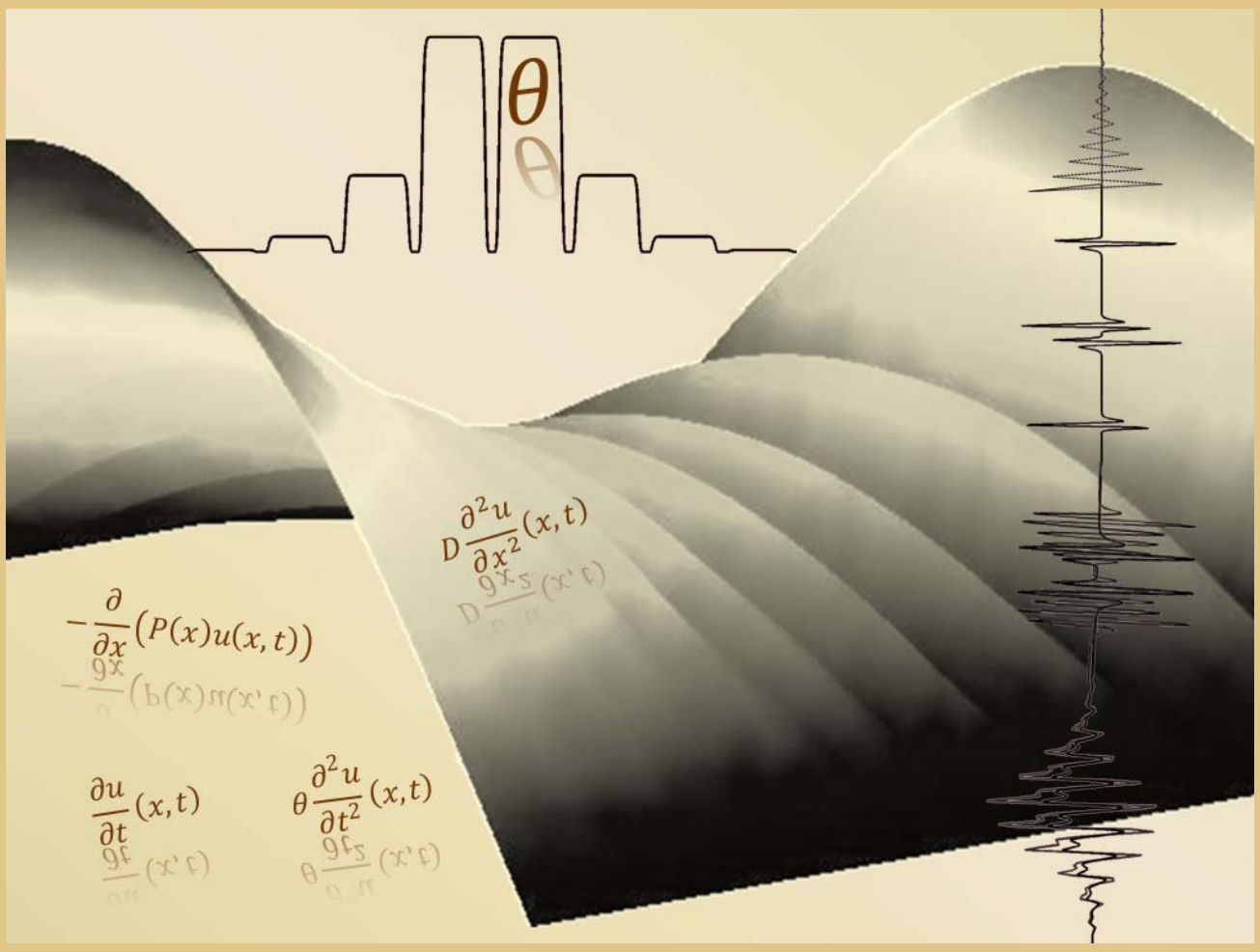

Cidália Alves das Neves

NUMERICAL METHODS FOR HYPERBOLIC DIFFUSIVE PROBLEMS

Tese de Doutoramento na área científica de Matemática, especialidade Matemática Aplicada, orientada pelo Professor Doutor Adérito Luís Martins Araújo e pela Professora Doutora Ercília Cristina Costa Sousa e apresentada ao Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

2013

· U C ·

UNIVERSIDADE DE COIMBRA

Cidália Alves das Neves

NUMERICAL METHODS FOR HYPERBOLIC DIFFUSIVE PROBLEMS

Coimbra

2013

· U C ·

UNIVERSIDADE DE COIMBRA

Tese de doutoramento apresentada ao Departamento de Matemática, Universidade de Coimbra, para obtenção do grau de Doutor em Matemática, especialidade de Matemática Aplicada, sob a orientação do Professor Doutor Adérito Luís Martins Araújo e da Professora Doutora Ercília Cristina Costa Sousa.

Doctoral thesis submitted to Department of Mathematics, University of Coimbra, for the degree of Ph.D. in Mathematics, specialty in Applied Mathematics, supervised by Professor Adérito Luís Martins Araújo and Professor Ercília Cristina Costa Sousa.

To my daughter Vitória

Acknowledgments

To my supervisors Professor Adérito Luís Martins Araújo and Professor Ercília Cristina Costa Sousa for their permanent availability, patience and unconditional support. Their enthusiasm with the small progresses obtained were an incentive to the elaboration of this thesis. I am deeply grateful for the dedication, transmission of knowledge and words of encouragement given throughout this research work.

To my family for all the support and care. A special thanks to my mother for her constant presence and my daughter Vitória for her love, kindness and understanding when I was not around.

To my colleagues and friends for their understanding, great fellowship, and especially for listening.

To Lisa for the design of the cover image.

To CMUC - Centre for Mathematics of the University of Coimbra and FCT - Foundation of Science and Technology, through the project research UTAustin/MAT/0066/2008, for financial and logistical support.

To program PROTEC and presidency of ISCAC for financial support and partial reduction of service.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE

QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

Resumo

O uso de equações diferenciais hiperbólicas de segunda ordem, na modelação de problemas difusivos, tem-se revelado útil em muitos ramos da ciência como a física, química, biologia e finanças. A condução de calor, a difusão de massa e a dinâmica dos fluidos são alguns exemplos pertencentes à vasta gama de temas abrangidos por este tipo de equações hiperbólicas. Contudo, a inclusão de um potencial não tem sido estudada de forma exaustiva, apesar da sua grande relevância em aplicações práticas como, por exemplo, a distribuição da concentração de massa em problemas de difusão.

