The heat equation

June 7, 2020

We consider the *heat equation* on an open spatial domain $\Omega \subseteq \mathbb{R}^d$ ($d \in \{1, 2, 3\}$) for a time interval $[0, T]$:

$$\frac{\partial u}{\partial t}(x,t) = c\Delta u(x,t) + f(x,t),\ (x,t) \in \Omega \times [0,T],$$

where $c > 0$ and $f : \Omega \times [0, T] \to \mathbb{R}$ are given and $u : \overline{\Omega} \times [0, T] \to \mathbb{R}$ is the unknown.

Here $u(x,t)$ can be interpreted as the temperature at the point $x \in \Omega$ and at the time $t \in [0, T]$. In this context, the constant c depends on conductivity, specific heat and density of the material in Ω and the function f takes into account sources and sinks of heat.

To obtain a problem with a unique solution, we need:

- a *boundary condition*: we consider the homogeneous boundary condition

  $$u(x,t) = 0,\ (x,t) \in \Gamma \times [0,T],$$

  where Γ is the boundary of Ω;

- an *initial condition*

  $$u(x,0) = u_0(x),\ x \in \Omega,$$

  where $u_0 : \Omega \to \mathbb{R}$.

1 The semi-discretization in space

Let us suppose that $\Omega = (0,1)^2$ is the unit square in $\mathbb{R}^2$.

In this case, by introducing the mesh $\mathbb{R}^2_h$, $h = \frac{1}{N}$ with N positive integer, and then Ω_h, Γ_h and $\overline{\Omega}_h$ as discretizations of Ω, Γ and $\overline{\Omega}$, respectively, the Laplacian Δ can be discretized by the five-point discretization Δ_h.

Thus, in the discrete problem, we look for a function $u_h : \overline{\Omega}_h \times [0, T] \to \mathbb{R}$ such that

$$\begin{aligned}
&\frac{\partial u_h}{\partial t}(x,t) = c\Delta_h u_h(x,t) + f(x,t),\ (x,t) \in \Omega_h \times [0,T],\\
&u_h(x,t) = 0,\ (x,t) \in \Gamma_h \times [0,T],\\
&u_h(x,0) = u_0(x),\ x \in \Omega_h.
\end{aligned}$$

This is an ODE. In particular, it is a system of $(N-1)^2$ scalar ODEs in $(N-1)^2$ unknown scalar functions: there is a scalar ODE for each point in Ω_h and there is an unknown scalar function

$$u_h(x,\cdot):[0,T]\to\mathbb{R}$$

for each point $x\in\Omega_h$ (recall that u_h is known in Γ_h).

Observe that a space point in the Poisson equation was denoted by (x,y) and, now, it is denoted by $x=(x_1,x_2)$.

The process of reducing an evolutionary PDE, like the heat equation, to a system of scalar ODEs by using a finite difference approximation of the spatial operator, as when the five-point discretization is used for approximating the Laplacian, is called the *semi-discretization in space* or *the method of lines.* The lines are the unknown functions of time

$$u_h(x,\cdot):[0,T]\to\mathbb{R},\ x\in\Omega_h.$$

We have a line for each point in Ω_h.

The method of lines is not a full discretization, since we still have to choose a numerical method for solving in time the system of scalar ODEs. In principle, any of the methods for the numerical solution of ODEs could be used to obtain a full discretization. We shall investigate some of the simplest possibilities, namely:

- the forward Euler method,
- the backward Euler method,
- the trapezoidal rule.

For simplicity, we consider one space dimension, where

$$\Omega=(0,1)\,,\Omega_h=\{h,2h,\ldots,(N-1)h\},\ \text{and}\ \ \Gamma_h=\{0,1\}.$$

To consider the one-dimensional case is simply a notational convenience. The analysis in two space dimensions is very similar.

In the one-dimensional case, the systems of scalar ODEs of the method of lines is

$$\begin{aligned}&\frac{\partial u_h}{\partial t}(x,t)=c\Delta_h u_h(x,t)+f(x,t)\,,\ (x,t)\in\Omega_h\times[0,T]\,,\\&u_h(0,t)=u_h(1,t)=0\\&u_h(x,0)=u_0(x)\,,\ x\in\Omega_h,\end{aligned}\tag{1}$$

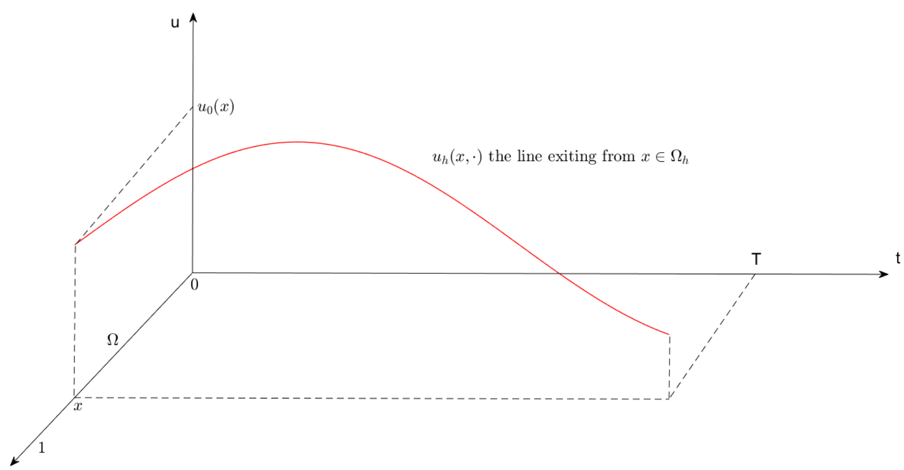
where

$$\Delta_h u_h(x,t)=\frac{u_h(x-h,t)-2u_h(x,t)+u_h(x+h,t)}{h^2}.$$

The ODE (1) is a system of $N-1$ scalar ODEs in the $N-1$ unknown scalar functions: there is a scalar ODE for each point in Ω_h and the unknown scalar functions are the lines

$$u_h(x,\cdot) : [0,T] \to \mathbb{R},\ x \in \Omega_h.$$

(see Figure 1).



2 The centered difference/forward difference method

We begin by considering a full discretization of the heat equation that corresponds to the forward Euler method for solving the system (1) of the method of lines.

Let $h = \frac{1}{N}$ be the spatial stepsize, N positive integer, and let $k = \frac{T}{M}$ be the time stepsize, M positive integer.

By setting

$$U_n^m := u_h(nh, mk), \quad n \in \{0, \ldots, N\} \text{ and } m \in \{0, \ldots, M\},$$

and

$$f_n^m := f(nh, mk),\ n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M\},$$

we consider

$$\begin{aligned}
&\frac{U_n^{m+1} - U_n^m}{k} = c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m \\
&n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\} \\
&U_0^m = 0 \text{ and } U_N^m = 0,\quad m \in \{0, \ldots, M\}, \\
&U_n^0 = u_0(nh),\ n \in \{1, \ldots, N-1\},
\end{aligned}$$

as a full discretization of the heat equation.

We call this full discretization *the centered difference/forward difference method* for the heat equation since, in

$$\frac{\partial u}{\partial t}(nh, mk) = c\Delta u(nh, mk) + f(nk, mk) = c\frac{\partial^2 u}{\partial x^2}(nh, mk) + f(nk, mk),$$

we are approximating the spatial derivative $\frac{\partial^2 u}{\partial x^2}(nh, mk)$ by the centered difference

$$\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2}$$

and the time derivative $\frac{\partial u}{\partial t}(nh, mk)$ by the forward difference

$$\frac{U_n^{m+1} - U_n^m}{k}.$$

In the following, we use the compact notation

$$U^m := \left(U_1^m, \ldots, U_{N-1}^m\right) \text{ and } f^m := \left(f_1^m, \ldots, f_{N-1}^m\right), \ m \in \{0, \ldots, M\}.$$

By writing the full discretization of the heat equation

$$\begin{gathered}\frac{U_n^{m+1} - U_n^m}{k} = c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m \\ n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\}\end{gathered}$$

as

$$\begin{gathered}U_n^{m+1} = U_n^m + k\left(c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m\right) \\ n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\}\end{gathered}$$

and then in compact form as

$$U^{m+1} = U^m + k\left(c\Delta_h U^m + f^m\right), \ m \in \{0, \ldots, M-1\},$$

where

$$\left(\Delta_h U^m\right)_n = \frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2}, \ n \in \{1, \ldots, N-1\},$$

we see that the centered difference/forward difference method is exactly the forward Euler method as applied to the system (1) of the method of lines: the forward Euler method for

$$y'(t) = F(t, y(t)) = c\Delta_h y(t) + (f(h, t), \ldots, f((N-1)h, t))$$

is

$$U^{m+1} = U^m + kF(mk, U^m) = U^m + k\left(c\Delta_h U^m + f^m\right).$$

Since the forward Euler method is an explicit method for ODEs, not any linear equation has to be solved to obtain U^{m+1} from U^m, $m \in \{0, \dots, M-1\}$. Indeed, the components of U^{m+1} can be obtained by the components of U^m as

$$\begin{aligned} U_n^{m+1} &= U_n^m + k\left(c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m\right) \\ &= \frac{ck}{h^2}U_{n-1}^m + \left(1 - 2\frac{ck}{h^2}\right)U_n^m + \frac{ck}{h^2}U_{n+1}^m + kf_n^m \\ &= \lambda U_{n-1}^m + (1-2\lambda)\,U_n^m + \lambda U_{n+1}^m + kf_n^m \end{aligned}$$

$$n \in \{1, \dots, N-1\},$$

where

$$\lambda := \frac{ck}{h^2}.$$

Observe that we start with

$$U^0 = \left(U_1^0, \dots, U_{N-1}^0\right) = \left(u_0(h), \dots, u_0\left((N-1)\,h\right)\right).$$

Exercise. In the physical interpretation of the heat equation, what are the dimensions of λ?

