International Journal of Applied Mathematics

Volume 37 No. 6 2024, 649–698
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v37i6.6

REVISITING FINITE DIFFERENCE SOLUTIONS FOR HEAT-TYPE EQUATIONS. PART I: DIFFUSIVE TRANSFER

G. Garrido [1,§], J. L. G. Pestaña [2]

[1] Departamento de Informática

Universidad de Jaén, Campus Las Lagunillas

23071 Jaén, SPAIN

[1,§] e-mail: ggluque@ujaen.es

[2] Departamento de Física

Universidad de Jaén, Campus Las Lagunillas

[2] e-mail: jlg@ujaen.es

Abstract

In this first paper of a series, we revisit all major systematic uncertainties that affect a complete and unbiased sample of five finite difference schemes for diffusion-like equations. In order to provide the coherent picture, unlike the existing way, we use as the key tenets both the reverse Taylor's analysis and the discrete Fourier's analysis, as well as the monotonicity analysis. For every type of scheme, their theoretical uncertainties are examined. A detailed graphical investigation is also provided and used to give a physical reinterpretation of the Courant-Friedrichs-Lewy-type condition. We find that no scheme considered in this study resolves the smaller length scales well. Furthermore, we present several numerical experiments on an equal footing corroborating our demonstrations and proving whether the accuracy of each scheme is impaired by the discontinuities in the data. A comparison with each other is made as well. Our results indicate that the simplest Schmidt's scheme is also preferred by experiments.

Received: June 28, 2024

MSC 2020: 35Q68, 35K05, 65M06, 65N22, 97N40

Key Words and Phrases: PDEs in connection with computer science; heat equation; finite difference methods for initial value and initial-boundary value problems involving PDEs; numerical solution of discretized equations for boundary value problems involving PDEs; numerical analysis (educational aspects)

1. Introduction

The old finite difference method is presently used in numerous compute resources and courses, not only in mathematics [1], but also in natural sciences [2], engineering [3, 4], and noticeably even in social sciences [5]. However, it is not easy to find the why similar schemes on a given partial differential equation provide dissimilar results, and even a given scheme on comparable partial differential equations offers disparate results. By relying on general diagnostic tools, this series aims to find it.

Here we plan to numerically integrate the full heat-like equation, which remains technically challenging [6]. To say the least, apart from its nonlinearities, this initial boundary value problem is a combination of two fundamentally different physical processes: one of them is associated with diffusion, in which case the underlying partial differential equation is parabolic; while the other one is associated with advection, in which case the underlying partial differential equation is hyperbolic. Thus, first we program an unified understanding of each one of them (which does not seem to be quite well-established) to achieve our prospect.

Although many authors have been numerically solving partial differential equations for a little over a century, since the first application of the finite difference method to them in 1910 [15], there have been relatively few systematic studies on the numerical approximations in the specialized literature and even fewer of these have been comprehensive investigations to the basic principles to the best of our knowledge. Also in textbooks [5, 4, 1, 2, 3] some practical issues are typically ignored. It is therefore timely, without loss of novelty, to understand in practice and test in depth all the difficulties and subtleties that inevitably appear in any numerical solution and specifically in the simpler ones. In fact, the goal of this series is therefore to understand the rich phenomenology in a priori similar results implying new theoretical analysis of the proposed procedures and simulations, specially in the case of the full equation but also in the other cases.

In the present paper, the first in the series, we finite difference integrate the simplest (Dirichlet, linear and unidimensional) parabolic differential equation using all at most three-level, low-order schemes. Also, we do it in the chronological sequence in which they were developed. While doing so, we explain all the methodological insights and concerns of the difference equation needed

to follow the very recent review by Sagaut et al. [6]. This is the position of our entry-level paper with respect to the state of art. We shall not consider related methods, but the same principles apply.

The paper is organized as follows. As the first paper in the series, Section 2 is devoted to some basics of finite difference approach, from a physical more than a mathematical perspective. This covers the key notions not only of consistency, stability, and accuracy, but also of convergence. With these conceptual underpinnings, in Section 3, we rigorously investigate all five pioneering schemes which our topical current understanding of parabolic differential equations is based on, namely the Richardson's, Schmidt's, Crank-Nicolson, Laasonen's and Du Fort-Frankel schemes. In Section 4 we implement all of these schemes. Then, as a proof of concept, we both discuss their performance properties and eventually compare them with each other. Finally, Section 5 presents a short summary and conclusive remarks. Motivated in particular by discontinuities in the data, A shows the exemplary initial value problem, as simple as it is very interesting, we apply in Section 4.

2. **Finite differencing basics**

In this section, we introduce the very powerful finite difference basic approach to solve partial differential equations and examine its error. To achieve this we follow several techniques: by a heuristic derivation first pioneered by Hirt [7], by extensive discrete Fourier analysis first pioneered by von Neumann (e.g., see [8]), and by direct calculation against monotonicity first pioneered by Godunov [9]. We shall investigate agreement among the different methods. The relative mathematics and its demonstration are kept to a bare simplest minimum in the review.

2.1. **Discrete geometry.**

2.1.1. **The grid function.** The primary solution quantity of any numerical method, and in particular the finite difference approach, is the real-valued scalar field T, which is a function of space x and time t; i. e. the temperature at any point in a two-dimensional domain is characterized by a function $T = T(x, t)$. In the discrete setting of the numerical method however, the function is retained only at a finite number of points in space and time and the function evaluated at the discrete point (x_j, t^n) is referred to as the grid function T_j^n; i.e. we usually replace spatial domain with N_x points with a uniform grid spacing Δx, whereas the time interval of interest is discretized with N_t points with a uniform time step Δt. Of course, it is therefore of consequence to use grids that are optimal for the significant time and length scale posed by the physics of the problems.

Then, even before we compute the numerical solution grid function, to fully understand its local errors incurred at the different time and length scales or,

in Fourier space, at different wavenumbers then we shall naturally think of the grid function as being formed as a superposition of its harmonic components:

$$\begin{aligned} T_j^n &= \sum_{\mu=min}^{max} A_0\, e^{i\,(k_\mu\, j\Delta x - w_\mu\, n\Delta t)} = \sum_{\mu=min}^{max} A_0\, e^{i\,k_\mu\,(j\Delta x - v\, n\Delta t)} \\ &= \sum_{\mu=min}^{max} A(n\Delta t)\, e^{i\,k_\mu\, j\Delta x}, \end{aligned} \tag{2.1}$$

where k_μ is the discrete wavenumber, w_μ is its angular frequency, v is the propagation speed and $A_0 = |A_0|\, e^{i\,\phi}$ is the initial complex amplitude. Eq. 2.1 is the complex notation describing the real waves in space and time:

$$\Re(T_j^n) = \sum_{\mu=min}^{max} |A_0|\, \cos\,(k_\mu\, x - w_\mu\, t + \phi), \tag{2.2}$$

with real amplitude $|A_0|$ and phase constant ϕ. However, which is particularly important here, in the rectilinear grid of size Δx only a finite number of harmonic waves can exist. On the one hand, because we need at least three points for the wave, namely for its amplitude, wavelength and phase shift, then the Fourier mode with the the shortest wavelength (referred to as the Nyquist wavelength) that is resolved by our grid is given by $2\,\Delta x$; i.e. $k_{max} = \frac{2\pi}{\lambda_{min}} = \frac{\pi}{\Delta x}$. On the other hand, because on the grid two wavenumbers k_1 and k_2 give the same values of $e^{i\,k\,j\Delta x}$ if $k_1\,\Delta x = k_2\,\Delta x + 2\pi$ (phenomenon referred to as aliasing), then on the grid we only can see $-\pi < k\Delta x \leq \pi$; i.e., taking the symmetry of the Fourier modes into account, the only relevant values of consequence to analyze in the problem are the following range of wavenumbers (named resolution requirements):

$$0 \leq k\Delta x \leq \pi. \tag{2.3}$$

The lowest value of k is 0 corresponding to the constant wave mode in Eq. 2.1. If $k < \frac{\pi}{2\Delta x}$ are called low wavenumbers (if $k\Delta x$ is small then waves have long wavelengths in comparison to Δx), which oscillates more slowly; and viceversa for $k\Delta x$ near π.

Finally, after we compute the numerical solution grid function, to measure global errors in it we shall naturally derive the grid function absolute error with respect to the l^2-norm either by:

$$||\Delta T||_2(t^n) = \sqrt{\Delta x \sum_{j=1}^{N_x} |T_j^n - T(x_j,\, t^n)|^2}, \tag{2.4}$$

which involves a summation only over all nodes in space and therefore remains a function of time; or, in a completely analogous way, by:

$$||\Delta T||_2 = \sqrt{\Delta x\, \Delta t \sum_{j=1}^{N_x} \sum_{n=1}^{N_t} |T_j^n - T(x_j,\, t^n)|^2}, \tag{2.5}$$

where $T(x_j,\, t^n)$ is the exact solution. We shall use both of these formulas in the series. This being said, when the exact solution is not known the error can still be equally computed by employing a third (o more) numerical evaluation of the solution.

2.1.2. **The difference quotients.** Of course, in the discrete setting, the finite difference approach proceeds by replacing the derivatives, in time as well as in space, in the differential equation with difference quotients based only on values of the grid function. The invaluable tool we follow to this purpose is the local Taylor-series expansion; a choice motivated by these series being suitable for understanding the terms neglected by themselves as well, as we shall show shortly. Each of these replacements is achieved as follows.

(1) Temporal differences

We may write the forward, one-sided first derivative of T at the point (x_j, t^n) by Taylor expanding the value T_i^{n+1} around this point,

$$T_j^{n+1} = T_j^n + \frac{\partial T}{\partial t}(x_j,\, t^n)\, \Delta t + \frac{1}{2}\frac{\partial^2 T}{\partial t^2}(x_j,\, t^n)\, (\Delta t)^2 + \mathcal{O}((\Delta t)^3), \tag{2.6}$$

to obtain (isolating the time derivate and dividing by Δt):

$$\frac{\partial T}{\partial t}(x_j,\, t^n) = \frac{T_j^{n+1} - T_j^n}{\Delta t} + \mathcal{O}(\Delta t). \tag{2.7}$$

Then by neglecting terms of higher order in this expression Eq. 2.7 may be used to formulate a finite difference approximation to the time rate of change of T. Furthermore, since we know T^0, such an approximation in time is the obvious choice. Note that to simplify expressions such as those in these equations we use the big-Oh notation (first used by Bachmann and popularized by Landau).

However, we may also choose to write centered first derivative of T at the point (x_j, t^n) by eliminating the second order terms by using the two Taylor expansions T_j^{n+1} and T_j^{n-1}, i.e. by subtracting from Eq. 2.6 the following Taylor expansion

$$T_j^{n-1} = T_j^n - \frac{\partial T}{\partial t}(x_j,\, t^n)\, \Delta t + \frac{1}{2}\frac{\partial^2 T}{\partial t^2}(x_j,\, t^n)\, (\Delta t)^2 + \mathcal{O}((\Delta t)^3), \tag{2.8}$$

to obtain:

$$\frac{\partial T}{\partial t}(x_j,\, t^n) = \frac{T_j^{n+1} - T_j^{n-1}}{2\, \Delta t} + \mathcal{O}((\Delta t)^2), \tag{2.9}$$

which by neglecting terms of higher order may also be used to formulate another more accurate finite difference approximation to the same first order time derivative.

(2) Spatial diferences

We may perform exactly the same calculations based on Taylor series expansion to arrive at finite difference approximation for the spatial derivatives, i.e. to construct either a centered finite difference approximation,

$$\frac{\partial T}{\partial x}(x_j,\, t^n) \approx \frac{T^n_{j+1} - T^n_{j-1}}{2\,\Delta x}, \tag{2.10}$$

or a forward approximation,

$$\frac{\partial T}{\partial x}(x_j,\, t^n) \approx \frac{T^n_{j+1} - T^n_{j}}{\Delta x}. \tag{2.11}$$

These two equations are examples of two-point derivative at position j.

Similarly, we may also construct approximations to higher order derivatives. For instance, we may write the centered second order derivative of T with respect to x by first eliminating the first order derivatives by using the two following Taylor expansions:

$$T^n_{j+1} = T^n_j + \frac{\partial T}{\partial x}(x_j,\, t^n)\,\Delta x + \frac{1}{2}\frac{\partial^2 T}{\partial x^2}(x_j,\, t^n)\,(\Delta x)^2 + \mathcal{O}((\Delta x)^3) \tag{2.12}$$

and

$$T^n_{j-1} = T^n_j - \frac{\partial T}{\partial x}(x_j,\, t^n)\,\Delta x + \frac{1}{2}\frac{\partial^2 T}{\partial x^2}(x_j,\, t^n)\,(\Delta x)^2 + \mathcal{O}((\Delta x)^3), \tag{2.13}$$

i.e. by adding Eqs. 2.12 and 2.13 to obtain:

$$\frac{\partial^2 T}{\partial x^2}(x_j,\, t^n) = \frac{T^n_{j+1} - 2\,T^n_j + T^n_{j-1}}{\Delta x^2} + \mathcal{O}((\Delta x)^2). \tag{2.14}$$

Then, again, by neglecting terms of higher order in this expression Eq. 2.14 may be used to formulate another finite difference approximation to the second order space derivative of T. This is a three-point derivative at position j.

2.2. **Error estimation.** The ultimate aim of any numerical method, so of the finite difference approach in this series, is the accuracy of the solution quantity. In order to achieve it, it is necessary that scheme satisfies three criteria as defined below and explained in the next four subsections.

- A numerical finite difference scheme is consistent if it becomes the corresponding partial differential equation as the grid size and time step approach zero.
- A difference approximation is stable if its error remains bounded.

- A difference equation is convergent if its solution approaches that of the partial differential equation as the grid size approaches zero.

2.2.1. **The consistency: Hirt's reverse Taylor analysis.** Once a finite difference approximation is constructed we have to analyze it with respect to its physical consistency.

Whether a finite difference approximation is based on Taylor series expansions, we already know that the consistency requirement is satisfied. However, even then we find very useful the technique introduced by Hirt [7] in deducing the consistency, as well as the order of accuracy, of any finite difference approximation. In fact, much more will be deduced by using the Hirt's analysis shortly.

The idea is to reduce the difference equation to an equation (indeed, containing an infinite number of partial derivatives) by expanding each of its terms in a Taylor series, and then eliminating time derivatives higher than first order and mixed time and space derivatives. In order to achieve it, we must raise and/or lower their indices by Taylor-expanding until all of them result in the same point. Only if we do this without the use of the original differential equation [10], then the equation obtained is called the modified equation associated with the difference scheme. Finally, the limit of the first several lowest-order terms appearing in the modified equation which are not in the original partial differential equation, when the grid size and time step go to zero, allows us to prove consistency.

Besides that, the lowest-order powers of the increments appearing in these terms is the local error of the algorithm.

2.2.2. **The stability: von Neumann's discrete Fourier analysis.** Additionally, once a finite difference approximation is designed we also have to analyze it with respect to its stability.

Even with nonlinear differential equations, we find extremely useful the technique introduced by von Neumann (e.g., see [8]) in deducing the stability of any finite difference approximation. In fact, much more will be deduced by using the von Neumann's analysis shortly.

The idea is that just one discrete complex Fourier component (cf. Eq. 2.1), $A(n\Delta t)\, e^{i\,k\,j\Delta x}$, unbounded within any given finite time span is enough to get a scheme to explode. Thus, the von Neumann's analysis consist of considering the behavior of the complex amplitude ratio or growth factor

$$G = \frac{A((n+1)\Delta t)}{A(n\Delta t)}, \tag{2.15}$$

that has to satisfy the following basic stability condition, in order to have both physical and numerical stability:

$$|G| \leq 1, \tag{2.16}$$

for the wavenumber range given by Eq. 2.3. Depending on the algorithm, the scheme's frequency content might vary, ranging from domination of low wavenumbers (low frequencies) up to high wavenumbers (high frequencies); which determines if it is appropriate to resolve all the time and length scales posed by the physics of the problem, i.e. the resolution requirements. We note that G can become less that 0, which means the numerical solution will decay but in an oscillatory fashion.