O objetivo principal desta tese consiste em desenvolver e estudar métodos numéricos para problemas hiperbólicos de segunda ordem que tomam em consideração a presença de um potencial. Em particular, pretende-se estudar como a variação do coeficiente de relaxamento temporal e o potencial afetam o comportamento da solução, nomeadamente quando se consideram tempos longos. Para isso, iremos começar por analisar diferentes métodos numéricos para o caso unidimensional, tais como um método de diferenças finitas do tipo Crank-Nicolson e um algoritmo baseado na transformada de Laplace combinado com diferentes estratégias de discretização espacial. Todos os algoritmos considerados são estudados quanto à sua consistência e estabilidade e são apresentados exemplos numéricos que, para além de ilustrar os resultados teóricos obtidos, comparam a sua eficiência. Entre as simulações numéricas realizadas, é de destacar uma aplicação interessante que modela a dinâmica de uma partícula Browniana na presença de um potencial periódico simétrico.

A investigação realizada revela as vantagens e desvantagens inerentes às

diferentes formulações. Em particular, quando há interesse no comportamento da solução para tempos longos, os métodos baseados na transformada de Laplace, a qual é combinada com métodos de diferenças finitas ou a formulação de volumes finitos, mostraram ser mais eficazes que o método de Crank-Nicolson. Contudo, dependendo da discretização espacial usada em problemas que contenham condições iniciais descontínuas, por vezes surgem oscilações numéricas em alguns testes numéricos. Para suprimir essa lacuna, procurou-se uma abordagem alternativa e usou-se a técnica de linearização seccionada na discretização espacial que, combinada com a transformada de Laplace, se revelou o método mais eficaz para os problemas considerados no caso unidimensional.

Os métodos usados no caso unidimensional são generalizados para o caso bidimensional. Contudo, o método que se revelou mais eficaz, pela sua natureza particular, não pode ser considerado. Como forma de melhorar a eficiência computacional do algoritmo de Crank-Nicolson desenvolvemos um método implícito de direção alternada. Esta abordagem, apesar de clássica para métodos de diferenças finitas, é uma inovação no contexto dos problemas hiperbólicos com derivadas parciais de primeira e segunda ordem, tanto no espaço como no tempo. Os resultados teóricos desenvolvidos na análise deste método constituem um importante contributo desta tese e mostram o grande potencial do algoritmo no tratamento numérico do problema que nos propusemos estudar.

Abstract

The use of second order hyperbolic differential equations in modeling diffusive problems, has shown to be useful in many branches of science such as physics, chemistry, biology and finance. The heat conduction, the mass diffusion and the fluid dynamics are some of the examples belonging to a wide range of subjects covered by these hyperbolic equations. However, the incorporation of a potential field has not been studied exhaustively, despite its great relevance in practical applications such as, for instance, the mass concentration distribution of diffusion problems.

The main purpose of this thesis is to develop and study numerical methods for second order hyperbolic problems that take into account the presence of a potential field. In particular, we intend to study how the coefficient of variation of the relaxation time and the potential affect the solution behavior, namely when long times are considered. To accomplish this, we will start to analyze different numerical methods for the one dimensional case such as, a finite difference method of Crank-Nicolson type and an algorithm based on Laplace transform combined with distinct spatial discretization strategies. Consistency and stability are studied for all the considered algorithms and numerical examples are presented to illustrate the theoretical results and, in addition, to compare their efficiency. Among the numerical simulations performed, we highlight an interesting application that models the dynamics of a Brownian particle in the presence of a symmetric periodic potential.

The research carried out reveals the advantages and disadvantages inherent in the different formulations. In particular, when there is an

interest in the behavior of solution for long times, the methods based on the Laplace transform, which is combined with finite difference methods or finite volume formulations, show to be more effective than the Crank-Nicolson method. However, depending on the spatial discretization used in problems that contain discontinuous initial conditions, sometimes numerical oscillations arise in numerical tests. To suppress this shortcoming, we seek an alternative approach and used the piecewise linearized technique in the spatial discretization which, combined with the Laplace transform, turns out to be the most effective method for the problems considered in one dimensional case.

The numerical methods applied in the one dimensional case are generalized in two dimensions. However, the most effective method, due to its specific nature, can not be considered. In order to improve the computational efficiency of the Crank-Nicolson algorithm, we develop an alternating direction implicit method. This approach, although classical for the finite difference methods, is an innovation in the context of hyperbolic problems, with partial derivatives of first and second order, in both space and time. The theoretical results developed in the analysis of the method are an important contribution of this thesis and show the great potential of the algorithm in the numerical treatment of the problem we proposed to study.

Contents

Preface

The research developed in this work has the purpose to bring some new features and contributions in the study of diffusion problems. It is organized in five chapters and we describe below a short summary of each chapter.

Despite the large number of numerical methods implemented to solve diffusion problems, only a few partial differential equations of hyperbolic type incorporate a potential field. In Chapter 1 we derive our model problem, which includes a potential field. Some fundamental concepts are defined to support the theoretical analysis performed in the remaining chapters. Also, a literature review is included, where the numerical approaches developed and most commonly used in recent years for similar problems are described.

Chapters 2 and 3 are concerned with the one dimensional problem. In Chapter 2 we present numerical solutions for the one dimensional second order hyperbolic equation. We apply the Crank-Nicolson method based on first order discretization in time and second order discretizations in space. The need to obtain results for very long times led us to seek other numerical methods presented in Chapter 3. They consist of first applying the Laplace transform to remove the time dependent terms in the governing equation. Three distinct schemes are considered for the spatial discretization: a finite difference scheme, a finite volume formulation and a piecewise linearized method. At last, the approximate solution is obtained by a numerical inverse Laplace transform. The inverse Laplace transform algorithm used in this work is based on a continued fraction approach described in the first section of Chapter 3. In order to compare the computational efficiency, performance and convergence of these three schemes, some numerical results

are obtained which provide interesting conclusions about all the methods. They are also compared with the Crank-Nicolson method described in the previous chapter. In addition to the numerical schemes implementation, a theoretical analysis of such schemes is carried out. We devote the last section of Chapter 3 to the numerical solution of the governing equation involving a symmetric periodic potential.