2.1 A numerical test

Consider the particular instance of the heat equation, where

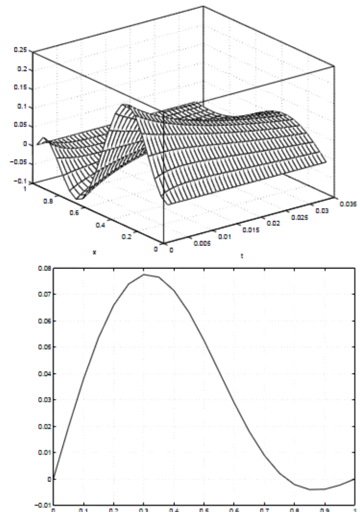
$$\begin{aligned} &c = 1 \\ &u_0\,(x) = \left(x - x^2\right)\left(x^2 + \sin 2\pi x\right),\ x \in (0,1)\,, \\ &f(x,t) = 0,\ (x,t) \in (0,1) \times [0,T], \\ &T = \frac{1}{30}. \end{aligned}$$

We use the centered difference/forward difference method with

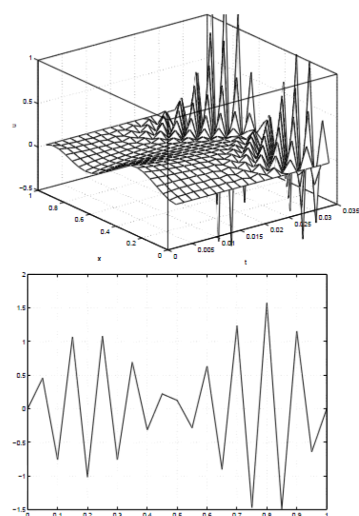
- $h = \frac{1}{20}$ $(N = 20)$ and $k = \frac{1}{1200}$ $(M = 40)$
- $h = \frac{1}{20}$ $(N = 20)$ and $k = \frac{1}{600}$ $(M = 20)$.

Since there are no sources of heat, the solution $u(x,t)$ is expected to become zero at any point in Ω as $t \to +\infty$, i.e. in the long-time the temperature inside of the body Ω becomes constant and equal to the temperature at the boundary.

The first computation with $h = \frac{1}{20}$ and $k = \frac{1}{1200}$ gives very reasonable results (and we could have extended it for a much longer time without problem). The second plot shows $U^M \approx u(\cdot, T)$.



The second computation with $h = \frac{1}{20}$ and $k = \frac{1}{600}$ becomes unreasonable after a few time steps. The second plot shows $U^M \approx u(\cdot, T)$.



Experimentation with this method shows that the good behavior is controlled by the value of $\lambda = \frac{ck}{h^2}$.

In the first computation, we have

$$\lambda = \frac{ck}{h^2} = \frac{\frac{1}{1200}}{\frac{1}{400}} = \frac{1}{3}$$

and in the second, we have

$$\lambda = \frac{ck}{h^2} = \frac{\frac{1}{600}}{\frac{1}{400}} = \frac{2}{3}.$$

Indeed, for $\lambda \leq \frac{1}{2}$ the computation proceeds reasonably, but for $\lambda > \frac{1}{2}$, the computed solution U^m, $m \in \{0, \ldots, M\}$, becomes oscillatory with an amplitude that grows exponentially as m increases.

Now, we try to explain analytically this fact.

2.2 Error analysis

Let u be the exact solution of the heat equation which is assumed sufficiently smooth. Let

$$u_n^m := u\left(nh, mk\right), \quad n \in \{0, \ldots, N\} \text{ and } m \in \{0, \ldots, M\}.$$

These values u_n^m should be compared with the corresponding values

$$U_n^m = u_h\left(nh, mk\right), \quad n \in \{0, \ldots, N\} \text{ and } m \in \{0, \ldots, M\},$$

introduced for the discrete solution u_h.

We introduce the *consistency error*

$$\varepsilon_n^{m+1} := \frac{u_n^{m+1} - u_n^m}{k} - c\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - f_n^m$$
$$n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\},$$

Let $n \in \{1, \ldots, N-1\}$ and $m \in \{0, \ldots, M-1\}$. We have previously seen that the centered difference scheme satisfies

$$\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - \frac{\partial^2 u}{\partial x^2}\left(nh, mk\right) = \frac{1}{24}\left(\frac{\partial^4 u}{\partial x^4}\left(\alpha, mk\right) + \frac{\partial^4 u}{\partial x^4}\left(\beta, mk\right)\right)h^2,$$

where $\alpha \in \left(\left(n-1\right)h, nh\right)$ and $\beta \in \left(nh, \left(n+1\right)h\right)$. So, we have

$$\left|\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - \frac{\partial^2 u}{\partial x^2}\left(nh, mk\right)\right| \leq \frac{1}{12}\max_{(x,t)\in\overline{\Omega}\times[0,T]}\left|\frac{\partial^4 u}{\partial x^4}\left(x, t\right)\right| h^2.$$

Exercise. Given a sufficiently smooth function $v(t)$ of one real variable t and $k > 0$, a finite difference approximating the first derivative $v'\left(t\right)$ is the *forward difference*

$$v'\left(t\right) \approx \frac{v\left(t+k\right) - v\left(t\right)}{k}.$$

Prove that

$$\frac{v\left(t+k\right) - v\left(t\right)}{k} = v'\left(t\right) + \frac{1}{2}v''\left(\gamma\right)k,$$

where $\gamma \in (t, t+k)$. Then, prove that

$$\left| \frac{u_n^{m+1} - u_n^m}{k} - \frac{\partial u}{\partial t}(nh, mk) \right| \leq \frac{1}{2} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| k.$$

Now, since

$$\begin{aligned}
\varepsilon_n^{m+1} &= \frac{u_n^{m+1} - u_n^m}{k} - c\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - f_n^m \\
&= \frac{u_n^{m+1} - u_n^m}{k} - \frac{\partial u}{\partial t}(nh, mk) \\
&\quad -c\left(\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - \frac{\partial^2 u}{\partial x^2}(nh, mk)\right) \\
&\quad + \underbrace{\frac{\partial u}{\partial t}(nh, mk) - c\frac{\partial^2 u}{\partial x^2}(nh, mk) - f_n^m}_{=0},
\end{aligned}$$

we have

$$\left|\varepsilon_n^{m+1}\right| \leq \frac{c}{12} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^4 u}{\partial x^4}(x,t) \right| h^2 + \frac{1}{2} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| k.$$

So we have obtained the *consistency result* for the centered difference/forward difference method:

$$\max_{\substack{n \in \{1,\ldots,N-1\} \\ m \in \{1,\ldots,M\}}} \left|\varepsilon_n^m\right| \leq \frac{c}{12} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^4 u}{\partial x^4}(x,t) \right| h^2 + \frac{1}{2} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| k. \tag{2}$$

Now, we introduce the *convergence error*

$$e_n^m := U_n^m - u_n^m, \; n \in \{0, \ldots, N\} \text{ and } m \in \{0, \ldots, M\}.$$

Since

$$\begin{aligned}
&\frac{U_n^{m+1} - U_n^m}{k} = c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m \\
&n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\} \\
&U_0^m = 0 \text{ and } U_N^m = 0, \quad m \in \{0, \ldots, M\}, \\
&U_n^0 = u_0(nh), \; n \in \{1, \ldots, N-1\},
\end{aligned}$$

$$\begin{aligned}
&\frac{u_n^{m+1} - u_n^m}{k} = c\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} + f_n^m + \varepsilon_n^{m+1} \\
&n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\} \\
&u_0^m = 0 \text{ and } u_N^m = 0, \quad m \in \{0, \ldots, M\}, \\
&u_n^0 = u_0(nh), \; n \in \{1, \ldots, N-1\},
\end{aligned}$$

by subtracting we see that the convergence error is the solution of the full discrete problem

$$\begin{aligned}
&\frac{e_n^{m+1}-e_n^m}{k} = c\frac{e_{n-1}^m - 2e_n^m + e_{n+1}^m}{h^2} - \varepsilon_n^{m+1},\\
&n \in \{1,\ldots,N-1\} \text{ and } m \in \{0,\ldots,M-1\},\\
&e_n^m = 0,\ n \in \{0,N\} \text{ and } m \in \{0,\ldots,M\},\\
&e_n^0 = 0,\ n \in \{1,\ldots,N-1\}.
\end{aligned}$$

where the function f is the opposite of the consistency error and zero initial condition holds.

