Specific techniques for parabolic problems

Contents

Chapter 1

Specific techniques for parabolic problems

1.1 Introduction

Parabolic PDEs behave much like initial-value ODE problems and, therefore, are somewhat easier to approach than hyperbolic or elliptic problems. Let's derive the canonical parabolic equation by considering a specific physical problem: unsteady thermal diffusion through a plane wall. Applying the energy conservation principle to a plane slab of thickness $\mathrm{d}x$, width B and height H, under the assumption of one-dimensional thermal diffusion, yields:

$$\frac{\partial}{\partial t}\left(\rho\, c\, T\, B\, H\, \mathrm{d}x\right) = -k\, \left.\frac{\partial T}{\partial x}\right|_{x} B\, H - \left[-k\, \left.\frac{\partial T}{\partial x}\right|_{x+\mathrm{d}x} B\, H\right]$$

Assuming constant thermophysical properties we are left with:

$$\frac{\partial T}{\partial t} = \alpha\, \frac{\partial^2 T}{\partial x^2} \tag{1.1}$$

where the *thermal diffusivity* α is defined as

$$\alpha \equiv \frac{k}{\rho\, c}$$

Let's assume that the temperatures of the wall's exposed faces are known:

$$T(x=0,t) = T_1; \qquad T(x=L,t) = T_2 \tag{1.2}$$

These equations provide two boundary conditions. The initial temperature distribution across the slab provides the initial condition:

$$T(x, t=0) = f(x) \tag{1.3}$$

Using non-dimensional quantities,

$$\vartheta \equiv \frac{T - T_1}{T_2 - T_1}; \quad y \equiv \frac{x}{L}; \quad \mathrm{Fo} \equiv \frac{t\,\alpha}{L^2}$$

the problem can be recast as

$$\frac{\partial \vartheta}{\partial \mathrm{Fo}} = \frac{\partial^2 \vartheta}{\partial y^2} \tag{1.4a}$$

$$\vartheta(y=0,\mathrm{Fo}) = 0; \qquad \vartheta(y=1,\mathrm{Fo}) = 1 \tag{1.4b}$$

$$\vartheta(y, \mathrm{Fo} = 0) = \frac{f(y) - T_1}{T_2 - T_1} \tag{1.4c}$$

Changing the dependent variable as

$$\Theta \equiv \vartheta - y$$

allows to enforce homogeneous boundary conditions:

$$\frac{\partial \Theta}{\partial \mathrm{Fo}} = \frac{\partial^2 \Theta}{\partial y^2} \tag{1.5a}$$

$$\Theta(y = 0, \mathrm{Fo}) = 0; \qquad \Theta(y = 1, \mathrm{Fo}) = 0 \tag{1.5b}$$

$$\Theta(y, \mathrm{Fo} = 0) = \frac{f(y) - T_1}{T_2 - T_1} - y =: g(y) \tag{1.5c}$$

Parameters for the dimensional problem:

$$\alpha; \quad L; \quad T_1; \quad T_2; \quad f(x); \quad x; \quad t$$

Parameters for the non-dimensional form:

$$g(y); \quad y; \quad \mathrm{Fo}$$

Problem (1.5) looks like an ODE-IVP in time and an ODE-BVP in space. For this reason, in attempting a numerical solution of it, we start by approximating the spatial derivative in much the same way as it is treated in direct methods for the numerical solution of ODE-BVPs. Using a second-order, central differencing scheme on a uniform grid yields a system of IVPs in time:

$$\frac{\mathrm{d}\Theta}{\mathrm{dFo}} = \frac{\Theta_{j+1} - 2\,\Theta_j + \Theta_{j-1}}{(\Delta y)^2}, \; j = 1, \ldots, N \tag{1.6a}$$

$$\Theta_0 = \Theta_{N+1} = 0 \;\; \forall \; \mathrm{Fo} > 0 \tag{1.6b}$$

$$\Theta_j = g(y_j), \; j = 1, \ldots, N \;\; \text{for Fo} = 0 \tag{1.6c}$$

where $\Delta y \equiv 1/(N+1)$.

In matrix form:

$$\frac{\mathrm{d}\boldsymbol{\Theta}}{\mathrm{dFo}} = \boldsymbol{A}\,\boldsymbol{\Theta} \tag{1.7}$$

$$\boldsymbol{\Theta}(\mathrm{Fo}=0) = \boldsymbol{g} \tag{1.8}$$

where

$$\boldsymbol{A} \equiv \frac{1}{(\Delta y)^2}\,\mathrm{Tr}\,(N;[1 \quad -2 \quad 1])$$

The boundary conditions (1.5b) are accounted for in matrix $\boldsymbol{A}$.

$\boldsymbol{A}$ is a real, symmetric, tridiagonal matrix with constant coefficients along the diagonals. According to Appendix A, it admits a set of real eigenvalues $\{\lambda_k\}_{k=1}^N$, with corresponding eigenvectors $\left\{\boldsymbol{\Theta}^{(k)}\right\}_{k=1}^N$:

$$\begin{aligned}\lambda_m &= \frac{1}{(\Delta y)^2}\left[-2+2\,\cos\left(\frac{m\,\pi}{N+1}\right)\right] \\ &= (N+1)^2\left[-2+2\,\cos\left(\frac{m\,\pi}{N+1}\right)\right]\end{aligned}$$

The eigenvalues are all negative with magnitude increasing with m. The lowest-magnitude eigenvalue is

$$\lambda_1 = (N+1)^2\left[-2+2\,\cos\left(\frac{\pi}{N+1}\right)\right] \longrightarrow -\pi^2 \text{ as } N \longrightarrow +\infty$$

The largest-magnitude eigenvalue is

$$\lambda_N = (N+1)^2\left[-2+2\,\cos\left(\frac{N\,\pi}{N+1}\right)\right] \approx -4\,(N+1)^2 \text{ for large } N$$

The corresponding eigenvectors are (see Appendix A):

$$\Theta_j^{(m)} = (-1)^j\,\sin\left(\frac{j\,m\,\pi}{N+1}\right),\ j=1,\,\ldots,\,N \tag{1.9}$$

The eigenvectors are mutually orthogonal w.r.t. the discrete inner product

$$(\{f_j\},\{g_i\}) \equiv \frac{2}{N+1}\sum_{j=1}^{N} f_j\,g_j$$

and form a basis of $\mathbb{R}^N$. The eigenvectors (1.9) form an orthonormal basis of $\mathbb{R}^N$:

$$\boldsymbol{\Theta}^{(k)} \cdot \boldsymbol{\Theta}^{(j)} = \delta_{j\,k}$$

What is important here is that we are allowed to write any arbitrary $\boldsymbol{\Theta} \in \mathbb{R}^N$ as a unique linear combination of the eigenvector's basis:

$$\boldsymbol{\Theta}(\mathrm{Fo}) = \sum_{j=1}^{N} c_j(\mathrm{Fo})\, \boldsymbol{\Theta}^{(j)}$$

Substituting in (1.7) yields:

$$\sum_{j=1}^{N} \dot{c}_j\ (\mathrm{Fo})\, \boldsymbol{\Theta}^{(j)} = \boldsymbol{A} \left(\sum_{k=1}^{N} \dot{c}_k\ (\mathrm{Fo})\, \boldsymbol{\Theta}^{(k)} \right) = \sum_{k=1}^{N} \dot{c}_k\ (\mathrm{Fo})\, \lambda_k\, \boldsymbol{\Theta}^{(k)}$$

Using orthogonality (or recalling that $\left\{\boldsymbol{\Theta}^{(k)}\right\}_{k=1}^{N}$ is a set of linearly independent vectors) yields:

$$\dot{c}_k = \lambda_k\, c_k;\ \ \forall\, k = 1,\, \ldots,\, N \Longrightarrow c_k(\mathrm{Fo}) = c_k(0)\, \mathrm{e}^{\lambda_k\, \mathrm{Fo}}$$

The constants $\{c_k(0)\}_{k=1}^{N}$ can be determined using the initial condition:

$$\boldsymbol{g} \cdot \boldsymbol{\Theta}^{(m)} = c_m(0)$$

Thus:

$$\boldsymbol{\Theta}(\mathrm{Fo}) = \sum_{m=1}^{N} \left[\boldsymbol{g} \cdot \boldsymbol{\Theta}^{(m)} \right] \boldsymbol{\Theta}^{(m)}\, \mathrm{e}^{\lambda_m\, \mathrm{Fo}} \tag{1.10}$$

This concludes the formal solution of the problem. Yet, this method is impractical whenever the eigenvalues and eigenvectors are unknown and can not be easily extended to non-linear problems. Nevertheless, it allows to study the behavior of numerical methods for parabolic PDEs.

The largest-magnitude expansion terms decay more rapidly, so that, in the considered case, the long-time behavior of the solution is controlled by the lower-index components. The spectral radius of the matrix tends to

$$\left| \frac{\lambda_M}{\lambda_1} \right| \approx \frac{4}{\pi^2}\ (N+1)^2 \tag{1.11}$$

showing that, for large N, the system of ODEs (1.7) is quite *stiff*: this is a key point in the selection of proper numerical methods for parabolic PDEs. Due to the stiffness of the semi-dicrete system of ODEs, the stability requirement will force explicit methods to take very small time-steps.

Example 1.1.1. *Let's attempt an explicit Euler, second-order finite-difference solution of the heat equation:*

$$\Psi_j^{n+1} - \Psi_j^n = \frac{\Delta Fo}{(\Delta y)^2} \left(\Psi_{j+1}^n - 2\,\Psi_j^n + \Psi_{j-1}^n \right)$$

The Euler method is stable (why? For what model equation?) as long as

$$|\lambda_l|\ \Delta Fo < 2 \ \ \forall l = 1, \ldots, m$$

The most stringent constraint is imposed by the largest-magnitude eigenvalue:

$$|\lambda_{max}|\ \Delta Fo < 2 \Longrightarrow 4\,\frac{\Delta Fo}{(\Delta y)^2} < 2$$

This, in turn, implies

$$\frac{\alpha\,\Delta t}{(\Delta y)^2} < \frac{1}{2}$$

The dimensionless group $\dfrac{\alpha\,\Delta t}{(\Delta y)^2}$ *can be interpreted as a ratio between the time-step and a diffusion time-scale. Notice that* $\Delta t \sim (\Delta y)^2$*, while the stability limit for advection, expressed in terms of the CFL number, provides* $\Delta t \sim \Delta x$*. Since diffusion is of major importance in high-gradient regions, as boundary- or shear-layers, the need of refining the mesh in these regions is likely to conflict with the stability limit imposed by diffusion, more than that imposed by advection.*

Remark 1.1.2. *From dimensional analysis, the numerical solution of the 1D diffusion equation can be recast as:*

$$\Psi = f\left(\Delta y,\, \alpha,\, \Delta t,\, L,\, \Psi_0\right) \Longrightarrow \frac{\Psi}{\Psi_0} = g\left(\frac{L}{\Delta y},\, \frac{\alpha\,\Delta t}{(\Delta y)^2}\right) = g\left(N,\, \frac{\alpha\,\Delta t}{(\Delta y)^2}\right)$$

Exercise 1.1.3. *Solve the 1D diffusion equation with the explicit-Euler / central finite-difference method AND with the explicit-Euler / compact 4th-order scheme. Consider the following initial solution*

$$\Psi(y, 0) = 1 - \cos\left(2\,\pi\,y\right)$$

and Dirichlet, homogeneous boundary conditions. Derive a-priori stability limits for the two methods (compute numerically the eigenvalues, when using the compact scheme).