On the other hand, the analysis of the next-to-leading order term of the modified equation is a complementary way to check this; i.e. the Hirt's method is another way to obtain whether the algorithm is stable [11, 12, 13, 14].

2.2.3. **The accuracy: more from Hirt's and von Neumann's analysis.** Furthermore, once a finite difference approximation has been sanctioned with respect to its consistency and stability as well as its order of accuracy we still can and should analyze it with respect to the structure of its local error whose order is the only thing we know about it so far. Based on these considerations, the goal will be to solve some unphysical problems that inevitably appear due to the numerical scheme used in itself, i.e., mainly, to introduce strategies to countered or correct for such systematic biases. In the fourth paper in the series, these aspects will be argued. Here, we quantify these systematic truncation errors making again use of both the Hirt's analysis and the von Neumann's analysis as follows.

(1) More from Hirt's analysis

As a prelude to how to use the Hirt's analysis to estimate these algorithmic errors, let us first apply a wave solution to each of the linear differential equations we are interested in and also to the equations involving the third derivative with respect to position. Indeed, as will be shown in this series, all of these equations (linear and nonlinear) admit wave mode solutions. Therefore, if we apply a continuous Fourier series (formed by replacing $j\Delta x$ by x and $n\Delta t$ by t in Eq. 2.1) to each of them, we shall find the following global behaviors:

(a) To the advection equation (the second paper in the series) involving the first x-derivative,

$$\frac{\partial T}{\partial t} = -a\frac{\partial T}{\partial x}, \tag{2.17}$$

where a is the speed along the x-axis, the general analytical solution becomes:

$$T(x,\, t) = \sum_{\mu=0}^{\infty} A_0\, e^{i\, k_\mu\, (x - a\, t)}, \tag{2.18}$$

i.e. the phase speed is $v = a$ or the frequency is $w_\mu = k_\mu\, a$. Therefore, if a is uniform all the waves propagate with the same velocity, and hence the solution is neither dissipative nor dispersive.

(b) To the Fourier diffusion equation (this paper) involving the second x-derivative,

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}, \tag{2.19}$$

where α is the diffusion coefficient, the analytical solution becomes:

$$T(x,\, t) = \sum_{\mu=0}^{\infty} A_0\, e^{i\, k_\mu\, (x + i\, \alpha\, k_\mu\, t)} = \sum_{\mu=0}^{\infty} A_0\, e^{i\, k_\mu\, x} e^{-\alpha\, k_\mu^2\, t}, \tag{2.20}$$

i.e. the frequency is $w_\mu = i\, k_\mu^2\, \alpha$, and hence the amplitude decreases exponentially as time increases. Therefore, the solution is dissipative and no dispersive. In fact, we also see that the part of the analytical solution associated with the shorter waves (higher wavenumbers) decreases faster than the part associated with the longer waves (lower wavenumbers). We search for the signature of this phenomenon, dubbed as the (not physical but) numerical stiffness after its discoverers [18], in this series.

(c) To the equation involving the third x-derivative,

$$\frac{\partial T}{\partial t} = -b \frac{\partial^3 T}{\partial x^3}, \tag{2.21}$$

the analytical solution becomes:

$$T(x,\, t) = \sum_{\mu=0}^{\infty} A_0\, e^{i\, k_\mu\, (x + b\, k_\mu^2\, t)}, \tag{2.22}$$

i.e. the phase speed is $v = -b\, k_\mu^2$ and the frequency is $w_\mu = -k_\mu^3\, b$. Therefore, waves of different wavelenghts propagate at different speeds, and hence the solution is no dissipative but dispersive. Especially, and not incidentally, these dispersive waves propagate in the the opposite direction of Eq. 2.18. In fact, the frequencies of the same equations but involving both even-order and odd-order x-derivatives have alternating signs.

(d) In that way, now if we apply a Fourier series solution to the advection-diffusion equation (the third paper in the series) involving both the first and the second x-derivatives

$$\frac{\partial T}{\partial t} = -a \frac{\partial T}{\partial x} + \alpha \frac{\partial^2 T}{\partial x^2}, \tag{2.23}$$

then we shall find that the analytical solution becomes:

$$T(x,\,t)=\sum_{\mu=0}^{\infty}A_0\,e^{i\,k_\mu\,(x+i\,\alpha\,k_\mu\,t-a\,t)}=\sum_{\mu=0}^{\infty}A_0\,e^{i\,k_\mu\,(x-a\,t)}e^{-\alpha\,k_\mu^2\,t}, \tag{2.24}$$

i.e. the frequency is $w_\mu=k_\mu\,a+i\,k_\mu^2\,\alpha$. Hence, the solution is dissipative and no dispersive (according to its constant coefficients). The same behavior as that for the diffusion equation.

(e) In the same way, the analytical solution to the equation in which first, second and third x-derivatives all occur,

$$\frac{\partial T}{\partial t}=-a\frac{\partial T}{\partial x}+\alpha\frac{\partial^2 T}{\partial x^2}-b\frac{\partial^3 T}{\partial x^3} \tag{2.25}$$

becomes:

$$\begin{aligned}T(x,\,t)&=\sum_{\mu=0}^{\infty}A_0\,e^{i\,k_\mu\,(x+i\,\alpha\,k_\mu\,t-a\,t+bk_\mu^2\,t)}\\&=\sum_{\mu=0}^{\infty}A_0\,e^{i\,k_\mu\,(x-a\,t+bk_\mu^2\,t)}e^{-\alpha\,k_\mu^2\,t},\end{aligned} \tag{2.26}$$

i.e. the propagation velocity is $c=a-bk_\mu^2$ and frequency is $w_\mu=k_\mu\,a-k_\mu^3\,b+i\,k_\mu^2\,\alpha$, and hence the solution is both dissipative and dispersive.

It is clear from all of this evidence that the second x-derivative is a dissipative term and the third is a dispersive one. In fact, all the even order derivatives are dissipative and all the odd order derivatives greater than one are dispersive, implying that in general the effect of dissipation dominates over dispersion. Thus, it is clear that any modified equation, i.e. any difference algorithm, is both dissipative and dispersive. Specifically, given that the frequency of a numerical wave component is always going to be a complex number, $w_{num}=\Re(w_{num})+i\,\Im(w_{num})$, then, by applying the discrete Fourier series grid solution of Eq. 2.1 to any modified equation up to third order in derivatives, the numerical solution becomes:

$$T_j^n=\sum_{\mu=min}^{max}A_0(n\,\Delta t)\,e^{i\,k_\mu\,(j\,\Delta x+\frac{\Re(w_{num})}{k}n\,\Delta t)}e^{-\Im(w_{num})\,n\,\Delta t)}, \tag{2.27}$$

with $\Re(w_{num})$ as well as $\Im(w_{num})$ depending on the wavelength (and generally $\Im(w_{num})\neq i\,k^2\,\alpha$); that is, we always find that the numerical solution presents both dissipative and dispersive errors. Well, in diffusion the propagation velocity is zero, $Re(w_{num})=0$; although of course there exists numerical dissipation. Even we may sometimes find false numerical modes in the positive x direction, as explained above.

In other words, despite the (continuous) differential equation is neither dissipative nor dispersive (propagates the initial condition, at the constant velocity, without reduction of its amplitude), the (discrete) difference equation is generally both diffusive and dispersive.

That said, note that contrary to the accuracy, the stability condition (Eq. 2.16) requires to have a greater dissipative error, by way of illustration.

Therefore, the Hirt's analysis already used also gives us a qualitatively simple explanation, which is competitive to the von Neumann's analysis, on the mechanisms involved in the dissipation and dispersion algorithmic errors, indicating whether the algorithm is physically reasonable. In fact, Eq. 2.27 summarizes the potential of the Hirt's error estimation. Having said all of this, its major handicap is that the Hirt's error estimation does not provide accurate results not even with linear differential equations. Notwithstanding this shortcoming, we conclude that the Hirt's analysis offers an average description of evolutionary behaviors of both amplitudes and phases that is appropriate at least qualitatively.

(2) More from von Neumann's analysis

Alternatively, from the point of view of the von Neumann's analysis, we can see that the time evolution of the numerical solution is fully contained in the growth factor, Eq. 2.15. Specifically, since w_{num} is a complex function, the growth factor can be converted from rectangular to polar form, i.e. separated in an numerical amplitude and a numerical phase:

$$G = |G| e^{-i\,\phi_{num}}, \tag{2.28}$$

in such a way that the modulus of G will influence the amplitude of the numerical solution, while its phase will influence the phase of the numerical solution, i.e. the relative error in amplitude (aka dissipation error) in one of the Fourier modes of the numerical solution of our difference equation is stated as being the ratio of the computed amplitude to the exact amplitude:

$$\delta A = \frac{|G| - e^{-\alpha\, k^2 \Delta t}}{e^{-\alpha\, k^2 \Delta t}}, \tag{2.29}$$

while its relative error in the phase (aka dispersion error) is stated as being by the ratio of the computed phase angle to the exact phase angle:

$$\delta\phi = \frac{\phi_{num} - v\, k\Delta t}{v\, k\Delta t}. \tag{2.30}$$

If $\delta\phi > 0$, referred to as a leading phase error, this means that the Fourier components of the numerical solution have a wave speed greater than that of the exact solution. Similarly, if $\delta\phi < 0$, referred to as a

lagging phase error, this means that the Fourier components of the numerical solution are decelerated with respect to that of the exact solution.

Thus, the von Neumann's analysis already used also gives us a quantitatively simple explanation, which is complementary to the Hirt's analysis, on the mechanisms involved in the dissipation and dispersion algorithmic errors, e.g. showing us even the high-frequency (i.e. oscillatory) and low-frequency (i.e. smooth) components separately. In fact, Eq. 2.28 summarizes the high potential of the von Neumann's error estimation.

2.2.4. **The convergence: Godunov's monotonicity analysis.** Lastly, we can also sometimes find another kind of error nonlocal, thereby impacting the rate of convergence. Both the unavoidable dissipation and dispersion errors are within the truncation error, even if there are no such mechanisms in the governing differential equation. Likewise, despite the underlying equation cannot never generate new maximum or minimum values over those contained in the initial or boundary conditions, then this is not necessarily true for a difference equation. Surprisingly, problems of this kind, with the risk of getting (bounded but) spurious oscillations in the numerical solutions, always appear with the second (and higher) order schemes when applied to the advection equation not only in presence of discontinuities of the analytical solution but also its derivatives, by way of illustration. Even more surprisingly, in numerical solutions to the diffusion equation monotonicity conditions may result in negative grid functions at some points.

It was Godunov [9] in 1959 who demonstrated that such wiggles are the consequence of the non-monotone behavior of the aforementioned second order algorithms. Indeed, after providing the formal definition of that a difference scheme will be monotone if no new extrema be created by it, i.e. the new solution must be contained within the same variation range of the solution at an earlier time step,

$$\sum_j |T_j^{n+1} - T_{j-1}^{n+1}| \leq \sum_j |T_j^n - T_{j-1}^n|, \tag{2.31}$$

he was able to prove that linear monotone schemes for the advection equation can be only first order accurate (although first order schemes are not always monotone), independently of the phase error.

Thus, still before implementing a scheme on the computer, we should analyze it with respect to its (preservation of) monotonicity. To do so, we shall therefore estimate the explicit conditions on a numerical scheme to satisfy this requirement using the criterion for which the new solution, T_j^{n+1}, is a convex combination (i.e. where all coefficients are non-negative and, accordingly, sum to unity). However, with implicit integrations the monotonicity property is

not sometimes so easy to check, the convex combination is far too crude for a reliable criterion and the examination of the negative increment in the Eq. 2.31 (by virtue of the absolute value) is required in these rare cases.

3. The difference equations

In this section we lay groundwork for detailed follow-up performance tests. We thoroughly revisit the finite difference equations for canonical Eq. 2.19 up to second order in the derivatives and also carefully examine the sources and form of their theoretical uncertainties, i.e. we answer analytically to all three aforementioned criteria. We highlight the conceptual differences between not only hierarchical and monolithic schemes but also notably two and three level hierarchical schemes. We also show for the first time the risk of spatial oscillations throughout the entire range of wavenumbers, that is to say we examine separately all the Fourier components of the spatial variation. We sort the exposition, motivated by technical considerations lacking in literature, by chronology and acknowledging their authors.

TABLE 1. Schemes analysed in this work. Note that n represents the temperature at the current time step whereas $n+1$ (colored in red) represents the new (future) temperature. Note also the temperature one time step in the past, $n-1$ (colored in violet). Common features and differences are visible.

Author	Algorithm
Richardson [15]	$T_j^{n+1} = T_j^{n-1} + 2\,F\,(T_{j-1}^n - 2\,T_j^n + T_{j+1}^n)$
Schmidt [16]	$T_j^{n+1} = F\,T_{j-1}^n + (1-2\,F)\,T_j^n + F\,T_{j+1}^n$
Crank-Nicolson [19]	$-F\,T_{j-1}^{n+1} + 2\,(1+F)\,T_j^{n+1} - F\,T_{j+1}^{n+1}$ $= F\,T_{j-1}^n + 2\,(1-F)\,T_j^n + F\,T_{j+1}^n$
Laasonen [21]	$-F\,T_{j-1}^{n+1} + (1+2\,F)\,T_j^{n+1} - F\,T_{j+1}^{n+1} = T_j^n$
Du Fort-Frankel [22]	$T_j^{n+1} = \left(\frac{1-2\,F}{1+2\,F}\right) T_j^{n-1} + \left(\frac{2\,F}{1+2\,F}\right)(T_{j-1}^n + T_{j+1}^n)$

3.1. The Richardson's 1910 three-time-level explicit scheme.

3.1.1. **Construction.** In 1910 Richardson [15], who pioneered the numerical study of partial differential equations and their potential role in physics, devised a scheme which was explicit in time and gave values at any time level in terms of values at the previous two time levels. To achieve this, he discretized both time and space derivatives by second order central differences, Eqs. 2.9 and 2.14 respectively, i.e.,

$$\frac{T_j^{n+1} - T_j^{n-1}}{2\,\Delta t} = \alpha \frac{T_{j-1}^n - 2\,T_j^n + T_{j+1}^n}{(\Delta x)^2}. \tag{3.1}$$

However, as a three-time-level bottom-up scheme that it is, the Richardson's scheme has the problem of setting the initial grid function value; i.e. we have to decide how start the process. The solution is to use any other single-step scheme, as shortly discussed.

For easy reference we reformulate Eq. 3.1 in the second column in Table 1 introducing, for convenience, the so-called dimensionless Fourier number:

$$F = \alpha \frac{\Delta t}{(\Delta x)^2}, \tag{3.2}$$

which, like the so-called dimensionless Courant-Friedrichs-Lewy number in solutions of hyperbolic partial differential equations (e.g. see second paper in the series), provides a measure for the spatiotemporal discretization.

3.1.2. **Consistency and order of accuracy.** We firstly analyze the consistency of the scheme using the Hirt's method. Substituting each value of the grid function at points other than point (x_j, t^n) in the scheme in a Taylor series around the value T_j^n at that point (x_j, t^n), i.e. substituting Eqs. 2.6, 2.12 and 2.13 in Eq. 3.1, gives

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = \alpha \frac{\partial^2 T}{\partial x^2} - \frac{(\Delta t)^2}{6} \frac{\partial^3 T}{\partial t^3} + \frac{\alpha\,(\Delta x)^2}{6} \frac{\partial^4 T}{\partial x^4} + \cdots . \tag{3.3}$$

We then see from that equation, which is not yet the modified equation, that its right hand side vanishes in the limit that the grid spacing Δx and timestep Δt approach zero and therefore the Richardson's scheme of calculation is consistent. In addition, this side goes to zero as the second power of Δt and the second power of Δx, implying that the scheme is of second order accuracy in both time and space as well. Indeed, this result further verifies what would otherwise be expected because the way this scheme is based on Taylor series expansions.

