TMA4212 Numerical solution of partial differential equations with finite difference methods

Solutions to Problem Set 3

Problem 1. We start by semidiscretizing the equation. We split the interval $0 \leq x \leq c$ into M_1 parts, each with length $h_1 = c/M_1$, and the interval $c \leq x \leq 1$ into M_2 parts, each with length $h_2 = (1-c)/M_2$. We then let $U_m(t)$, $0 \leq m \leq M = M_1 + M_2$, as usual denote the approximation to the exact solution (so $U_m(t) \approx u(mh_1, t)$ for $0 \leq m \leq M_1$; $U_m(t) \approx u(c+(m-M_1)h_2, t)$ for $M_1 < m \leq M$). We form the solution vector

$$U(t) \equiv [U_1(t), U_2(t), \ldots, U_{M-1}(t)]^T,$$

and seek a matrix A so that

$$\partial_t U(t) = AU(t) + g(t).$$

For points not on the boundary $x = c$ we use the normal central difference approximation,

$$\partial_t U_m = \frac{a_1}{h_1^2}\delta_x^2 U_m, \quad 1 \leq m \leq M_1 - 1$$
$$\partial_t U_m = \frac{a_2}{h_2^2}\delta_x^2 U_m, \quad M_1 + 1 \leq m \leq M - 1.$$

The problem is $\partial_t U_{M_1}$. We must use the two conditions

$$u^1(c,t) = u^2(c,t), \ t > 0, \tag{1}$$
$$\lambda_1 u_x^1(c,t) = \lambda_2 u_x^2(c,t), \ t > 0. \tag{2}$$

In the above we used a second order approximation, so we would like to keep doing this. To achieve this we introduce two additional "dummy" points which we will subsequently eliminate. We introduce $U^*_{M_1+1}$ as an approximation to $u^1(c + h_1, t)$, i.e. the continuation of the u^1-solution into area 2. Then we have a second order approximation to $u_x^1(c,t)$ given by

$$u_x^1(c,t) = \frac{U^*_{M_1+1} - U_{M_1-1}}{2h_1} + \mathcal{O}(h_1^2).$$

Similarly, we introduce $U^*_{M_1-1} \approx u^2(c - h_2, t)$, yielding

$$u_x^2(c,t) = \frac{U_{M_1+1} - U^*_{M_1-1}}{2h_2} + \mathcal{O}(h_2^2).$$

Due to continuity, U_{M_1} should be the same from the left and right. Using central differences and our dummy points, we can now find two expressions for $\partial_t U_{M_1}$,

$$\partial_t U_{M_1} = \frac{a_1}{h_1^2}\left[U^*_{M_1+1} - 2U_{M_1} + U_{M_1-1}\right] = \frac{a_2}{h_2^2}\left[U_{M_1+1} - 2U_{M_1} + U^*_{M_1-1}\right] \tag{3}$$

These two expressions give *one* equation to decide $U^*_{M_1\pm1}$, while (2) gives another.

$$\frac{\lambda_1}{2h_1}\left[U^*_{M_1+1} - U_{M_1-1}\right] = \frac{\lambda_2}{2h_2}\left[U_{M_1+1} - U^*_{M_1-1}\right].$$

So we have two equations with two unknowns, and after some effort we arrive at

$$U^*_{M_1+1} = \frac{2a_2h_1^2\lambda_2 U_{M_1+1} + 2\lambda_2(a_1h_2^2 - a_2h_1^2)U_{M_1} + h_2(a_2h_1\lambda_1 - a_1h_2\lambda_2)U_{M_1-1}}{h_2(a_2h_1\lambda_1 + a_1h_2\lambda_2)}.$$

Inserting this into (3) we get (after some more effort),

$$\partial_t U_{M_1} = \frac{2a_1 a_2}{h_1 h_2} \frac{\lambda_2 h_1 U_{M_1+1} - (\lambda_2 h_1 + \lambda_1 h_2) U_{M_1} + \lambda_1 h_2 U_{M_1-1}}{\lambda_2 a_1 h_2 + \lambda_1 a_2 h_1}. \tag{4}$$

If all parameters are equal in area 1 and 2 we get the ordinary central difference. This completes the matrix A, with the following structure:

$$A = \begin{bmatrix} \begin{bmatrix} & A_1 & \end{bmatrix} & & \\ & [v_1 \; v_2 \; v_3] & \\ & & \begin{bmatrix} & A_2 & \end{bmatrix} \end{bmatrix} \tag{5}$$

where v_1, v_2, og v_3 is given by (4), so that $\partial_t U_{M_1} = v_1 U_{M_1-1} + v_2 U_{M_1} + v_3 U_{M_1+1}$. The matrix A_1 is the $(M_1-1) \times M_1$-matrix,

$$A_1 = \frac{a_1}{h_1^2} \begin{bmatrix} -2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & -2 & 1 & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & & & 1 & -2 & 1 \end{bmatrix},$$

while A_2 is the $(M_2-1) \times M_2$-matrix,

$$A_2 = \frac{a_2}{h_2^2} \begin{bmatrix} 1 & -2 & 1 & & & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 1 & -2 & 1 \\ 0 & \cdots & \cdots & 0 & 1 & -2 \end{bmatrix}.$$