Remark 1.1.4. *According to the aforementioned remarks, explicit methods might not be the most reasonable choice for solving parabolic PDEs except, perhaps, when high accuracy is sought and, therefore, small time-steps must be used anyway. Also, notice that* $\lambda_{max} = 4\,(N+1)^2$ *is a consequence of the adopted spatial discretization method. Should more accurate spatial discretization methods be used,* λ_{max} *would become even larger.*

1.2 Physical interpretation of the *stability Fourier number*

We have already recognized that the stability limits of discretization methods for the 1D heat equation can be expressed in terms of the dimensionless group Fo_Δ, henceforth referred to as the *stability Fourier number*:

$$\mathrm{Fo}_\Delta = \frac{\alpha\,\Delta t}{(\Delta x)^2} \tag{1.12}$$

In this section, we aim to provide a physical interpretation for Fo_Δ. To this end, let us consider a specific problem: a rod of length L, made of a homogeneous, isotropic material with thermal diffusivity α, is initially at uniform temperature T_0. The rod is thermally insulated at its right end, while for $t > 0$ its left end stays connected with a thermal source at constant temperature $T_1 > T_0$. Heat diffuses progressively throughout the rod, so that, after a sufficiently long elapsed time, the bar attains a nearly uniform temperature T_1. How long does it take in order that the *thermal disturbance* propagates through the bar?

The aforementioned diffusion problem is described by the following PDE problem:

$$\frac{\partial T}{\partial t} = \alpha\,\frac{\partial^2 T}{\partial x^2} \tag{1.13a}$$

$$T(x,0) = T_0 \tag{1.13b}$$

$$T(0,t) = T_1 \tag{1.13c}$$

$$\left.\frac{\partial T}{\partial x}\right|_{(L,t)} = 0 \tag{1.13d}$$

In terms of the dimensionless quantities

$$\tau \equiv \frac{\alpha\, t}{L^2}; \quad \vartheta \equiv \frac{T - T_0}{T_1 - T_0}; \quad z \equiv \frac{x}{L} \tag{1.14}$$

the PDE problem is recast as:

$$\frac{\partial \vartheta}{\partial \tau} = \frac{\partial^2 \vartheta}{\partial z^2} \tag{1.15a}$$

$$\vartheta(z, 0) = 0 \tag{1.15b}$$

$$T(0, \tau) = 1 \tag{1.15c}$$

$$\left.\frac{\partial \vartheta}{\partial z}\right|_{(1,\tau)} = 0 \tag{1.15d}$$

We attempt a finite-difference solution on a uniform grid, endowing only three nodes: node $i = 1$ is located on the left boundary, node $i = 2$ is located in the middle of the rod, node $i = 3$ lies on the right boundary. Using the boundary conditions,

$$\vartheta(0, \tau) = 1 \Longrightarrow \vartheta_1 = 1; \qquad \left.\frac{\partial \vartheta}{\partial z}\right|_{(1,\tau)} = 0 \Longrightarrow \vartheta_3 = \vartheta_2$$

yields the following IVP for node $i = 2$:

$$\begin{aligned} &\frac{\mathrm{d}\vartheta_2}{\mathrm{d}\tau} = 4\,(1 - \vartheta_2) \\ &\vartheta_2(0) = 0 \end{aligned} \tag{1.16}$$

which yields the time dependence of ϑ_2 as:

$$\vartheta_2(\tau) = 1 - \mathrm{e}^{-4\,\tau} \tag{1.17}$$

The tangent line to the graph of ϑ_2 in $\tau = 0$ attains a value of 1.0 for $\tau = 0.25 =: \tau_R$. Thus, τ_R may be considered as a dimensionless-time scale for the

propagation of the thermal disturbance throughout the bar. In dimensional terms:

$$\tau_R = \frac{\alpha\, t_R}{L^2} = 0.25 \Longrightarrow t_r = 0.25\, \frac{L^2}{\alpha} = \frac{(L/2)^2}{\alpha}$$

t_R can be interpreted as the time required so that a thermal disturbance propagates through a distance $L/2$.

The stability Fourier number can now be interpreted as the ratio between two time intervals: the time-step Δt and the *diffusion time scale* $(\Delta x)^2/\alpha$, required so that a thermal disturbance travels across a single computational cell. It is rather intuitive that the time-step must be shorter than the diffusion time scale: otherwise, the computational cell would not *feel* the passage of the disturbance.

1.3 Von Neumann analysis

Let us consider the heat equation in an unbounded domain:

$$\frac{\partial u}{\partial t} = \alpha\, \frac{\partial^2 u}{\partial x^2} \tag{1.18}$$

$$u(x,0) = h(x) \tag{1.19}$$

Applying the Fourier Transform (in space) yields

$$\frac{\partial \widehat{u}}{\partial t} = -\alpha\, v^2\, \widehat{u} \tag{1.20}$$

$$\widehat{u}(v,0) = \widehat{h}(u) \tag{1.21}$$

where

$$\widehat{u}(v,t) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} u(x,t)\, \mathrm{e}^{-i\,v\,x}\, \mathrm{d}x$$

and

$$u(x,t) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} \widehat{u}(v,t)\, \mathrm{e}^{i\,v\,x}\, \mathrm{d}v$$

yielding

$$\widehat{u}(v,t) = \widehat{h}(v)\, \mathrm{e}^{-\alpha\, v^2\, t} \tag{1.22}$$

Then,

$$u(x,t) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} \widehat{h}(v)\, \mathrm{e}^{-\alpha\, v^2\, t}\, \mathrm{e}^{i\, v\, x}\, \mathrm{d}v \tag{1.23}$$

A similar approach is used to figure out the solution of the semi-disctrete system of equations resulting from a central-difference discretization of (1.18):

$$\frac{\partial u}{\partial t} = \alpha\, \frac{u_{j+1} - 2\, u_j + u_{j-1}}{(\Delta x)^2} \tag{1.24}$$

In this case we assume (M+1)-periodicity in space so that $\{u_j\}_{j=0}^{M}$ can be represented by a Discrete Fourier Transform:

$$\widehat{u}_k(t) = \sum_{j=0}^{M} u_j(t)\, \mathrm{e}^{\,-i\, \dfrac{2\,\pi\, j\, k}{M+1}}$$

$$u_j(t) = \frac{1}{M+1} \sum_{k=0}^{M} \widehat{u}_k(t)\, \mathrm{e}^{\,i\, \dfrac{2\,\pi\, j\, k}{M+1}}$$

The following solution is readily found:

$$\widehat{u}_k(t) = \widehat{u}_k(0)\, \mathrm{e}^{-\lambda\, t} \tag{1.25}$$

with

$$\lambda = \frac{2\,\alpha}{(\Delta x)^2} \left[1 - \cos\left(\frac{2\,\pi\, k}{M+1}\right)\right] \tag{1.26}$$

Equation (1.25) provides an alternative physical interpretation of the *resolution* error induced by the central difference approximation. Namely, let us recast (1.25) as:

$$\widehat{u}_k(t) = \widehat{u}_k(0)\, \mathrm{e}^{-\alpha_{num}\, \widehat{k}^2\, t} \tag{1.27}$$

where

$$\Delta x \equiv \frac{L}{M+1}; \quad \widehat{k} \equiv \frac{2\,\pi\, k}{(M+1)\,\Delta x}; \quad \widetilde{k} \equiv \widehat{k}\,\Delta x = \frac{2\,\pi\, k}{M+1}$$

and

$$\alpha_{num} \equiv 2\,\alpha\,\frac{1-\cos\left(\widetilde{k}\right)}{\widetilde{k}^2}$$

The k-th Fourier mode of the exact solution is instead

$$\widehat{u}_k(t) = \widehat{u}_k(0)\,\mathrm{e}^{-\alpha\,\widehat{k}^2\,t} \tag{1.28}$$

Comparing (1.22) with (1.28) provides the following interpretation of the resolution error induced by the central-difference approximation of the second-derivative in the heat equation: the physical diffusivity α is replaced by an *artificial* diffusivity α_{num}. The ratio α_{num}/α is shown in figure 1.1. The

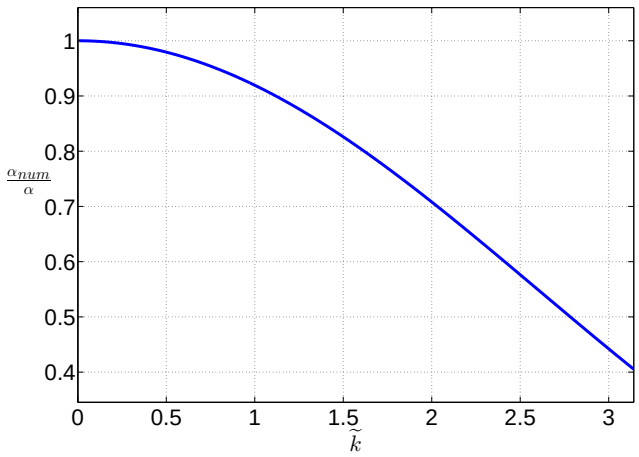

Figure 1.1: Numerical diffusivity.

highest-wavenumber modes are damped less efficiently in the numerical solution than in the exact solution: the numerical diffusivity α_{num} depends on the modified wavenumber $\widetilde{k}$, whereas the physical diffusivity in constant. A favorable feature of the central difference scheme is that, as evident from figure 1.1, the numerical diffusivity is always positive and, therefore, the all Fourier modes are progressively damped (i.e., no counter-diffusion occurs).