The next one is a *stability result* for the centered difference/forward difference method.

Theorem 1 *If* $\lambda \leq \frac{1}{2}$, *then*

$$\max_{\substack{n\in\{1,\ldots,N-1\}\\ m\in\{0,\ldots,M\}}} |e_n^m| \leq T \max_{\substack{n\in\{1,\ldots,N-1\}\\ m\in\{1,\ldots,M\}}} |\varepsilon_n^m|$$

Proof. Suppose $\lambda \leq \frac{1}{2}$. Since

$$\begin{aligned}
&\frac{e_n^{m+1}-e_n^m}{k} = c\frac{e_{n-1}^m - 2e_n^m + e_{n+1}^m}{h^2} - \varepsilon_n^{m+1}\\
&n \in \{1,\ldots,N-1\} \text{ and } m \in \{0,\ldots,M-1\},
\end{aligned}$$

we obtain

$$\begin{aligned}
e_n^{m+1} &= \frac{ck}{h^2}e_{n-1}^m + \left(1 - 2\frac{ck}{h^2}\right)e_n^m + \frac{ck}{h^2}e_{n+1}^m - k\varepsilon_n^{m+1}\\
&= \lambda e_{n-1}^m + (1-2\lambda)\,e_n^m + \lambda e_{n+1}^m - k\varepsilon_n^{m+1}\\
n &\in \{1,\ldots,N-1\} \text{ and } m \in \{0,\ldots,M-1\}
\end{aligned}$$

and then

$$\begin{aligned}
&\left|e_n^{m+1}\right| \leq \lambda\left|e_{n-1}^m\right| + (1-2\lambda)\left|e_n^m\right| + \lambda\left|e_{n+1}^m\right| + k\left|\varepsilon_n^{m+1}\right|\\
&n \in \{1,\ldots,N-1\} \text{ and } m \in \{0,\ldots,M-1\}.
\end{aligned} \tag{3}$$

Here we are using that fact that $\lambda \leq \frac{1}{2}$.

By introducing the vectors

$$e^m := \left(e_1^m,\ldots,e_{N-1}^m\right),\ m \in \{0,\ldots,M\},$$

and the vectors

$$\varepsilon^{m+1} := \left(\varepsilon_1^{m+1},\ldots,\varepsilon_{N-1}^{m+1}\right),\ m \in \{0,\ldots,M-1\},$$

by (3) we obtain, since $e_0^m = e_N^m = 0$,

$$\begin{aligned}
&\|e^{m+1}\|_{L^\infty(\Omega_h)} \\
&\leq \lambda\|e^m\|_{L^\infty(\Omega_h)} + (1-2\lambda)\,\|e^m\|_{L^\infty(\Omega_h)} + \lambda\|e^m\|_{L^\infty(\Omega_h)} + k\|\varepsilon^{m+1}\|_{L^\infty(\Omega_h)} \\
&= \|e^m\|_{L^\infty(\Omega_h)} + k\|\varepsilon^{m+1}\|_{L^\infty(\Omega_h)},\ m \in \{0,\ldots,M-1\}\,.
\end{aligned}$$

Since $e^0 = 0$, we have

$$\begin{aligned}
&\|e^m\|_{L^\infty(\Omega_h)} \leq \|e^0\|_{L^\infty(\Omega_h)} + k\|\varepsilon^1\|_{L^\infty(\Omega_h)} + \cdots + k\|\varepsilon^m\|_{L^\infty(\Omega_h)} \\
&= k(\|\varepsilon^1\|_{L^\infty(\Omega_h)} + \cdots + \|\varepsilon^m\|_{L^\infty(\Omega_h)}),\ m \in \{0,\ldots,M\}\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
\max_{\substack{n\in\{1,\ldots,N-1\}\\ m\in\{0,\ldots,M\}}} |e_n^m| &= \max_{m\in\{0,\ldots,M\}} \|e^m\|_{L^\infty(\Omega_h)} \\
&\leq kM \max_{m\in\{1,\ldots,M\}} \|\varepsilon^m\|_{L^\infty(\Omega_h)} \\
&= T \max_{\substack{n\in\{1,\ldots,N-1\}\\ m\in\{1,\ldots,M\}}} |\varepsilon_n^m|.
\end{aligned}$$

■

As in case of the Poisson equation,

consistency and stability $\Rightarrow$ convergence

and we have the *convergence result* for the centered difference/forward difference method: if $\lambda \leq \frac{1}{2}$, then

$$\begin{aligned}
&\max_{\substack{n\in\{1,\ldots,N-1\}\\ m\in\{0,\ldots,M\}}} |e_n^m| \\
&\leq T\left(\frac{c}{12}\max_{(x,t)\in\overline{\Omega}\times[0,T]}\left|\frac{\partial^4 u}{\partial x^4}(x,t)\right| h^2 + \frac{1}{2}\max_{(x,t)\in\overline{\Omega}\times[0,T]}\left|\frac{\partial^2 u}{\partial t^2}(x,t)\right| k\right).
\end{aligned} \tag{4}$$

The method is convergent of order two with respect to h and of order one with respect to k.

Since we have stability and then convergence as long as the condition

$$\lambda = \frac{ch^2}{k} \leq \frac{1}{2}, \text{ i.e. } k \leq \frac{h^2}{2c},$$

is satisfied, the centered difference/forward difference method is said *conditionally stable* and *conditionally convergent*. With the centered difference /forward difference method, we are not free to choose independently h and k.

2.3 Fourier analysis

Up to now, we have given our results by using for the space discretization the norm L^∞ on $L(\Omega_h)$.

Another very useful way to analyze stability and convergence of the centered difference/forward difference method is to use the Fourier analysis, as we have done for the Poisson equation. In this analysis, we use the discrete L^2 norm $\|\cdot\|_h$ on $L(\Omega_h)$. Recall that functions of $L(\Omega_h)$ are extended to $\overline{\Omega}_h$ giving zero values in Γ_h.

Consider the discrete problem relating convergence error and consistency error:

$$\begin{aligned}
&\frac{e_n^{m+1}-e_n^m}{k} = c\frac{e_{n-1}^m - 2e_n^m + e_{n+1}^m}{h^2} - \varepsilon_n^{m+1},\\
&\text{i. e. } e_n^{m+1} = e_n^m + kc\frac{e_{n-1}^m - 2e_n^m + e_{n+1}^m}{h^2} - k\varepsilon_n^{m+1},\\
&n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M-1\},\\
&e_n^m = 0,\ n \in \{0,N\} \text{ and } m \in \{0,\dots,M\}\\
&e_n^0 = 0,\ n \in \{1,\dots,N-1\}.
\end{aligned}$$

By using the vectors

$$e^m = \left(e_1^m,\dots,e_{N-1}^m\right),\ m \in \{0,\dots,M\},$$

and

$$\varepsilon^{m+1} = \left(\varepsilon_1^{m+1},\dots,\varepsilon_{N-1}^{m+1}\right),\ m \in \{0,\dots,M-1\},$$

we obtain the recursive equation

$$e^{m+1} = e^m + kc\Delta_h e^m - k\varepsilon^{m+1} = R(kc\Delta_h)e^m - k\varepsilon^{m+1},\ m \in \{0,\dots,M-1\}, \tag{5}$$

where $\Delta_h : L(\Omega_h) \to L(\Omega_h)$ is the discrete Laplacian and $R(kc\Delta_h) : L(\Omega_h) \to L(\Omega_h)$ is given by

$$R(kc\Delta_h) = I + kc\Delta_h.$$

Since the linear operator Δ_h (which is a $(N-1)\times(N-1)$ matrix) has eigenvalues

$$\lambda_{n,h} = -\frac{4\sin^2\left(\frac{\pi n h}{2}\right)}{h^2},\ n \in \{1,\dots,N-1\},$$

with relevant eigenvectors the discrete functions $\phi_{n,h}$, $n \in \{1,\dots,N-1\}$, given by

$$\phi_{n,h}(x) = \sin(\pi n x),\ x \in \Omega_h,$$

the linear operator $R(kc\Delta_h)$ has eigenvalues

$$R(kc\lambda_{n,h}) = 1 + kc\lambda_{n,h}$$

with relevant eigenvectors $\phi_{n,h}$, $n \in \{1,\dots,N-1\}$.

Now, we introduce the following lemma.

Lemma 2 *Let S be an analytic function whose domain contains the eigenvalues $\lambda_{n,h}$, $n \in \{1,\ldots,N-1\}$, of Δ_h. For the linear operator $S(\Delta_h) : L(\Omega_h) \to L(\Omega_h)$, whose eigenvalues are $S(\lambda_{n,h})$ with relevant eigenvectors $\phi_{n,h}$, $n \in \{1,\ldots,N-1\}$, we have*

$$\|S(\Delta_h)\| = \rho(S(\Delta_h)),$$

where $\|S(\Delta_h)\|$ is the operator norm of $S(\Delta_h)$ relevant to the norm $\|\cdot\|_h$ and

$$\rho(S(\Delta_h)) = \max_{n\in\{1,\ldots,N-1\}} |S(\lambda_{n,h})|$$

is the spectral radius of $S(\Delta_h)$.