The numerical methods applied in previous chapters are extended in Chapter 4 for two spatial dimensions. These methods do not support the computational effort needed to return the final solution, for a large number of discretization points, due to the huge size of the systems obtained. Hence, to overcome the difficulty of solving such systems with direct solutions, an alternating direction implicit method is derived. The results obtained in the end of Chapter 4 show the great efficiency and good performance of the method.

Chapter 5 contains the main conclusions and comments of this work. Also, we point out some possible ways to follow in future research.

Chapter 1

A second order hyperbolic equation

Partial differential equations describing diffusive processes constitute the basis of many models in several fields such as [18, 41, 44, 49, 72, 85, 87, 90, 95]. In this chapter we essentially model the problem that underpins the work of the remaining chapters. In Section 1.1 we derive a second order hyperbolic equation that describes a diffusion process in the presence of a potential field. The variation of the parameters involved in the equation of our study makes it more embracing, since it allows the application to a wider range of diffusion problems. Among all the problems considered, we focus our attention on the example studied in [4] which includes a periodic potential field. To perform the theoretical analysis of the numerical methods we define some fundamental concepts in Section 1.2.

Section 1.3 is a literature review where we present several studies for similar equations, mainly in one and two dimensions. Many authors have been proposed distinct numerical solutions for the resolution of such models and we emphasize the particular techniques developed in each method.

In the end of the chapter, Section 1.4, we include the original contribution of this thesis in the development of numerical methods to solve hyperbolic diffusion problems.

1.1 Model problem

The hyperbolic equation of diffusive nature that we are going to study, can be derived from the Kramers equation [23, 24, 34, 80], which describes the Brownian motion in a potential, given by

$$\frac{\partial f}{\partial t} + \frac{p}{m}\frac{\partial f}{\partial x} - \frac{dV}{dx}\frac{\partial f}{\partial p} = \gamma\frac{\partial}{\partial p}(pf) + mk_BT\gamma\frac{\partial^2 f}{\partial p^2}, \tag{1.1}$$

where x is the space variable, t is the time, $V(x)$ is the potential field, γ is a friction parameter, m is the mass, $f(x,p,t)$ is the probability density function for the position component x and momentum component p of a Brownian particle, k_B is the Boltzmann's constant and T is the temperature of the fluid. We also assume the boundary conditions are

$$\lim_{p\to\pm\infty} f(x,p,t) = \lim_{p\to\pm\infty}\frac{\partial f}{\partial p}(x,p,t) = 0.$$

We start to take moments of p over equation (1.1). This procedure consists of multiplying the equation by various powers of p and then integrating over p. Taking the zeroth moment of momentum p we get

$$\int_{-\infty}^{+\infty}\frac{\partial f}{\partial t}\,dp + \int_{-\infty}^{+\infty}\frac{p}{m}\frac{\partial f}{\partial x}\,dp - \int_{-\infty}^{+\infty}\frac{dV}{dx}\frac{\partial f}{\partial p}\,dp$$

$$= \int_{-\infty}^{+\infty}\gamma\frac{\partial}{\partial p}(pf)\,dp + \int_{-\infty}^{+\infty}mk_BT\gamma\frac{\partial^2 f}{\partial p^2}\,dp.$$

The number density of particles at position x is given by

$$\int_{-\infty}^{+\infty} f(x,p,t)\,dp = u\,(x,t)\,.$$

Therefore, we can write the first term as

$$\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} f\,dp = \frac{\partial u}{\partial t}\,(x,t)\,.$$

Defining the current density by

$$j\,(x,t) = \int_{-\infty}^{+\infty}\frac{p}{m}f(x,p,t)\,dp$$

the second term becomes

$$\frac{\partial}{\partial x}\int_{-\infty}^{+\infty}\frac{p}{m}f\,dp = \frac{\partial j}{\partial x}\,(x,t)\,.$$

Also, for the third term we have

$$\int_{-\infty}^{+\infty} \frac{dV}{dx}\frac{\partial f}{\partial p}\,dp = \frac{dV}{dx}\int_{-\infty}^{+\infty} \frac{\partial f}{\partial p}\,dp = \frac{dV}{dx}\,[f]_{-\infty}^{+\infty} = 0.$$

The first term of the second member is

$$\int_{-\infty}^{+\infty} \gamma\frac{\partial}{\partial p}(pf)\,dp = \gamma\left(\int_{-\infty}^{+\infty} f\,dp + \int_{-\infty}^{+\infty} p\frac{\partial f}{\partial p}\,dp\right).$$

Integrating by parts we obtain

$$\gamma\left(\int_{-\infty}^{+\infty} f\,dp + [fp]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f\,dp\right) = 0.$$

The last term of the equation is

$$mk_BT\gamma\int_{-\infty}^{+\infty} \frac{\partial}{\partial p}\left(\frac{\partial f}{\partial p}\right)\,dp = mk_BT\gamma\left[\frac{\partial f}{\partial p}\right]_{-\infty}^{+\infty} = 0$$

and, therefore, the zeroth moment equation becomes

$$\frac{\partial u}{\partial t}\,(x,t) + \frac{\partial j}{\partial x}\,(x,t) = 0. \tag{1.2}$$

Taking the first moment of equation (1.1) we can write

$$\int_{-\infty}^{+\infty} p\frac{\partial f}{\partial t}\,dp + \int_{-\infty}^{+\infty} \frac{p^2}{m}\frac{\partial f}{\partial x}\,dp - \int_{-\infty}^{+\infty} \frac{dV}{dx}p\frac{\partial f}{\partial p}\,dp$$

$$= \int_{-\infty}^{+\infty} \gamma p\frac{\partial}{\partial p}(pf)\,dp + \int_{-\infty}^{+\infty} mk_BT\gamma p\frac{\partial^2 f}{\partial p^2}\,dp. \tag{1.3}$$

The first term leads to

$$\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} pf\,dp = m\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} \frac{p}{m}f\,dp = m\frac{\partial j}{\partial t}\,(x,t)\,.$$