Exercise 1.1. *Derive and interpret the numerical dissipation for the case when the second derivative in the heat equation is approximated as*

$$\left.\frac{\delta^2}{\delta x^2}\right|_j = \frac{f_j - 2\,f_{j-1} + f_{j-2}}{(\Delta x)^2}$$

1.3.1 *Oscillatory* instability for the heat equation

Oscillatory instability for the heat equation

Consider the simple IVP:

$$\begin{cases} \dot{y} = -\alpha y \\ y(0) = y_0 \end{cases}$$

Explicit - Euler scheme yields:

$$\frac{y^{n+1}}{y^n} = 1 - \alpha \Delta t$$

The method becomes unstable for $\alpha \Delta t > 2$.
Nevertheless, consider what happens for $\alpha \Delta t > 1$:

$$\alpha \Delta t > 1 \Rightarrow \frac{y^{n+1}}{y^n} < 0$$

Thus, while the exact solution shows an exponential decay from y_0 towards 0, for $\alpha \Delta t > 1$ the numerical solution, though progressively decaying, develops an unphysical behavior: it changes sign from one time-step to the next.

The explicit Euler / central difference approximation of the heat equation

$$\begin{cases} \frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} & x \in (0, L) \\ u(0,t) = u(L,t) = 0 \\ u(x,0) = f(x) \end{cases}$$

can be cast as:

$$\vec{u} \equiv (u_1 \quad \dots \quad u_N)^T$$

$$\vec{u}^{n+1} = \frac{\alpha \Delta t}{(\Delta x)^2} [A] \vec{u}^n + \vec{u}^n =$$

$$[A] = \mathrm{Toe}(N, 1, -2, 1)$$

The eigenvalues:

$$\lambda_k = -2 + 2 \cos\left(\frac{k\pi}{N+1}\right), \; k = 1 \dots N$$

The largest-magnitude eigenvalue is:

$$\lambda_N = -2 + 2 \cdot \cos\left(\frac{N\pi}{N+1}\right) \simeq -4 \quad (\text{for large } N)$$

The method is stable when

$$\left| 1 + \frac{\alpha \Delta t}{\Delta x^2} \lambda_k \right| \leq 1 \Rightarrow |\lambda_k| \frac{\alpha \Delta t}{\Delta x^2} \leq 2 \quad \forall k$$

and develops oscillatory instability in time whenever

$$1 + \frac{\alpha \Delta t}{(\Delta x)^2} \lambda_k < 0 \quad \text{for some } k.$$

The most restrictive condition holds for $\lambda_N \simeq -4$, yielding

$$1 - 4\frac{\alpha \Delta t}{(\Delta x)^2} < 0 \Rightarrow \frac{\alpha \Delta t}{(\Delta x)^2} > \frac{1}{4}$$

Notice that the corresponding eigenvector is:

$$v_j^{(N)} = \sin\left(\frac{N \pi j}{N+1}\right), \quad j = 1 \ldots N$$

This eigenvector changes sign at each grid point:

$$v_{j+1}^{(N)} = \sin\left(\frac{N \pi j}{N+1}\right)\cos\left(\frac{N \pi}{N+1}\right) + \cos\left(\frac{N \pi j}{N+1}\right)\sin\left(\frac{N \pi}{N+1}\right)$$

$$\simeq -\sin\left(\frac{N \pi j}{N+1}\right) \quad (N \text{ large})$$

Thus, the component $\bar{v}^{(N)}$ of the numerical solution oscillates both in space and in

true while it decays towards zero

(provided $\frac{\alpha \Delta t}{(\Delta x)^2} < \frac{1}{2}$).

1.4 Fourier-Galerkin solution of the heat equation

Fourier - Galerkin solution of the heat equation

$$\begin{cases} \dfrac{\partial u}{\partial t} = \alpha \dfrac{\partial^2 u}{\partial x^2} & x \in (0,L) \qquad (1a) \\ u(0,t) = u(L,t) = 0 & \qquad (1b) \\ u(x,0) = f(x) & x \in (0,L) \qquad (1c) \end{cases}$$

(f(x) a media nulla su [0,L] ed f(0) = f(L) = 0)

ad es. doppia tenda...

We seek an approximate solution in the "trial space" S_{2N}, the set of all sine functions of degree $\leq N$:

$$S_{2N} \equiv \left\{ \sin\left(\frac{2\pi K x}{L}\right), -N \leq K \leq N-1, K \in \mathbb{Z} \right\}$$

Thus:

$$\hat{u}(x,t) = \sum_{-N}^{N-1} {}_K \, C_K(t) \sin(2\pi K x/L) \qquad (2)$$

Notice that $\hat{u}(x,t)$ automatically satisfies the boundary conditions (1b). In the Fourier - Galerkin method, the space of test functions is coincident with the space of trial functions.

At this point it is worth recalling that S_{2N} is an orthonormal space w.r.t. the inner product (of $L_2(0,L)$):

$$\langle f, g \rangle \equiv \frac{2}{L} \int_0^L f(x) g(x) dx$$

The residual of (1a) with the approximation (2) is:

$$R(\hat{u}) = \frac{\partial \hat{u}}{\partial t} - \alpha \frac{\partial^2 \hat{u}}{\partial x^2} = \sum_{-N}^{N} {}_K \left(\dot{C}_K + 4\pi^2 K^2 \alpha C_K \right) \sin\left(\frac{2\pi K x}{L}\right)$$

Then, following the method of weighted residuals:

$$(3)\quad 0 = \left\langle R(\hat{u}), \sin\left(\frac{2\pi m x}{L}\right)\right\rangle = \dot{c}_m + \frac{4\pi^2 m^2}{L^2} c_m, \quad -N \le m \le N-1$$

Thus, we are left with a set of independent IVPs, that are easily solved:

$$c_m(t) = c_m(0)\, e^{-\frac{4\pi^2 m^2 \alpha}{L^2} t} \qquad (3)$$

The coefficients $\{c_m(0)\}_{m=-N}^{N}$ are found using (1c):

$$\hat{f}(x) \equiv \sum_{k=-N}^{N-1} b_k \sin\left(\frac{2\pi k x}{L}\right)$$

$$\Rightarrow \quad b_k = \left\langle f(x), \sin\left(\frac{2\pi k x}{L}\right)\right\rangle$$

Then, by (1c) and linear independence of the considered sine functions, it is found that

$$c_m(0) \equiv b_m$$

Eventually:

$$\hat{u}(x,t) = \sum_{k=-N}^{N-1} b_k\, e^{-\frac{4\pi^2 k^2 \alpha}{L^2} t}$$

Now, let's solve (3) by the explicit Euler method:

$$c_m^{n+1} - c_m^n = -\frac{4\pi^2 m^2 \alpha}{L^2} \Delta t\, c_m^n$$

The amplification factor $r_m \equiv \left| \frac{C_m^{n+1}}{C_m^n} \right|$ is:

$$r_m = \left| 1 - \frac{4\pi^2 m^2 \alpha}{L^2} \Delta t \right|,$$

Stability of the method requires:

$$r_m \leq 1 \quad \forall m : |m| \leq N$$

$$\Rightarrow \frac{4\pi^2 m^2 \alpha}{L^2} \Delta t \leq 2, \quad -N \leq m \leq N-1$$

The most restrictive condition corresponds to $m = -N$:

$$\frac{4\pi^2 N^2}{L^2} \alpha \Delta t \leq 2$$

Define the grid-spacing Δx as:

$$\Delta x \equiv \frac{L}{2N} \qquad (4)$$

to get

$$\left(\frac{\alpha \Delta t}{\Delta x^2} \right) \leq \frac{2}{\pi^2} \simeq 0,203 \qquad (5)$$

According to Nyquist's sampling theorem, Δx, defined as in eq. (4), corresponds to the largest grid spacing that can be used to "resolve"

(i.e, identify) a Fourier mode of order N.
The stability limit (5) is about 2.5 times more stringent than the corresponding one, when the central difference scheme is used to approximate the second-order spatial derivative.

Remark: the $K=0$ mode is identically zero and must be disregarded.

1.5 Pseudo-spectral solution of the heat equation

Pseudo-spectral methods are collocation methods, i.e., the residual of the PDE is enforced to be zero at a set of *collocation* nodes. Let's consider the same problem faced already, i.e.

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \alpha \frac{\partial^2 u}{\partial x^2} \quad x \in [0, L] \\
u(x,t) &= 0 \quad t \geq 0 \\
u(L,t) &= 0 \quad t \geq 0 \\
u(x,0) &= f(x) \quad x \in [0, L]
\end{aligned}$$

We seek for an approximation $\widehat{u}(x,t)$ to the exact solution, in the form of a truncated sine series expansion:

$$\widehat{u}(x,t) = \sum_{k=1}^{N} c_k(t) \, \sin\left(\frac{2\,\pi\,k\,x}{L}\right)$$

The expansion inherently satisfies the boundary conditions. Since $\widehat{u}$ is inherently odd and zero-mean, we assume that the same properties hold for the initial condition $f(x)$.

The residual of the PDE reads:

$$R(\widehat{u}) = \frac{\partial \widehat{u}}{\partial t} - \alpha \frac{\partial^2 \widehat{u}}{\partial x^2} = \sum_{k=1}^{N} \left[\dot{c}_k\,(t) + \frac{4\,\pi^2\,k^2}{L^2} c_k(t)\right] \, \sin\left(\frac{2\,\pi\,k\,x}{L}\right)$$

In collocation methods, the residual is requested to be zero at the collocation nodes, which, in the present case, we define as

$$x_j \equiv \frac{j\,L}{N+1}, \; j = 1, \, \ldots, \, N$$

The conditions on the residual result in the following system of ODEs:

$$0 = \sum_{k=1}^{N} \left[\dot{c}_k\,(t) + \frac{4\,\pi^2\,k^2}{L^2} c_k(t)\right] \, \sin\left(\frac{2\,\pi\,k\,x_j}{L}\right), \; j = 1, \, \ldots, \, N \qquad (1.29)$$

These equations can be de-coupled by introducing the following, discrete inner product:

$$\langle f, g \rangle \equiv \frac{2}{N+1} \, \sum_{j=0}^{N} f(x_j)\, g(x_j)$$

which, for sine functions, can be re-defined as

$$\langle f, g \rangle \equiv \frac{2}{N+1} \sum_{j=1}^{N} f(x_j)\, g(x_j)$$

The following orthogonality conditions do hold:

$$\left\langle \sin\left(\frac{2\,\pi\,k\,x_j}{L}\right),\, \sin\left(\frac{2\,\pi\,m\,x_j}{L}\right) \right\rangle = \delta_{m\,k}$$

Taking the inner product of (1.29) with all the sine functions in the expansion yields:

$$\dot{c_k}\,(t) + \frac{4\,\pi^2\,k^2}{L^2} c_k(t) = 0,\ k = 1, \ldots, N$$

This is the same set of ODEs derived by the spectral-Galerkin method and its numerical solution is subject to the same stability constraints already mentioned. There is a minor difference in the calculation of the initial conditions, though:

$$f(x_j) = \sum_{k=1}^{N} c_k(0)\, \sin\left(\frac{2\,\pi\,k\,x_j}{L}\right) \Longrightarrow \left\langle \sin\left(\frac{2\,\pi\,m\,x}{L}\right),\, f(x) \right\rangle = c_m(0),\ m = 1,\, \ldots,\, N$$

That is, $c_0(t)$ is calculated via the discrete inner product rather via the L_2 inner product.