Proof. Given $f \in L(\Omega_h)$ with discrete Fourier series

$$f = \sum_{n=1}^{N-1} a_n \phi_{n,h},$$

we can easily compute the discrete Fourier series of $S(\Delta_h) f$: we have

$$\begin{aligned} S(\Delta_h) f &= S(\Delta_h)\left(\sum_{n=1}^{N-1} a_n\phi_{n,h}\right) = \sum_{n=1}^{N-1} a_n S(\Delta_h)\phi_{n,h} \\ &= \sum_{n=1}^{N-1} a_n S(\lambda_{n,h})\phi_{n,h}. \end{aligned}$$

By the discrete Parseval's identity, we conclude that

$$\begin{aligned} &\|S(\Delta_h) f\|_h^2 = \sum_{n=1}^{N-1} |a_n S(\lambda_{n,h})|^2 \|\phi_{n,k}\|_h^2 \\ &= \sum_{n=1}^{N-1} |a_n|^2 |S(\lambda_{n,h})|^2 \|\phi_{n,k}\|_h^2 \le \rho(S(\Delta_h))^2 \sum_{n=1}^{N-1} |a_n|^2 \|\phi_{n,k}\|_h^2 \\ &= \rho(S(\Delta_h))^2 \|f\|_h^2 \end{aligned}$$

and then

$$\|S(\Delta_h) f\|_h \le \rho(S(\Delta_h)) \|f\|_h.$$

Moreover, for $f = \phi_{\overline{n},h}$, where $\overline{n} \in \{1,\ldots,N-1\}$ is such that

$$\rho(S(\Delta_h)) = |S(\lambda_{\overline{n},h})|,$$

By the Parseval's identity it is clear that

$$\|S(\Delta_h) f\|_h = \rho(S(\Delta_h)) \|f\|_h.$$

We conclude that

$$\|S(\Delta_h)\| = \max_{f\in L(\Omega_h)\setminus\{0\}} \frac{\|S(\Delta_h) f\|_h}{\|f\|_h} = \rho(S(\Delta_h)).$$

■

Exercise. Give another proof of the previous lemma by taking into account that Δ_h is a symmetric matrix and then $S(\Delta_h)$ is a symmetric matrix. Use the fact that the spectral norm (i.e. the 2-norm) of a symmetric matrix is the spectral radius.

By the recursive equation (5), we obtain

$$e^m = k \sum_{i=1}^{m} R\left(kc\Delta_h\right)^{m-i}\left(-\varepsilon^i\right),\ m \in \{0, \ldots, M\}. \tag{6}$$

Exercise. Prove that (6) is the solution of (5).

Now, we are ready to give a stability result for the centered difference/forward difference method.

Theorem 3 *If*

$$\rho\left(R\left(kc\Delta_h\right)\right) \leq 1,$$

then

$$\max_{m\in\{0,\ldots,M\}} \|e^m\|_h \leq T \max_{i\in\{1,\ldots,M\}} \|\varepsilon^i\|_h.$$

Proof. By using the norm $\|\cdot\|_h$ on $L\left(\Omega_h\right)$ in (6), we have, for $m \in \{0, \ldots, M\}$,

$$\begin{aligned}
\|e^m\|_h &= \left\| k \sum_{i=1}^{m} R\left(kc\Delta_h\right)^{m-i}\left(-\varepsilon^i\right)\right\|_h \leq k \sum_{i=1}^{m} \left\| R\left(kc\Delta_h\right)^{m-i}\left(-\varepsilon\right)^i\right)\right\|_h \\
&\leq k \sum_{i=1}^{m} \|R\left(kc\Delta_h\right)\|^{m-i} \|\varepsilon^i\|_h.
\end{aligned}$$

Under the assumption $\rho\left(R\left(kc\Delta_h\right)\right) \leq 1$, the previous Lemma 2, as applied to the function $S(z) = R(kcz)$, says that

$$\|R\left(kc\Delta_h\right)\| \leq 1$$

and then

$$\|e^m\|_h \leq k \sum_{i=1}^{m} \|\varepsilon^i\|_h \leq km \max_{i\in\{1,\ldots,m\}} \|\varepsilon^i\|_h.$$

Now,

$$\max_{m\in\{0,\ldots,M\}} \|e^m\|_h \leq \underbrace{kM}_{=T} \max_{i\in\{1,\ldots,M\}} \|\varepsilon^i\|_h$$

follows. ■

The condition

$$\rho\left(R\left(kc\Delta_h\right)\right) \leq 1$$

in the previous theorem means

$$|R(kc\lambda_{n,h})| = |1 + kc\lambda_{n,h}| = \left|1 - \frac{4kc}{h^2}\sin^2\left(\frac{\pi nh}{2}\right)\right| \leq 1$$
for all $n \in \{1, \ldots, N-1\}$

i.e.

$$-1 \leq 1 - \frac{4kc}{h^2}\sin^2\left(\frac{\pi nh}{2}\right) \leq 1$$
for all $n \in \{1, \ldots, N-1\}$

i.e.

$$\frac{4kc}{h^2}\sin^2\left(\frac{\pi nh}{2}\right) \leq 2$$
for all $n \in \{1, \ldots, N-1\}$.

This holds for all positive integer N if and only if

$$\frac{4kc}{h^2} \leq 2 \text{ i.e. } \lambda = \frac{kc}{h^2} \leq \frac{1}{2}.$$

So the centered difference/forward difference method is stable if $\lambda \leq \frac{1}{2}$. We have seen that the same result holds for the L^∞ norm on $L(\Omega_h)$.

Next theorem explores the situation $\lambda > \frac{1}{2}$.

Theorem 4 *Let $\beta > \frac{1}{2}$. Consider discretizations with $\lambda \geq \beta$. There exists a consistency error $\left(\varepsilon^1, \ldots, \varepsilon^M\right)$ such that*

$$\lim_{M,N\to\infty} \frac{\max\limits_{m\in\{0,\ldots,M\}} \|e^m\|_h}{\max\limits_{i\in\{1,\ldots,M\}} \|\varepsilon^i\|_h} = +\infty.$$

Proof. We have, for $n \in \{1, \ldots, N-1\}$,

$$R(kc\lambda_{n,h}) = 1 - \frac{4kc}{h^2}\sin^2\left(\frac{\pi nh}{2}\right) = 1 - 4\lambda\sin^2\left(\frac{\pi nh}{2}\right) \leq 1 - 4\beta\sin^2\left(\frac{\pi nh}{2}\right).$$

Thus, fixed $\gamma > \frac{1}{2}$ such that $\gamma < \beta$, for N sufficiently large, we obtain

$$\beta\sin^2\left(\frac{\pi nh}{2}\right) \geq \gamma$$

for some $n \in \{1, \ldots, N-1\}$ and so

$$R(kc\lambda_{n,h}) \leq 1 - 4\gamma < -1.$$

Then, for N sufficiently large, we have

$$\rho\left(R\left(kc\Delta_h\right)\right) \geq 4\gamma - 1 > 1.$$

Observe that

$$\rho\left(R\left(kc\Delta_h\right)\right) = \left|R\left(kc\lambda_{N-1,h}\right)\right| \text{ if } \rho\left(R\left(kc\Delta_h\right)\right) > 1.$$

In fact, since

$$\lambda_{N-1,h} < \cdots < \lambda_{1,h} < 0$$

and $R(x) = 1 + x$ is an increasing function of x, we have

$$R(kc\lambda_{N-1,h}) < \cdots < R(kc\lambda_{1,h}) < 1 = R(0).$$

Therefore, if $\rho\left(R\left(kc\Delta_h\right)\right) > 1$, the maximum modulus among $R(kc\lambda_{n,h})$, $n \in \{1, \ldots, N-1\}$, is obtained for a negative $R(kc\lambda_{n,h})$ and then for the most negative $R(kc\lambda_{N-1,h})$.