We furthermore provide here the evolution equation with only space derivatives, i.e. the modified equation. In order to achieve it, we firstly find expressions for $\frac{\partial^3 T}{\partial t^3}$; which, for this case, turns out to be zero. This implies that the modified equation associated with the Richardson's scheme, the equation of the grid function from Richardson's difference equation, is:

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = \alpha \frac{\partial^2 T}{\partial x^2} + \mathcal{O}((\Delta x)^2), \tag{3.4}$$

which is one of two ways to next not only calculate the local error but also estimate the stability (see below).

3.1.3. **Stability.** We secondly analyze the stability using the von Neumann's method. Substituting a term of Eq. 2.1 into each term in Eq. 3.1 we get the equation for the growth factor (Eq. 2.15),

$$G = \frac{1}{G} - 4\,F + 2\,F\,(e^{-i\,k\,\Delta x} + e^{i\,k\Delta x}) = \frac{1}{G} - 4\,F\,[1 - \cos{(k\Delta x)}], \tag{3.5}$$

whose solutions are

$$G_{\pm} = -2\,F[1-\cos(k\Delta x)] \pm \sqrt{1+[2\,F(1-\cos(k\Delta x))]^2}. \tag{3.6}$$

That is to say, given its nature of three time levels, in a given instant each real valued growth factor associated with the Richardson's scheme has two temporal modes present in the numerical solution. In other words, as a three time level that it is, this scheme has two temporal modes. Note, the solution G_+ corresponds to the physical mode because for well resolved components ($k\Delta x \ll 1$) is as it should be; and viceversa (see Fig. 1):

$$\lim_{k\Delta x\to 0} G_{\pm} = \pm 1. \tag{3.7}$$

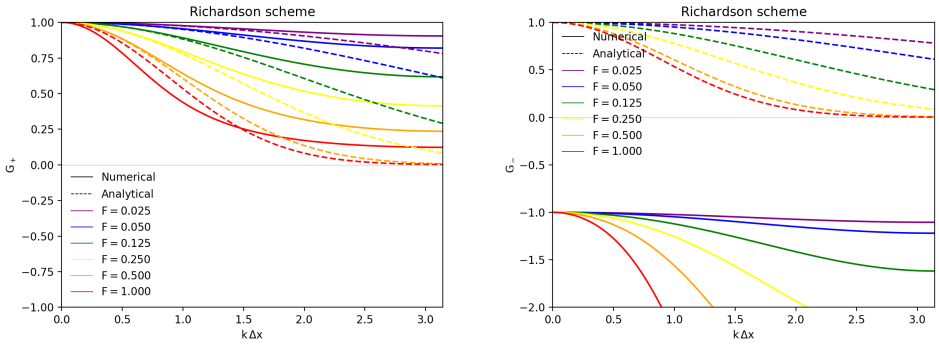


FIGURE 1. The physical (left panel) and computational (right panel) growth factors of the Richardson's scheme, Eq. 3.6 vs. frequency in radians, for specific F values as indicated in the legend. The growth factor of the analytic solution, Eq. 2.20 vs. frequency in radians, for the same F values are distinguished by dashed line colors. We show the stability region, Eq. 2.16, and the value "0" that represents the oscillation threshold.

However, the most important of what happens to the Richardson's growth factor is that the magnitude of its spurious mode takes the form

$$|G_-| = 2\,F[1-\cos(k\Delta x)] + \sqrt{1+[2\,F(1-\cos(k\Delta x))]^2} \geq 1, \tag{3.8}$$

which does not satisfy Eq. 2.16, implying that the Richardson's scheme is unconditionally unstable for the diffusion equation. In fact, this in turn tells us all about Eq. 3.1: it is a no go scheme which can not be applied to diffusion dominated problems.

3.2. **The Schmidt's 1924 single-step explicit scheme.**

3.2.1. **Construction.** In 1924 Schmidt [16] proposed a second order approximation for the spatial derivative but forward in time, Eqs. 2.7 and 2.14 respectively, i.e.

$$\frac{T_j^{n+1}-T_j^n}{\Delta t}=\alpha\frac{T_{j-1}^n-2T_j^n+T_{j+1}^n}{(\Delta x)^2}, \tag{3.9}$$

which is also reformulated in Table 1 using Eq. 3.2.

3.2.2. **Consistency and order of accuracy.** We firstly analyze the consistency of the scheme using the Hirt's method. Substituting each value of T at points other than point (x_j,t^n) in the scheme in a Taylor series around the value T_j^n at that point (x_j,t^n), i.e. we substitute Eqs. 2.6 and 2.13 in Eq. 3.9, gives

$$\frac{\partial T}{\partial t}-\alpha\frac{\partial^2 T}{\partial x^2}=-\frac{\Delta t}{2}\frac{\partial^2 T}{\partial t^2}-\frac{(\Delta t)^2}{6}\frac{\partial^3 T}{\partial t^3}-\frac{(\Delta t)^3}{24}\frac{\partial^4 T}{\partial t^4}+\frac{\alpha\,(\Delta x)^2}{12}\frac{\partial^4 T}{\partial x^4}+\cdots. \tag{3.10}$$

We then see from that equation, which is not yet the modified equation, that the right hand side vanishes when $\Delta t\to 0$ and $\Delta x\to 0$ and therefore the Schmidt's scheme of calculation is consistent. In addition, this side goes to zero as the first power of Δt and the second power of Δx, implying that the scheme is of first order accuracy in time and second order accuracy in space as well. Indeed, this result further verifies what would otherwise be expected because the way this scheme is based on Taylor series expansions.

We furthermore provide here the evolution equation with only space derivatives, i.e. the modified equation. In order to achieve it, we firstly find expressions for $\frac{\partial^2 T}{\partial t^2}$, $\frac{\partial^3 T}{\partial t^3}$ and $\frac{\partial^4 T}{\partial t^4}$, and secondly (be careful with this) for $\frac{\partial^3 T}{\partial x\partial t^2}$, $\frac{\partial^3 T}{\partial x^2\partial t}$ and $\frac{\partial^4 T}{\partial x^2\partial t^2}$, by differentiating Eq. 3.10; all of which we use systematically to eliminate the time derivatives in it. This implies that the modified equation associated with the Schmidt's scheme, the equation of the grid function from Schmidt's difference equation, is:

$$\frac{\partial T}{\partial t}-\alpha\frac{\partial^2 T}{\partial x^2}=\frac{(\Delta x)^4}{12\,\Delta t}(F-6\,F^2)\frac{\partial^4 T}{\partial x^4}+\mathcal{O}((\Delta x)^3), \tag{3.11}$$

which is one of two ways to next not only calculate the local error but also estimate the stability.

3.2.3. **Stability.** We secondly analyze the stability using the von Neumann's method. Substituting a term of Eq. 2.1 into each term in Eq. 3.9 we get the following real valued, growth factor associated with the Schmidt's scheme:

$$G=1-2\,F+F\,(e^{-i\,k\,\Delta x}+e^{i\,k\Delta x})=1-2\,F\,[1-\cos{(k\Delta x)}], \tag{3.12}$$

which must to satisfy Eq. 2.16, $-1\le G\le 1$. Since $0\le 1-\cos{(k\Delta x)}\le 2$, then $G\le 1$; but $-1\le G$ only if we restrict to

$$F\le\frac{1}{2}, \tag{3.13}$$

which, like the so-called (after its discoverers [17]) Courant-Friedrichs-Lewy condition in solutions of hyperbolic partial differential equations (e.g. see second paper in the series), provides critical information about how to discretize the space and time variables of the difference equation. Indeed, this inequality states that the time step must go to zero at least as fast as $(\Delta x)^2$, which is certainly very restrictive for fine spatial resolution and hence numerically expensive for large-scale computations. In fact, this condition for stability (not for accuracy) has a qualitatively different physical meaning from that of the better understood Courant-Friedrichs-Lewy condition, which is shortly discussed. We anticipate (see Eq. 3.12) that the shorter and faster waves are most prone to instability, implying stiffness. We also foresee that when $k\Delta x = \pi$ then:

$$\lim_{k\Delta x \to \pi} G = 1 - 4\,F, \tag{3.14}$$

which means that high frequency oscillations ($G < 0$) appear when F is above 0.25. The comparison of Fig. 2 shows it quite well. As a result, this Courant-Friedrichs-Lewy-type condition in explicit solutions of parabolic partial differential equations is of stiff rather than kinematic origin (see second paper in the series). Eqs. 2.19 and 2.23 are stiff [18].

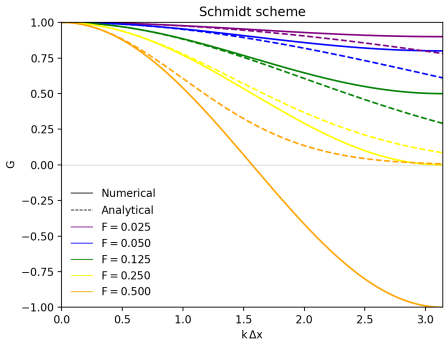


FIGURE 2. The same as Fig. 1 but for the Schmidt's scheme, Eq. 3.12.

Additionally, the Hirt's method is another way to obtain this. We may think that if the first (even) term on the right hand of modified equation 3.11, which acts as a diffusion term, has to be positive for T_j^n to be damped in time, must have $F \leq 1/6$. However, this is only a partial condition. The case is actually quite subtle because this first (even) term on the right hand of Eq. 3.11 is of the same order as the wavenumber in the large wavenumber limit. Thus, in this situation the necessary and sufficient condition for stability is

[10, Eq. 5.11]:

$$\frac{8\Delta t}{(\Delta x)^4}\left[\frac{(\Delta x)^2}{3}(\alpha)-\Delta t(\alpha^2)-\left(\frac{(\Delta x)^4}{12\,\Delta t}(F-6\,F^2)\right)\right]\geq 0, \tag{3.15}$$

which implies certainly the same as Eq. 3.13, $F \leq 1/2$.

3.2.4. **Accuracy.** So far, we know the truncation error is of first order accuracy in time and second order accuracy in space. Now, thirdly, again making use of the Hirt's and von Neumann's analysis we then investigate its behavior in Fourier space; while noting that the absence of odd-order derivatives in any diffusion difference equation indicates that, in this paper, we do not need to consider the error related with timing or phase.

On the one hand, by applying the discrete Fourier series grid solution of Eq. 2.1 to the Schmidt's modified equation 3.11 including terms up to the next-to-leading order, the damping term of the numerical solution becomes (see Eq. 2.27):

$$\exp\left\{\left[-\alpha+\frac{1}{12}(F-6\,F^2)\,\frac{(\Delta x)^4}{\Delta t}k^2\right]k^2\,\Delta t\right\}. \tag{3.16}$$

Like its analytical counterpart (see Eq. 2.20), we again notice that there is no dissipation for $k=0$ and it will be greater at higher wavenumbers. We now present the search for stiffness in the left panel of Fig. 3 by comparing Eq. 3.16 to Eq. 2.20 in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F values.

In this panel we observe that there is three possibilities for the numerical damping factor: the case when $F<\frac{1}{6}$ where the Schmidt's low wavelengths dissipate more slowly that those in the the analytical solution do (hereafter underdamped regimen); the case threshold when $F=\frac{1}{6}$ (black line in the left panel of Fig. 3) where the Schmidt's solution dissipate just as fast as that the analytical solution does (hereafter critically damped regimen); and the case when $F>\frac{1}{6}$ where Schmidt's low wavelengths dissipate faster that those in the analytical solution do (hereafter overdamped regimen). In the interesting threshold when $F=\frac{1}{6}$ the next-to-leading damping coefficient is zero, and for progressively higher values of F, depending on the frequency of the Fourier mode, shift gradually to an anomalous overdamped process, which causes the stiffness. Next, a more in-depth explanation of this effect is found.

On the other hand, since the exact growth factor in one of the Fourier modes is equal to $\exp(-\alpha\,k^2\,\Delta t)$, the Schmidt's relative amplitude error, Eq. 2.29, is given by

$$\delta A=[1-2\,F\,(1-\cos\,(k\Delta x))]\,e^{\alpha\,k^2\,\Delta t}-1. \tag{3.17}$$

In the right panel of Fig. 3 we show this relative error in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F values.

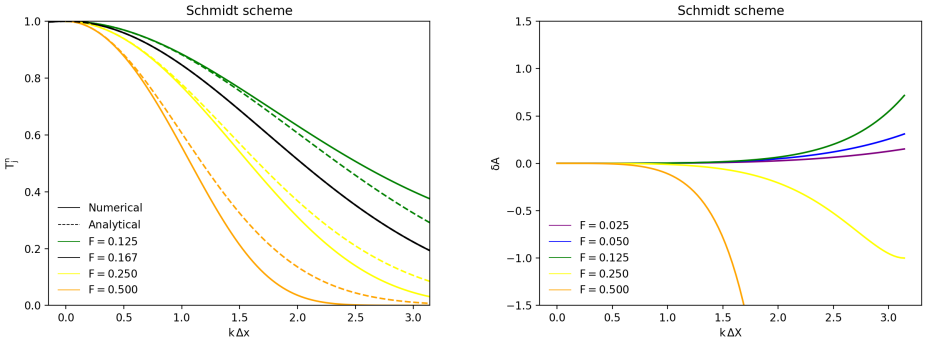


FIGURE 3. Schmidt's dissipation error. The left panel shows the decays of the numerical solutions (approximated to the five-order derivatives); while the right panel shows the relative amplitude error as a function of phase angle or frequency, in radians, for specific Fourier number values. The overlapping black solid and dashed lines in the left panel show the (numerical and semi-analytical) solutions corresponding to critical $F = \frac{1}{6}$. We conclude that the damping factor offers an average description of evolutionary behaviors of amplitudes that is valid at least qualitatively.

In this right panel we observe that there is three Fourier number ranges where the Schmidt's amplitude decreases to different degrees relative to the corresponding ones in the analytical solution: $F < \frac{1}{6}$ range where Schmidt's low wavelengths dissipate more slowly that those in the the analytical solution do (in agreement with the underdamped regimen from Eq. 3.16); $\frac{1}{6} < F \leq 0.25$ range where they achieve overdamping (in agreement with the overdamped regimen from Eq. 3.16); and $0.25 < F \leq 0.5$ range where we find that $\delta A < -1$ (i.e. the ratio of the numerical amplitude to the analytical amplitude is negative) for them, which means that their numerical growth factor changes its sign from one time level to the next and, consequently, they will oscillate until eventually decay completely (we are still in the overdamped regimen). In fact, all aforementioned spurious low wavelengths are those that go unstable and as we observe they already appear with the marginally stable value or limit of the stiff regime $F = 0.5$ and even lower in the near-critical regime: the relative error is smaller than -1, then T_j^n can be considered completely erroneous since the error is larger in magnitude as the exact growth factor. In this overdamped regime, where the difference equation become stiff, all the low wavelength modes explode oscillatorily, $G < -1$ or $|G| > 1$ in Fig. 2, if the time step is larger than that of condition for stability in Eq. 3.13.

3.2.5. **Monotonicity.** Finally, looking at Eq. 3.9 in Table 1 we see that the coefficients of the new solution are 1, $1-2\,F$ and F, i.e. all of them are positive in the stability region, Eq. 3.13. Hence, the Schmidt's scheme is monotone and preserves extrema (i.e. structure) over a very long time.

3.3. The Crank-Nicolson 1947 single-step semi-explicit scheme.

3.3.1. **Construction.** In 1947 Crank and Nicolson [19] achieved a forward in time scheme using, to say the least, the average of the scheme formerly proposed by Schmidt and the Laasonen's scheme described below. Hence, this scheme is not based on Taylor series expansions, implying that we do not know apriori whether the consistency requirement is satisfied or not. In this way they obtained the scheme:

$$\frac{T_j^{n+1}-T_j^n}{\Delta t}=\alpha\left(\frac{1}{2}\right)\left(\frac{T_{j-1}^{n+1}-2T_j^{n+1}+T_{j+1}^{n+1}}{(\Delta x)^2}+\frac{T_{j-1}^{n}-2T_j^{n}+T_{j+1}^{n}}{(\Delta x)^2}\right), \tag{3.18}$$

which is again reformulated in Table 1 using Eq. 3.2.