The second term is

$$\frac{\partial}{\partial x}\int_{-\infty}^{+\infty} \frac{p^2}{m}f\,dp.$$

In order to obtain an equation for $u\,(x,t)$, Das introduced in [23] the approximation

$$f\,(x,p,t) \simeq u\,(x,t)\,\phi(p),$$

where $\phi(p)$ is the Maxwellian distribution function for one component of momentum for a particle [11], that is,

$$\phi(p) = \sqrt{\frac{1}{2\pi mK_BT}}e^{-\frac{p^2}{2mK_BT}}.$$

On the other hand,

$$\int_{-\infty}^{+\infty} \frac{p^2}{m}\phi(p)\,dp = \int_{-\infty}^{+\infty} \frac{p^2}{m}\sqrt{\frac{1}{2\pi m K_B T}}e^{-\frac{p^2}{2mK_BT}}\,dp$$

$$= -\frac{1}{m}\sqrt{\frac{1}{2\pi m K_B T}}mK_BT\int_{-\infty}^{+\infty} pe^{-\frac{p^2}{2mK_BT}}\left(-\frac{p}{mK_BT}\right)\,dp.$$

Integrating by parts results in

$$-\sqrt{\frac{1}{2\pi m K_B T}}K_BT\left(\left[pe^{-\frac{p^2}{2mK_BT}}\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{p^2}{2mK_BT}}\,dp\right).$$

Since

$$\lim_{p\to\pm\infty} pe^{-\frac{p^2}{2mK_BT}} = 0,$$

we get

$$\sqrt{\frac{1}{2\pi m K_B T}}K_BT\int_{-\infty}^{+\infty} e^{-\frac{p^2}{2mK_BT}}\,dp.$$

With the change of variable

$$\varphi = \sqrt{\frac{1}{2mK_BT}}p$$

and using the standard integral

$$\int_{-\infty}^{+\infty} e^{-\varphi^2}\,d\varphi = \sqrt{\pi},$$

then

$$\sqrt{\frac{1}{2\pi m K_B T}}K_BT\sqrt{2mK_BT}\int_{-\infty}^{+\infty} e^{-\varphi^2}\,d\varphi = K_BT.$$

Therefore, the second term of equation (1.3) is

$$\frac{\partial}{\partial x}\int_{-\infty}^{+\infty}\frac{p^2}{m}f\,dp = \frac{\partial u}{\partial x}\int_{-\infty}^{+\infty}\frac{p^2}{m}\phi(p)\,dp = K_BT\frac{\partial u}{\partial x}(x,t).$$

Integrating by parts the third term gives

$$\frac{dV}{dx}\int_{-\infty}^{+\infty} p\frac{\partial f}{\partial p}\,dp = \frac{dV}{dx}\left([fp]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f\,dp\right) = -\frac{dV}{dx}u(x,t).$$

The first term of the second member of equation (1.3) is

$$\gamma\int_{-\infty}^{+\infty} p\left(f + p\frac{\partial f}{\partial p}\right)\,dp = \gamma\int_{-\infty}^{+\infty} pf\,dp + \gamma\int_{-\infty}^{+\infty} p^2\frac{\partial f}{\partial p}\,dp.$$

After integrating by parts, this is the same as

$$\gamma \int_{-\infty}^{+\infty} pf\, dp + \gamma \left[fp^2\right]_{-\infty}^{+\infty} - \gamma \int_{-\infty}^{+\infty} 2pf\, dp$$

$$= -\gamma \int_{-\infty}^{+\infty} pf\, dp = -m\gamma \int_{-\infty}^{+\infty} \frac{p}{m} f\, dp = -m\gamma j\,(x,t)\,.$$

The last term of equation (1.3) is

$$mk_BT\gamma \int_{-\infty}^{+\infty} p\frac{\partial^2 f}{\partial p^2}\, dp \;=\; mk_BT\gamma \left(\left[p\frac{\partial f}{\partial p} \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{\partial f}{\partial p}\, dp \right)$$

$$= \; -mk_BT\gamma \left[f\right]_{-\infty}^{+\infty} = 0.$$

Finally, the first moment equation becomes

$$m\frac{\partial j}{\partial t}\,(x,t) + K_BT\frac{\partial u}{\partial x}\,(x,t) + \frac{dV}{dx}u\,(x,t) = -m\gamma j\,(x,t)\,. \tag{1.4}$$

Defining the diffusion coefficient $D = K_BT/m\gamma$, (1.4) turns into the following equation

$$j\,(x,t) = -\left(D\frac{\partial u}{\partial x}\,(x,t) + \frac{1}{m\gamma}\frac{dV}{dx}u\,(x,t) \right) - \frac{1}{\gamma}\frac{\partial j}{\partial t}\,(x,t)\,. \tag{1.5}$$

The derivation of the zeroth moment equation (1.2) with respect to t gives

$$\frac{\partial^2 u}{\partial t^2}\,(x,t) + \frac{\partial^2 j}{\partial t \partial x}\,(x,t) = 0$$

which implies that

$$\frac{\partial^2 j}{\partial t \partial x}\,(x,t) = -\frac{\partial^2 u}{\partial t^2}\,(x,t)\,. \tag{1.6}$$

The derivation of the first moment equation (1.5) with respect to x leads us to

$$\frac{\partial j}{\partial x}\,(x,t) = -D\frac{\partial^2 u}{\partial x^2}\,(x,t) - \frac{1}{m\gamma}\frac{\partial}{\partial x}\left[\frac{dV}{dx}(x)u\,(x,t)\right] - \frac{1}{\gamma}\frac{\partial^2 j}{\partial x \partial t}\,(x,t)\,. \tag{1.7}$$

From (1.2), (1.6) and (1.7) we have

$$-\frac{\partial u}{\partial t}\,(x,t) = -D\frac{\partial^2 u}{\partial x^2}\,(x,t) - \frac{1}{m\gamma}\frac{\partial}{\partial x}\left[\frac{dV}{dx}(x)u\,(x,t)\right] + \frac{1}{\gamma}\frac{\partial^2 u}{\partial t^2}\,(x,t)$$

which permit us to obtain equation

$$\frac{1}{\gamma}\frac{\partial^2 u}{\partial t^2}\,(x,t) + \frac{\partial u}{\partial t}\,(x,t) = D\frac{\partial^2 u}{\partial x^2}\,(x,t) + \frac{1}{m\gamma}\frac{\partial}{\partial x}\left[\frac{dV}{dx}(x)u\,(x,t)\right]. \tag{1.8}$$