1.6 Homework

Derive the PDE modelling unsteady water flooding through a saturated, permeable soil. Water is considered compressible:

$$\rho_w = \rho_w(p_w); \qquad \gamma \equiv \frac{1}{\rho_w} \frac{\mathrm{d}\rho_w}{\mathrm{d}p_w}$$

In this example, a rigid porous matrix is assumed. Darcy's law relates the water flow through the porous media with the applied pressure gradient:

$$\boldsymbol{q}''_w = -\frac{k}{\mu} \nabla \left(p_w - \rho_w\, \boldsymbol{g} \cdot \boldsymbol{x}\right)$$

Given a surface S within the porous matrix, the water flow rate through the surface is given by

$$Q_w = \int_S \boldsymbol{q}''_w \cdot \mathrm{d}\boldsymbol{S}$$

You must use two main concepts:

1. water flows throughout the boundaries of a control volume under the action of a pressure gradient
2. when pressure increases, water is squeezed: so, a larger mass of water can be stored within the pores of the porous media

Then, suggest an approach for solving the 1D-PDE with, say, Dirichlet boundary conditions.

Eventually, implement the proposed algorithm in a computer program and simulate a flow through a porous medium.

1.7 More accurate time-stepping schemes for the heat equation

1.7.1 Leapfrog method

Let's start by applying the Leapfrog time-stepping scheme to the heat equation:

$$\frac{u^{n+1} - u^{n-1}}{2\,\Delta t} = \alpha\,\frac{\delta u^n}{\delta x^2} \tag{1.30}$$

Applying the von Neumann analysis with the second-order, central difference scheme for $\delta^2/\delta x^2$ yields a difference equation, for the Fourier mode with modified wavenumber $\widehat{k}$:

$$\sigma^2 + 4\,\beta\,\sigma\,\left(1 - cos\left(\widehat{k}\right)\right) - 1 = 0 \tag{1.31}$$

with general solution

$$\sigma^n = A\,\lambda_+^n + B\,\lambda_-^n \tag{1.32}$$

where

$$\lambda_\pm = -2\,\beta\,\left(1 - \cos\left(\widehat{k}\right)\right) \pm \sqrt{1 + 4\,\beta^2\,\left(1 - \cos\left(\widehat{k}\right)\right)^2} \in \mathbb{R}$$

Using the explicit Euler method as starting scheme, we end up with:

$$\sigma^1 = \sigma^0\,\left(1 - 2\,\beta\,\left(1 - \cos\left(\widehat{k}\right)\right)\right)$$

The constants A and B are derived from the initial conditions:

$$\begin{aligned} \sigma^0 &= A + B \\ \sigma^1 &= A\,\lambda_+ + B\,\lambda_- \end{aligned}$$

It is worth noticing that $|\lambda^-| \geq 1$ for any $\widehat{k}$ and any β, showing that the method is unconditionally unstable, unless the starting time-stepping scheme and the initial solution are chosen in such a way to annihilate B.

1.7.2 Dufort-Frankel method

Dufort and Frankel tried to *cure* the Leapfrog method by replacing the second difference $\delta^2 u^n/\delta x^2$ with:

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{x_j,t} \approx \frac{u_{j+1}^n - (u_j^{n+1} + u_j^{n-1}) + u_{j-1}^n}{(\Delta x)^2}$$

yielding the following approximation for the heat equation:

$$(1+2\,\beta)\;u_j^{n+1} = (1-2\,\beta)\;u_j^{n-1} + (2\,\beta)\;u_{j-1}^n + (2\,\beta)\;u_{j+1}^n \tag{1.33}$$

Carrying out the von Neumann analysis highlights some peculiarities of the method:

- The method is absolutely stable.
- Considering for simplicity the highest modified wavenumber, $\widehat{k} = \pi$, yields the following two roots for the amplification factor σ:

  $$\sigma^{\pm} = \frac{-2\beta \pm 1}{1+2\,\beta}$$

  Thus, the magnitude of σ^+ is smaller than unity, while σ^- equals -1, irrespective of the value of β. This shows that one of the components of the solution is not damped at all and changes sign from step to step. This, in turn, suggests that the method is not convergent, in general.
- For $\beta > 0.5$ also σ^+ becomes negative, and both components of the solution switch sign from a time-step to the next.

The dependence of u_j^{n+1} from the values of u at the neighboring locations and previous time-steps is shown in figure 1.2: it is not what would be expected for a parabolic equation.

The modified PDE associated to the heat equation via the Dufort-Frankel method is:

$$\frac{\partial u}{\partial t} - \alpha\,\frac{\partial^2 u}{\partial x^2} = -\frac{(\Delta t)^2}{6}\,\frac{\partial^3 u}{\partial t^3} + \frac{\alpha\,(\Delta x)^2}{12}\,\frac{\partial^4 u}{\partial x^4} - \alpha\left(\frac{\Delta t}{\Delta x}\right)^2\,\frac{\partial^2 u}{\partial t^2} - \frac{\alpha\,(\Delta t)^4}{12\,(\Delta x)^2}\,\frac{\partial^4 u}{\partial t^4} + \ldots \tag{1.34}$$

Equation (1.34) is noteworthy in several respects:

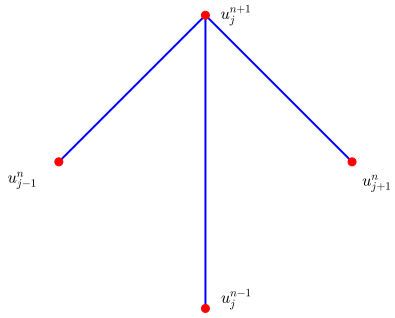

Figure 1.2: Computational molecule for the Dufort-Frankel scheme.

1. It shows that the method is inconsistent unless, besides Δt and Δx tending to zero, also $\Delta t/\Delta x$ tends to zero.

2. Whenever Δt and Δx are small but, at the same time, $(\Delta t/\Delta x)^2$ is not, the modified PDE takes on a hyperbolic character:

$$\left(\frac{\Delta t}{\Delta x}\right)^2 \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} \approx -\frac{1}{\alpha}\frac{\partial u}{\partial t}$$

1.8 Towards a cost-effective solution of multi-dimensional, parabolic problems

1.8.1 Stability limit for the explicit solution of the two-dimensional heat equation on a rectangular domain

Let's consider the two-dimensional heat equation on a rectangular domain, with Dirichlet boundary conditions on $\partial\Omega$:

$$\begin{aligned} u_t &= \alpha\left(u_{xx} + u_{yy}\right) && \boldsymbol{x} \in \Omega \equiv [a,b]\times[c,d];\ t>0 \\ u &= \gamma(\boldsymbol{x},t) && \boldsymbol{x} \in \partial\Omega,\ t>0 \\ u &= \phi(\boldsymbol{x}) && \boldsymbol{x} \in \Omega\cup\partial\Omega,\ t=0 \end{aligned} \tag{1.35}$$

With suitable changes of dependent variable it is readily recognized that the solution of (1.35) can be obtained by combining a particular solution of a stationary Poisson problem with homogeneous boundary conditions with

the solution of a homogeneous heat equation with homogeneous boundary conditions:

$$\begin{cases} \Delta u_1 = 0 & \text{in } \Omega \\ u_1 = \gamma & \text{on } \partial\Omega \end{cases} \tag{1.36a}$$

$$\begin{cases} \dfrac{\partial u_2}{\partial t} - \alpha\,\Delta u_2 = 0 & \text{in } \Omega \\ u_2 = 0 & \text{on } \partial\Omega \\ u_2 = \phi - u_1 & \text{at } t = 0 \end{cases} \tag{1.36b}$$

Therefore, we focus the following analysis to the heat equation with homogeneous Dirichlet conditions on the boundary.

Consider an explicit-Euler / second-order central difference discretization. The eigenvectors of the two-dimensional, difference operator satisfying homogeneous Dirichlet boundary conditions are obtained as tensor product of the corresponding eigenvectors for the 1D operator:

$$v_{j\,k}^{(m\,n)} = \sin\left(\frac{m\,\pi\,j}{M}\right)\,\sin\left(\frac{n\,\pi\,k}{N}\right),\; j = 1,\,\ldots,\,M,\quad k = 1,\,\ldots,\,N$$

They are mutually orthogonal w.r.t. the conventional inner product in $\mathbb{R}^{M\times N}$. Expanding the finite-difference solution as a linear conbination of eigenvectors and substituting in the finite-difference equations yields the amplification factor $\rho^{(m,n)}$ for the (m,n) component as:

$$\rho^{(m,n)} = 1 - 2\,\left[\widehat{\beta}_x + \widehat{\beta}_y\right]$$

where

$$\widehat{\beta}_x \equiv \beta_x\,\left(1 - \cos\left(\frac{m\,\pi}{M}\right)\right);\quad \widehat{\beta}_y \equiv \beta_y\,\left(1 - \cos\left(\frac{n\,\pi}{N}\right)\right)$$

and

$$\beta_x \equiv \frac{\alpha\,\Delta t}{(\Delta x)^2};\quad \beta_y \equiv \frac{\alpha\,\Delta t}{(\Delta y)^2}$$

The most stringent stability condition $\left|\rho^{(m,n)}\right| \leq 1$ applies for $m = M$ and $n = N$ and, for M, N large, yields:

$$\beta_x + \beta_y \leq \frac{1}{2}$$

If $\Delta x = \Delta y$ the stability requirement is $\alpha\,\Delta t/(\Delta x)^2 \leq 1/4$, which is more restrictive than in one dimension. Thus, there is an even greater motivation to study implicit methods in two dimensions.