However, as a scheme that does not proceed in a hierarchical, bottom-up fashion, the Crank-Nicolson scheme gives a (tridiagonal) system of equations to solve for all the values of T_i^{n+1} simultaneously and we must to resort to standard matrix equation solvers (e.g., see [20]), which is complicated to implement for parallel execution. Specifically, the Crank-Nicolson scheme will be referred to as a semi-explicit scheme. Besides, this monolithic nature of several much-used nonexplicit schemes implies that any anomaly strongly affects the entire solution; which, depending on the initial conditions, is not necessarily one more attractive feature for diffusion (see next Section).

3.3.2. **Consistency and order of accuracy.** Eq. 3.18 now poses a significant challenge to deriving consistency because we must reduce the number of indexes in it before raising and/or lowering them. Thus, in this situation we reformulate Eq. 3.18 once more so it would be suitable to derive an equation with derivatives only. Specifically, we rewrite Eq. 3.18 as $T_j^{n+1}=T_j^n+F\,T_{j-1}^{n+1/2}+F\,T_{j+1}^{n+1/2}-2\,F\,T_j^{n+1/2}$. Next, substituting each value of T at points other than point $(x_j,t^{n+1/2})$ in this equation in a Taylor series around the value T_j^n at that point $(x_j,t^{n+1/2})$, gives

$$\frac{\partial T}{\partial t}-\alpha\frac{\partial^2 T}{\partial x^2}=\frac{(\Delta t)^2}{24}\frac{\partial^3 T}{\partial t^3}+\frac{\alpha\,(\Delta x)^2}{12}\frac{\partial^4 T}{\partial x^4}+\cdots. \tag{3.19}$$

We then see from that equation, which is not yet the modified equation, that the right hand side vanishes when $\Delta t\to 0$ and $\Delta x\to 0$ and therefore the Crank-Nicolson scheme of calculation is consistent. In addition, this side goes to zero as the second power of Δt and the second power of Δx, implying that the scheme is of second order accuracy in both time and space.

We furthermore provide here the evolution equation with only space derivatives, i.e. the modified equation. In order to achieve it, we firstly find expressions for $\frac{\partial^3 T}{\partial t^3}$, and secondly (be careful with this) for $\frac{\partial^3 T}{\partial x^2 \partial t}$, by differentiating Eq. 3.19; all of which we use systematically to eliminate the time derivatives in it. This implies that the modified equation associated with the Crank-Nicolson scheme, the equation of the grid function from Crank-Nicolson difference equation, is:

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = F \, \frac{(\Delta x)^4}{12 \, \Delta t} \frac{\partial^3 T}{\partial x^4} + \mathcal{O}((\Delta x)^3), \tag{3.20}$$

which is one of two ways to next not only calculate the local error but also estimate the stability.

3.3.3. **Stability.** We secondly now analyze the stability using the von Neumann's method. Substituting a term of Eq. 2.1 into each term in Eq. 3.18 we get the following real valued, growth factor associated with the Crank-Nicolson scheme:

$$-1 \leq G = \frac{2 - 2\,F\,(1 - \cos{(k\Delta x)})}{2 + 2\,F\,(1 - \cos{(k\Delta x)})} \leq 1, \tag{3.21}$$

which satisfies Eq. 2.16 and demonstrates that no restrictions are put on the resolution using the Crank-Nicolson scheme, which have the advantage of being able to represent smooth solutions (see below). We anticipate (see Eq. 3.21) that the shorter waves are now not prone to instability, not implying stiffness. However, we also foresee that when $k\Delta x = \pi$ then:

$$\lim_{k\Delta x \to \pi} G = \frac{2 - 4\,F}{2 + 4\,F}, \tag{3.22}$$

which means now that the solution is now oscillatory, $G < 0$, when F is above 0.5 (see Fig. 4); i.e. the low (oscillatory) wavelengths are propagated as weakly damped oscillations in time and therefore may persist for a long time (see Fig. 4), thereby implying (not a condition or constraint but) a prominent shortcoming.

Additionally, the Hirt's method is another way to obtain this. The even, first term on the right hand of modified equation Eq. 3.20, which acts as a diffusion term (there is no dispersive derivatives, as already said), is both not of the same order as k_{max} and always positive; implying that the Crank-Nicolson scheme is indeed unconditionally stable.

3.3.4. **Accuracy.** We thirdly investigate how accurate the numerical solution is using the Hirt's and von Neumann's complementary analysis.

On the one hand, by applying the discrete Fourier series grid solution of Eq. 2.1 to the Crank-Nicolson modified equation 3.20 including terms up to the next-to-leading order, the damping term of the numerical solution becomes

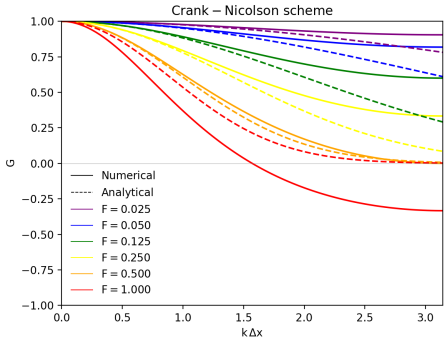


FIGURE 4. The same as Fig. 1 but for the Crank-Nicolson scheme, Eq. 3.21.

(see Eq. 2.27):

$$\exp\left\{\left[-\alpha+\frac{1}{12}F\,\frac{(\Delta x)^4}{\Delta t}k^2\right]k^2\,\Delta t\right\}. \tag{3.23}$$

Like its analytical counterpart, we again notice that there is no dissipation for $k=0$ and it will be greater at higher wavenumbers. We present the search for stiffness in the left panel of Fig. 5 by comparing Eq. 3.23 to Eq. 2.20 in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F values. We now do not observe stiffness because the next-to-leading term in Eq. 3.23 is indeed anti-damping. In all cases, we observe that the Crank-Nicolson low wavelengths dissipate more slowly that those in the the analytical solution do, implying there is no overdamped or stiff regime. However, we do observe that for progressively higher values of F the Crank-Nicolson dissipation is closer to the one from the analytical solution; i.e. the Crank-Nicolson grid function has just one underdamped regimen. Next, a more in-depth explanation of this effect is found.

On the other hand, since the exact growth factor in one of the Fourier modes is equal to $\exp(-\alpha\,k^2\,\Delta t)$, the Crank-Nicolson relative amplitude error, Eq. 2.29, is given by

$$\delta A=\frac{\left[2-2\,F\,(1-\cos\left(k\Delta x\right))\right]e^{\alpha\,k^2\,\Delta t}}{2+2\,F\,(1-\cos\left(k\Delta x\right))}-1, \tag{3.24}$$

In the right panel of Fig. 5 we show this relative error in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F values, where we observe the marked preference for F near 0.5. Indeed, for $F<0.5$ the low wavelengths are underdamped (in agreement with the underdamped regimen from Eq. 3.23) while for $F\geq 0.5$ they undergo an oscillating damping (in agreement with Eq. 3.22), giving rise to an oscillatory behavior, $\delta A<-1$. Only for $F<0.5$ the spurious oscillating modes are completely dissipated, implying

low computational efficiency. These undesired oscillations, scarcely recognized or underestimated in the literature [19], are accordingly due to instability involved in the explicit half step of the not purely implicit Crank-Nicolson method.

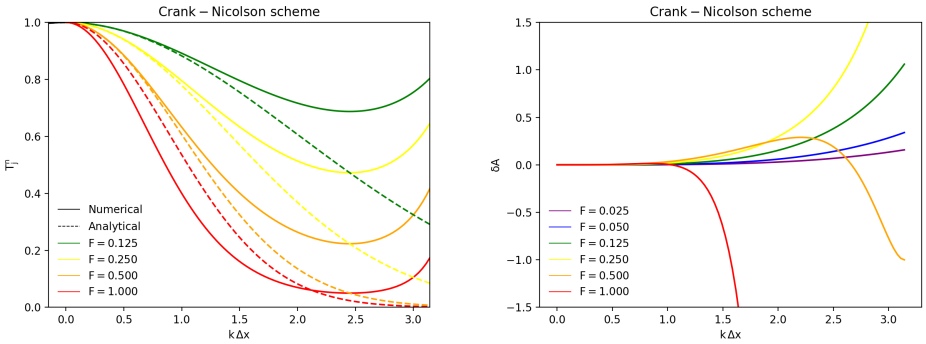


FIGURE 5. The same as Fig. 3 but for the Crank-Nicolson scheme.

3.3.5. **Monotonicity.** Finally, looking at Eq. 3.18 in Table 1 we see that the coefficients of the new solution are $-0.5\,F$, $1+F$, F and $1-F$. Hence, the Crank-Nicolson scheme is nonmonotone, i.e. it also may produce spurious oscillations.

3.4. The Laasonen's 1949 single-step implicit scheme.

3.4.1. **Construction.** In 1949 Laasonen [21] used the same approximation as Schmidt did, but evaluating the space derivative forwards in time, at time step $n+1$ instead of at time step n; i.e.,

$$\frac{T_j^{n+1}-T_j^n}{\Delta t}=\alpha\frac{T_{j-1}^{n+1}-2T_j^{n+1}+T_{j+1}^{n+1}}{(\Delta x)^2}. \tag{3.25}$$

Table 1 emphasizes that Eq. 3.25 is purely time-implicit. Therefore, as a non-explicit scheme that it is, the solution at the next time level is computed from the present time level by solving the tridiagonal system of equations that Eq. 3.25 gives (e.g., see [20]). The same as the Crank-Nicolson scheme.

3.4.2. **Consistency and order of accuracy.** We firstly analyze the consistency of the scheme using the Hirt's method. Substituting each value of T at points other than point (x_j,t^n) in the scheme in a Taylor series around the value T_j^n at that point (x_j,t^n), i.e. we substitute Eqs. 2.6 and 2.13 in Eq. 3.25, gives

$$\frac{\partial T}{\partial t}-\alpha\frac{\partial^2 T}{\partial x^2}=\frac{\Delta t}{2}\frac{\partial^2 T}{\partial t^2}+\frac{(\Delta t)^2}{6}\frac{\partial^3 T}{\partial t^3}+\frac{\alpha\,(\Delta t)^3}{24}\frac{\partial^4 T}{\partial t^4}+\frac{\alpha\,(\Delta x)^2}{12}\frac{\partial^4 T}{\partial x^4}+\cdots. \tag{3.26}$$

We then see from that equation, which is not yet the modified equation, that the right hand side vanishes when $\Delta t \to 0$ and $\Delta x \to 0$ and therefore the Laasonen's scheme of calculation is consistent. In addition, this side goes to zero as the first power of Δt and the second power of Δx, implying that the scheme is of first order accuracy in time and second order accuracy in space as well. Indeed, this result further verifies what would otherwise be expected because the way this scheme is based on Taylor series expansions.

We furthermore provide here the evolution equation with only space derivatives, i.e. the modified equation. In order to achieve it, we firstly find expressions for $\frac{\partial^2 T}{\partial t^2}$, $\frac{\partial^3 T}{\partial t^3}$ and $\frac{\partial^4 T}{\partial t^4}$, and secondly (be careful with this) for $\frac{\partial^3 T}{\partial x^2 \partial t}$ and $\frac{\partial^4 T}{\partial x^2 \partial t^2}$, by differentiating Eq. 3.26; all of which we use systematically to eliminate the time derivatives in it. This implies that the modified equation associated with the Laasonen's scheme, the equation of the grid function from Laasonen's difference equation, is:

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = \frac{(\Delta x)^4}{12\,\Delta t}(F + 6\,F^2)\frac{\partial^4 T}{\partial x^4} + \mathcal{O}((\Delta x)^4), \tag{3.27}$$

which is one of two ways to next not only calculate the local error but also estimate the stability.

3.4.3. **Stability.** We secondly now analyze the stability using the von Neumann's method. Substituting a term of Eq. 2.1 into each term in Eq. 3.25 we get the following real valued, growth factor associated with the Laasonen's scheme:

$$-1 \leq G = \frac{1}{1 + 2\,F\,(1 - \cos{(k\Delta x)})} \leq 1, \tag{3.28}$$

which satisfies Eq. 2.16 and demonstrates that no restrictions are put on the resolution using the Laasonen's scheme, which have the advantage of being able to represent smooth solutions (see below). We also anticipate (see Eq. 3.28) that the shorter waves are neither prone to instability, not implying stiffness. The same as using the Crank-Nicolson scheme. However, quite the opposite of what has been found in Crank-Nicolson scheme, we now foresee that when $k\Delta x = \pi$ then G is very small:

$$\lim_{k\Delta x \to \pi} G = \frac{1}{1 + 4\,F}, \tag{3.29}$$

which means that the Laasonen's scheme never produces oscillations in time, thereby implying a unique feature (see Fig. 6 compared to Figs. 2 and 4). As a consequence, the Laasonen's numerical solution will be indeed smoother than expected, e.g., from the Crank-Nicolson scheme.

Additionally, the Hirt's method is another way to obtain this. The even, first term on the right hand of modified equation Eq. 3.27, which acts as a diffusion term (there is no dispersive derivatives), is both not of the same order

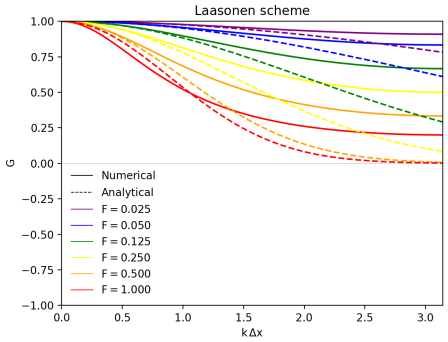


FIGURE 6. The same as Fig. 1 but for the Laasonen's scheme, Eq. 3.28.

as k_{max} and always positive; implying that the Laasonen's scheme is indeed unconditionally stable.

In fact, this property of being stable turns out to be true for all purely implicit schemes. That said, if not for stability, for accuracy we cannot use a large time step. To clearly see it, note the ratio of the Laasonen's numerical damping rate (from Eq. 2.28, $e^{-iw_{num}\Delta t} = |G|$) to the exact damping rate (Eq. 2.20) is

$$\begin{aligned} \frac{w_{num}}{w} &= -\frac{\ln\left[1 + 2\,F\,(1 - \cos\left(k\Delta x\right))\right]}{Fk^2(\Delta x)^2} \\ &= -1 + \left(\frac{1+6F}{12}\right)(k\Delta x)^2 + \mathcal{O}((k\Delta x)^4); \end{aligned} \tag{3.30}$$

which show us that for, e.g., $F = 1.5$ the error on the damping rate is about the exact damping rate (Eq. 3.30 is on average about 1).

3.4.4. **Accuracy.** We thirdly investigate how accurate the numerical solution is using the Hirt's and von Neumann's complementary analysis.

On the one hand, by applying the discrete Fourier series grid solution of Eq. 2.1 to the Laasonen's modified equation 3.27 including terms up to the next-to-leading order, the damping term of the numerical solution becomes (see Eq. 2.27):

$$\exp\left\{\left[-\alpha + \frac{1}{12}(F + 6\,F^2)\,\frac{(\Delta x)^4}{\Delta t}k^2\right]k^2\,\Delta t\right\}. \tag{3.31}$$

Like its analytical counterpart, we once again notice that there is no dissipation for $k = 0$ and it will be greater at higher wavenumbers. We present the search for stiffness in the left panel of Fig. 7 by comparing Eq. 3.31 to Eq. 2.27 in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F

values. We now do not observe stiffness. All next-to-leading terms in Eq. 3.31 are indeed anti-damping. In all cases, we observe that the Laasonen's low wavelengths dissipate more slowly (indeed, too slowly) that those in the the analytical solution do, and do so equally for progressively higher values of F. The Laasonen's grid function has also just one underdamped regimen. Next, a more in-depth explanation of this effect is found.