Quite recently, the instantaneous velocity of a Brownian particle has been experimentally investigated [47, 54, 77], providing an additional motivation for studying (1.8). There is also another paper [10] which models transport of ions in insulating media through a hyperbolic diffusion equation of the type (1.8). We can also write equation (1.8) as

$$\theta \frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(P(x)u(x,t)\right) + D\frac{\partial^2 u}{\partial x^2}(x,t), \tag{1.9}$$

where u is the mass concentration, $\theta = 1/\gamma \in]0,1]$ is the parameter that measures the propagation speed of the mass wave and can be regarded as the relaxation time of the mass flux; function P is defined as

$$P = -\frac{1}{m\gamma}\frac{dV}{dx}(x),$$

where $V(x)$ is the potential field. Note that for $\theta = 0$, equation (1.9) is the classical parabolic convection diffusion equation.

The extension of equation (1.9) in two dimensions is the following hyperbolic equation, which includes also diffusion and a potential field $V(x,y)$, defined in a rectangular domain $\Omega \subset \mathbb{R}^2$,

$$\begin{aligned}\theta \frac{\partial^2 u}{\partial t^2}(x,y,t) + \frac{\partial u}{\partial t}(x,y,t) = &- \frac{\partial}{\partial x}(P(x,y)u(x,y,t)) - \frac{\partial}{\partial y}(Q(x,y)u(x,y,t)) \\ &+ D\frac{\partial^2 u}{\partial x^2}(x,y,t) + D\frac{\partial^2 u}{\partial y^2}(x,y,t),\end{aligned} \tag{1.10}$$

with $(x,y) \in \Omega$, $t > 0$ and

$$(P,Q) = -\frac{1}{m\gamma}\left(\frac{\partial V}{\partial x}(x,y), \frac{\partial V}{\partial y}(x,y)\right).$$

1.2 Fundamental concepts

In this section, we present some definitions that will be used in the remaining chapters of this thesis. We define consistency, stability and recall the Lax equivalence theorem. The consistency role is to measure how well a difference equation approximates the partial differential equation. On the other hand, the main idea in analyzing the stability of a difference scheme

is to bound the growth of errors caused by perturbations of the input data or rounding errors introduced during the computation.

Following [81], let us consider a linear (initial-)boundary value problem

$$\mathcal{L}u = f \tag{1.11}$$

defined on some domain D with boundary ∂D, where $\mathcal{L}$ is a linear operator $\mathcal{L} : \mathcal{U} \longrightarrow \mathcal{F}$ that has bounded inverse, $\mathcal{L}^{-1} : \mathcal{F} \longrightarrow \mathcal{U}$, with $\mathcal{U}$ and $\mathcal{F}$ appropriate spaces, such as Banach spaces. In other words, we will assume that (1.11) is uniquely solvable for every $f \in \mathcal{F}$ and well-posed. In order to approximately compute the solution u of (1.11) given the data f, we need to specify a set of points $D_\Delta \subset D \cup \partial D$ that is called the grid (or mesh). Note that we are assuming that the given data can be initial and/or boundary conditions. Let us define a linear normed space $\mathcal{U}_\Delta$ of all discrete functions defined on the grid D_Δ and let u_Δ be the restriction of the continuous solution u on the grid.

Since neither the continuous exact solution not its restriction on the grid are known, we need to consider a numerical method to compute u_Δ approximately. For that purpose, let us consider a finite difference scheme obtained by replacing the continuous derivatives in $\mathcal{L}$ by appropriate differences, given by the system of equations

$$\mathcal{L}_\Delta U_\Delta = f_\Delta \tag{1.12}$$

with respect to the unknown function $U_\Delta \in \mathcal{U}_\Delta$. This method should be such that the approximate solution U_Δ converges to the exact solution u_Δ as the grid is refined, according to the following definition.

Definition 1.2.1. *Let u_Δ be the restriction of the problem (1.11) on the grid D_Δ and U_Δ the approximate solution obtained by (1.12). The approximate solution is convergent to the exact solution if and only if*

$$\|u_\Delta - U_\Delta\|_{\mathcal{U}_\Delta} \longrightarrow 0, \quad \textit{as } \Delta \longrightarrow 0, \tag{1.13}$$

where $\|\cdot\|_{\mathcal{U}_\Delta}$ represents a norm on $\mathcal{U}_\Delta$. If $p > 0$ is the largest integer such that

$$\|u_\Delta - U_\Delta\|_{\mathcal{U}_\Delta} \leq c\Delta^p,$$

with c a constant independent of the grid parameter Δ, we say that the convergence rate is $\mathcal{O}(\Delta^p)$ or that the global error $u_\Delta - U_\Delta$ has order p with respect to the grid parameter Δ in the chosen norm $\|\cdot\|_{\mathcal{U}_\Delta}$.

Remark 1.2.1. *Note that the way we define the convergence for finite-difference schemes differs from the traditional definition in vector spaces. When $\Delta \longrightarrow 0$ the number of nodes in the grid D_Δ will increase, and so the dimension of the space $\mathcal{U}_\Delta$. To overcome this drawback, $\mathcal{U}_\Delta$ shall be interpreted as a sequence of spaces on increasing dimension parameterized by Δ and the limit (1.13) as a limit of the sequence of norms in vector spaces that have increasing dimensions.*

The construction of a convergent scheme (1.12) is usually done in two steps: first we obtain a scheme that is consistent with the problem; then we must verify that the chosen scheme is stable. We will now define these two concepts.

Let us start with consistency. To define this concept rigorously, we should start by introducing a norm in the linear space $\mathcal{F}_\Delta$ that contains the right hand side f_Δ of (1.12). Similarly to $\mathcal{U}_\Delta$, $\mathcal{F}_\Delta$ should be interpreted as a sequence of spaces of increasing dimension parameterized by Δ.