For a consistency error such that

$$-\varepsilon^i = \begin{cases} \alpha\phi_{N-1,h} \text{ if } i = 1 \\ 0 \text{ if } i \in \{2, \ldots, M\}, \end{cases}$$

where $\alpha \in \mathbb{R}$, we have, for $m \in \{1, \ldots, M\}$,

$$\begin{aligned} e^m &= k\sum_{i=1}^{m} R\left(kc\Delta_h\right)^{m-i}\left(-\varepsilon^i\right) = kR\left(kc\Delta_h\right)^{m-1}\alpha\phi_{N-1,h} \\ &= kR\left(kc\lambda_{N-1,h}\right)^{m-1}\alpha\phi_{N-1,h} = kR\left(kc\lambda_{N-1,h}\right)^{m-1}\left(-\varepsilon^1\right). \end{aligned}$$

Then

$$\begin{aligned} \|e^m\|_h &= k\left|R\left(kc\lambda_{N-1,h}\right)\right|^{m-1}\|\varepsilon^1\|_h \\ &\geq k\left(4\gamma - 1\right)^{m-1}\|\varepsilon^1\|_h \\ &= k\left(4\gamma - 1\right)^{m-1}\max_{i\in\{1,\ldots,M\}}\|\varepsilon^i\|_h \end{aligned}$$

and then

$$\max_{m\in\{0,\ldots,M\}}\|e^m\|_h \geq \|e^M\|_h \geq \underbrace{k}_{=\frac{T}{M}}\left(4\gamma - 1\right)^{M-1}\max_{i\in\{1,\ldots,M\}}\|\varepsilon^i\|_h.$$

Thus, we have

$$\lim_{M,N\to\infty}\frac{\max\limits_{m\in\{0,\ldots,M\}}\|e^m\|_h}{\max\limits_{i\in\{1,\ldots,M\}}\|\varepsilon^i\|_h} = +\infty.$$

■

We conclude that a stability estimate cannot be given when we consider discretizations such that $\lambda \geq \beta$, where $\beta > \frac{1}{2}$: we cannot have a bound

$$\max_{m\in\{0,\dots,M\}} \|e^m\|_h \leq C \max_{i\in\{1,\dots,M\}} \|\varepsilon^i\|_h$$

where C is a constant independent of the discretization, valid for any consistence error $\left(\varepsilon^1,\dots,\varepsilon^M\right)$.

So the centered difference/forward difference method is stable if and only if $\lambda \leq \frac{1}{2}$.

Exercise. Prove that the " only if " part also holds for the L^∞ norm on $L(\Omega_h)$.

We can give another interpretation of our stability analysis. The formula

$$e^m = k\sum_{i=1}^{m} R\left(kc\Delta_h\right)^{m-i}\left(-\varepsilon^i\right),\ m\in\{0,\dots,M\},$$

shows that every consistency error ε^i, $i\in\{1,\dots,M\}$, is propagated to the convergence error e^{i+p}, $p\in\{0,1,\dots,M-i\}$, by

$$R\left(kc\Delta_h\right)^p\left(-k\varepsilon^i\right).$$

By considering the discrete Fourier series of

$$\varepsilon^i = \sum_{n=1}^{N-1} a_n\phi_{n,h},$$

we obtain

$$R\left(kc\Delta_h\right)^p\left(-k\varepsilon^i\right) = \sum_{n=1}^{N-1} R\left(kc\lambda_{n,h}\right)^p\left(-ka_n\phi_{n,h}\right).$$

So, we see that the term $a_n\phi_{n,h}$, $n\in\{1,\dots,N-1\}$, in the discrete Fourier series of ε^i is propagated to the convergence error e^{i+p} by

$$R\left(kc\lambda_{n,h}\right)^p\left(-ka_n\phi_{n,h}\right).$$

We can observe that:

- The terms

  $$a_n\phi_{n,h},\ n\in\{1,\dots,N-1\}\ \text{such that}\ \left|R\left(kc\lambda_{n,h}\right)\right| > 1,$$

  propagate to e^{i+p} by an exponential growth with p. Since $R\left(kc\lambda_{n,h}\right) = 1 + kc\lambda_{n,h} < 1$ and $\left|R\left(kc\lambda_{n,h}\right)\right| > 1$, we have $R\left(kc\lambda_{n,h}\right)$ negative and so the exponential growth is oscillatory, as we have seen in the numerical test in case of instability.

- The terms

$$a_n \phi_{n,h},\ n \in \{1, \dots, N-1\} \text{ such that } |R(kc\lambda_{n,h})| \leq 1,$$

  does not grow with p and the terms

$$a_n \phi_{n,h},\ n \in \{1, \dots, N-1\} \text{ such that } |R(kc\lambda_{n,h})| < 1,$$

  are exponentially damped with p, in the propagation to e^{i+p}.

- If

$$\rho(R(kc\Delta_h)) > 1, \text{ i.e. } |R(kc\lambda_{n,h})| > 1 \text{ for some } n \in \{1, \dots, N-1\},$$

  then, for a large p the propagation of $a_{N-1}\phi_{N-1,h}$ dominates (whenever $a_{N-1}\phi_{N-1,h} \neq 0$) the propagation of $a_n\phi_{n,h}$, $n \in \{1, \dots, N-2\}$. In fact, the maximum of $|R(kc\lambda_{n,h})|$, $n \in \{1, \dots, N-1\}$, is obtained for $n = N-1$.

Under the condition $\lambda \leq \frac{1}{2}$, we have the following convergence result for the centered difference/forward difference method: we have

$$\max_{m \in \{0,\dots,M\}} \|e^m\|_h \leq T \max_{m \in \{1,\dots,M\}} \|\varepsilon^m\|_h \leq T \max_{m \in \{1,\dots,M\}} \|\varepsilon^m\|_{L^\infty(\Omega_h)}$$

and then (recall (2))

$$\max_{m \in \{0,\dots,M\}} \|e^m\|_h$$
$$\leq T \left(\frac{c}{12} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^4 u}{\partial x^4}(x,t) \right| h^2 + \frac{1}{2} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| k \right).$$

Observe that the same bound has been obtained for $\max_{m \in \{0,\dots,M\}} \|e^m\|_{L^\infty(\Omega_h)}$ (see (4)).

Exercise. Assume $\lambda = \frac{ck}{h^2} = \frac{1}{2}$. Suppose that numbers A and B such that

$$\max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^4 u}{\partial x^4}(x,t) \right| \leq A \quad \text{and} \quad \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| \leq B$$

are known. Use the previous result for determining a spatial stepsize h such that

$$\max_{m \in \{0,\dots,M\}} \|e^m\|_h \leq \text{TOL},$$

where TOL is a given tolerance. Then, give an estimate of the number of flops (i.e arithmetic operations) for obtaining the discrete solution.

3 The centered difference/backward difference method

We consider now a different time discretization for the heat equation, namely we consider the backward Euler method, rather than the forward Euler method, for solving the system of ODEs of the method of lines.

Consider the full discretization of the heat equation

$$\begin{aligned}
&\frac{U_n^{m+1}-U_n^m}{k}=c\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}+f_n^{m+1},\\
&n\in\{1,\ldots,N-1\} \text{ and } m\in\{0,\ldots,M-1\}\\
&U_n^m=0,\ n\in\{0,N\} \text{ and } m\in\{0,\ldots,M\}\\
&U_n^0=u_0(nh),\ n\in\{1,\ldots,N-1\}.
\end{aligned}$$

This full discretization is called the *centered difference/backward difference method* since, in

$$\frac{\partial u}{\partial t}(nh,(m+1)k)=c\frac{\partial^2 u}{\partial x^2}(nh,(m+1)k)+f(nk,(m+1)k),$$

we are approximating the spatial derivative $\frac{\partial^2 u}{\partial x^2}(nh,(m+1)k)$ by the centered difference

$$\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}$$

and the time derivative $\frac{\partial u}{\partial t}(nh,(m+1)k)$ by the backward difference

$$\frac{U_n^{m+1}-U_n^m}{k}=\frac{U_n^m-U_n^{m+1}}{-k}.$$

By using the notation

$$U^m:=\left(U_1^m,\ldots,U_{N-1}^m\right) \text{ and } f^m:=\left(f_1^m,\ldots,f_{N-1}^m\right),\ m\in\{0,\ldots,M\},$$

we can write the full discretization

$$\begin{aligned}
&\frac{U_n^{m+1}-U_n^m}{k}=c\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}+f_n^{m+1}\\
&n\in\{1,\ldots,N-1\} \text{ and } m\in\{0,\ldots,M-1\}
\end{aligned}$$

as

$$\begin{aligned}
&U_n^{m+1}=U_n^m+k\left(c\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}+f_n^{m+1}\right)\\
&n\in\{1,\ldots,N-1\} \text{ and } m\in\{0,\ldots,M-1\}
\end{aligned}$$

and then in compact form as

$$U^{m+1} = U^m + k\left(c\Delta_h U^{m+1} + f^{m+1}\right)$$
$$m \in \{0, \dots, M-1\}.$$

We see that the centered difference/backward difference method is the backward Euler method as applied to the system (1) of the method of lines: the backward Euler method for

$$y'(t) = F(t, y(t)) = c\Delta_h y(t) + (f(h,t), \dots, f((N-1)h, t))$$

is

$$U^{m+1} = U^m + kF((m+1)k, U^{m+1}) = U^m + k\left(c\Delta_h U^{m+1} + f^{m+1}\right).$$

Since the backward Euler method is an implicit method for ODEs, a linear system has to be solved to obtain U^{m+1} from the vector U^m, $m \in \{0, \dots, M-1\}$. The vector U^{m+1} is obtained from U^m by solving the tridiagonal linear system of dimension $d = N-1$:

$$-\lambda U_{n-1}^{m+1} + (1+2\lambda) U_n^{m+1} - \lambda U_{n+1}^{m+1} = U_n^m + kf_n^{m+1}, \ n \in \{1, \dots, N-1\},$$

where $\lambda = \frac{ck}{h^2}$ and $U_0^{m+1} = U_N^{m+1} = 0$. These equations are obtained by rewriting the equations

$$U_n^{m+1} = U_n^m + k\left(c\frac{U_{n-1}^{m+1} - 2U_n^{m+1} + U_{n+1}^{m+1}}{h^2} + f_n^{m+1}\right)$$
$$n \in \{1, \dots, N-1\}.$$

Since the matrix of this system is strictly diagonally dominant, it is nonsingular and so existence and uniqueness for the solution U^{m+1} is proved.

