta\mu \left(u_{j+1}^{m+1} + u_{j-1}^{m+1} \right) + (1-\theta)\mu \left(u_{j+1}^m + u_{j-1}^m \right)$$
$$+ \left[1 - 2(1-\theta)\mu \right] u_j^m + \Delta t T_j^m.$$

Define the **global error**, that is the discrepancy at a mesh-point between the exact solution and its numerical approximation, by

$$e_j^m := u(x_j, t_m) - U_j^m.$$

Define the **global error**, that is the discrepancy at a mesh-point between the exact solution and its numerical approximation, by

$$e_j^m := u(x_j, t_m) - U_j^m.$$

It then follows that

$$e_0^{m+1} = 0, \ e_J^{m+1} = 0, \ e_j^0 = 0, \quad j = 0, \dots, J,$$

and

$$(1+2\theta\mu) e_j^{m+1} = \theta\mu \left(e_{j+1}^{m+1} + e_{j-1}^{m+1} \right) + (1-\theta)\mu \left(e_{j+1}^m + e_{j-1}^m \right) + [1-2(1-\theta)\mu] e_j^m + \Delta t T_j^m.$$

Define the **global error**, that is the discrepancy at a mesh-point between the exact solution and its numerical approximation, by

$$e_j^m := u(x_j, t_m) - U_j^m.$$

It then follows that

$$e_0^{m+1} = 0, \ e_J^{m+1} = 0, \ e_j^0 = 0, \quad j = 0, \dots, J,$$

and

$$(1+2\theta\mu) e_j^{m+1} = \theta\mu \left(e_{j+1}^{m+1} + e_{j-1}^{m+1} \right) + (1-\theta)\mu \left(e_{j+1}^m + e_{j-1}^m \right) + [1-2(1-\theta)\mu] e_j^m + \Delta t T_j^m.$$

We define,

$$E^m = \max_{0 \le j \le J} |e_j^m|$$
 and $T^m = \max_{0 \le j \le J} |T_j^m|$.

$$\theta \mu \geq 0,$$
 $(1-\theta)\mu \geq 0,$ $1-2(1-\theta)\mu \geq 0,$

we have that

$$(1+2\theta\mu)E^{m+1} \le 2\theta\mu E^{m+1} + E^m + \Delta t T^m.$$

$$\theta \mu \geq 0,$$
 $(1-\theta)\mu \geq 0,$ $1-2(1-\theta)\mu \geq 0,$

we have that

$$(1+2\theta\mu)E^{m+1} \le 2\theta\mu E^{m+1} + E^m + \Delta t T^m.$$

Hence,

$$E^{m+1} \leq E^m + \Delta t \ T^m.$$

$$\theta\mu \geq 0, \qquad (1-\theta)\mu \geq 0, \qquad 1-2(1-\theta)\mu \geq 0,$$

we have that

$$(1+2\theta\mu)E^{m+1} \le 2\theta\mu E^{m+1} + E^m + \Delta t T^m.$$

Hence,

$$E^{m+1} \leq E^m + \Delta t \ T^m.$$

As $E^0 = 0$, upon summation,

$$E^{m} \leq \Delta t \sum_{n=0}^{m-1} T^{n}$$

$$\leq m \Delta t \max_{0 \leq n \leq m-1} T^{n}$$

$$\leq T \max_{0 \leq m \leq M} \max_{1 \leq j \leq J-1} |T_{j}^{m}|,$$

$$\theta \mu \ge 0,$$
 $(1-\theta)\mu \ge 0,$ $1-2(1-\theta)\mu \ge 0,$

we have that

$$(1+2\theta\mu)E^{m+1} \le 2\theta\mu E^{m+1} + E^m + \Delta t T^m.$$

Hence,

$$E^{m+1} \leq E^m + \Delta t \ T^m.$$

As $E^0 = 0$, upon summation,

$$E^{m} \leq \Delta t \sum_{n=0}^{m-1} T^{n}$$

$$\leq m \Delta t \max_{0 \leq n \leq m-1} T^{n}$$

$$\leq T \max_{0 \leq m < M} \max_{1 \leq j \leq J-1} |T_{j}^{m}|,$$

which then implies that

$$\max_{0 \leq j \leq J} \max_{0 \leq m \leq M} |u(x_j, t_m) - U_j^m| \leq T \max_{1 \leq j \leq J-1} \max_{0 \leq m \leq M} |T_j^m|.$$

Recall that the consistency error of the θ -scheme is

$$T_j^m = \left\{ egin{array}{ll} \mathcal{O}\left((\Delta x)^2 + (\Delta t)^2
ight) & ext{for } \theta = 1/2, \\ \mathcal{O}\left((\Delta x)^2 + \Delta t
ight) & ext{for } \theta
eq 1/2. \end{array}
ight.$$

Recall that the consistency error of the θ -scheme is

$$T_j^m = \begin{cases} \mathcal{O}\left((\Delta x)^2 + (\Delta t)^2\right) & \text{for } \theta = 1/2, \\ \mathcal{O}\left((\Delta x)^2 + \Delta t\right) & \text{for } \theta \neq 1/2. \end{cases}$$