1.9 Alternate Direction Implicit methods

An ADI method (in 2D) is a two-step iteration process that alternately updates the column and row spaces of an approximate solution to $\boldsymbol{AX} - \boldsymbol{XB} = \boldsymbol{C}$, where $\boldsymbol{X}$, $\boldsymbol{C} \in \mathbb{R}^{M\times N}$, $\boldsymbol{A} \in \mathbb{R}^{M\times M}$, $\boldsymbol{B} \in \mathbb{R}^{N\times N}$.

1.9.1 Peaceman - Rachford method

The Peaceman - Rachford method (PRM) is a predictor-corrector method for the solution of parabolic equations. In the following, the PRM is applied for the numerical solution of the unsteady, two-dimensional heat equation on a Cartesian grid, which is uniform along each independet coordinate direction. The mesh spacings along x and y are denoted by h_x and h_y, respectively. The time-step is denoted by Δt. When discrete, space-differenting operators are not specified, they are denoted as $\delta^2/\delta x^2$ and $\delta^2/\delta y^2$.

Predictor step The predictor step may be interpreted as a mixed Backward-Euler / Forward-Euler time-stepping scheme, over a step length $\Delta t/2$ from t_n to $t_{n+1/2}$:

$$u_{ij}^{n+1/2} - u_{ij}^{n} = \alpha \frac{\Delta t}{2} \left[\frac{\delta^2 u_{ij}^{n+1/2}}{\delta x^2} + \frac{\delta^2 u_{ij}^{n}}{\delta y^2} \right] \tag{1.37}$$

Corrector step The corrector step may be interpreted as a mixed Backward-Euler / Forward-Euler time-stepping scheme, over a step length $\Delta t/2$ from $t_{n+1/2}$ to t_{n+1}. The Backward- and Forward- time-stepping schemes are applied to the discrete, second-difference operators with inverse order with respect to the predictor step.

$$u_{ij}^{n+1} - u_{ij}^{n+1/2} = \alpha \frac{\Delta t}{2} \left[\frac{\delta^2 u_{ij}^{n+1/2}}{\delta x^2} + \frac{\delta^2 u_{ij}^{n+1}}{\delta y^2} \right] \tag{1.38}$$

The PRM can be interpreted as a combination of a mid-point time-stepping scheme for $\delta^2/\delta x^2$ and a trapezoidal time-stepping scheme for $\delta^2/\delta y^2$. Indeed, summing up equations (1.37) and (1.38) yields:

$$u_{ij}^{n+1} - u_{ij}^{n} = \alpha\, \Delta t \frac{\delta^2 u_{ij}^{n+1/2}}{\delta x^2} + \alpha \frac{\Delta t}{2} \left[\frac{\delta^2 u_{ij}^{n+1}}{\delta y^2} + \frac{\delta^2 u_{ij}^{n}}{\delta y^2} \right] \tag{1.39}$$

This would suggest that the PRM is second-order accurate in time.

The issue of boundary conditions

The question arises, about the proper boundary conditions to apply to $u^{n+1/2}$. Indeed, $u^{n+1/2}$ is a rather poor ($o(\Delta t)$) approximation to the exact solution at $t_n + \Delta t/2$. We can single-out $u^{n+1/2}$ from equations (1.37) and (1.38). To this end, let's get $\delta^2 u^{n+1/2}/\delta x^2$ from (1.37) and substitute it into (1.38), to get:

$$\begin{aligned} u_{i\,j}^{n+1/2} &= \frac{u_{i\,j}^{n+1} + u_{i\,j}^{n}}{2} - \alpha \frac{\Delta t}{4} \left[\frac{\delta^2 u_{i\,j}^{n+1}}{\delta y^2} - \frac{\delta^2 u_{i\,j}^{n}}{\delta y^2} \right] \\ &= \frac{1}{2} \left[1 + \alpha \frac{\Delta t}{2} \frac{\delta^2}{\delta y^2} \right] u_{i\,j}^{n} + \frac{1}{2} \left[1 - \alpha \frac{\Delta t}{2} \frac{\delta^2}{\delta y^2} \right] u_{i\,j}^{n+1} \end{aligned} \tag{1.40}$$

Let's define the arithmetic mean of $u_{i\,j}$ between t_n and t_{n+1},

$$\overline{u}_{i\,j} := \frac{u_{i\,j}^{n+1} + u_{i\,j}^{n}}{2}$$

and the *jump* in $u_{i\,j}$ between t_n and t_{n+1},

$$\Delta u_{i\,j} := u_{i\,j}^{n+1} - u_{i\,j}^{n}$$

Equation (1.40) can be rewritten as

$$u_{i\,j}^{n+1/2} = \overline{u}_{i\,j} - \alpha \frac{\Delta t}{2} \frac{\delta^2 \Delta u}{\delta y^2} \tag{1.41}$$

Equation (1.41) provides proper boundary conditions for the intermediate field $u^{n+1/2}$. Nevertheless, these conditions are rarely applied due to the inherent complexity of evaluating the second derivatives at the boundaries. In addition notice that, when the boundary conditions are independent of time, (1.41) is equivalent to the *standard* boundary condition

$$u_{i\,j}^{n+1/2} = u_{i\,j}^{n+1} \tag{1.42}$$

Von Neumann stability analisys

Let's assume that the heat equation is solved on a rectangular mesh, composed of $(N_x+1)\times(N_y+1)$ nodes. Periodic boundary conditions are enforced between opposite boundaries: thus, $u_{i\,j} = u_{i+N_x\,j+N_y} \forall\ i,\ j$. The following DFT representation can be used:

$$u_{m\,n} = \frac{1}{N_y\,N_y} \sum_{p=0}^{N_x-1} \sum_{q=0}^{N_y-1} U_{p\,q}\, \mathrm{e}^{i\,\frac{2\pi\,p\,m}{N_x}}\, \mathrm{e}^{i\,\frac{2\pi\,q\,n}{N_y}} \tag{1.43}$$

where

$$U_{p\,q} = \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} u_{m\,n}\, \mathrm{e}^{-i\,\frac{2\pi\,p\,m}{N_x}}\, \mathrm{e}^{-i\,\frac{2\pi\,q\,n}{N_y}} \tag{1.44}$$

The following sets of Fourier coefficients are identified:

$$\begin{aligned} u_{r\,s}^{n} &\longleftrightarrow U_{p\,q}^{n} \\ u_{r\,s}^{n+1/2} &\longleftrightarrow U_{p\,q}^{n+1/2} \\ u_{r\,s}^{n+1} &\longleftrightarrow U_{p\,q}^{n+1} \end{aligned}$$

The following orthogonality property holds:

$$(\ ,\ )_{N_x,N_y} := \frac{1}{N_x\,N_y} \sum_{r=0}^{N_x-1} \sum_{s=0}^{N_y-1} \mathrm{e}^{i\,\frac{2\pi\,p\,r}{N_x}}\, \mathrm{e}^{i\,\frac{2\pi\,q\,s}{N_y}} = \delta_{p,\,q} \tag{1.45}$$

Substituting the DFT expansions into (1.37) and (1.38) while using (1.45) yields:

For the predictor step:

$$0 = U_{p\,q}^{n+1/2} - U_{p\,q}^{n} + \beta_x \left[1 - \cos\left(\frac{2\pi\,p}{N_x}\right)\right] U_{p\,q}^{n+1/2} + \beta_y \left[1 - \cos\left(\frac{2\pi\,q}{N_y}\right)\right] \tag{1.46}$$

or, equivalently,

$$U_{p\,q}^{n+1/2} = U_{p\,q}^{n}\, \frac{1 - \widehat{\beta}_y}{1 + \widehat{\beta}_x} \tag{1.47}$$

where

$$\begin{aligned} \widehat{\beta}_x &\equiv \beta_x \left[1 - \cos\left(\frac{2\pi\,p}{N_x}\right)\right] \\ \widehat{\beta}_y &\equiv \beta_y \left[1 - \cos\left(\frac{2\pi\,q}{N_y}\right)\right] \end{aligned}$$

Defining the modified wavenumbers $\widehat{k}_x$ and $\widehat{k}_y$ as

$$\widehat{k}_x := \frac{2\,\pi\,p\,\Delta x}{N_x\,\Delta x}; \qquad \widehat{k}_y := \frac{2\,\pi\,q\,\Delta y}{N_y\,\Delta x}$$

we end up with

$$\begin{aligned}\widehat{\beta}_x &\equiv \beta_x \left[1-\cos\left(\widehat{k}_x\right)\right]\\ \widehat{\beta}_y &\equiv \beta_y \left[1-\cos\left(\widehat{k}_y\right)\right]\end{aligned}$$

For the corrector step:

$$U_{p\,q}^{n+1} = U_{p\,q}^{n+1/2}\,\frac{1-\widehat{\beta}_x}{1+\widehat{\beta}_y} \tag{1.48}$$

Amplification factor for the Peaceman - Rachford method Gathering (1.47) and (1.48) yields:

$$U_{p\,q}^{n+1} = U_{p\,q}^{n}\,\frac{1-\widehat{\beta}_x}{1+\widehat{\beta}_y}\,\frac{1-\widehat{\beta}_y}{1+\widehat{\beta}_x} \tag{1.49}$$

Thus, the amplification factor for the numerical solution is:

$$\rho\left(\widehat{k}_x,\widehat{k}_y\right) := \frac{U_{p\,q}^{n+1}}{U_{p\,q}^{n}} = \frac{1-\widehat{\beta}_x}{1+\widehat{\beta}_y}\,\frac{1-\widehat{\beta}_y}{1+\widehat{\beta}_x} \tag{1.50}$$

The magnitude of the amplification factor is invariably smaller than unity. Higher wavenumbers, corresponding to higher $\widehat{\beta}$ for fixed β, are damped more significantly than lower wavenumbers. The amplification factor vanishes for $\widehat{\beta}_x = 1$ or $\widehat{\beta}_y = 1$.

The amplification factor is represented in figure 1.3 with the (rather reasonable) assumption $\widehat{\beta}_x = \widehat{\beta}_y \equiv \widehat{\beta}$.