Since the matrix of the system is tridiagonal, the amount of flops for solving this system of dimension $d = N-1$ by gaussian elimination is $O(d) = O(N-1)$.

In the two-dimensional case, the matrix of the system is not tridiagonal, but a band matrix with bandwidth $w = O(N-1)$. So, the amount of flops for solving this system of dimension $d = (N-1)^2$ by gaussian elimination is not $O(d) = O\left((N-1)^2\right)$, but $O\left(dw^2\right) = O\left((N-1)^4\right)$.

Exercise. What about the amount of flops for solving the system by gaussian elimination in the three-dimensional case?

3.1 Error Analysis

We define the consistency error as

$$\varepsilon_n^{m+1} := \frac{u_n^{m+1} - u_n^m}{k} - c\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} - f_n^{m+1},$$
$$n \in \{1, \dots, N-1\} \text{ and } m \in \{0, \dots, M-1\}.$$

where the u-values are the values of the exact solution.

Let $n \in \{1, \ldots, N-1\}$ and $m \in \{0, \ldots, M-1\}$. The centered difference scheme satisfies

$$\begin{aligned}
&\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} - \frac{\partial^2 u}{\partial x^2}(nh, (m+1)k) \\
&= \frac{1}{24}\left(\frac{\partial^4 u}{\partial x^4}(\alpha, (m+1)k) + \frac{\partial^4 u}{\partial x^4}(\beta, (m+1)k)\right) h^2,
\end{aligned}$$

where $\alpha \in ((n-1)\,h, nh)$ and $\beta \in (nh, (n+1)\,h)$. So, we have

$$\left|\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} - \frac{\partial^2 u}{\partial x^2}(nh, (m+1)k)\right| \leq \frac{1}{12} \max_{(x,t)\in\overline{\Omega}\times[0,T]} \left|\frac{\partial^4 u}{\partial x^4}(x,t)\right| h^2.$$

Exercise. Given a sufficiently smooth function $v(t)$ of one real variable t and $k > 0$, a finite difference approximating the first derivative $v'(t)$ is the *backward difference*

$$v'(t) \approx \frac{v(t) - v(t-k)}{k} = \frac{v(t-k) - v(t)}{-k}.$$

Prove that

$$\frac{v(t) - v(t-k)}{k} = v'(t) - \frac{1}{2} v''(\gamma)\, k,$$

where $\gamma \in (t-k, t)$. Then, prove that

$$\left|\frac{u_n^{m+1} - u_n^m}{k} - \frac{\partial u}{\partial t}(nh, (m+1)k)\right| \leq \frac{1}{2} \max_{(x,t)\in\overline{\Omega}\times[0,T]} \left|\frac{\partial^2 u}{\partial t^2}(x,t)\right| k.$$

Now, since

$$\begin{aligned}
\varepsilon_n^{m+1} &= \frac{u_n^{m+1} - u_n^m}{k} - c\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} - f_n^{m+1} \\
&= \frac{u_n^{m+1} - u_n^m}{k} - \frac{\partial u}{\partial t}(nh, (m+1)\,k) \\
&\quad - c\left(\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} - \frac{\partial^2 u}{\partial x^2}(nh, (m+1)\,k)\right) \\
&\quad + \underbrace{\frac{\partial u}{\partial t}(nh, (m+1)\,k) - c\frac{\partial^2 u}{\partial x^2}(nh, (m+1)\,k) - f_n^{m+1}}_{=0},
\end{aligned}$$

we obtain

$$|\varepsilon_n^{m+1}| \leq \frac{c}{12} \max_{(x,t)\in\overline{\Omega}\times[0,T]} \left|\frac{\partial^4 u}{\partial x^4}(x,t)\right| h^2 + \frac{1}{2} \max_{(x,t)\in\overline{\Omega}\times[0,T]} \left|\frac{\partial^2 u}{\partial t^2}(x,t)\right| k.$$

So, as in case of the centered difference/forward difference method, we obtain for the centered difference/backward difference method

$$\max_{\substack{n\in\{1,\dots,N-1\}\\ m\in\{1,\dots,M\}}} |\varepsilon_n^m| \leq \frac{c}{12} \max_{(x,t)\in\overline{\Omega}\times[0,T]} \left|\frac{\partial^4 u}{\partial x^4}(x,t)\right| h^2 + \frac{1}{2} \max_{(x,t)\in\overline{\Omega}\times[0,T]} \left|\frac{\partial^2 u}{\partial t^2}(x,t)\right| k.$$

Now, we consider the convergence error

$$e_n^m = U_n^m - u_m^n,\ n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M\}.$$

Since

$$\begin{aligned}
&\frac{U_n^{m+1} - U_n^m}{k} = c\frac{U_{n-1}^{m+1} - 2U_n^{m+1} + U_{n+1}^{m+1}}{h^2} + f_n^{m+1}\\
&n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M-1\}\\
&U_0^m = 0 \text{ and } U_N^m = 0,\quad m \in \{0,\dots,M\},\\
&U_n^0 = u_0(nh),\ n \in \{1,\dots,N-1\},
\end{aligned}$$

$$\begin{aligned}
&\frac{u_n^{m+1} - u_n^m}{k} = c\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} + f_n^{m+1} + \varepsilon_n^{m+1}\\
&n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M-1\}\\
&u_0^m = 0 \text{ and } u_N^m = 0,\quad m \in \{0,\dots,M\},\\
&u_n^0 = u_0(nh),\ n \in \{1,\dots,N-1\},
\end{aligned}$$

we obtain the full discrete problem

$$\begin{aligned}
&\frac{e_n^{m+1} - e_n^m}{k} = c\frac{e_{n-1}^{m+1} - 2e_n^{m+1} + e_{n+1}^{m+1}}{h^2} - \varepsilon_n^{m+1}\\
&n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M-1\}\\
&e_0^m = 0 \text{ and } e_N^m = 0,\quad m \in \{0,\dots,M\},\\
&e_n^0 = 0,\ n \in \{1,\dots,N-1\}.
\end{aligned}$$

By using

$$e^m = \left(e_1^m,\dots,e_{N-1}^m\right),\ m \in \{0,\dots,M\},$$

and

$$\varepsilon^{m+1} = \left(\varepsilon_1^{m+1},\dots,\varepsilon_{N-1}^{m+1}\right),\ m \in \{0,\dots,M-1\},$$

we can rewrite this full discrete problem as

$$e^{m+1} = e^m + kc\Delta_h e^{m+1} - k\varepsilon^{m+1},\ m \in \{0,\dots,M-1\},$$

i.e.

$$(I - kc\Delta_h)\, e^{m+1} = e^m - k\varepsilon^{m+1},\quad m \in \{0,\dots,M-1\},$$

i.e.

$$e^{m+1} = R\left(kc\Delta_h\right) e^m - kR\left(kc\Delta_h\right) \varepsilon^{m+1}, \quad m \in \{0, \ldots, M-1\},$$

where

$$R\left(kc\Delta_h\right) = \left(I - kc\Delta_h\right)^{-1}. \tag{7}$$

Observe that the linear operator $R\left(kc\Delta_h\right)$ has the eigenvalues

$$R\left(kc\lambda_{n,h}\right) = \frac{1}{1 - kc\lambda_{n,h}}, \; n \in \{1, \ldots, N-1\},$$

and since the eigenvalues $\lambda_{n,h}$, $n \in \{1, \ldots, N-1\}$, of Δ_h are negative, we have

$$0 < R\left(kc\lambda_{n,h}\right) < 1, \; n \in \{1, \ldots, N-1\}.$$

Exercise. Explain why $I - kc\Delta_h$ (whose inverse appears in (7)) is invertible.

Exercise. Prove the stability result for the centered difference/backward difference method

$$\max_{m \in \{0,\ldots,M\}} \|e^m\|_h \leq T \max_{i \in \{1,\ldots,M\}} \|\varepsilon^i\|_h,$$

which is valid without any assumption of a relation between h and k. For this reason, the centered difference/backward difference method is called *unconditionally stable*.

By using the previous unconditional stability result, we have the unconditional convergence result for the centered difference/backward difference method

$$\max_{m \in \{0,\ldots,M\}} \|e^m\|_h \leq T \max_{i \in \{1,\ldots,M\}} \|\varepsilon^i\|_h \leq T \underbrace{\max_{i \in \{1,\ldots,M\}} \|\varepsilon^i\|_{L^\infty(\Omega_h)}}_{= \max_{\substack{n \in \{1,\ldots,N-1\} \\ m \in \{1,\ldots,M\}}} |\varepsilon_n^m|}$$

$$\leq T \left(\frac{c}{12} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^4 u}{\partial x^4}(x,t) \right| h^2 + \frac{1}{2} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| k \right).$$

Exercise. Use the previous convergence result for determining stepsizes h and k such that

$$\max_{m \in \{0,\ldots,M\}} \|e^m\|_h \leq \mathrm{TOL},$$

where TOL is a given tolerance. Put both spatial error bound

$$T \frac{c}{12} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^4 u}{\partial x^4}(x,t) \right| h^2$$

and time error bound

$$T \frac{1}{2} \max_{(x,t) \in \overline{\Omega} \times [0,T]} \left| \frac{\partial^2 u}{\partial t^2}(x,t) \right| k$$

equal to $\frac{\mathrm{TOL}}{2}$. Then, give an estimate $O(\mathrm{TOL}^{-p})$ of the number of flops for obtaining the discrete solution.