The whole of A is tridiagonal, and if all parameters are equal for the two areas, we see that the boundary fades away, and we remain with our ordinary central difference matrix.

We then discretize time as well, so that U_m^n is an approximation to $U_m(nk)$. If we use the θ-method, we get the method

$$(1 - \theta k A) U^{n+1} = [1 + (1-\theta) k A]\, U^n + g^n, \tag{6}$$

where g^n handles the boundary conditions, and is given, in this case, by

$$g^n = \theta k \begin{bmatrix} \frac{a_1}{h_1^2} g_0((n+1)k) \\ 0 \\ \vdots \\ 0 \\ \frac{a_2}{h_2^2} g_1((n+1)k) \end{bmatrix} + (1-\theta) k \begin{bmatrix} \frac{a_1}{h_1^2} g_0(nk) \\ 0 \\ \vdots \\ 0 \\ \frac{a_2}{h_2^2} g_1(nk) \end{bmatrix}. \tag{7}$$

Problem 2. Below we give a MATLAB program to implement the method (6) with the conditions given in problem 1. The program is constructed much the same as in problem set 2. Notice how easy it is to construct the matrix A in (5) from the relevant parts. The solution is plotted in figure

1, using $\theta = 1$ (Backwards Euler). In figure 2 the solution for $t = 0.02$ is plotted, and we see that the derivative is continuous over $x = 0.5$, as it should be (try $\lambda_1 \neq \lambda_2$, to see what happens!). In figure 3 and 4 the time development for the solution in the point $x = 0.75$ is plotted, computed with $\theta = 1$ and $\theta = 0.5$ (Crank–Nicolson) respectively, and we can see quite clearly the oscillations typical for the latter method. You can try Forward Euler ($\theta = 0$) yourself using $ka_i/h_i^2 > 0.5$ (outside the stability area)—not very good

The relatively large change during the first time step in figure 3 can be attributed to the fact that the system is forced into an "uncomfortable" initial condition. Note that the way the constants a_i are involved, indicates that an optimal choice, numerically speaking, should have $h_1^2/h_2^2 = a_1/a_2$.

The MATLAB *program:*

```
a1=1; a2=100; l1=1; l2=1; c=0.5;  % Constants
M1=20; M2=20; k=0.01; N=20; t=1;  % Accuracy

h1=c/M1; h2=(1-c)/M2; M=M1+M2;

tid=[0:k:N*k];
x=[0:h1:c c+h2:h2:1];

U0=zeros(1,N);            % Homogeneous boundary cond.
UM=zeros(1,N);
Ustart=zeros(M-1,1);      % Initial condition
for i=1:M1
  Ustart(i)=4*(i*h1)*(1-i*h1);
end
for i=1:M2-1
  Ustart(M1+i)=4*(c+i*h2)*(1-(c+i*h2));
end
U0start=0;
UMstart=0;

d=ones(M,1);

% Generating the matrix A

A1=a1/h1^2*spdiags([d -2*d d],-1:1,M1-1,M1);
A2=a2/h2^2*spdiags([d -2*d d],0:2,M2-1,M2);
vv=2*a1*a2/(h1*h2*(a1*l2*h2+a2*l1*h1));
v= vv*[l1*h2 -(l2*h1+l1*h2) l2*h1] ;
A=[A1 zeros(M1-1,M2-1);
 zeros(1,M1-2) v zeros(1,M2-2);
 zeros(M2-1,M1-1) A2];
```

```
% Matrices involved in the theta-method

B=speye(M-1)-t*k*A;
C=speye(M-1)+(1-t)*k*A;

U=zeros(M-1,N);

% First step

c=C*Ustart;

% Boundary conditions:
% (could be removed, as they are homogeneous)

c(1)=c(1)+t*k*a1/h1^2*U0(1)+(1-t)*k*a2/h2^2*U0start;
c(M-1)=c(M-1)+t*k*a1/h1^2*UM(1)+(1-t)*k*a2/h2^2*UMstart;
U(:,1)=B\c;

for n=1:N-1                   % The rest of the time steps
  c=C*U(:,n);
  c(1)=c(1)+t*k*a1/h1^2*U0(n+1)+(1-t)*k*a2/h2^2*U0(n);
  c(M-1)=c(M-1)+t*k*a1/h1^2*UM(n+1)+(1-t)*k*a2/h2^2*UM(n);
  U(:,n+1)=B\c;
end

% Plotting

Utot=[[U0start U0];[Ustart U];[UMstart UM]];
mesh(tid,x,Utot);
view(130,50);
```