The amplification factor for the exact solution is:

$$\widetilde{\rho}(\widehat{k}_x,\widehat{k}_y) = \mathrm{e}^{-\widehat{k}_x^2\,\beta_x}\;\mathrm{e}^{-\widehat{k}_y^2\,\beta_y} = 1-\beta_x\,\widehat{k}_x^2-\beta_y\,\widehat{k}_y^2+\frac{1}{2}\left[\beta_x\,\widehat{k}_x^2+\beta_y\,\widehat{k}_y^2\right]^2+\ldots+\frac{1}{2}\left(\beta_y\,\widehat{k}_y^2\right)^2 \tag{1.51}$$

The Taylor expansion of ρ is instead:

$$\rho(\widehat{k}_x,\widehat{k}_y) = 1-\beta_x\,\widehat{k}_x^2-\beta_y\,\widehat{k}_y^2+\frac{1}{12}\,\beta_x\,\widehat{k}_x^4+\frac{1}{12}\,\beta_y\,\widehat{k}_y^4+\ldots \tag{1.52}$$

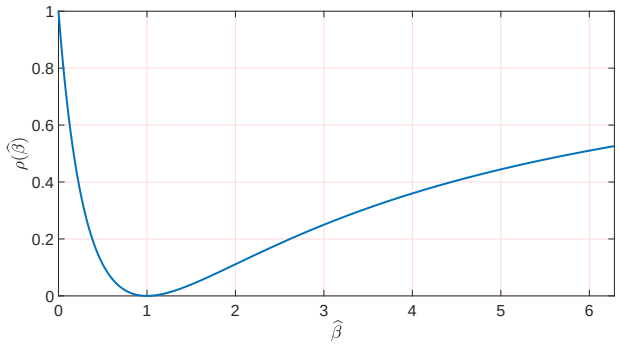

Figure 1.3: Aplification factor for the Peaceman-Rachford method.

Thus:

$$\begin{aligned}
\rho - \widetilde{\rho} \;=\;& \frac{\beta_x{}^3\,\widehat{k}_x^8}{16} - \frac{\beta_x{}^3\,\widehat{k}_x^6}{12} + \frac{\beta_x{}^2\,\beta_y\,\widehat{k}_x^6\,\widehat{k}_y^2}{12} + \frac{\beta_x{}^2\,\beta_y\,\widehat{k}_x^4\,\widehat{k}_y^4}{24} + \frac{\beta_x{}^2\,\widehat{k}_x^8}{160} \\
-\;& \frac{\beta_x{}^2\,\widehat{k}_x^6}{12} + \frac{\beta_x\,\beta_y{}^2\,\widehat{k}_x^4\,\widehat{k}_y^4}{24} + \frac{\beta_x\,\beta_y{}^2\,\widehat{k}_x^2\,\widehat{k}_y^6}{12} + \frac{\beta_x\,\beta_y\,\widehat{k}_x^6\,\widehat{k}_y^2}{360} \\
+\;& \frac{\beta_x\,\beta_y\,\widehat{k}_x^4\,\widehat{k}_y^4}{144} - \frac{\beta_x\,\beta_y\,\widehat{k}_x^4\,\widehat{k}_y^2}{12} + \frac{\beta_x\,\beta_y\,\widehat{k}_x^2\,\widehat{k}_y^6}{360} - \frac{\beta_x\,\beta_y\,\widehat{k}_x^2\,\widehat{k}_y^4}{12} \\
-\;& \frac{\beta_x\,\widehat{k}_x^{10}}{1814400} + \frac{\beta_x\,\widehat{k}_x^8}{20160} - \frac{\beta_x\,\widehat{k}_x^6}{360} + \frac{\beta_x\,\widehat{k}_x^4}{12} + \frac{\beta_y{}^3\,\widehat{k}_y^8}{16} - \frac{\beta_y{}^3\,\widehat{k}_y^6}{12} \\
+\;& \frac{\beta_y{}^2\,\widehat{k}_y^8}{160} - \frac{\beta_y{}^2\,\widehat{k}_y^6}{12} - \frac{\beta_y\,\widehat{k}_y^{10}}{1814400} + \frac{\beta_y\,\widehat{k}_y^8}{20160} - \frac{\beta_y\,\widehat{k}_y^6}{360} + \frac{\beta_y\,\widehat{k}_y^4}{12}
\end{aligned} \tag{1.53}$$

This shows that, for fixed wavenumbers, the method is only first-order accurate in time.

1.9.2 Approximate-Factorization method

Let's apply an implicit-Euler discretization in time of the two-dimensional heat equation:

$$u^{n+1} - u^n = \alpha\,\Delta t\,\left[\frac{\delta^2 u^{n+1}}{\delta x^2} + \frac{\delta^2 u^{n+1}}{\delta y^2}\right] \tag{1.54}$$

Consider the following approximate factorization:

$$\left(1 - \alpha\,\Delta t\,\frac{\delta}{\delta x^2}\right)\,\left(1 - \alpha\,\Delta t\,\frac{\delta}{\delta y^2}\right)\,u^{n+1} = u^{n+1} - \alpha\,\Delta t\,\left[\frac{\delta}{\delta x^2} + \frac{\delta}{\delta y^2}\right] + o\left((\Delta t)^2\right) \tag{1.55}$$

Then, let's solve an approximate form of the heat equation:

$$\left(1 - \alpha\,\Delta t\,\frac{\delta}{\delta x^2}\right) u^* \;=\; u^n \tag{1.56}$$

$$\left(1 - \alpha\,\Delta t\,\frac{\delta}{\delta y^2}\right) u^{n+1} \;=\; u^* \tag{1.57}$$

What are the boundary conditions for u^*? Answ: they can be derived from (1.57).

Von Neumann analysis

$$u_{j,l} = \frac{1}{M\,N} \sum_{k_x=0^{M-1}} \sum_{k_y=0}^{N-1} \widehat{u}_{k_x,\,k_y}\, \exp i\,\frac{2\pi j k_x}{M}\, \exp i\,\frac{2\pi l k_y}{N}$$

$$\frac{\delta u}{\delta x^2} = \frac{1}{M\,N} \sum_{k_x=0^{M-1}} \sum_{k_y=0}^{N-1} \frac{\beta(k_x)}{(\Delta x)^2} \widehat{u}_{k_x,\,k_y}\, \exp i\,\frac{2\pi j k_x}{M}\, \exp i\,\frac{2\pi l k_y}{N}$$

with

$$\beta(k_x) \equiv -2\,\left[1 - \cos\left(\frac{2\,\pi\,k_x}{M}\right)\right]$$

Introducing the *modified* wavenumber:

$$\widehat{k}_x := \frac{2\,\pi\,k_x}{M}$$

yields:

$$\beta(\widehat{k}_x) \equiv -2\,\left[1 - \cos\widehat{k}_x\right]$$

Substituting into (1.56) and (1.57):

$$\begin{aligned}
\left[1 - \frac{\alpha\,\Delta t}{(\Delta x)^2}\,\beta\left(\widehat{k}_x\right)\right] \widehat{u}^*_{\widehat{k}_x,\widehat{k}_x} &= \widehat{u}^n_{\widehat{k}_x,\widehat{k}_x} \\
\left[1 - \frac{\alpha\,\Delta t}{(\Delta x)^2}\,\beta\left(\widehat{k}_x\right)\right] \widehat{u}^{n+1}_{\widehat{k}_x,\widehat{k}_x} &= \widehat{u}^*_{\widehat{k}_x,\widehat{k}_x}
\end{aligned}$$

The amplification factor turns out to be:

$$\widehat{\rho}\left(\widehat{k}_x, \widehat{k}_y, \frac{\alpha\,\Delta t}{(\Delta x)^2}, , \frac{\alpha\,\Delta t}{(\Delta y)^2}\right) := \frac{\widehat{u}^{n+1}_{\widehat{k}_x,\widehat{k}_x}}{\widehat{u}^n_{\widehat{k}_x,\widehat{k}_x}} = \frac{1}{\left[1 - \frac{\alpha\,\Delta t}{(\Delta x)^2}\,\beta(\widehat{k}_x)\right]\,\left[1 - \frac{\alpha\,\Delta t}{(\Delta y)^2}\,\beta(\widehat{k}_y)\right]}$$

Let's compare the Taylor series expansions of the numerical ($\widehat{\rho}$) and exact ($\widehat{\rho}^*$) solutions. As for the numerical solution, assuming $\Delta x = \Delta y$:

$$\widehat{\rho}\left(\widehat{k}_x, \widehat{k}_y, \frac{\alpha\,\Delta t}{(\Delta x)^2}, , \frac{\alpha\,\Delta t}{(\Delta y)^2}\right) = 1 - D\left[\widehat{k}_x^2 + \widehat{k}_y^2 - \frac{\widehat{k}_x^4 + \widehat{k}_y^4}{12}\right] + D^2\left[\widehat{k}_x^4 + \widehat{k}_x^2\,\widehat{k}_y^2 + \widehat{k}_y^4\right] + \ldots$$

$$\widehat{\rho}^*\left(\widehat{k}_x, \widehat{k}_y, \frac{\alpha\,\Delta t}{(\Delta x)^2}, , \frac{\alpha\,\Delta t}{(\Delta y)^2}\right) = 1 - D\left[\widehat{k}_x^2 + \widehat{k}_y^2\right] + \frac{D^2}{2}\left[\widehat{k}_x^2 + \widehat{k}_y^2\right]^2 + \ldots$$

Thus, the Approximate-Factorization method is only first-order accurate in time.... (convergence error).

1.10 Finite-volume solution for the 2D heat equation on a Cartesian, non-uniform grid

The two-dimensional heat equation is solved on a Cartesian, non-uniform mesh by a cell-centred, second-order accurate both in space and time, finite-volume method. The Peacemann-Rachford ADI method is used to efficiently decouple the solution along the two coordinate directions. The *ghost-cell* method is used to enforce the boundary conditions. At each time-step, the coordinate directions treated implicitly within each ADI sub-step are interchanged. A co-located mesh arrangement is used. The algorithm has been implemented both in Fortran90 and in Matlab.

The diffusive fluxes across the cell's boundaries are approximated as follows:

$$\left.\frac{\delta u}{\delta x}\right|_w = \frac{u_{i,j} - u_{i-1,j}}{\delta x_{i-1}}$$

$$\left.\frac{\delta u}{\delta x}\right|_e = \frac{u_{i+1,j} - u_{i,j}}{\delta x_i}$$

$$\left.\frac{\delta u}{\delta y}\right|_s = \frac{u_{i,j} - u_{i,j-1}}{\delta y_{j-1}}$$

$$\left.\frac{\delta u}{\delta y}\right|_n = \frac{u_{i,j+1} - u_{i,j}}{\delta y_j}$$

Let's write down the algebraic equations to be solved during each sub-step.