4 The Crank-Nicolson method

If we use the trapezoidal rule to discretize in time the system of ODEs of the method of lines, we get the *Crank-Nicolson method*:

$$\begin{aligned}
\frac{U_n^{m+1}-U_n^m}{k} &= c\frac{1}{2}\left(\frac{U_{n-1}^m-2U_n^m+U_{n+1}^m}{h^2}+\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}\right) \\
&\quad +\frac{1}{2}\left(f_n^m+f_n^{m+1}\right) \\
&= \frac{1}{2}\left(\underbrace{c\frac{U_{n-1}^m-2U_n^m+U_{n+1}^m}{h^2}+f_n^m}_{\text{Centered difference/forward difference method}}\right. \\
&\qquad \left. + \underbrace{c\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}+f_n^{m+1}}_{\text{Centered difference/backward difference method}}\right)
\end{aligned}$$

$$\begin{aligned}
& n\in\{1,\dots,N-1\} \text{ and } m\in\{0,\dots,M-1\}, \\
& U_n^m=0,\ n\in\{0,N\} \text{ and } m\in\{0,\dots,M\}, \\
& U_n^0=u_0(nh),\ n\in\{1,\dots,N-1\}.
\end{aligned}$$

This full discretization is also called the *centered difference/centered difference method* since, in

$$\frac{\partial u}{\partial t}\left(nh,\left(m+\frac{1}{2}\right)k\right)=c\frac{\partial^2 u}{\partial x^2}\left(nh,(m+\frac{1}{2})k\right)+f\left(nk,\left(m+\frac{1}{2}\right)k\right),$$

we are approximating:

- $\frac{\partial^2 u}{\partial x^2}\left(nh,(m+\frac{1}{2})k\right)$ by the average

$$\frac{1}{2}\left(\frac{U_{n-1}^m-2U_n^m+U_{n+1}^m}{h^2}+\frac{U_{n-1}^{m+1}-2U_n^{m+1}+U_{n+1}^{m+1}}{h^2}\right)$$

  of the centered difference approximations of $\frac{\partial^2 u}{\partial x^2}(nh,mk)$ and $\frac{\partial^2 u}{\partial x^2}(nh,(m+1)k)$;

- $\frac{\partial u}{\partial t}\left(nh,\left(m+\frac{1}{2}\right)k\right)$ by the centered difference

$$\frac{U_n^{m+1}-U_n^m}{k};$$

- $f\left(nh,\left(m+\frac{1}{2}\right)k\right)$ by the average

$$\frac{1}{2}\left(f_n^{m+1}+f_n^m\right)$$

  of the values $f(nh,mk)$ and $f(nh,(m+1)k)$.

By using the usual notation

$$U^m := \left(U_1^m, \ldots, U_{N-1}^m\right) \text{ and } f^m := \left(f_1^m, \ldots, f_{N-1}^m\right), \ m \in \{0, \ldots, M\},$$

we can write the full discretization

$$\begin{aligned} \frac{U_n^{m+1} - U_n^m}{k} &= c\frac{1}{2}\left(\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + \frac{U_{n-1}^{m+1} - 2U_n^{m+1} + U_{n+1}^{m+1}}{h^2}\right) \\ &\quad + \frac{1}{2}\left(f_n^m + f_n^{m+1}\right) \\ n \in \{1, \ldots, N-1\} & \text{ and } m \in \{0, \ldots, M-1\} \end{aligned}$$

as

$$\begin{aligned} U_n^{m+1} = U_n^m + k\frac{1}{2}\Bigg(& c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m \\ & + c\frac{U_{n-1}^{m+1} - 2U_n^{m+1} + U_{n+1}^{m+1}}{h^2} + f_n^{m+1}\Bigg) \\ n \in \{1, \ldots, N-1\} & \text{ and } m \in \{0, \ldots, M-1\} \end{aligned}$$

and then in compact form as

$$\begin{aligned} & U^{m+1} = U^m + k\frac{1}{2}\left(c\Delta_h U^m + f^m + c\Delta_h U^{m+1} + f^{m+1}\right) \\ & m \in \{0, \ldots, M-1\}. \end{aligned}$$

We see that the centered difference/centered difference method (the Crank-Nicolson method) is the trapeziodal rule as applied to the system (1) of the method of lines: the trapezoidal rule for

$$y'(t) = F(t, y(t)) = c\Delta_h y(t) + (f(h,t), \ldots, f((N-1)h, t))$$

is

$$\begin{aligned} U^{m+1} &= U^m + k\frac{1}{2}\left(F(mk, U^m) + F((m+1)k, U^{m+1})\right) \\ &= U^m + k\frac{1}{2}\left(c\Delta_h U^m + f^m + c\Delta_h U^{m+1} + f^{m+1}\right). \end{aligned}$$

Since the trapezoidal rule is an implicit method for ODEs, a linear system has to be solved to obtain U^{m+1} from U^m, $m \in \{0, \ldots, M-1\}$. The vector $U^{m+1}$is obtained from the vector U^m by solving the tridiagonal linear system

$$\begin{aligned} & -\frac{1}{2}\lambda U_{n-1}^{m+1} + (1+\lambda) U_n^{m+1} - \frac{1}{2}\lambda U_{n+1}^{m+1} \\ & = \frac{1}{2}\lambda U_{n-1}^m + (1-\lambda) U_n^m + \frac{1}{2}\lambda U_{n+1}^m + \frac{1}{2}k\left(f_n^m + f_n^{m+1}\right) \\ & n \in \{1, \ldots, N-1\}. \end{aligned}$$

where $\lambda = \frac{ck}{h^2}$ and $U_0^{m+1} = U_N^{m+1} = 0$. These equations are obtained by rewriting the equations

$$\begin{aligned} U_n^{m+1} = U_n^m + k\frac{1}{2}\Bigg(& c\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + f_n^m \\ & + c\frac{U_{n-1}^{m+1} - 2U_n^{m+1} + U_{n+1}^{m+1}}{h^2} + f_n^{m+1}\Bigg) \\ n \in \{1, \ldots, N-1\}. & \end{aligned}$$

The matrix of this system is strictly diagonally dominant. So, it is non-singular and so existence and uniqueness for the solution U^{m+1} is proved.

4.1 Error Analysis

The consistency error is given by

$$\begin{aligned} \varepsilon_n^{m+1} := & \frac{u_n^{m+1} - u_n^m}{k} - c\frac{1}{2}\left(\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - \frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2}\right) \\ & - \frac{1}{2}\left(f_n^m + f_n^{m+1}\right) \\ n \in \{1, \ldots, & N-1\} \text{ and } m \in \{0, \ldots, M-1\}. \end{aligned}$$

For $n \in \{1, \ldots, N-1\}$ and $m \in \{0, \ldots, M-1\}$, we have

$$\begin{aligned} \varepsilon_n^{m+1} = & \frac{u_n^{m+1} - u_n^m}{k} - c\frac{1}{2}\left(\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} + \frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2}\right) \\ & - \frac{1}{2}\left(f_n^{m+1} + f_n^m\right) \\ = & \frac{u_n^{m+1} - u_n^m}{k} - \frac{\partial u}{\partial t}\left(nh, \left(m + \frac{1}{2}\right)k\right) \\ & - c\frac{1}{2}\left(\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} - \frac{\partial^2 u}{\partial x^2}(nh, mk)\right) \\ & - c\frac{1}{2}\left(\frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2} - \frac{\partial^2 u}{\partial x^2}(nh, (m+1)k)\right) \\ & - c\left(\frac{1}{2}\left(\frac{\partial^2 u}{\partial x^2}(nh, mk) + \frac{\partial^2 u}{\partial x^2}(nh, (m+1)k)\right) - \frac{\partial^2 u}{\partial x^2}\left(nh, \left(m + \frac{1}{2}\right)k\right)\right) \\ & - \left(\frac{1}{2}\left(f_n^m + f_n^{m+1}\right) - f\left(nh, \left(m + \frac{1}{2}\right)k\right)\right) \\ & + \underbrace{\frac{\partial u}{\partial t}\left(nh, \left(m + \frac{1}{2}\right)k\right) - c\frac{\partial^2 u}{\partial x^2}\left(nh, \left(m + \frac{1}{2}\right)k\right) - f\left(\left(nh, \left(m + \frac{1}{2}\right)k\right)\right)}_{=0}. \end{aligned}$$

Exercise. Given a sufficiently smooth function $v(t)$ of one real variable t and $k > 0$, a finite difference approximating the first derivative $v'(t)$ is the *central difference*

$$v'(t) \approx \frac{v(t+k) - v(t-k)}{2k}.$$

Prove that

$$\frac{v(t+k) - v(t-k)}{2k} = v'(t) + \frac{1}{12}\left(v'''(\gamma) + v'''(\delta)\right)k^2,$$

where $\gamma \in (t, t+k)$ and $\delta \in (t-k, t)$. Then, prove that

$$\max_{\substack{n\in\{1,\dots,N-1\}\\ m\in\{1,\dots,M\}}} |\varepsilon_n^m| \leq Ch^2 + Dk^2$$

for some constants $C, D \geq 0$ which depends on the maximum absolute values on $\overline{\Omega} \times [0,T]$ of partial derivates of u and f.