Definition 1.2.2. *The difference scheme (1.12) is said to be consistent with the problem (1.11) if and only if for any sufficiently smooth $u \in \mathcal{U}$ we have*

$$\|\mathcal{L}u_\Delta - \mathcal{L}_\Delta u_\Delta\|_{\mathcal{F}_\Delta} \longrightarrow 0, \quad \text{as } \Delta \longrightarrow 0, \tag{1.14}$$

where $\|\cdot\|_{\mathcal{F}_\Delta}$ represents a norm on $\mathcal{F}_\Delta$ and

$$\mathcal{T}_\Delta = \mathcal{L}u_\Delta - \mathcal{L}_\Delta u_\Delta$$

is called the local truncation error. If $p > 0$ is the largest integer such that

$$\|\mathcal{T}_\Delta\|_{\mathcal{U}_\Delta} \leq c\Delta^p,$$

with c a constant independent of the grid parameter Δ, we say that the consistency rate is $\mathcal{O}(\Delta^p)$ or that scheme is accurate of order p with respect to the grid parameter Δ in the chosen norm $\|\cdot\|_{\mathcal{F}_\Delta}$ with respect to the given problem.

We now consider the definition of stability.

Definition 1.2.3. *The finite-difference scheme (1.12) is called stable if there is an Δ_0 such that for any $\Delta < \Delta_0$ and $f_\Delta \in \mathcal{F}_\Delta$ it is unique solvable and the solution U_Δ satisfies*

$$\|U_\Delta\|_{\mathcal{U}_\Delta} \leq c\|f_\Delta\|_{\mathcal{F}_\Delta}, \tag{1.15}$$

with c a constant independent of the grid parameter Δ and of f_Δ.

A similar definition of stability is the following: there is an Δ_0 such that for any $\Delta < \Delta_0$, the inverse operators $\mathcal{L}_\Delta^{-1}$ exist and are bounded uniformly, that is,

$$\|\mathcal{L}_\Delta^{-1}\| \leq c,$$

where $\|.\|$ is the norm in the space of linear operators $\mathcal{L}_\Delta^{-1} : \mathcal{F}_\Delta \longrightarrow \mathcal{U}_\Delta$, with c a constant independent of the grid parameter Δ.

According to the previous definition, stability is an intrinsic property of the scheme. The formulation of this property does not involve any direct relation to the original problem.

Remark 1.2.2. *Note that, if we consider in the problem (1.11) homogeneous Dirichlet boundary conditions, the term $\|f_\Delta\|_{\mathcal{F}_\Delta}$ only involves the initial condition.*

The property of stability is formulated independently of either consistency or convergence. The following theorem establishes a fundamental relation between consistency, stability and convergence [81].

Theorem 1.2.1. ***[Lax Equivalence Theorem]*** *A consistent finite difference scheme (1.12) for the well-posed problem in (1.11) is convergent if and only if it is stable. Moreover, the converging rate coincides with the consistency rate of the scheme.*

1.3 Literature review

In this section we present some of the numerical methods developed in recent years to solve linear second order hyperbolic equations, in one and two dimensions, similar to equations (1.9) and (1.10).

Standard finite difference methods are known to be one of the first techniques applied for solving these partial differential equations. They are still used extensively in many practical computations due to their inherent simplicity. Even though these methods are very effective for solving various kinds of partial differential equations, the conditional stability of explicit finite difference procedures and the need for a big computational effort in implicit finite difference schemes to obtain an accurate numerical solution, require further research. Therefore, new difference schemes are constantly being presented, some of them featuring high order accuracy.

The design of higher order accurate finite difference methods for the second order hyperbolic equations is challenging. When we employ higher order standard schemes, which need a large number of mesh points near the boundaries, some of the difficulties arise in treating the approximations of the discrete points, requiring many times the introduction of fictitious points. Therefore, the second order methods that do not require the use of fictitious points are common methods to solve these equations [28, 67, 71]. Another difficulty associated with the numerical methods of hyperbolic problems is the observation of numerical oscillations that appear in the vicinity of sharp discontinuities [14, 15, 82].

In what follows, we present an overview of the different approaches that have been appearing recently.

1.3.1 One dimensional linear hyperbolic equation

We start to review some of the numerical methods presented in the literature to solve problems involving equations similar to equation (1.9). Among the numerical methods implemented to solve second order linear hyperbolic equations, different strategies have been developed for the

solution of the telegraph equation [25, 27, 29, 35, 39, 42, 59, 66, 68, 71, 76].

The telegraph equation does not include a potential field and it has an additional reaction term or source term. This equation can represent a damped wave equation and models many reaction-diffusion problems such as, the electrical voltage, the propagation of electrical signals in a cable of transmission line and wave phenomena, mentioned in [25, 68, 76] . Also, the pulsate blood flow in arteries and the one dimensional random motion of bugs along a hedge [25, 68], the propagation of acoustic waves in Darcy-type porous media and the parallel flows of viscous Maxwell fluids [25, 68, 64] are some of the phenomena governed by the telegraph equation.

Let us now describe some of the approaches known to find a numerical solution for the telegraph equation. To solve an initial boundary value problem involving a damped wave equation that models heat conduction, a finite difference scheme was constructed in [65]. It consists of applying the standard centered difference approximation to the second order derivative in time while the first order time derivative is approximated by a combination of forward Euler and centered difference quotients. For the second order spatial derivative a Dufort-Frankel approximation was considered, that is,

$$\frac{\partial^2 u}{\partial x^2}(x,t) \approx \frac{U_{i-1}^n - (U_i^{n+1} + U_i^{n-1}) + U_{i+1}^n}{\Delta x^2},$$

where the mesh points are given by $x_i = i\Delta x$, $i = 0, \ldots, N$, with $\Delta x = 1/N$, N is a positive integer, $t_n = n\Delta t$ with Δt being the time increment and U_i^n denotes the approximate solution to $u(x_i, t_n)$. The numerical method has accuracy of order $\mathcal{O}(\Delta t + \Delta x^2)$. A von Neumann stability analysis was carried out to conclude this scheme is conditionally stable. The method involves three levels in time and depends on the parameter θ. We note that the authors developed an earlier scheme [64] that only required two levels, t_n and t_{n+1}, because they assumed $\theta \ll \Delta t$ which caused the loss of one time level and the method became independent of the parameter θ.