Problem 3. Choose a k so that

$$|U_k| = \max_{0 \le m \le M} |U_m|.$$

Then, by the triangle inequality ($r > 0$)

$$|v_k| = |(1+2r)U_k - r(U_{k-1} + U_{k+1})| \ge \Big| |(1+2r)U_k| - r\,|U_{k-1} + U_{k+1}| \Big| \ge |(1+2r-2r)U_k| = |U_k|,$$

since $|U_{k\pm1}| \le |U_k|$ giving $|U_{k-1} + U_{k+1}| \le 2|U_k|$. Thus $|v_k|$ is greater than all $|U_i|$, and we have

$$\max_{0 \le m \le M} |U_m| \le \max_{1 \le m \le M-1} |v_m| \tag{8}$$

Backwards Euler is given by

$$(1 - r\delta_x^2)U_m^{n+1} = U_m,$$

with truncation error

$$(1 - r\delta_x^2)u_m^{n+1} - u_m^n = \mathcal{O}(k^2 + kh^2).$$

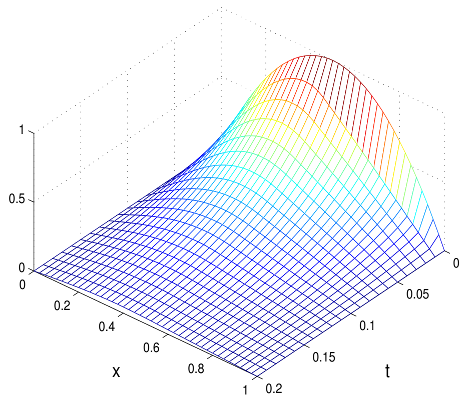

Figure 1: Numerical solution wrt. x and t.

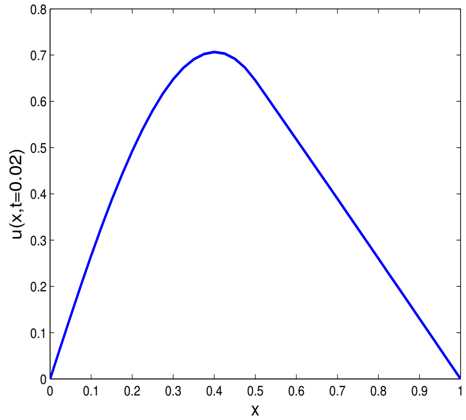

Figure 2: A cross section for $t = 0.02$.

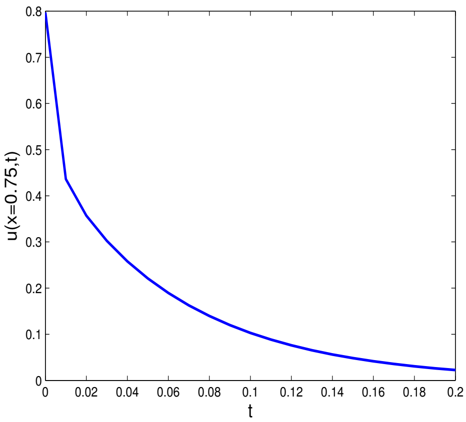

Figure 3: A cross section for $x = 0.75$.

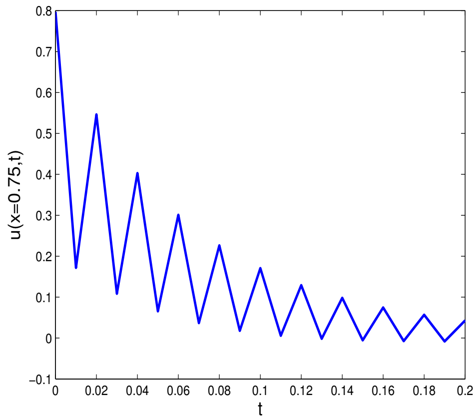

Figure 4: A cross section for $x = 0.75$ with Crank–Nicolson.