So, unlike the centered difference/forward difference method and the centered difference/backward difference method, the Crank-Nicolson method has consistency order two with respect to time stepsize k, as the consistency order with respect to the spatial stepsize h.

Now, we consider the convergence error

$$e_n^m = U_n^m - u_m^n,\ n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M\}.$$

Since

$$\begin{aligned}
&\frac{U_n^{m+1} - U_n^m}{k} = c\frac{1}{2}\left(\frac{U_{n-1}^m - 2U_n^m + U_{n+1}^m}{h^2} + \frac{U_{n-1}^{m+1} - 2U_n^{m+1} + U_{n+1}^{m+1}}{h^2}\right)\\
&\qquad\qquad + \frac{1}{2}\left(f_n^m + f_n^{m+1}\right)\\
&n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M-1\},\\
&U_0^m = 0 \text{ and } U_N^m = 0,\quad m \in \{0,\dots,M\}\\
&U_n^0 = u_0(nh),\ n \in \{1,\dots,N-1\},
\end{aligned}$$

$$\begin{aligned}
&\frac{u_n^{m+1} - u_n^m}{k} = c\frac{1}{2}\left(\frac{u_{n-1}^m - 2u_n^m + u_{n+1}^m}{h^2} + \frac{u_{n-1}^{m+1} - 2u_n^{m+1} + u_{n+1}^{m+1}}{h^2}\right)\\
&\qquad\qquad + \frac{1}{2}\left(f_n^m + f_n^{m+1}\right) + \varepsilon_n^{m+1}\\
&n \in \{1,\dots,N-1\} \text{ and } m \in \{0,\dots,M-1\}\\
&u_0^m = 0 \text{ and } u_N^m = 0,\quad m \in \{0,\dots,M\}\\
&u_n^0 = u_0(nh),\ n \in \{1,\dots,N-1\},
\end{aligned}$$

we have the full discrete problem

$$\frac{e_n^{m+1}-e_n^m}{k} = c\frac{1}{2}\left(\frac{e_{n-1}^m - 2e_n^m + e_{n+1}^m}{h^2} + \frac{e_{n-1}^{m+1} - 2e_n^{m+1} + e_{n+1}^{m+1}}{h^2}\right) - \varepsilon_n^{m+1}$$
$$n \in \{1, \ldots, N-1\} \text{ and } m \in \{0, \ldots, M-1\}$$
$$e_0^m = 0 \text{ and } e_N^m = 0, \quad m \in \{0, \ldots, M\}$$
$$e_n^0 = 0, \ n \in \{1, \ldots, N-1\}.$$

By using

$$e^m = \left(e_1^m, \ldots, e_{N-1}^m\right), \ m \in \{0, \ldots, M\},$$

and

$$\varepsilon^{m+1} = \left(\varepsilon_1^{m+1}, \ldots, \varepsilon_{N-1}^{m+1}\right), \ m \in \{0, \ldots, M-1\},$$

we can rewrite this full discrete problem as

$$e^{m+1} = e^m + \frac{1}{2}kc\left(\Delta_h e^m + \Delta_h e^{m+1}\right) - k\varepsilon^{m+1}, \ m \in \{0, \ldots, M-1\},$$

i.e.

$$\left(I - \frac{1}{2}kc\Delta_h\right)e^{m+1} = \left(I + \frac{1}{2}kc\Delta_h\right)e^m - k\varepsilon^{m+1}, \quad m \in \{0, \ldots, M-1\},$$

i.e.

$$e^{m+1} = R\left(kc\Delta_h\right)e^m - k\left(I - \frac{1}{2}kc\Delta_h\right)^{-1}\varepsilon^{m+1}, \quad m \in \{0, \ldots, M-1\},$$

where

$$R\left(kc\Delta_h\right) = \left(I - \frac{1}{2}kc\Delta_h\right)^{-1}\left(I + \frac{1}{2}kc\Delta_h\right).$$

The linear operator $R\left(kc\Delta_h\right)$ has the eigenvalues

$$R\left(kc\lambda_{n,h}\right) = \frac{1 + \frac{1}{2}kc\lambda_{n,h}}{1 - \frac{1}{2}kc\lambda_{n,h}}, \ n \in \{1, \ldots, N-1\}.$$

Exercise. Prove that

$$\left|R\left(kc\lambda_{m,h}\right)\right| \leq 1, \ m \in \{1, \ldots, N-1\}.$$

Exercise. Prove the stability result for the Crank-Nicolson method

$$\max_{m\in\{0,\ldots,M\}} \|e^m\|_h \leq T \max_{i\in\{1,\ldots,M\}} \|\varepsilon^i\|_h,$$

which is valid without any assumption of a relation between h and k. The Crank-Nicolson method is unconditionally stable.

By using the previous unconditional stability result, we have the unconditional convergence result for the Crank-Nicolson method

$$\max_{m\in\{0,\ldots,M\}} \|e^m\|_h \leq T \max_{i\in\{1,\ldots,M\}} \|\varepsilon^i\|_h \leq T \underbrace{\max_{i\in\{1,\ldots,M\}} \|\varepsilon^i\|_{L^\infty(\Omega_h)}}_{=\max_{\substack{n\in\{1,\ldots,N-1\}\\ m\in\{1,\ldots,M\}}} |\varepsilon_n^m|}$$

$$\leq T\left(Ch^2 + Dk^2\right).$$

The order of convergence of the Crank-Nicolson method is two, both in space and time.

Exercise. Use the previous convergence result for determining stepsizes h and k such that

$$\max_{m\in\{0,\ldots,M\}} \|e^m\|_h \leq \text{TOL},$$

where TOL is a given tolerance. Put both spatial error bound TCh^2 and time error bound TDk^2 equal to $\frac{\text{TOL}}{2}$. Then, give an estimate $O(\text{TOL}^{-p})$ of the number of flops for obtaining the discrete solution.

Exercise. For the two and three-dimensional cases and for the centered difference/forward difference, centered difference/backward difference and Crank-Nicolson methods, give estimates $O(\text{TOL}^{-p})$ of the number of flops for obtaining a discrete solution such that

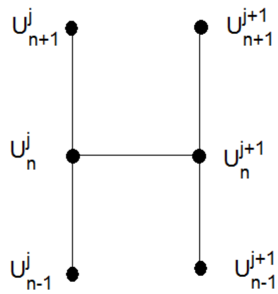
$$\max_{m\in\{0,\ldots,M\}} \|e^m\|_h \leq \text{TOL}.$$

In all situations considers stepsizes h and k such that both spatial error bound and time error bound are equal to $\frac{\text{TOL}}{2}$.

5 The three numerical methods for the heat equation: a summary

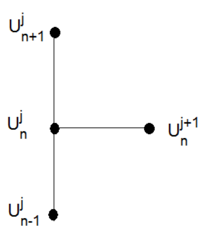
In the next tables, we give the basic information about the three methods that we have introduced for the heat equation.

The first column of each table contains the *stencil* of the finite difference, namely a picture with the values of the discrete solution involved in the equation for the space index n and the time index m (j in the tables and in the figure below). Here is the stencil for the Crank-Nicolson method:

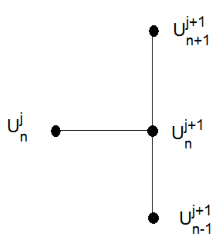


These values are geometrically arranged in a space-time grid with the vertical axis as space and the horizontal axis as time. Points that are involved in a finite difference in space are vertically connected and points that are involved in a finite difference in time are horizontally connected.

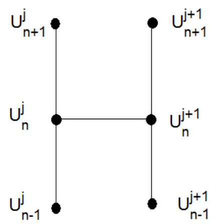
Centered difference/forward difference method

Stencil	Explicit/Implicit	Order	Stability
	Explicit	$O\left(h^2\right) + O\left(k\right)$	conditionally stable: $\frac{ck}{h^2} \leq \frac{1}{2}$

Centered difference/backward difference method

Stencil	Explicit/Implicit	Order	Stability
	Implicit	$O\left(h^2\right) + O\left(k\right)$	unconditionally stable

Crank-Nicolson method (centered difference/centered difference method)

Stencil	Explicit/Implicit	Order	Stability
	Implicit	$O\left(h^2\right) + O\left(k^2\right)$	unconditionally stable