On the other hand, since the exact growth factor in one of the Fourier modes is equal to $\exp(-\alpha\, k^2\, \Delta t)$, the Laasonen's relative amplitude error, Eq. 2.29, is given by

$$\delta A = \frac{e^{\alpha\, k^2\, \Delta t}}{1 + 2\, F\, (1 - \cos\,(k \Delta x))} - 1, \tag{3.32}$$

In the right panel of Fig. 7 we show this error in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F values, where we observe that almost all the modes, but especially those of high frequency (say, the worst modes), are excesively under-damped (in agreement with the underdamped regimen from Eq. 3.31) and, consequently, the Laasonen's scheme is going to lose accuracy. In fact, in agreement with Eq. 3.30, we now observe that the Laasonen's scheme becomes too little dissipative and little or no scale selective, which indeed can lead to some loss of measurable Fourier modes and a great reduction in sensitivity. The Laasonen's scheme, otherwise but the same as the Crank-Nicolson scheme, suffers from low computational efficiency.

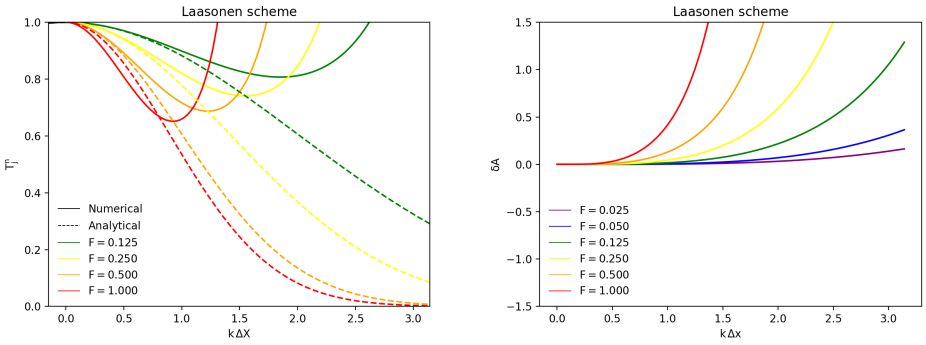


FIGURE 7. The same as Fig. 3 but for the Laasonen's scheme.

3.4.5. **Monotonicity.** Finally, looking at Eq. 3.25 in Table 1 we see that the coefficients of the new solution are $-F$, $1 + 2\, F$ and 1, respectively. Hence, the Laasonen's scheme is nonmonotone, i.e. it also may produce spurious oscillations.

3.5. The Du Fort-Frankel 1953 three-time-level explicit scheme.

3.5.1. **Construction.** In 1953 Du Fort and Frankel [22] (who were working with the wave equation) proposed the following scheme with the aim of overcoming the stability problem due to the Richardson approach we found in Eq. 3.8 by taking the time average of T_j^n at (x_j, t^n), i.e.,

$$\frac{T_j^{n+1} - T_j^{n-1}}{2\,\Delta t} = \alpha \frac{T_{j-1}^n - (T_j^{n+1} + T_j^{n-1}) + T_{j+1}^n}{(\Delta x)^2}. \tag{3.33}$$

Hence, the Du Fort and Frankel explicit scheme (see Table 1) still has the problem of setting the initial grid function value, i.e. another single-step scheme is needed; and is also not based on Taylor series expansions, which implies that we do not know apriori whether the consistency requirement is satisfied or not.

3.5.2. **Consistency and order of accuracy.** We firstly analyze the consistency of the scheme using Hirt's method. Substituting each value of T at points other than point (x_j, t^n) in the scheme in a Taylor series around the value T_j^n at that point (x_j, t^n), i.e. we substitute Eqs. 2.6 and 2.13 in Eq. 3.33, gives

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = -\alpha \frac{(\Delta t)^2}{(\Delta x)^2} \frac{\partial^2 T}{\partial t^2} - \frac{(\Delta t)^2}{6} \frac{\partial^3 T}{\partial t^3} - \frac{\alpha}{12} \frac{(\Delta t)^4}{(\Delta x)^2} \frac{\partial^3 T}{\partial x^3} + \alpha \frac{(\Delta x)^2}{12} \frac{\partial^4 T}{\partial t^4} + \cdots . \tag{3.34}$$

We then see from that equation (which is not yet the modified equation) that, due to the first term on the right hand side, the Du Fort-Frankel scheme of calculation is consistent only if $\Delta t \to 0$ and $\Delta x \to 0$ are taken such that $\frac{\Delta t}{\Delta x} \to 0$. In other words,

(1) if a time step $\Delta t \sim (\Delta x)^2$ is used we then achieve second order accuracy (on space and consequently in time as well);
(2) however, for any value of $\frac{\Delta t}{\Delta x}$ the solution is not quite parabolic but nearly hyperbolic (i.e. it converges to that of a wave equation). We illustrate the different effects associated to this conflicting region, such as complex amplitudes, coalescence and divergence in the next subsections.

That is to say, the Du Fort-Frankel scheme is conditionally consistent, i.e. we must now to be careful in choosing the time step to assure the consistency of the scheme (if we want to use it, as suggested for its other virtues). We note that the reason for its conditional consistency is that Du Fort and Frankel changed the method after they (actually Richardson) made the Taylor expansions in the usual way, i.e. an apparently small technical difference turn out to have profound consequences for the way in which a difference scheme creates the numerical solution.

We furthermore provide here the evolution equation with only space derivatives, i.e. the modified equation. In order to achieve it, we firstly find expressions for $\frac{\partial^2 T}{\partial t^2}$, $\frac{\partial^3 T}{\partial t^3}$ and $\frac{\partial^4 T}{\partial t^4}$, and secondly (be careful with this) for $\frac{\partial^3 T}{\partial x \partial t^2}$, $\frac{\partial^3 T}{\partial x^2 \partial t}$ and $\frac{\partial^4 T}{\partial x^2 \partial t^2}$, by differentiating Eq. 3.34; all of which we use systematically to eliminate the time derivatives in it. This implies that the modified equation associated with the Du Fort-Frankel scheme, the equation of the grid function from Du Fort-Frankel difference equation, is:

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = \frac{(\Delta x)^4}{12\,\Delta t}(F - 12\,F^3)\frac{\partial^4 T}{\partial x^4} + \mathcal{O}((\Delta x)^4), \tag{3.35}$$

which is one of two ways to next not only calculate the local error but also estimate the stability.

3.5.3. **Stability.** We secondly now analyze the stability using the von Neumann's method. Substituting a term of Eq. 2.1 into each term in Eq. 3.33 we get the following equation for the growth factor associated with the Du Fort-Frankel scheme:

$$\begin{aligned} G &= \left(\frac{1-2\,F}{1+2\,F}\right)\frac{1}{G} + \left(\frac{2\,F}{1+2\,F}\right)(e^{-i\,k\,\Delta x} + e^{i\,k\Delta x}) \\ &= \left(\frac{1-2\,F}{1+2\,F}\right)\frac{1}{G} - \frac{4\,F\,\cos(k\Delta x)}{1+2\,F}. \end{aligned} \tag{3.36}$$

Whence the two solutions for the Du Fort-Frankel growth factor are:

$$G_\pm = \frac{2\,F\,\cos(k\Delta x) \pm \sqrt{1-(2\,F\,\sin(k\Delta x))^2}}{1+2\,F}; \tag{3.37}$$

which means that now spurious solutions exist which can contaminate the solution. The same as the Richardson's solution and for the same reason: both of them are three level schemes. This fact makes the three (or more) level schemes such as the Du Fort-Frankel scheme difficult to understand. We offer a more in-depth interpretation of this problem in the next subsection concerning accuracy.

Here, as regards stability, the growth factor magnitudes are such that

(1) either $1-(2\,F\,\sin(k\Delta x))^2 \geq 0$ which implies $|G_\pm| \leq \frac{1+2\,F\,|\cos(k\Delta x)|}{1+2\,F} \leq 1$, which satisfies Eq. 2.16 and thus no parabolic stability restriction on the time step appears in this case;
(2) or $1-(2\,F\,\sin(k\Delta x))^2 < 0$ which implies $|G_\pm| = \frac{2\,F-1}{2\,F+1} < 1$, i.e. both cases satisfy Eq. 2.16.

Therefore, the Du Fort-Frankel scheme is unconditionally stable with the diffusion equation, as intended by its authors. The Du Fort-Frankel scheme avoids the Richardson's instability. We point out that the shorter waves are most

prone to instability (see Eq. 3.36), implying stiffness. We also point out that when $k\Delta x = \pi$ then G_- is not very small:

$$\lim_{k\Delta x\to\pi} G_+ = \frac{1-2\,F}{1+2\,F}, \tag{3.38a}$$

$$\lim_{k\Delta x\to\pi} G_- = -1, \tag{3.38b}$$

which means that the Du Fort-Frankel scheme produces high frequency oscillations in time (see upper panels of Fig. 10 below). The same as the Schmidt's scheme (see Fig. 2).

Additionally, the Hirt's method is another way to obtain this. Although this is now actually quite subtle, the same as analyzing the Schmidt's scheme. We may now think that if the first (even) term on the right hand of Eq. 3.35, which acts as a diffusion term, has to be positive for T_j^n to be damped in time, must have $F \leq \sqrt{\frac{1}{12}}$. However, this in not the case. This is only a partial condition because this first (even) term on the right hand of Eq. 3.35 is of the same order as the wavenumber in the large wavenumber limit. Thus, in this case the necessary and sufficient condition for stability is [10, Eq. 5.11]:

$$\frac{8\Delta t}{(\Delta x)^4}\left[\frac{(\Delta x)^2}{3}(\alpha) - \Delta t(\alpha^2) - \left(\frac{(\Delta x)^4}{12\,\Delta t}(F-12\,F^3)\right)\right] \geq 0, \tag{3.39}$$

which is always is positive, implying the same as Eq. 3.37; i.e. the Du Fort-Frankel scheme is indeed unconditionally stable.

That being said, if not for stability, for consistency we cannot use a large time step. Therefore, we are with the same restriction as the one for the Schmidt's scheme. To clearly see it, Fig. 8 shows the region $1-(2\,F\,\sin(k\Delta x))^2 < 0$ where the discrete Fourier amplitudes are a complex quantity (white region in the Fig. 8), i.e. where we have a hyperbolic Du Fort-Frankel scheme. Increasing F above 0.5 induces an imaginary part to the growth factor for certain mid frequencies.

3.5.4. **Accuracy.** We thirdly investigate how accurate the numerical solution is using the Hirt's and von Neumann's complementary analysis.

On the one hand, by applying the discrete Fourier series grid solution of Eq. 2.1 to the Du Fort-Frankel modified equation 3.35 including terms up to the next-to-leading order, the damping term of the numerical solution becomes (see Eq. 2.27):

$$\exp\left\{\left[-\alpha + \frac{1}{12}(F-12\,F^3)\,\frac{(\Delta x)^4}{\Delta t}k^2\right]k^2\,\Delta t\right\}. \tag{3.40}$$

Like its analytical counterpart, we again notice that there is no dissipation for $k=0$ and it will be greater at higher wavenumbers. We present the search for stiffness in the Fig. 9 by comparing Eq. 3.40 to Eq. 2.27 in all the Fourier modes on the grid (Eq. 2.3) evaluated for specific F values.

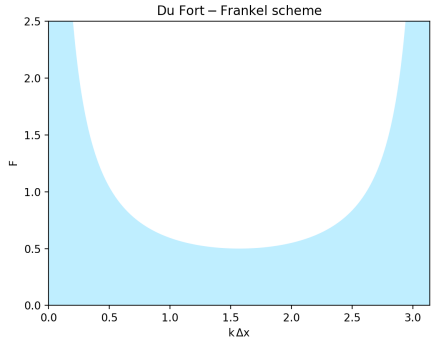


FIGURE 8. Consistency region (by blue colour) of the Du Fort-Frankel scheme.

We observe that there is three possibilities for the numerical damping factor: the underdamped regimen when $F < \sqrt{\frac{1}{12}}$ where the Du Fort-Frankel low wavelengths dissipate less rapidly that those in the the analytical solution do; the critically damped regimen when $F = \sqrt{\frac{1}{12}}$ (black line in Fig. 9) where the Du Fort-Frankel solution dissipate just as fast as that the analytical solution does; and the overdamped regimen when $F > \sqrt{\frac{1}{12}}$ where Du Fort-Frankel low wavelengths dissipate faster that those in the the analytical solution do. In the interesting threshold when $F = \sqrt{\frac{1}{12}}$ the next-to-leading damping coefficient is zero, and for progressively higher values of F, depending on the frequency of the Fourier mode, shift gradually to the stiff limit. The same anomalous overdamping process as that for the Schmidt's stiffness, which appears in Fig. 9. At that point, however, Du Fort-Frankel modes do not explode but yield an incorrect solution thanks to the computational modes. Next, a more in-depth explanation of this now somewhat complicated effect is found.

On the other hand, since the exact growth factor in one of the Fourier components is equal to $\exp(-\alpha\, k^2\, \Delta t)$, the Du Fort-Frankel relative amplitude error, Eq. 2.29, is given by

$$\delta A_{\pm} = \frac{e^{\alpha\, k^2\, \Delta t}}{1 + 2\, F} \left(2\, F\, \cos\left(k \Delta x\right) \pm \sqrt{1 - \left(2\, F\, \sin\left(k \Delta x\right)\right)^2} \right) - 1, \qquad (3.41)$$

for each solution from Eq. 3.37. In fact, the exact analysis of the amplification factors here is now more difficult because of the presence of these two solutions. To do so, we consider the two consistency regions and we first resort to the use of the growth factors for gaining a greater insight.

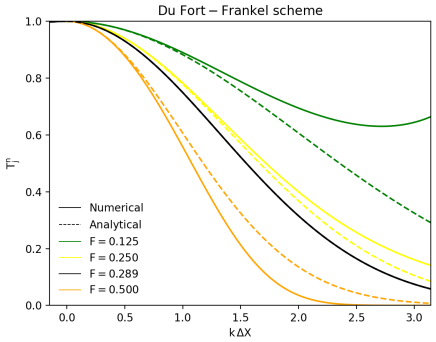


FIGURE 9. The exponential suppression factor of the Du Fort-Frankel grid function. The overlapping black solid and dashed lines show the (numerical and semi-analytical) solutions corresponding to critical $F = \sqrt{1/12}$.

(1) Setting the inconsistent region aside for the moment, then the numerical solution for each Fourier mode (see Eq. 2.15) becomes

$$T_j^n = (c_1\,|G_+^{(1)}|^n + c_2\,|G_-^{(1)}|^n)\,e^{i\,k\,j\Delta x}, \tag{3.42}$$

where c_1 and c_2 are two as yet unknown constants and the growth factors in this region, hereafter referred to as $G_\pm^{(1)}$, are real (see Eq. 3.37). In other words, in a given instant both modes are present in the numerical solution independently of each other and they both, we have already discussed, satisfy the equation Eq. 2.16; i.e. we obtain two families of linearly coupled modes for every extra time level. Fig. 10 shows that $G_+^{(1)}$ approaches the analytic solution $+1$ when $\Delta t \to 0$, and is referred as the physical mode; while $G_-^{(1)}$ approaches -1, and is referred to as the computational mode which thereby will decay in a oscillatory fashion.