To solve the telegraph equation we can find in [39] the alternating group explicit method, which is second order accurate in time and space and also conditionally stable. First, a three level implicit formula was constructed by

using centered differences to approximate the first order and second order derivatives in time and, for the second order spatial derivative, centered differences combined with a weighting factor were considered. Secondly, the implicit equations derived from the finite difference discretization were split into explicit equations using an intermediate time level.

Another implicit three level difference scheme of order $\mathcal{O}(\Delta t^2 + \Delta x^2)$ was implemented in [66] and, in this case, we obtain an unconditionally stable difference scheme by a von Neumann stability analysis. To accomplish this, two terms were added to the final discrete equation, which did not affect the second order accuracy of the scheme. Then, the Gauss-elimination method was used to solve the resulting linear system of equations. In [67] the same author developed a similar numerical method for the telegraph equation, with the same properties, but for some variable coefficients in the diffusive and reactive terms.

A quite different method was applied in [42] for the telegraph equation with a source term. A semi-discretization technique was implemented, where the second order derivative in space was approximated by centered differences. The matrix form of the difference scheme obtained involves an exponential function in t. The need to approximate this exponential function led the author to use Padé approximations of order [1,1] and [2,2] to get the numerical solution. The Padé approximation of order [m,n] consists of approximating a given function f by the rational function

$$R(x) = \frac{a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m}{1 + b_1 x + b_2 x^2 + \cdots + b_n x^n} \tag{1.16}$$

that satisfies $f(0) = R(0)$, $f'(0) = R'(0)$, $\ldots$, $f^{(m+n)}(0) = R^{(m+n)}(0)$. Two explicit difference schemes were obtained with different accuracy orders of $\mathcal{O}(\Delta t^3 + \Delta x^2)$ and $\mathcal{O}(\Delta t^5 + \Delta x^2)$, associated with the approximations of order [1,1] and [2,2], respectively. The unconditional stability was concluded in both methods by analyzing the eigenvalues of the matrices.

An alternative numerical method introduced in [29] to solve the one dimensional hyperbolic telegraph equation involves a collocation method. After the discretization of time derivatives with finite differences, the

collocation method was considered by using thin plate splines radial basis functions: assuming there are a total of $N-2$ interpolation points, $u(x,t_n)$ can be approximated by

$$u^n(x) \simeq \sum_{j=1}^{N-2} \lambda_j^n \varphi(r_j) + \lambda_{N-1}^n x + \lambda_N^n, \tag{1.17}$$

where the radial basis function is defined by $\varphi(r_j) = r_j^4 \log(r_j)$ with $x = (x_1, x_2, \ldots, x_N)$ and $r_j = \|x - x_j\|$ the Euclidean norm. To determine the interpolation coefficients $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ the collocation method was used by applying (1.17) at every point x_i, $i = 1, 2, \ldots, N-2$. The method was then completed by joining the additional conditions

$$\sum_{j=1}^{N-2} \lambda_j^n = \sum_{j=1}^{N-2} \lambda_j^n x_j = 0$$

and was written in a matrix form in order to use the LU factorization. The scheme works similarly to the finite difference methods, although it is a meshless method. A mesh free method does not require a mesh to discretize the domain of the problem under consideration, that is, the approximate solution is constructed entirely based on a set of scattered nodes. Later on, this technique of employing the collocation method and approximating directly the solution using thin plate splines radial basis functions was extended by the authors in [30], in the two dimensional telegraph equation with variable diffusive and reactive coefficients.

In [27] a method that has Chebyshev cardinal functions is presented for the solution of telegraph equation. The Chebyshev cardinal functions of order N in $[-1, 1]$ are defined by

$$C_j(t) = \frac{P_{N+1}(t)}{P'_{N+1}(t_j)(t - t_j)},\ j = 1, 2, \ldots, N+1,$$

where $P_{N+1}(t) = \cos\left((N+1)\arccos(t)\right)$, and t_j are the zeros of $P_{N+1}(t)$. The method used the shifted Chebyshev cardinal functions on interval $[0, 1]$ after a change of variable. Then, function $u(x,t)$ was expanded in terms of double Chebyshev cardinal functions on interval $[0,1] \times [0,1]$, that is,

$$u(x,t) = \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} u(x_i, t_j)\, C_i(t) C_j(x) = \Phi_N^T(t) U \Phi_N(x), \tag{1.18}$$

for $\Phi_N(.) = [C_1(.), C_2(.), \ldots, C_{N+1}(.)]^T$ and $U = [u(x_i, t_j)]$, $i, j = 1, 2, \ldots, N+1$ the unknown matrix to be determined. The numerical method was obtained after differentiating $\Phi_N(.)$, substituting in the main equation the function $u(x,t)$ by (1.18) and manipulating some algebraic equations.

Another approach is presented in [25] based on the boundary integral equation technique and the dual reciprocity method (DRM). The DRM method uses a technique where the domain integral is converted to an equivalent boundary integral by using suitable approximation functions [51]. First, the telegraph equation was restated as an integral equation by the following identity

$$\int_a^b \left[u_{tt} + 2\mathcal{A}u_t + \mathcal{B}^2 u - u_{xx} - f(x,t)\right] \omega \, dx = 0, \tag{1.19}$$

where the weight function ω was chosen to be

$$\omega(x,\xi) = \frac{1}{2}|x-\xi| \qquad \text{and} \qquad \omega_x(x,\xi) = \frac{1}{2} sgn(x-\xi),$$

x is a field point, ξ is a source point and sgn denotes the signum function defined by

$$sgn\,(x) = \begin{cases} -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ 1, & \text{if } x > 0 \end{cases}.$$

The interval $[a,b]$ was partitioned in $N-1$ subintervals with N source points: $a = x_1 < x_2 < \cdots < x_{N-1} < x_N = b$. Then, the dual reciprocity method was implemented: the time derivatives and the inhomogeneous terms were interpolated by radial basis functions. Three different types of radial basis functions were used for interpolation: linear, cubic and thin plate spline radial basis functions. This gave rise to a system with linear differential equations containing the unknown functions of time $u(x_i,t)$ and $\frac{\partial u}{\partial x}(x_i,t)$, $i = 2, 3 \ldots, N-1$, after applying the Dirichlet boundary conditions. The system was solved by considering backward and centered differences to approximate the first derivative and second derivative in time, respectively. In the end, the Crank-Nicolson method [19] was applied to obtain the approximations to the unknown variables. The main advantage of the method is that, since the telegraph equation was integrated over the boundary of the

domain, the discretization and computation of the solution only took place on the boundary. As an application of the telegraph equation an example that involves dispersive wave propagation was included.