For the explicit/implicit Euler schemes, for which

$$T_j^m = \mathcal{O}\left((\Delta x)^2 + \Delta t\right),$$

one has the following bound on the global error:

$$\max_{0 \le j \le J} \max_{0 \le m \le M} |u(x_j, t_m) - U_j^m| \le \text{Const.} \left((\Delta x)^2 + \Delta t \right),$$

Recall that the consistency error of the θ -scheme is

$$T_j^m = \begin{cases} \mathcal{O}\left((\Delta x)^2 + (\Delta t)^2\right) & \text{for } \theta = 1/2, \\ \mathcal{O}\left((\Delta x)^2 + \Delta t\right) & \text{for } \theta \neq 1/2. \end{cases}$$

For the explicit/implicit Euler schemes, for which

$$T_j^m = \mathcal{O}\left((\Delta x)^2 + \Delta t\right),\,$$

one has the following bound on the global error:

$$\max_{0 \le j \le J} \max_{0 \le m \le M} |u(x_j, t_m) - U_j^m| \le \text{Const.} \left((\Delta x)^2 + \Delta t \right),$$

while for the Crank-Nicolson scheme, which has consistency error

$$T_j^m = \mathcal{O}\left((\Delta x)^2 + (\Delta t)^2\right),\,$$

one has

$$\max_{0 \leq j \leq J} \max_{0 \leq m \leq M} |u(x_j, t_m) - U_j^m| \leq \text{Const.} \left((\Delta x)^2 + (\Delta t)^2 \right).$$

Finite difference approximation in two space-dimensions

Consider the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, \qquad (x, y) \in \Omega := (a, b) \times (c, d), \ t \in (0, T],$$

subject to the initial condition

$$u(x, y, 0) = u_0(x, y),$$
 $(x, y) \in [a, b] \times [c, d],$

and the Dirichlet boundary condition

$$u|_{\partial\Omega} = B(x, y, t), \qquad (x, y) \in \partial\Omega, \ \ t \in (0, T],$$

where $\partial\Omega$ is the boundary of Ω .

Finite difference approximation in two space-dimensions

Consider the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial v^2}, \qquad (x, y) \in \Omega := (a, b) \times (c, d), \ t \in (0, T],$$

subject to the initial condition

$$u(x, y, 0) = u_0(x, y),$$
 $(x, y) \in [a, b] \times [c, d],$

and the Dirichlet boundary condition

$$u|_{\partial\Omega} = B(x, y, t), \qquad (x, y) \in \partial\Omega, \ \ t \in (0, T],$$

where $\partial\Omega$ is the boundary of Ω .

We begin by considering the explicit Euler finite difference scheme for this problem.

The explicit Euler scheme

Let

$$\delta_x^2 U_{ij} := U_{i+1,j} - 2U_{ij} + U_{i-1,j},$$

and

$$\delta_y^2 U_{ij} := U_{i,j+1} - 2U_{ij} + U_{i,j-1}.$$

The explicit Euler scheme

Let

$$\delta_x^2 U_{ij} := U_{i+1,j} - 2U_{ij} + U_{i-1,j},$$

and

$$\delta_y^2 U_{ij} := U_{i,j+1} - 2U_{ij} + U_{i,j-1}.$$

Let, further, $\Delta x := (b-a)/J_x$, $\Delta y := (d-c)/J_y$, $\Delta t := T/M$, and define

$$x_i = a + i\Delta x,$$
 $i = 0, ..., J_x,$
 $y_j = c + j\Delta y,$ $j = 0, ..., J_y,$
 $t_m = m\Delta t,$ $m = 0, ..., M.$

The explicit Euler finite difference scheme for the unsteady heat equation on the space-time domain $\overline{\Omega} \times [0, T]$ is then:

$$\frac{U_{ij}^{m+1}-U_{ij}^m}{\Delta t}=\frac{\delta_x^2U_{ij}^m}{(\Delta x)^2}+\frac{\delta_y^2U_{ij}^m}{(\Delta y)^2},$$

for
$$i = 1, ..., J_x - 1$$
, $j = 1, ..., J_y - 1$, $m = 0, 1, ..., M - 1$,

The explicit Euler finite difference scheme for the unsteady heat equation on the space-time domain $\overline{\Omega} \times [0, T]$ is then:

$$\frac{U_{ij}^{m+1}-U_{ij}^{m}}{\Delta t}=\frac{\delta_{x}^{2}U_{ij}^{m}}{(\Delta x)^{2}}+\frac{\delta_{y}^{2}U_{ij}^{m}}{(\Delta y)^{2}},$$

for $i=1,\ldots,J_x-1,\,j=1,\ldots,J_y-1,\,m=0,1,\ldots,M-1,$ subject to the initial condition

$$U_{ij}^0 = u_0(x_i, y_j), \quad i = 0, \dots, J_x, \ j = 0, \dots, J_y,$$

The explicit Euler finite difference scheme for the unsteady heat equation on the space-time domain $\overline{\Omega} \times [0, T]$ is then:

$$\frac{U_{ij}^{m+1} - U_{ij}^{m}}{\Delta t} = \frac{\delta_{x}^{2} U_{ij}^{m}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m}}{(\Delta y)^{2}},$$

for $i=1,\ldots,J_x-1,\,j=1,\ldots,J_y-1,\,m=0,1,\ldots,M-1$, subject to the initial condition

$$U_{ij}^0 = u_0(x_i, y_j), \quad i = 0, \dots, J_x, \ j = 0, \dots, J_y,$$

and the boundary condition

 $U_{ij}^m = B(x_i, y_j, t_m)$, at the boundary mesh points, for $m = 1, \dots, M$.