The upper left panel of Fig. 10 shows that Du Fort-Frankel G_+ is comparable with analytic G (upper panel) except for low waves where the Du Fort-Frankel damping is slightly different than that of the exact solution in agreement with the regimes from Eq. 3.40. At the same time, the middle right panel has the opposite behavior; i.e., on the contrary, there is very little damping for the spurious mode of low waves, especially the $2\,\Delta x$ wave (see Eq. 2.3), which is completely undamped and just will oscillate with a period of $2\,\Delta t$ because G_- changes its sign every time step. Since we cannot hope to eliminate the computational mode completely, we shall therefore have both a weaker

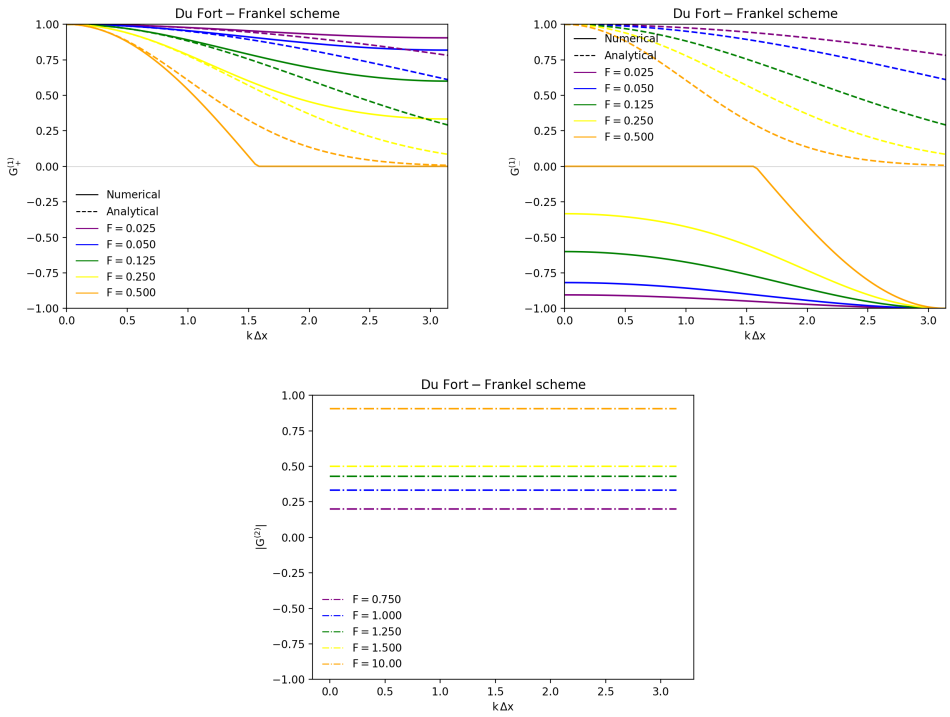


FIGURE 10. Upper: the physical (left panel) and computational (right panel) growth factors of the Du Fort-Frankel scheme using $F \leq 0.5$ vs. frequency in radians, for specific F values. Lower: the coalesced growth factors of the Du Fort-Frankel scheme using $F > 0.5$ vs. frequency in radians, for specific F values.

damping of the low waves and a stronger damping of the high waves as the result of the superposition of modes in each Fourier component.

Using upper left and right panels of Fig. 10 we also observe that for the critical value $F = 0.5$ the amplitude for each time step is completely independent of that of the immediately preceding one. In fact, interestingly, Eq. 3.33 becomes $T_i^{n+1} = \frac{1}{2}(T_{i-1}^n + T_{i+1}^n)$, i.e. a two time level scheme; specifically the Schmidt's scheme, Eq. 3.9. Furthermore, we find/observe that the amplitude of the shortest wave (the $2\,\Delta x$ wave) will remain constant forever if $T_{i-1}^n = T_{i+1}^n$, as usual; and then decays to zero for the $4\,\Delta x$ wave ($k\,\Delta x = \frac{\pi}{2}$ in the Fig. 10).

Let now the upper panel of the Fig. 11 show Eq. 3.41 in the real region. Indeed, the upper left panel shows that almost all physical low wavelenghts are underdamping, except the $F = 0.5$ case we have just

discussed. Meanwhile, the upper right panel shows that virtually all computational low wavelenghts give rise to numerical instabilities. So the superposition of panels certainly results in the regimes shown in complementary Fig. 9, as expected.

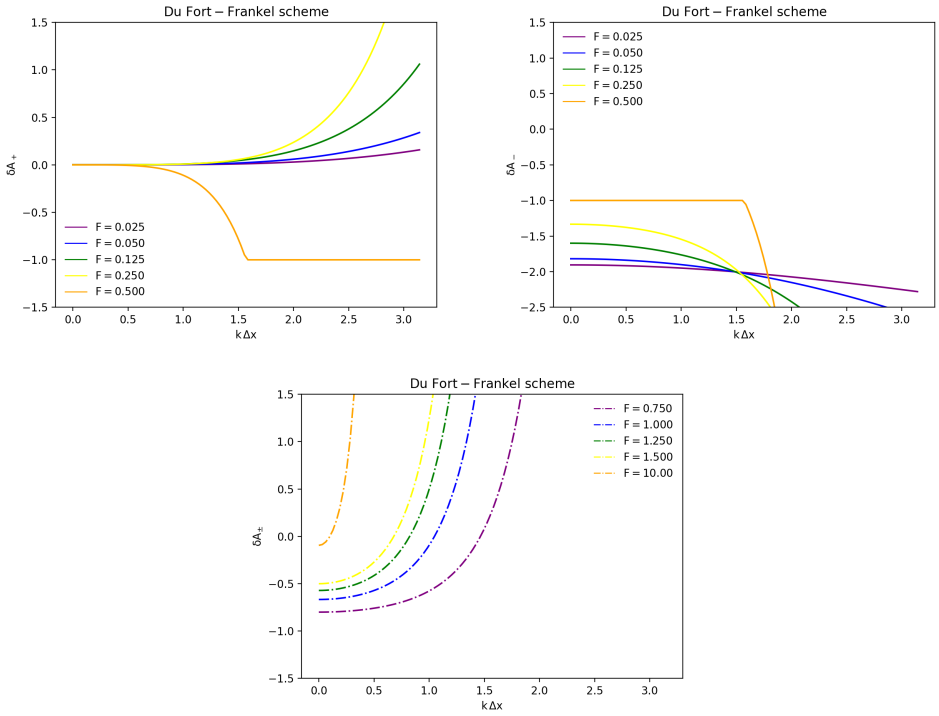


FIGURE 11. Upper: the relative amplitude error of the physical (left panel) and computational (right panel) modes using $F \leq 0.5$ vs. frequency in radians, for specific F values. Lower: the relative amplitude error of the coalesced modes using $F > 0.5$ vs. frequency in radians, for specific F value.

(2) In addition, regarding the behavior of the solution in the region not discussed so far, then the numerical solution for each Fourier mode becomes (their phase errors do exactly cancel)

$$T_j^n = [c_1 + (-1)^n\, c_2]\, |G^{(2)}|^n\, e^{i\,\phi_+^{(2)}\, n\Delta t}\, e^{i\,k\,j\Delta x}, \tag{3.43}$$

where the now growth factors are complex conjugates, thus having the same modulus, $|G_+^{(2)}| = |G_-^{(2)}| = |G^{(2)}|$, and whose respective phases are given by $\phi_\pm^{(2)} = \arctan\left(\pm\frac{(\sqrt{2F\sin k\Delta x)^2 - 1}}{2F\cos k\Delta x}\right)$ and thus after each

time step, this time, phase is shifted for those wavenumbers which satisfy $1-(2\,F\,\sin(k\Delta x))^2<0$ (see Eq. 3.43). In other words, $G_+^{(2)}$ and $G_-^{(2)}$ coalesce into one which corresponds to the inconsistency damped mode, i.e. the wave that does propagate rather than becomes diffused. Therefore, the Du Fort-Frankel stable scheme is qualitatively dissimilar to the exact solution in this region.

In fact, in lower panel of the Fig. 10 the growth factor, $|G^{(2)}|$, is plotted against the phase angle. Here, the damping in the complex region is found to be an increasing function of Fourier number whereas it is independent of frequency; which, although it avoids the Schmidt's instability, is a reason for divergence and we still wish to avoid it.

Let now the lower panel of the Fig. 11 show Eq. 3.41 in the case for which the amplitude is a complex quantity. Although the numerical solution is stable it never approaches to the analytic one because, as we can observe, the low waves do not decay at all, contrarily to what happens in the analytic solution (see the black line on the upper panel of the Fig. 10); in excellent agreement with complementary Fig. 9.

3.5.5. **Monotonicity.** Finally, looking at Eq. 3.33 in Table 1 we see that the coefficients of the new solution are $1+2\,F$, $1-2\,F$ and $2\,F$, i.e. all of them are positive in the $F\leq 0.5$ consistency region. Hence, the Du Fort-Frankel scheme is conditionally monotone.

4. The numerical tests

The critical properties of the depending on time instantaneous grid function for a basic example and proxy for the presence of unsmooth initial conditions are explored in this experimental section, as testbed for detailed fundamentals and practice quantitatively predicted in the previous Section. In particular, equipped with a semi-analytical solution, we critically assess the performance, in terms of stability, accuracy, and also convergence, of different numerical solution schemes to solve the diffusion. Comparisons with each other are finally commented.

More specifically, the illustrative application we use is the Fourier 1807 semi-analytical solution of the representative and challenging Cauchy problem described in the A on bounded domain size $L=1$, initial conditions $T_0=0$, boundary conditions $T_c=100$ and diffusivity $\alpha=0.5$ arbitrary units (we will omit these units in the rest of the paper for brevity), which deals with evolving discontinuities in solution quantity. So, by choosing a mean best-fit resolution, the full parameter space of every discretization model, summarized in Table 2, is validated (or otherwise) and compared to exact solution and other numerical solutions with special emphasis on graphical exposition.

Table 2: Parameter space of the schemes analysed in this work, including the unstable Richardson's scheme.

Author	Order	Stability	Next-to-leading accuracy		Oscillations
			Dissipation coefficient	Dispersion coefficient	
Richardson [15]	$\mathcal{O}((\Delta t)^2 + (\Delta x)^2)$	Unstable	$-F\,\frac{(\Delta x)^2}{\Delta t}$	0	No
Schmidt [16]	$\mathcal{O}(\Delta t + (\Delta x)^2)$	$F \leq 0.5$	$+\frac{1}{12}(F - 6\,F^2)\,\frac{(\Delta x)^4}{\Delta t}$	0	No
Crank-Nicolson [19]	$\mathcal{O}((\Delta \mathrm{t})^2 + (\Delta \mathrm{x})^2)$	Stable	$+\frac{1}{12}F\,\frac{(\Delta x)^4}{\Delta t}$	0	Yes
Laasonen [21]	$\mathcal{O}(\Delta t + (\Delta x)^2)$	Stable	$+\frac{1}{12}(F + 6\,F^2)\,\frac{(\Delta x)^4}{\Delta t}$	0	Yes
Du Fort-Frankel [22]	$\mathcal{O}(\Delta t)^2$	Stable	$+\frac{1}{12}(F - 12\,F^3)\,\frac{(\Delta x)^4}{\Delta t}$	0	Maybe

We remind that included in our sample is a no go condition using the Richardson's scheme and some problematic strictures with respect to the usage of both the Crank-Nicolson and the Du Fort-Frankel schemes, as well as the inability to use a large time step in all cases (for the parabolic version of the Courant-Friedrichs-Lewy condition in Schmidt's case, accuracy in non-explicit cases and consistency in Du Fort-Frankel case, respectively). In this context and with all these central questions in mind, to gain a better understanding of different subtleties creating their signatures, before anything else, we demonstrate how to perform frequentist scheme comparison.

4.1. **Resolution tests.** The present tests aim to constrain the Δt parameter by performing a frequentist error minimization either opening or, in order to resolve the initial condition well, keeping fixed a sufficiently small grid spacing $\Delta x = 10^{-1}$ units.

To this end, Fig. 12 shows the residuals' behavior obtained for the non unstable schemes, namely, Schmidt's, Du Fort-Frankel, Crank-Nicolson and Laasonen's schemes. Here, we use the simplest Schmidt's scheme to compute the lowest time level required by the Du Fort-Frankel three-level approximation. The codes naturally take into account the initial condition either in the first iteration, in case of the first row explicit schemes, or in the constant matrix, in other cases.

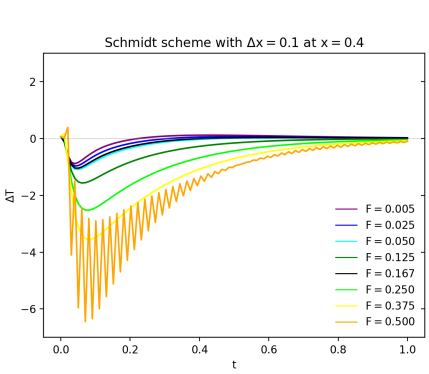
In doing so, we reveal five results of utmost importance. First, the stable schemes produce higher error with the use of larger values of F. This would seem in apparent contradiction to unconditionally stable non-explicit schemes at first glance.

Second, whilst using explicit schemes the error increases monotonically at larger time step, with non-explicit schemes we minimize the error (which, in contrast, we find to evolve with a constant value) using F near around 0.25. These results are robust with the adoption of finer spatial resolution. In addition, the residuals of the non-explicit schemes are remarkably positive using smaller Fourier numbers.

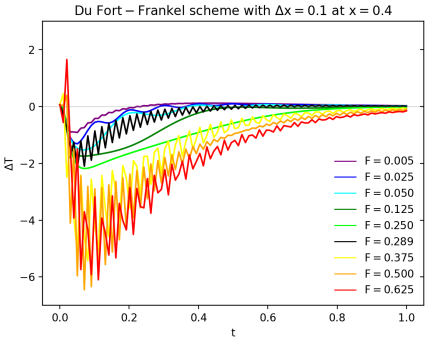
Third, the schemes of order 1 in time, namely, Schmidt's and Laasonen's schemes, are just as exact as the schemes of order 2 in time of the respective same types, namely, Du Fort-Frankel and Crank-Nicolson schemes respectively. This, again, seems very odd.

Fourth, in the explicit schemes there is a specific F value, corresponding to the black lines on the upper panels each, which interestingly does not follow its respective trend.

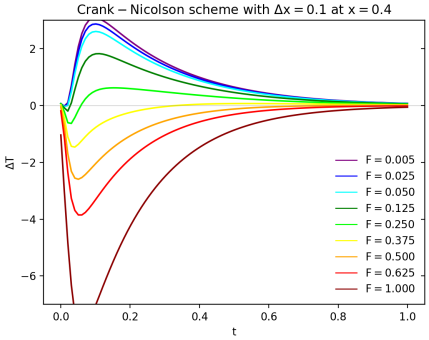
Fifth, the Schmidt's scheme starts to wildly oscillate for $F = 0.5$; and moreover, surprisingly at first glance, it does it exactly the same way as Du Fort-Frankel scheme does for the same value of F. Furthermore, the Du Fort-Frankel scheme still fluctuates considerably using F values both larger and smaller than about 0.25.



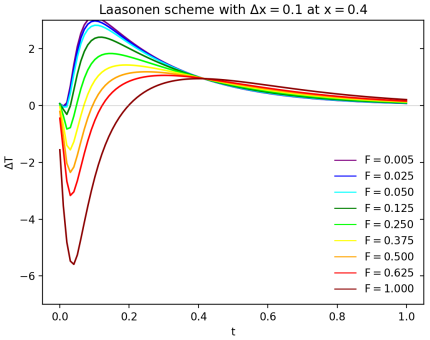
(A) Conditionally stable Schmidt method



(B) Conditionally consistent Du Fort-Frankel method



(C) Unconditionally stable Crank-Nicolson method



(D) Unconditionally stable Laasonen method

FIGURE 12. Errors of different methods for some values of the dimensionless Fourier number, Eq. 3.2 (i.e. different discretisations). We show both the Schmidt's and Du Fort-Frankel F thresholds by their respective black line.

Here we offer the following five respective explanations.

- First, the solution to this puzzle is the mixture of short-wavelength oscillations we already have quantified in Section 3 except for the Laasonen's case where the underdamped regimen is extremely large (see Eq. 3.30). Indeed, this result is a consequence of the behavior illustrated in Figs. 3, 5, 7 and 11 each. In addition to this, even though the diffusive heat transfer introduces natural damping to Godunov's non-physical oscillations and they thus do not significantly impact, in the second order schemes (including Du Fort-Frankel scheme as given in

Table 2) they, however, can leave observable remnants in the residuals, and potentially explain part of this puzzle.

- Second, there is a twofold explanation for this result. On the one hand, the monotonic evolution of the Schmidt's and Du fort-Frankel error is the remarkable consequence of the opposite dissipative behaviors above and below the critically damped regimen for each of these two schemes (left panels in Figs. 3 and 11); which does not happen in Eqs. 3.20 and 3.27.