Sub-step 1 The x direction is treated implicitly. For any cell (i, j):

$$\left(u_{i,j}^{n+1/2} - u_{i,j}^{n}\right) \Delta x_i \, \Delta y_j = \frac{\alpha \, \Delta t}{2} \left[\left. \frac{\delta u}{\delta x} \right|_e - \left. \frac{\delta u}{\delta x} \right|_w \right]^{n+1/2} \Delta y_j + \frac{\alpha \, \Delta t}{2} \left[\left. \frac{\delta u}{\delta y} \right|_n - \left. \frac{\delta u}{\delta y} \right|_s \right]^{n} \Delta x_i \tag{1.58}$$

A set of N independent linear, tridiagonal, algebraic $M \times M$ systems must be solved, one for each $j = 1, \ldots, N$. The boundary conditions on the west and east boundaries are used to calculate $u_{0,j}$ and $u_{M+1,j}$. Considering Dirichlet conditions applied on the west boundary $x = x_w$, $u(x_w, y, t) = f_w(y)$, the equation $(i = 1, j)$ can be modified as follows:

$$AW_{1,j} \, u_{0,j} + AP_{1,j} \, u_{1,j} + AE_{1,j} \, u_{2,j} = b_{1,j} \tag{1.59}$$

$$\frac{u_{0,j} + u_{1,j}}{2} \approx f_w(yc_j) \Longrightarrow u_{0,j} = -u_{1,j} + 2 \, f_w(yc_j)$$

where yc_j denotes the $y-$coordinate of the $j-$th row of cells. Substituting into (1.59):

$$(AP_{1,j} - AW_{1,j}) \, u_{1,j} + AE_{1,j} \, u_{2,j} = b_{1,j} + 2 \, AW_{1,j} \, f_w(yc_j)$$

Similarly, the boundary condition

$$-\alpha \left. \frac{\partial u}{\partial x} \right|_{x_w} = q_w(yc_j)$$

yields:

$$-\alpha \, \frac{u_{1,j} - u_{0,j}}{\delta x_0} \approx q_w(yc_j) \Longrightarrow u_{0,j} = u_{1,j} - q_w(yc_j) \frac{\delta x_0}{\alpha}$$

$$(AP_{1,j} + AW_{1,j}) \, u_{1,j} + AE_{1,j} \, u_{2,j} = b_{1,j} + AW_{1,j} \, q_w(yc_j) \, \frac{\delta x_0}{\alpha}$$

Sub-step 2 The y direction is treated implicitly. For any cells (i, j):

$$\begin{aligned} \left(u_{i,j}^{n+1} - u_{i,j}^{n+1/2}\right) \Delta x_i \, \Delta y_j \;\; = \;\; & \frac{\alpha \, \Delta t}{2} \left[\left. \frac{\delta u}{\delta y} \right|_n - \left. \frac{\delta u}{\delta y} \right|_s \right]^{n+1} \Delta x_i \\ + \;\; & \frac{\alpha \, \Delta t}{2} \left[\left. \frac{\delta u}{\delta x} \right|_e - \left. \frac{\delta u}{\delta x} \right|_w \right]^{n+1/2} \Delta y_j \end{aligned} \tag{1.60}$$

A set of M independent linear, tridiagonal, algebraic $N \times N$ systems must be solved, one for each $i = 1, \ldots, M$. The boundary conditions on the south

and north boundaries are used to calculate $u_{i,0}$ and $u_{i,N+1}$. Considering Dirichlet conditions applied on the south boundary $y = y_s$, $u(x, y_s, t) = f_s(x)$, the equation $(i, j = 1)$ can be modified as follows:

$$AS_{i,1}\, u_{i,0} + AP_{i,1}\, u_{i,1} + AN_{i,1}\, u_{i,2} = b_{i,1} \tag{1.61}$$

$$\frac{u_{i,0} + u_{i,1}}{2} \approx f_s(xc_i) \Longrightarrow u_{i,0} = -u_{i,1} + 2\, f_s(xc_i)$$

Substituting into (1.61):

$$(AP_{i,1} - AS_{i,1})\, u_{i,1} + AN_{i,1}\, u_{i,2} = b_{i,1} + 2\, AS_{i,1}\, f_s(xc_i)$$

Similarly, the boundary condition

$$-\alpha \left.\frac{\partial u}{\partial y}\right|_{y_s} = q_s(xc_i)$$

yields:

$$-\alpha \frac{u_{i,1} - u_{i,0}}{\delta y_0} \approx q_s(xc_i) \Longrightarrow u_{i,0} = u_{i,1} - q_s(xc_i)\frac{\delta y_0}{\alpha}$$

$$(AP_{i,1} + AS_{i,1})\, u_{i,1} + AN_{i,1}\, u_{i,2} = b_{i,1} + AS_{i,1}\, q_s(xc_i)\, \frac{\delta y_0}{\alpha}$$

A Matlab implementation...
Some results in figure 1.4.

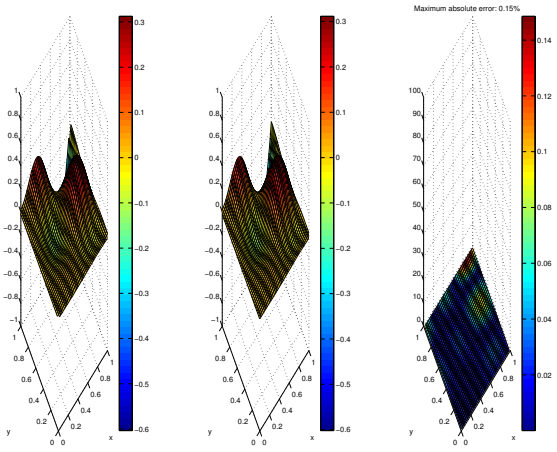

Figure 1.4: From left to right: numerical solution, analytical solution, absolute error.

Appendix A

Some facts about tridiagonal matrices

An $M \times M$ tridiagonal matrix with constant, real coefficients a, b, c along the lower, main and upper diagonals, respectively, is denoted by $\mathrm{Tr}\,(M : a, b, c)$. The first issue we aim to face is the calculation of the determinant of such a tridiagonal matrix and to figure out under which conditions the matrix is singular.

A.1 Recursion relation for $\det\left[\mathbf{Tr}\,(M : a, b, c)\right]$

Let D_M denote the determinant of the considered tridiagonal matrix, i.e.

$$D_M \equiv \det\left[\mathrm{Tr}\,(M : a, b, c)\right]$$

Thus:

$$\begin{aligned}
D_M &= b \det\left[\underbrace{\begin{pmatrix} b & c & 0 & \cdots & 0 \\ a & b & c & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & a & b & c \\ 0 & \cdots & 0 & a & b \end{pmatrix}}_{\mathrm{Tr}(M-1:a,b,c)}\right] - c \det\left[\underbrace{\begin{pmatrix} a & c & 0 & \cdots & 0 \\ 0 & b & c & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & a & b & c \\ 0 & \cdots & 0 & a & b \end{pmatrix}}_{\in \mathbb{R}^{(M-1)\times(M-1)}}\right] \\
&= b \det\left[\mathrm{Tr}\,(M-1 : a, b, c)\right] - c\,a \det\left[\underbrace{\begin{pmatrix} b & c & 0 & \cdots & 0 \\ a & b & c & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & a & b & c \\ 0 & \cdots & 0 & a & b \end{pmatrix}}_{\mathrm{Tr}(M-2:a,b,c)}\right] \\
&= b \det\left[\mathrm{Tr}\,(M-1 : a, b, c)\right] - c\,a \det\left[\mathrm{Tr}\,(M-2 : a, b, c)\right] \\
&= b\,D_{M-1} - c\,a\,D_{M-2}
\end{aligned}$$

Therefore, D_M can be computed by the recursion relation

$$D_M = b\,D_{M-1} - c\,a\,D_{M-2} \tag{A.1a}$$

initiated with the conditions

$$D_0 = 1; \qquad D_1 = b \tag{A.1b}$$

or

$$D_{-1} = 0; \qquad D_0 = 1 \tag{A.1c}$$

The recurrence relation (A.1) is an initial-value problem associated with a second-order, linear difference equation (A.1a).

The validity of the recurrence relation (A.1a) can be verified using the Matlab code reported below.

```
% Use the recurrence relation derived in the notes to
% calculate the determinant of a tridiagonal matrix of
% order M and constant entries along the diagonals.

M = 15; a = 1; b = -2; c = 1;

% Form the tridiagonal matrix
ld = a*ones(M-1,1); md = b*ones(M,1); ud = c*ones(M-1,1);
A = diag(ld,-1) + diag(md,0) + diag(ud,1);

% Compute the determinant using the built-in Matlab function "det"
detA = det(A);

% Compute the determinant using the recursive relation
Dold = 0; D = 1;
for j=1:M
  Doldold = Dold; Dold = D;
  D = b*Dold-a*c*Doldold;
end;
```

A.2 Conditions yielding $D_M = 0$

We face the problem of figuring out the conditions under which $D_M = 0$ by solving the difference equation (A.1a), then imposing the additional constraint $D_M = 0$. A solution to (A.1a) is sought in the form

$$D_j \equiv \vartheta^j, \quad \vartheta \in \mathbb{C}$$

Substituting this tentative solution into (A.1a) yields

$$\vartheta^j - b\,\vartheta^{j-1} + a\,c\,\vartheta^{j-2} = 0$$

which gives

$$\vartheta^2 - b\,\vartheta + a\,c = 0 \Longrightarrow \vartheta_\pm = \frac{b \pm \sqrt{\Delta}}{2}$$

with

$$\Delta \equiv b^2 - 4\,a\,c$$

The general solution of (A.1) is thus:

$$D_j = \alpha\,\vartheta_-^j + \beta\,\vartheta_+^j \qquad \text{for } \Delta \neq 0$$

$$D_j = \alpha\,j\,\vartheta^j + \beta\,\vartheta^j \qquad \text{for } \Delta = 0$$

where α, $\beta \in \mathbb{C}$ are constants to be determined by enforcing the initial conditions.