The implicit Euler scheme

Let
$$\Delta x:=(b-a)/J_x$$
, $\Delta y:=(d-c)/J_y$, $\Delta t:=T/M$, and define
$$x_i=a+i\Delta x, \qquad i=0,\dots,J_x, \\ y_j=c+j\Delta y, \qquad j=0,\dots,J_y, \\ t_m=m\Delta t, \qquad m=0,\dots,M.$$

The implicit Euler scheme

Let
$$\Delta x:=(b-a)/J_x$$
, $\Delta y:=(d-c)/J_y$, $\Delta t:=T/M$, and define
$$x_i=a+i\Delta x, \qquad i=0,\dots,J_x, \\ y_j=c+j\Delta y, \qquad j=0,\dots,J_y, \\ t_m=m\Delta t, \qquad m=0,\dots,M.$$

The implicit Euler finite difference scheme for the problem is then

$$\frac{U_{ij}^{m+1} - U_{ij}^{m}}{\Delta t} = \frac{\delta_{x}^{2} U_{ij}^{m+1}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m+1}}{(\Delta y)^{2}},$$

for
$$i = 1, ..., J_x - 1$$
, $j = 1, ..., J_y - 1$, $m = 0, 1, ..., M - 1$,

subject to the initial condition

$$U_{ij}^0 = u_0(x_i, y_j), \quad i = 0, \dots, J_x, \ j = 0, \dots, J_y,$$

and the boundary condition

$$U_{ij}^{m+1} = B(x_i, y_j, t_{m+1}),$$
 at the boundary mesh points, for $m = 0, \dots, M-1$.

The θ -scheme

Let $\Delta x := (b-a)/J_x$, $\Delta y := (d-c)/J_y$, $\Delta t := T/M$, and, for $\theta \in [0,1]$, consider the finite difference scheme

$$\frac{U_{ij}^{m+1} - U_{ij}^{m}}{\Delta t} = (1 - \theta) \left(\frac{\delta_{x}^{2} U_{ij}^{m}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m}}{(\Delta y)^{2}} \right) + \theta \left(\frac{\delta_{x}^{2} U_{ij}^{m+1}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m+1}}{(\Delta y)^{2}} \right)$$

for
$$i = 1, ..., J_x - 1$$
, $j = 1, ..., J_y - 1$, $m = 0, 1, ..., M - 1$,

The θ -scheme

Let $\Delta x := (b-a)/J_x$, $\Delta y := (d-c)/J_y$, $\Delta t := T/M$, and, for $\theta \in [0,1]$, consider the finite difference scheme

$$\frac{U_{ij}^{m+1} - U_{ij}^{m}}{\Delta t} = (1 - \theta) \left(\frac{\delta_{x}^{2} U_{ij}^{m}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m}}{(\Delta y)^{2}} \right) + \theta \left(\frac{\delta_{x}^{2} U_{ij}^{m+1}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m+1}}{(\Delta y)^{2}} \right)$$

for $i=1,\ldots,J_x-1,\,j=1,\ldots,J_y-1,\,m=0,1,\ldots,M-1,$ subject to the initial condition

$$U_{ij}^0 = u_0(x_i, y_j), \quad i = 0, \dots, J_x, \ j = 0, \dots, J_y,$$

The θ -scheme

Let $\Delta x := (b-a)/J_x$, $\Delta y := (d-c)/J_y$, $\Delta t := T/M$, and, for $\theta \in [0,1]$, consider the finite difference scheme

$$\frac{U_{ij}^{m+1} - U_{ij}^{m}}{\Delta t} = (1 - \theta) \left(\frac{\delta_{x}^{2} U_{ij}^{m}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m}}{(\Delta y)^{2}} \right) + \theta \left(\frac{\delta_{x}^{2} U_{ij}^{m+1}}{(\Delta x)^{2}} + \frac{\delta_{y}^{2} U_{ij}^{m+1}}{(\Delta y)^{2}} \right)$$

for $i=1,\ldots,J_x-1, j=1,\ldots,J_y-1, m=0,1,\ldots,M-1$, subject to the initial condition

$$U_{ij}^0 = u_0(x_i, y_j), \quad i = 0, \dots, J_x, \ j = 0, \dots, J_y,$$

and the boundary condition

$$U_{ij}^{m+1} = B(x_i, y_j, t_{m+1}),$$
 at the boundary mesh points, for $m = 0, \dots, M-1$.