  On the other hand, the counterintuitive evolution of the two non-explicit schemes we also find here has a number of surprising characteristics. It is the consequence of the high specific frequency underdamping above F near around 0.25 we can observe in the left panel both of Fig. 5 and in Fig. 7. In fact, for this reason the residuals of the non-explicit schemes are positive or opposite-sign using the smaller Fourier numbers. Yet, although uncertainties in both schemes increase with increasing F, the Crank-Nicolson oscillations are less accurate than the Laasonen's excessive underdamping; with these two effects being responsible for the correlation, respectively.
- Third, of note, all the schemes are second order accuracy because all of them are second order in space (as given in second column of Table 2) and the order in space leads to the order in time under these values of the discretization.
- Fourth, all the schemes are fourth order dissipative algorithms. Nonetheless, we observe in either Table 2 or Eqs. 3.11 and 3.35 that there are two schemes where these terms cancel for a specific value of F each, namely, $F = \frac{1}{6}$ in the case of the Laasonen's scheme and $F = \sqrt{\frac{1}{12}}$ in the Du Fort-Frankel scheme; which makes them more exact.
- Fifth, the still stable Schmidt's scheme with $F = 0.5$, Eq. 3.9, becomes $T_i^{n+1} = \frac{1}{2}(T_{i-1}^n + T_{i+1}^n)$; which obstructs the physical damping of our Dirichlet boundary conditions since the two neighboring grid points have the same temperature during the first time step. Moreover, the Du Fort-Frankel scheme with $F = 0.5$, Eq. 3.33, becomes exactly the same.

  Furthermore, concerning the Du Fort-Frankel consistency region Eq. 3.42 demonstrates the oscillations we found using $F \ll 0.5$ while Eq. 3.43 does the same above the inconsistency threshold.

Consequently, based on this graphical statistical evidence and at the same time with the awareness of both the Schmidt's stability limitation and the Du Fort-Frankel consistency limitation, we agree to carefully choose the simultaneous most accurate discretization of $F = 0.150$ from visual inspection of Fig. 12, i.e. a time increment $\Delta t = 3 \times 10^{-3}$ of almost two-and-a-half orders of magnitude smaller than the acceptable spatial resolution of $\Delta x = 10^{-1}$ units;

under which all the schemes are second order accuracy. In fact, resolving the abrupt initial and boundary conditions require time to reach described accuracy, and potentially explains our choice. In other words, this subsection explains why the precision of the solution does not only depend on the numerical resolution, but also on the temperature quantity gradient.

The advantage of this natural approach is that all schemes can be used on an equal footing. The caveat is that individual peculiarities are averaged out. Nonetheless, as previously anticipated, Fig. 12 illustrates that the stability benefit of the non-explicit schemes, consisting only in that we may use (and save computing time with) a large Δt parameter, does not keep against this difference equation. Here, we present the explanation that the non-explicit schemes stabilize the high frequency stiff oscillations by so much anti-damping which leads to the reduction of their accuracy rendering their viability extremely limited.

4.2. **Stability tests.** We are now able to extend that Section in three directions. We firstly demonstrate the stability of different schemes.

4.2.1. **Unstable scheme.** We compute the rod's warming in the unit domain $[0, 1]$ using the Richardson's approximation evolving its required first time step solution from the simplest Schmidt scheme. We used the same option to initialize the Du Fort-Frankel scheme, which likewise lacks its lowest time level.

In proof of agreement with Section 3.1, the upper panel of Figure 13 shows certainly that the Richardson's scheme amplifies errors and diverges very rapidly. The onset of instability occurs before $t = 0.005$ and then only get worse. The size of the prediction error, measured in RMSE or standard deviation of the residuals (lower panel of Fig. 13), is about two orders of magnitude greater than acceptable even at this early stages. We also find that it makes no difference with any other choice of the grid size. Therefore, we cannot use the Richardson's scheme to solve any parabolic partial differential equation (see third paper in the series).

4.2.2. **Conditionally stable, explicit scheme.** In the upper left panel of Figure 12 we cannot present the Schmidt's solution using $F > 0.5$, which reveals its conditional stability, in agreement with Section 3.2.

4.2.3. **Stable, explicit scheme.** In the upper right panel of Figure 12 we can present the Du Fort-Frankel solution using $F > 0.5$, which reveals the unconditional stability of this scheme. Even so, in practice we wish to avoid this complex region, in accordance with Section 3.5.

4.2.4. **Stable, non-explicit schemes.** The lower panels of Fig. 12 reveal the unconditional stability of both the Crank-Nicolson and Laasonen's schemes; in agreement with Sections 3.3 and 3.4, respectively.

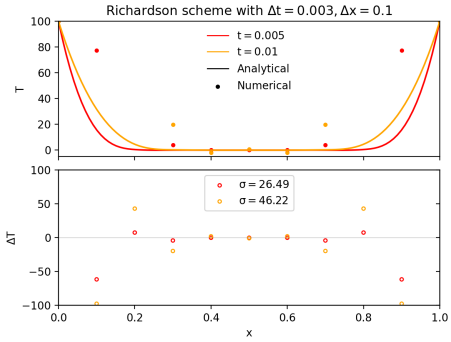


FIGURE 13. The Richardson (1910) method. Top panel: $T_i^{0.005}$ (red open circles: ○) and $T_i^{0.01}$ (orange circles: ○) for the Richardson method with $\Delta x = 10^{-1}$ and $\Delta t = 10^{-3}$. The corresponding solid lines reproduce $T(x, 0.005)$ (violet) and $T(x, 0.01)$ (blue) from Eq. A.2. Bottom panel shows residual temperatures for the mesh points. The legend in the lower panel displays the Root Mean Squared Error, RMSE, for each time.

4.3. **Accuracy tests.** We secondly demonstrate the accuracy of different schemes. To achieve this, we present in Fig. 14 the results we obtain. For the sake of easing comparisons, the left panels of Fig. 14 illustrate the temperature profiles at several selected times, and, on the other hand, the right panels of Fig. 14 illustrate the time evolution of the temperature at two illuminating locations.

By doing so, our investigation reveals four important results. First, it looks like the Du Fort-Frankel right panel has twice as many residual curves as the others.

Second, at the locations near the extremes of the rod the residuals are both anomalously high/large and positive using the non-explicit schemes.

Third, the non-explicit schemes also reach T values greater than T_c in the middle of the rod's warming time.

Fourth, concerning the quantitative comparison of the performances of the schemes, the Schmidt's scheme is the most accurate scheme; followed by the Du Fort-Frankel scheme.

Here we offer the following four respective explanations.

- First, this is what it looks like but it is the wave-optical effect created due to the tiny time step. Given the tiny temporal discretization, odd grid point residuals look like a different curve than the corresponding to the even points. Indeed, the Du Fort-Frankel solution is oscillating

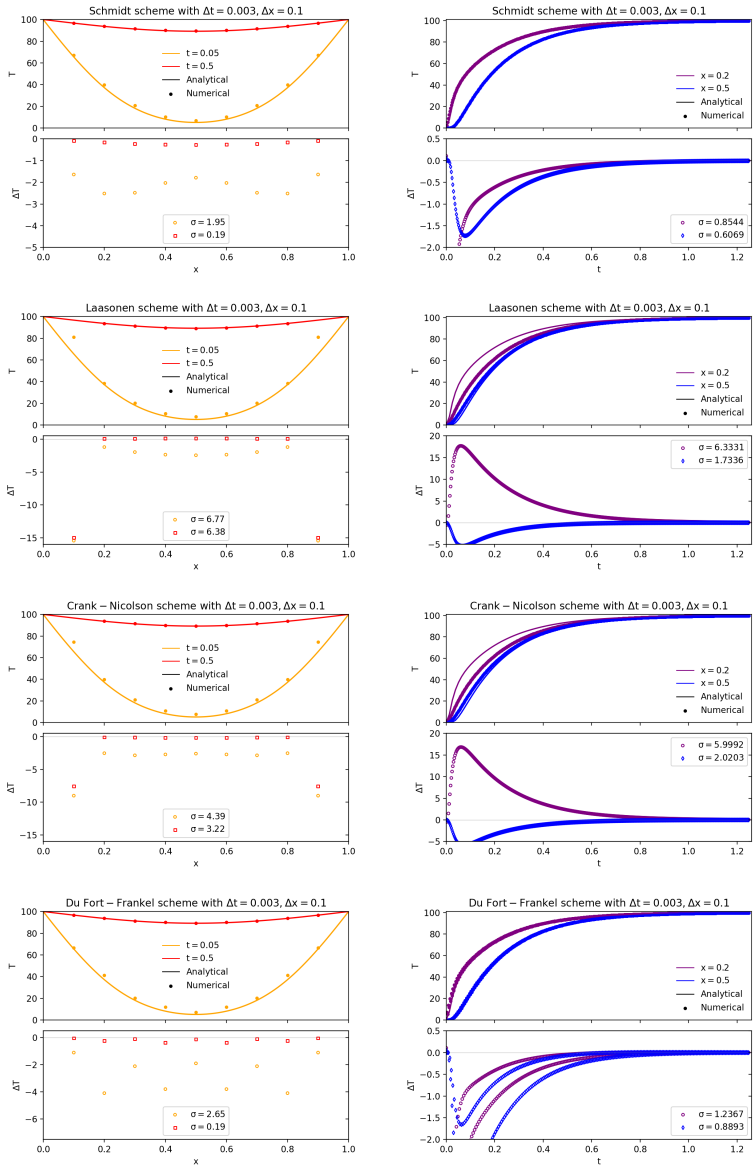


FIGURE 14. The Schmidt, Du Fort-Frankel, Crank-Nicolson and Laasonen schemes. Left panels: The same of Figure 13 but for $T_i^{0.05}$ (orange circles: ○) and $T_i^{0.5}$ (red circles: ○). Right panels: $T_{0.2}^n$ (violet circles: ○) and $T_{0.5}^n$ (blue circles: ○) as a evolving function of time until the end of the simulations; then the solid curves reproduce $T(0.2,\, t)$ (violet) and $T(0.5,\, t)$ (blue) from Eq. A.2, respectively.

at neighboring nodes due to the sign flippins of the computational mode; which is clearly seen at the Du Fort-Frankel left panel.

- Second, this is exactly what we have just found in Fig. 12. The non-explicit schemes becomes excessively under-damped, especially for the low waves, i.e. tend to reduce high-frequency (i.e., oscillatory) components of error rapidly but reduce low-frequency (i.e., smooth) components of error much more slowly; which is going to produce poor asymptotic rate of convergence. However, now in the right panels of Fig. 14 we present the location $x = 0.2$ units where such counterintuitive behavior is one order of magnitude greater than this shown in Fig. 12. Anyway, what is happening is especial given the discontinuous initial-boundary condition (see A) and also explains our next third finding.
- Third, given that the non-explicit schemes must be solved simultaneously then any irregularities of the data influence the entire solution every time step and thus they quickly smooth out them, which is altogether in good agreement with the natural damping of the dissipative transfer by diffusion. That said, when the initial-boundary conditions are discontinuous it is wiser to delay their influence as the explicit schemes indeed do. In fact, the comparison among violet circles (for shorter timescales) in the left panels of Fig. 14 provides the clearest picture where the disagreement with analytical solution is exacerbated in the nonexplicit schemes which evolved later. Likewise, the $T^n_{0.2}$ (violet circles in Fig. 14), $T^n_{0.4}$ (let's say green curves in Fig. 12) and $T^n_{0.5}$ (blue circles in Fig. 14) solutions for both Laasonen's and Crank-Nicolson schemes also show how the boundary influence decreases and disappears as we move away from it.
- Fourth, this result can be well explained by both the used initial data, and the chosen discretization. Indeed, this is because the explicit schemes are more suitable for our test case, and because of the Du Fort-Frankel fluctuations. In fact, Section 3 allowed us to anticipate the Crank-Nicolson oscillations, the Laasonen's extremely low dissipation and the Du Fort-Frankel numerical instabilities. We now show that the Schmidt's scheme not only is always better suited to our initial value problem but it presents also neither spurious waves nor excessive damping (attenuation) of the high resolution (low frequencies or large wavelength) waves, resulting in a much more preferred scheme.

4.4. **Convergence tests.** Lastly, we explore the error convergence of different schemes. In Section 2.2 we already have explained that consistency means that the error at each time step goes to zero as the grid is refined and in Section 3 we have estimated the rate that this one-step error goes to zero. Here we

investigate global (not local over one time step) rate of convergence, i.e. the overall approximation error.

Before the task, we re-present in the right panel of Fig. 15 the RMSE errors of both explicit and non-explicit, stable schemes presented as blue circles in the right panels of Fig. 14; i.e. this right panel shows the errors in the smooth time-evolution of the solution quantity derived for the middle of the rod. As we have just discussed, it is unsurprising that the performance of schemes varies greatly, and what this right panel shows is now understood. At the same time, importantly, we re-present in the left panel of Fig. 15 the RMSE errors of both explicit and non-explicit, stable schemes presented as red circles in the left panels of Fig. 14; i.e. this left panel shows the errors in the non-smooth middle of the time-evolution of the solution quantity derived for all the locations. As we have just discussed, it is also understood that the susceptibility of schemes to the initial conditions varies greatly, and this left panel verifies that the implicit results have gotten much worse especially where the influence of sufficiently large initial perturbations is much more evident. Nevertheless, the diffusion differential equation naturally puts any temperature irregularities right; i.e. it emends computational one-step errors. Therefore we note that the order of accuracy is not compromised when we compute the rate of convergence in the specific problem of diffusive transfer, something which is not always guaranteed (see second paper in the series).

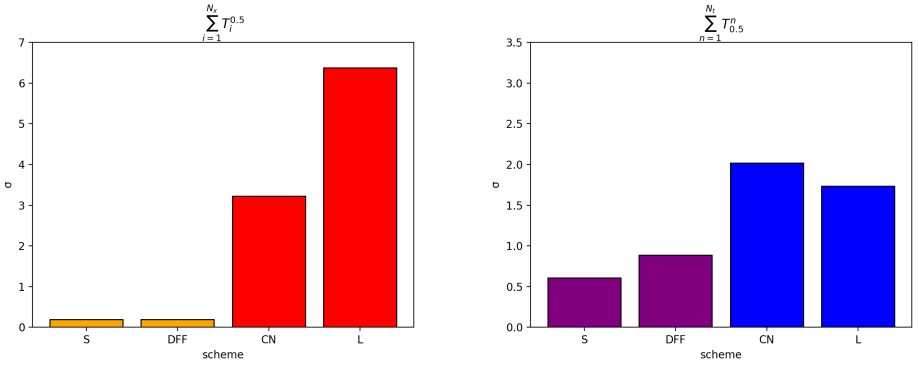


FIGURE 15. Bar plots with the RMSE errors. The left panel corresponds to all positions 0.5 units after starting; while the right panel corresponds to the central position of 0.5 units all the time.

To investigate the matter further, in this paper we compute l^2-norm data, Eq. 2.4, based on the true error using the semi-analytical solution obtained in A, for two resolutions whose ratio is 2 to indirectly infer the order of convergence. The reason behind this is that the usage of explicit schemes in difference

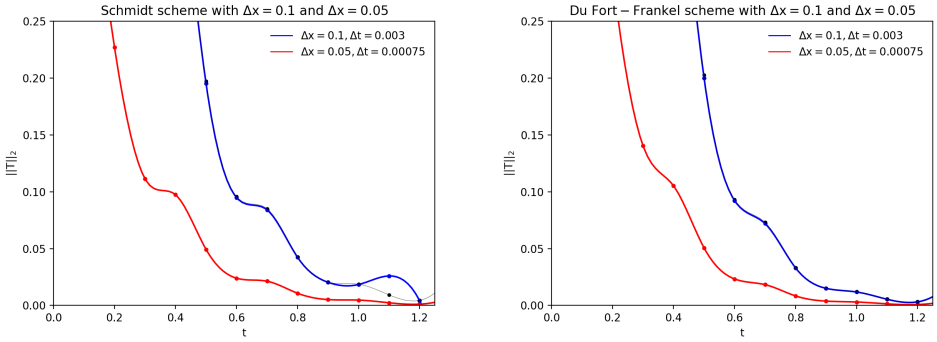
diffusion equations is too time-consuming to compute the rate of convergence. That is true also even for the usage of implicit schemes in diffusion problems (e.g., see Fig. 12). Thus, in exactly this way Fig. 16 shows that (when the spatial resolution increases by a factor of 2) then l^2-norm error decrease by a factor of 4. Our analysis thereby confirm, using all schemes, that they all actually are quadratic, something which is not before theoretically proven herein.