A.2.1 The case $\Delta = 0$

In this case,

$$\vartheta = \frac{b}{2}$$

and

$$b = 2\,a\,c$$

Thus,

$$D_M = \alpha\,M\,\vartheta^M + \beta\,\vartheta^M$$

Using (A.1b) yields

$$\alpha = \beta = 1$$

and, in turn,

$$D_M = (M+1)\left(\frac{b}{2}\right)^M$$

It is evident that, whenever $\Delta = 0$, it follows that $D_M = 0$ iff $b = 0$ and $a\,c = 0$, i.e., the matrix has at most an upper or a lower non-zero diagonal. This is not an interesting case in the applications considered in these notes.

A.2.2 The case $\Delta \neq 0$

Let us consider the case $\Delta > 0$ first. Enforcing the initial conditions (A.1b) yields

$$\alpha = -\frac{\vartheta_-}{\sqrt{\Delta}}; \qquad \beta = \frac{\vartheta_+}{\sqrt{\Delta}}$$

$$D_M = \frac{\vartheta_+^{M+1} - \vartheta_-^{M+1}}{\sqrt{\Delta}}$$

Thus, asking for $D_M = 0$ implies $\vartheta_+^{M+1} = \vartheta_-^{M+1}$, a condition that can not be fulfilled when $\Delta > 0$ (recall that $\vartheta_\pm$ are purely real for $\Delta > 0$). Thus, $D_M \neq 0$ whenever $\Delta > 0$.

As for the case $\Delta < 0$, let us first notice that

$$b^2 - 4\,a\,c < 0 \Longrightarrow a\,c > 0$$

The roots of the characteristic polynomial of the difference equation are now

$$\vartheta_+ = \frac{b + i\,\gamma}{2}; \qquad \vartheta_- = \frac{b - i\,\gamma}{2}$$

where

$$\gamma \equiv \sqrt{|\Delta|}$$

The initial conditions (A.1b) yield

$$\alpha = i\,\frac{\vartheta_-}{\gamma}; \qquad \beta = -i\,\frac{\vartheta_+}{\gamma}$$

Again, the singularity condition gives

$$\vartheta_+^{M+1} = \vartheta_-^{M+1} \Longrightarrow \left(\frac{\vartheta_+}{\vartheta_-}\right)^{M+1} = 1 \tag{A.2}$$

Since we have to deal with complex numbers, equation (A.2) requires that

$$\left(\frac{\vartheta_+}{\vartheta_-}\right) = \mathrm{e}^{i\,\varphi}, \quad \varphi = \frac{2\,m\,\pi}{M+1} \quad \forall 0 \leq m \leq M$$

The case $m = 0$ has to be rejected as it implies $\vartheta_+ = \vartheta_-$ and, therefore, $\Delta = 0$. We are left with:

$$\frac{\vartheta_+}{\vartheta_-} = \frac{b + i\,\gamma}{b - i\,\gamma} = \frac{b^2 - \gamma^2 + i\,2\,b\,\gamma}{b^2 + \gamma^2} = \cos\left(\frac{2\,m\,\pi}{M+1}\right) + i\,\sin\left(\frac{2\,m\,\pi}{M+1}\right),\ m = 1, \ldots, M$$

(A.3)

Equating the real parts on both sides of (A.3) results in

$$2\,b^2 - 4\,a\,c = 4\,a\,c\,\cos\left(\frac{2\,m\,\pi}{M+1}\right) = 4\,a\,c\,\left[2\,\left(\cos\left(\frac{m\,\pi}{M+1}\right)\right)^2 - 1\right]$$

yielding

$$b = \pm 2\,\sqrt{a\,c}\,\cos\left(\frac{m\,\pi}{M+1}\right),\ m = 1,\,\ldots,\,M \tag{A.4}$$

Notice that, due to the presence of the cosine factor in (A.4), the $\pm$ sign is unnecessary:

$$b = 2\,\sqrt{a\,c}\,\cos\left(\frac{m\,\pi}{M+1}\right),\ m = 1,\,\ldots,\,M \tag{A.5}$$

A.3 Eigenvalues and eigenvectors of $\mathrm{Tr}\,(M : a, b, c)$

The eigenvalues λ_m of $\mathrm{Tr}\,(M : a, b, c)$ are solutions, if any, of

$$\det\left[\mathrm{Tr}\,(M : a, b - \lambda_m, c)\right] = 0 \tag{A.6}$$

We already know that if equation (A.6) is satisfied by real eigenvalues, they must satisfy the inequality $\Delta < 0$, i.e.

$$(b - \lambda_m)^2 - 4\,a\,c < 0$$

Thus, a necessary condition for $\mathrm{Tr}\,(M : a, b - \lambda_m, c)$ having real eigenvalues is $a\,c > 0$.

Using the result (A.5) yields

$$\lambda_m = b - 2\,\sqrt{a\,c}\,\cos\left(\frac{m\,\pi}{M+1}\right),\ m = 1,\,\ldots,\,M \tag{A.7}$$

The eigenvalues can be either complex, when $a\,c < 0$, or purely real when $a\,c > 0$. When $a\,c < 0$, there are M distinct eigenvalues appearing in complex-conjugate pairs. When $a\,c > 0$ there are M distinct real eigenvalues. In both cases, there exists a basis of eigenvectors (of either $\mathbb{R}^M$, considered as a $\mathbb{R}$-vector space, or of $\mathbb{C}^M$, considered as a $\mathbb{C}$-vector space) and $\mathrm{Tr}\,(M : a, b, c)$ can be diagonalized.

As an application, let us calculate the eigenvalues of Tr $(M : 1, -2, 1)$, a finite difference operator arising from the FTCS discretization of the one dimensional heat equation:

$$\lambda_m = -2 \left[1 + \cos\left(\frac{m\,\pi}{M+1} \right) \right], \; m = 1, \ldots, M \tag{A.8}$$

It is interesting to notice that the eigenvalues of maximum and minimum magnitude are, respectively,

$$\lambda_{max} = \lambda_1 \longrightarrow 4 \text{ for } M \longrightarrow +\infty$$

and

$$\lambda_{min} = \lambda_M \longrightarrow 0 \text{ as } \left(\frac{\pi}{M+1} \right)^2 \text{ for } M \longrightarrow +\infty$$

Thus, the spectral radius of Tr $(M : 1, -2, 1)$ increases quadratically with M[1]:

$$\frac{|\lambda_{max}|}{|\lambda_{min}|} = \frac{\lambda_1}{\lambda_M} = \frac{-2 \left[1 + \cos\left(\frac{\pi}{M+1} \right) \right]}{-2 \left[1 + \cos\left(\frac{M\,\pi}{M+1} \right) \right]} \longrightarrow \frac{4}{\pi^2} \, (M+1)^2 \tag{A.9}$$

Let us compute the eigenvectors of Tr $(M : a, b, c)$. This task is readily accomplished by simply applying the definition of right eigenvectors:

$$\text{Tr}\,(M : a, b, c)\; \boldsymbol{v}^{(m)} = \lambda_m \, \boldsymbol{v}^{(m)}$$

yielding

$$a\, v_{j-1}^{(m)} + b\, v_j^{(m)} + c\, v_{j+1}^{(m)} = \lambda_m \, v_j^{(m)}, \; j = 1, \ldots, M \longrightarrow \sqrt{\frac{c}{a}}\, v_{j-1}^{(m)} + 2 \cos\left(\frac{m\,\pi}{M+1} \right) v_j^{(m)} + \sqrt{\frac{a}{c}}\, v_{j+1}^{(m)} = 0 \tag{A.10}$$

Equation (A.10) holds for $2 \leq j \leq M-1$ but has been extended to apply also for $j = 1$ and $j = M$ by assuming $v_0^{(m)} = 0 = v_{M+1}^{(m)}$, providing two boundary conditions for the difference equation (A.10). The associated characteristic polynomial is

$$\sqrt{\frac{c}{a}}\, z^2 + 2 \cos\left(\frac{m\,\pi}{M+1} \right) z + \sqrt{\frac{a}{c}} = 0$$

[1] Comment for me: I verified numerically this limit. It's okay.

with roots:

$$z_{1,2} = \frac{-\cos\left(\frac{m\,\pi}{M+1}\right) \pm \sqrt{\left(\cos\left(\frac{m\,\pi}{M+1}\right)\right)^2 - 1}}{\sqrt{\frac{c}{a}}} = -\sqrt{\frac{a}{c}}\left[\cos\left(\frac{m\,\pi}{M+1}\right) \pm i\,\sin\left(\frac{m\,\pi}{M+1}\right)\right]$$

A general solution of the difference equation is thus:

$$v_j^{(m)} = \left(-\sqrt{\frac{a}{c}}\right)^j \left[A\,\cos\left(\frac{m\,\pi}{M+1}\right) + B\,\sin\left(\frac{m\,\pi}{M+1}\right)\right], \qquad A,\, B \in \mathbb{C}$$

The boundary condition $v_0^{(m)} = 0$ yields $A = 0$, while the condition $v_{M+1}^{(m)} = 0$ is automatically satisfied. Thus, the eigenvectors are

$$v_j^{(m)} = \left(-\sqrt{\frac{a}{c}}\right)^j \sin\left(\frac{j\,m\,\pi}{M+1}\right), \quad j = 1,\,\ldots,\,M \tag{A.11}$$

The eigenvectors given by Eq. (A.11) are not normalized w.r.t the Euclidean norm.

Bibliography

J. Anderson. *Computational Fluid Dynamics*. Computational Fluid Dynamics: The Basics with Applications. McGraw-Hill Education, 1995. ISBN 9780070016859.

P.K. Kundu, I.M. Cohen, and D.R. Dowling. *Fluid Mechanics*. Elsevier Science, 2012. ISBN 9780123821003. URL https://books.google.it/books?id=iUo_4tsHQYUC.

K S Lele. Compact finite difference schems with spectral-like resolution. *J COMP PHYS*, 103:16–42, 1992.

M. Piller and E. Stalio. Finite-volume compact schemes on staggered grids. *J. Comp. Phys.*, 197(1):299 – 340, June 2004.

M. Piller and E. Stalio. Compact finite volume schemes on boundary-fitted grids. *J. Comp. Phys.*, 227(9):4736–4762, April 2008.

S.B. Pope. *Turbulent Flows*. Cambridge University Press, 2000. ISBN 9780521598866. URL https://books.google.it/books?id=HZsTw9SMx-0C.

J R Shewchuk. An introduction the conjugate gradient method without the agonizing pain. Technical report, School of Computer Science, Carnegie Mellon University, 1994.