The discretization error $z_m^n = u_m^n - U_m^n$ then satisfies

$$(1 - r\delta_x^2) z_m^{n+1} = z_m^n + \mathcal{O}(k^2 + kh^2). \tag{9}$$

If we introduce $Z^n = \max_{0 \le m \le M} |z_m^n|$, it follows from (8) and (9) that

$$Z^{n+1} \le Z^n + A(k^2 + kh^2)$$

for a constant A (when k and h are small). By repeated application of this inequality, as well as $Z^0 = 0$ ($U_m^0 = u_m^0$ using the initial condition), we get

$$Z^n \le Z^{n-1} + A(k^2 + kh^2) \le Z^{n-2} + 2A(k^2 + kh^2) \le \cdots \le Z^0 + nkA(k + h^2) = tA(k + h^2).$$

So we see that the error $Z^n \to 0$ when $k, h \to 0$ so that the time $t = nk$ is held constant. In other words, Backwards Euler converges for arbitrary $r > 0$.

Problem 4. We apply Euler's method on the equation $u_t = u_{xx} + \gamma u$. By introducing central differences, we get

$$\frac{1}{k}\left(U_m^{n+1} - U_m^n\right) = \frac{1}{h^2}\delta_x^2 U_m^n + \gamma U_m^n, \text{ i.e. } U_m^{n+1} = \left(1 + \gamma k + r\delta_x^2\right) U_m^n,$$

with $r = k/h^2$. In matrix form, this becomes

$$
\begin{aligned}
U^{n+1} &= AU^n + g^n, \\
A &= \begin{bmatrix}
1+\gamma k-2r & r & 0 & \cdots & \cdots & 0 \\
r & 1+\gamma k-2r & r & & & \vdots \\
0 & \ddots & \ddots & \ddots & & \vdots \\
\vdots & & \ddots & \ddots & \ddots & 0 \\
\vdots & & & r & 1+\gamma k-2r & r \\
0 & \cdots & \cdots & 0 & r & 1+\gamma k-2r
\end{bmatrix}.
\end{aligned}
$$

A necessary condition for stability is $\rho(A) \leq 1 + \mu k$ (where μ is independent k and h). Since A is symmetrical, this will also be sufficient. We know the eigenvalues of A (see the note), and they are given as

$$\lambda_m = 1 + \gamma k - 2r + 2r\cos\left(\frac{\pi m}{M}\right) = 1 + \gamma k - 4r\sin^2\left(\frac{\pi m}{2M}\right), \qquad m = 1, 2, \ldots, M-1.$$

Since

$$s^2 \equiv \sin^2\left(\frac{\pi}{2M}\right) \leq \sin^2\left(\frac{\pi m}{2M}\right) \leq \sin^2\left(\frac{\pi(M-1)}{2M}\right) = \cos^2\left(\frac{\pi}{2M}\right) \equiv c^2,$$

we have

$$\rho(A) = \max_{1\leq m\leq M-1}\left|1 + \gamma k - 4r\sin^2\left(\frac{\pi m}{2M}\right)\right| = \max\left\{|1+\gamma k - 4rs^2|, |1+\gamma k - 4rc^2|\right\}.$$

We define $F = 1+\gamma k-4rs^2$ and $G = 1+\gamma k-4rc^2$ and demand that $|F| \leq 1+\mu k$ and $|G| \leq 1+\mu k$. As h gets small, we get $s = \sin(\pi h/2) \to \pi h/2$ and $c \to 1$. Thus we have $F > 0$ small enough k, and we require

$$
\begin{aligned}
F &\leq 1+\mu k \;\Rightarrow\; \gamma \leq \mu + 4s^2/h^2 \;\Rightarrow\; \text{Satisfied if } \mu = \mu_1 = \max\{0, \gamma - \pi^2\} \text{ (for small } h\text{)}, \\
G &\leq 1+\mu k \;\Rightarrow\; \gamma - 4c^2/h^2 \leq \mu \;\Rightarrow\; \text{OK for } \mu = \mu_1, \\
G &\leq -1-\mu k \;\Rightarrow\; \left[4c^2/h^2 - (\mu+\gamma)\right]k \leq 2 \;\Rightarrow\; k \leq \frac{2}{4c^2/h^2 - (\mu+\gamma)} = \frac{1}{2}\,\frac{h^2}{c^2-(\mu+\gamma)h^2/4}
\end{aligned}
$$

In the last transition we divide by something positive on both sides of the inequality, as $c/h \to \infty$ when $h \to 0$. So we can give the following stability requirements:

$$k \leq \begin{cases} \dfrac{1}{2}\,\dfrac{h^2}{c^2 - (2\gamma - \pi^2)h^2/4} & \text{for } \gamma > \pi^2 \\[2ex] \dfrac{1}{2}\,\dfrac{h^2}{c^2 - \gamma h^2/4} & \text{for } \gamma \leq \pi^2 \end{cases}$$

We see that when $h \to 0$, these reduce to $k/h^2 \leq \frac{1}{2}$.