Note this finding proves the Lax-Richtmyer theorem [23] for diffusion-like equations, despite it only concerns well-posed problems; as a matter of fact, the governing differential equation and not the difference schemes is its cause (see second paper in the series).

Finally, although we have focused our study to understand and test both accuracy of each one of the five low order finite difference schemes and the fact that they always reach desired convergence rate as well as their stabilities, the convergence studies carried out and showed in Fig. 16 (which agrees with Fig. 15) also lead to better ability to discriminate between schemes. Besides we find performance achieved with the Schmidt's scheme to be comparable to or better than performance with rest schemes in our sample, the Schmidt's scheme has the advantage of being more tractable numerically. In conclusion, the Schmidt's nonoscillatory scheme of finite difference method provides the most successful description of diffusion equation of the A, and thus this result should be taken into consideration during the integration of the full heat-like equation subjected to some instabilities (which we shall discuss, e.g., in third paper in the series). Nonetheless, before doing this work, in the second paper in the series we have to explore under which circumstances such effectiveness holds in the transfer or transport by advection.

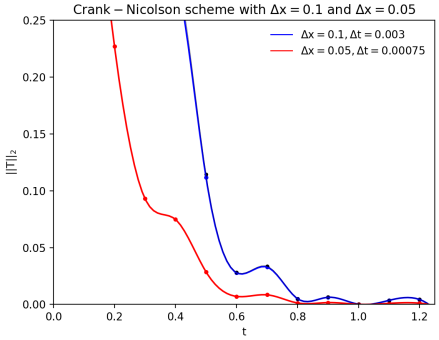
5. **Concluding remarks**

The overarching goal of this first expository work in our series paper is to enable the reader to completely grasp the finite difference approach to the diffusive transport while making the advanced development on the topic and beyond accesible. To achieve this goal, we select and discuss a complete and unbiased sample of schemes for the diffusion parabolic partial differential equation constituted by all those up to second-order in both time and space; which are of topical current interest. We also put to precisely test such sample considering the Fourier problem in the presence of initial discontinuous perturbations; which appear in a wide variety of physical contexts. In other words, different schemes include both single-step and three-time-level, both explicit or hierarchical and non-explicit or monolithic, both consistent and inconsistent, both stable and unstable, both first-order accuracy and second-order accuracy, and both convergent and divergent; while each scheme includes discretizations which obviously are both reliable or robust and inaccurate or uncertain. In

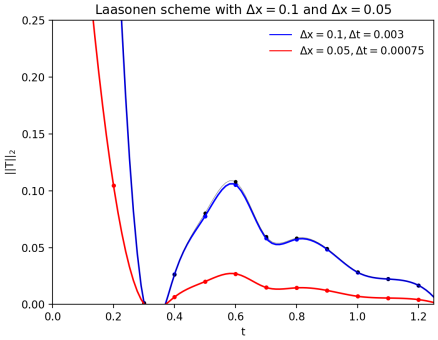


(A) The conditionally stable, Schmidt method

(B) The conditionally consistent Du Fort-Frankel method



(C) The unconditionally stable Crank-Nicolson method



(D) The unconditionally stable Laasonen method

FIGURE 16. Convergence of the error plot for the heating from its ends. The black line is the 2-norm data as a function of time using the resolution $\Delta x = 0.05$ multiplied by 4.

fact, we perform a complete study of each behavior and its implication on the error estimation of each scheme. Moreover, we empirically validate as well as illustrate all of this phenomenology by computation.

To be more precise, the classical Richardson's [15], Schmidt's [16], Crank-Nicolson [19], Laasonen's [21] and Du Fort-Frankel [22] schemes for a diffusion type transport equation have been coherently revised at low and (which is a vital aspect to capture the correct behavior of the system in the regions where the space scale is very small) high wavenumbers by exploiting the power of both the reverse Taylor's analysis [7] and the discrete Fourier's analysis [8], even though these associated formalisms are not physically equivalent. In fact, this distinction is important because we demonstrate these high mode

wavenumbers (smallest scales) always remain poorly constrained. At the same time, the Courant-Friedrichs-Lewy-type condition have been interpreted as the transition to stiffness. For more clarity, we homogeneously test all the schemes on an equal spatiotemporal discretization by performing an a priori minimization best-fit. In fact, we highlight the benefits of this framework. In addition, we use all of them to probe the Lax-Richtmyer theorem [23] and also the Godunov's theorem [9], despite we find the latter has in practice no significant effect on diffusive transport. Our main conclusions are summarized below.

(1) The pioneering Richardson's numerical scheme for approximating the solution to diffusion linear equation is a one step three time level scheme which unfortunately presents a unstable computational mode, and therefore is invalid for diffusive transfer in general. We even provide numerical simulations of a test case based on this scheme.
(2) The simplest Schmidt's numerical scheme for diffusion differential equation can be too prohibitive because its tight stability region and we highlight its critically damped solution, at approximately $F = 1/6$, dividing overdamped from underdamped or stiff regimes in a linear stability analysis; with the limit of the latter explaining the Courant-Friedrichs-Lewy-type condition at $F = 1/2$. Such processes, generally speaking, can be only observed in explicit solutions of diffusion-like equations. This scheme also is the preferred approximation when apply it to a discontinuous initial-value problem for the linear diffusion equation.
(3) The innovative Crank-Nicolson numerical scheme for the diffusion equation presents only one underdamped regime, explaining its stability. Nevertheless, we show that Crank-Nicolson solution exhibits a prominent limitation, namely, it also presents low-wavelength spatial oscillations with increasing F. Such effects, generally speaking, can be only observed in non-explicit solutions of diffusion-like equations. Besides, because starting from the discontinuous initial-boundary conditions the monolithic approximation based on this scheme evolves slowly, the discontinuous initial-value problem investigated disfavours this scheme, whereas the hierarchical approximations are very consistent with the exact solution.
(4) The curious Laasonen's numerical scheme also presents only one regime, which is excessively underdamped for the differential equation analyzed. Indeed, as a non-explicit scheme, it has no underdamped regimen but suffers from low numerical efficiency. Compared to the Crank-Nicolson scheme, its slowly evolving early solution around the leap at endpoints of the domain is still disfavoured.

(5) The fascinating Du Fort-Frankel numerical scheme for diffusion presents a stable computational mode, a critically damped regimen at approximately $F = \sqrt{1/12}$ and a surprising consistency region. In fact, we show that at the limit of its stiff regime, $F = 1/2$, the exceptional, unobservable (imaginary) computational mode recovers the stability but leads to incorrect solutions. Compared to the Schmidt's scheme we hence show that its wider stability region is not advantageous because its consistency region, though its performance is still a reliable approximation for the numerical resolution and for the test problem this paper investigates.

This has however been a challenging project which contains numerous different scenarios. Indeed, the focus of the present historical review is in unfamiliar variety of physical effects and numerical artifacts only little explored in the original literature rather than not necessarily desirable high accuracy or data systematics. Overall, our findings, we believe, should help researchers entering the field, while contributing to the ongoing efforts to refine our quantitative interpretation of the second order spatial derivative. We expect the fourth paper of this series to have a non-linear, stronger mathematical formulation and finally touch briefly whether the scheme destroys the inherent physical structure of the underlying problem or not (referred to as the conservation properties). In second and third papers of the series we shall continue working the theory of difference schemes with the physical help of the Fourier decomposition and the equivalent mathematical clarity of the theory of differential equations.

Competing interests statement.

We have no competing interests.

Appendix A. The Fourier's 1807 semi-analytical solution to the Fourier's 1807 problem as a test case

In this paper we choose as an academic test-case study for diffusion the classical one dimensional initial boundary value problem on an interval in $\mathbb{R}$, first formulated by Fourier [24] in 1807 in an unpublished monograph (influenced by Laplace). Specifically, an insulated rod of length given is suddenly heated from its ends, which is modeled as symmetric Dirichlet boundary conditions, i.e.,

$$\begin{cases} \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} & x \in (0,\, L),\, t > 0, \\ T(x,\, 0) = T_0 & x \in (0,\, L), \\ T(0,\, t) = T_c = T(L,\, t) & t \geq 0. \end{cases} \tag{A.1}$$

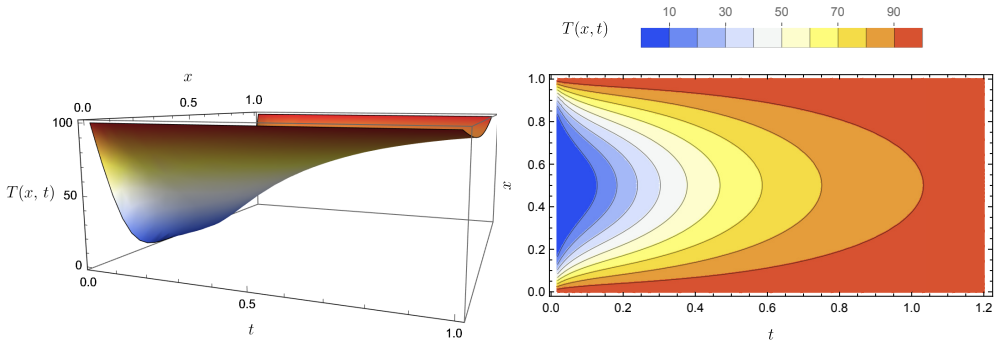
That is, the boundary temperature jumps from 0 to T_0 at $t = 0$. All the other surfaces of the rod are insulated such that a one dimensional model is

appropriate. Heat cannot escape, and since we supply heat at $x = 0$, all of the material will eventually be warmed up to the temperature T_0, i.e. this heat spreads out spatially as time increases. In fact, the Fourier heat equation not only governs heat flow, but all sorts of diffusion processes where some quantity flows from regions of higher to lower concentration (e.g., mass diffusion). What is more, considerations of boundary conditions play a key role in the empirical equivalence (or otherwise) of numerical solutions.

This benchmark problem, prototype of parabolic differential equation, guarantees the suitability of the formulation of the semi-analytical solution in trigonometric series [24]. In particular, by using the method of separation of variables, looking for the solution in the full Fourier series form and applying the boundary conditions, then we obtain:

$$T(x,t) = T_c + 2(T_0 - T_c)\sum_{n=1}^{\infty} \frac{1-(-1)^n}{n\pi} \sin\left(\frac{n\pi x}{L}\right) e^{-\alpha(n\pi/L)^2 t} \tag{A.2}$$

which certainly is a series of Fourier modes that will be dissipated. Additionally, we represent this tree-dimensional semi-analytical function in Figure 17 to graphically visualize the details of the rod's warming which have to be simulated by the numerical solutions.



(A) 3D Plot of $T(x,t)$ (B) Contour plot of $T(x,t)$

FIGURE 17. Representation of the semi-analytical temperature distribution given by Eq. A.2. The pseudo-colours indicate the temperature values. The contour lines are drawn at specific intervals of 10 units.

In second and third papers of the series another favorable initial perturbations will also be addressed.

Acknowledgments

We thank Tapan K. Sengupta for his comments and insightful conversations.

References

[1] R.J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations, Steady State and Time Dependent Problems*, SIAM (2007).

[2] L. Peter Røed, *Atmospheres and Oceans on Computers*, Springer (2019).

[3] T.K. Sengupta, *High Accuracy Computing Methods: Fluid Flows and Wave Phenomena*, Cambridge University Press (2013).

[4] N.D. Katopodes, *Free-Surface Flow Computational Methods*, Elsevier (2019).

[5] D.J. Duffy, *Finite Difference Methods in Financial Engineering*, Wiley (2006).

[6] P. Sagaut, V.K. Suman, P. Sundaram, M.K. Rajpoot, Y.G. Bhumkar, S. Sengupta, A. Sengupta and T.K. Sengupta, Global spectral analysis: Review of numerical methods, *Computers and Fluids*, **261** (2023), 105915; doi: 10.1016/j.compfluid.2023.105915

[7] C.W. Hirt, Heuristic stability theory for finite-difference equations, *Journal of Computational Physics*, **2**, No 4 (1968), 339–355; doi: 10.1016/0021-9991(68)90041-7

[8] J. von Neumann and R.D. Richtmyer, On the Numerical Solution of Partial Differential Equations of Parabolic Type, *Los Alamos Technical Reports*, **LA–657** (1947).

[9] S.K. Godunov, Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics, *Matematičeskij Sbornik*, **47**, No 89 (1959), 271–306.

[10] R.F. Warming and B.J. Hyett, The Modified Equation Approach to the Stability and Accuracy Analysis of Finite-Difference Methods, *Journal of Computational Physics*, **14**, No 2 (1974), 159–179; doi: 10.1016/0021-9991(74)90011-4

[11] Y.-K. Kwok, Stability analysis of six-point finite difference schemes for the constant coefficient convective-diffusion equation, *Computers Math. Applic.*, **23**, No 12 (1992), 3–11; doi: 10.1016/0898-1221(92)90088-Y

[12] I. Winnicki, J. Jasinski and S. Pietrek, New approach to the Lax-Wendroff modified differential equation for linear and nonlinear advection, *Numer. Methods Partial Differential Eq.*, **35** (2019), 2275–2304; doi: 10.1002/num.22412

[13] M. Karam, J.C. Sutherland and T. Saad, PyModPDE: A python software for modified equation analysis, *SoftwareX*, **12** (2020), 100541; doi: 10.1016/j.softx.2020.100541

[14] T. Bodnár, P. Fraunié and K. Kozel, Modified equation for a class of explicit and implicit schemes solving one-dimensional advection problem, *Acta Polytechnica*, **61** (2021), 49–58; doi: 10.14311/AP.2021.61.0049

[15] L.F. Richardson, The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam, *Philosophical Transactions of the Royal Society of London, Series A*, **210** (1910), 307–357; doi: 10.1098/rsta.1911.0009

[16] E. Schmidt, Über die Anwendung der Differenzenrechnung auf technische Anheiz- und Abkühlungsprobleme, *Beiträge zur Technischen Mechanik und Technischen Physik: August Föppl zum Siebzigsten Geburtstag am 25. Januar 1924* (1924), 179–189; doi: 10.1007/978-3-642-51983-3_19

[17] R. Courant, K. Friedrichs and H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik, *Mathematische Annalen*, **100** (1928), 32–74; doi: 10.1147/rd.112.0215

[18] C.F. Curtiss and J.O. Hirschfelder, Integration of Stiff Equations, *Proc. Natl. Acad. Sci.*, **38** (1952), 235–243; doi: 10.1073/pnas.38.3.235

[19] J. Crank and P. Nicolson, A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, *Mathematical Proceedings of the Cambridge Philosophical Society*, **43** (1947), 50–67; doi: 10.1017/S0305004100023197

[20] L.H. Thomas, *Elliptic Problems in Linear Difference Equations over a Network*, Watson Sci. Comput. Lab. Rept., Columbia University (1949).

[21] P. Laasonen, Über eine Methode zur Lösung der Wärmeleitungsgleichung, *Acta Mathematica*, **81** (1949), 309–317; doi: 10.1007/BF02395025

[22] E.C. Du Fort and S.P. Frankel, Stability conditions in the numerical treatment of parabolic differential equations, *Mathematical Tables and Other Aids to Computation*, **7**, No 43 (1953), 135–152; doi: 10.2307/2002754

[23] P.D. Lax and R.D. Richtmyer, Survey of the stability of linear finite difference equations, *Communications on Pure and Applied Mathematics*, **9**, No 2 (1956), 267–293; doi: 10.1002/cpa.3160090206

[24] J.B.J. Fourier, *Joseph Fourier, 1768–1830*, MIT Press (1972).