A differential quadrature method was employed in both time and space directions in the work presented in [76]. Compared with methods such as the finite difference and finite element methods, the differential quadrature method requires less computer time and storage. The essence of this new differential quadrature method is that a partial derivative of a function at a grid point is approximated by a weighted linear sum of the function values at all given discrete points. The weighting coefficients are determined using the Lagrange interpolation polynomial, which leads to the polynomial-based differential quadrature method (PDQ). For a specific function f, the first and the second order derivatives can be approximated at a grid point x_i by PDQ approach as

$$f_x(x_i) = \sum_{j=1}^{N} \omega_j^{(1)}(x_i) f(x_j) \qquad \text{and} \qquad f_{xx}(x_i) = \sum_{j=1}^{N} \omega_j^{(2)}(x_i) f(x_j),$$

where $i, j = 1, 2, \ldots, N$, N is the number of grid points in the whole domain. The weighting coefficients $\omega_j^{(1)}(x_i), \omega_j^{(2)}(x_i)$ are defined as

$$\omega_j^{(1)}(x_i) = \frac{M^{(1)}(x_i)}{(x_i - x_j) M^{(1)}(x_j)},\ i \neq j, \qquad \omega_i^{(1)}(x_i) = -\sum_{j=1, j\neq i}^{N} \omega_j^{(1)}(x_i),$$

$$\omega_j^{(2)}(x_i) = 2\omega_j^{(1)}(x_i) \left(\omega_i^{(1)}(x_i) - \frac{1}{x_i - x_j} \right), i \neq j, \quad \omega_i^{(2)}(x_i) = -\sum_{j=1, j\neq i}^{N} \omega_j^{(2)}(x_i),$$

where

$$M^{(1)}(x_j) = \prod_{k=1, k\neq j}^{N} (x_j - x_k).$$

When the function f is approximated by a Fourier series expansion, the weighting coefficients $\omega_j^{(1)}(x_i), \omega_j^{(2)}(x_i)$ are different yielding the Fourier-based differential quadrature method (FDQ). The weighting coefficients depend only on the grid spacing. A Gauss-Chebyshev-Lobatto grid points in space direction was considered, whereas equally spaced and also Gauss-Chebyshev-Lobatto grid points were used in time direction. The resulting system of

algebraic equations was solved by the least square method. The use of differential quadrature method in time direction, which also discretizes the given initial condition $\frac{\partial u}{\partial t}(x,0)$, provides the solution at any time level without an iteration between two time levels. This numerical procedure requires very small number of grid points in space directions and appropriate number of time grid points for reaching a certain time level.

With the purpose to achieve higher accuracy methods to solve telegraph equation, higher order compact (HOC) difference methods were developed in different works [21, 35, 59, 71, 93].

It is known that

$$\frac{\partial^2 u_i}{\partial x^2} = \delta_x^2 u_i + \mathcal{O}(\Delta x^2) = \frac{u_{i-1} - 2u_i + u_{i+1}}{\Delta x^2} + \mathcal{O}(\Delta x^2),$$

where δ_x^2 is the second order centered difference operator with respect to x, gives a second order approximation to the second order derivative. On the other hand, making a Taylor series expansion over u_{i+1} and u_{i-1} the following relation holds

$$\delta_x^2 u_i = \frac{\partial^2 u_i}{\partial x^2} + \frac{\Delta x^2}{12}\frac{\partial^4 u_i}{\partial x^4} + \mathcal{O}(\Delta x^4),$$

yielding the compact finite difference operator defined by

$$\frac{\partial^2 u_i}{\partial x^2} = \frac{\delta_x^2}{1 + \frac{\Delta x^2}{12}\delta_x^2} u_i + \mathcal{O}(\Delta x^4), \tag{1.20}$$

which has fourth order accuracy and it is used to approximate the second order derivative term in space. The application of this compact difference operator generates fourth order compact finite difference schemes. These schemes have the advantage of high accuracy to approximate the second order derivatives and keeping the desirable tridiagonal nature of the finite difference equations. As they also consume less memory space, this is one more reason that justifies the renewed interest, in recent years, in the development and application of these high order compact finite difference methods for the numerical solution of partial differential equations. They have been extensively applied in different fields such as fluid dynamics,

quantum mechanics and heat transfer. We describe some of these methods below.

For solving the telegraph equation the method introduced in [71] was based on a collocation approach for the time component. First, the second order spatial derivative was discretized by a fourth order compact difference scheme. Then, at each spatial grid point, the solution was approximated by a polynomial in time. The collocation points were used to obtain the unknown coefficients of the polynomial. The linear system obtained by applying this procedure for each grid point was solved by Gauss elimination method with partial pivoting. The numerical tests showed that for low values of final time low values of the polynomial degree are suitable. For large values of final time it is efficient to increase the polynomial degree in collocation approach to obtain higher accurate results. Besides the fourth order accuracy in space, the method is also unconditionally stable. The same procedure was used by the authors for the two dimensional hyperbolic equation. A compact finite difference approximation of fourth order for discretizing spatial derivatives and a collocation method for the time component were combined in [28].

Another high order and unconditionally stable method can be found in [35], where a three level compact difference scheme of $\mathcal{O}(\Delta t^2 + \Delta x^4)$ was proposed for solving telegraph equation. At first, two fourth order difference formulas were derived by Taylor series expansions for the first and second order derivatives in time. A fourth order compact finite difference operator was used to approximate the second order derivative in space. An additional term was added in order to obtain the unconditional stability.

A numerical method which is also unconditionally stable, second order accurate in time and fourth order accurate in space, can be found in [59] for the numerical solution of telegraph equation. The method developed is a three level implicit difference scheme based on quartic spline interpolation in space direction and finite difference discretization, by using Taylor series expansion, in time direction. A von Neumann analysis was used to prove the unconditional stability of the scheme.

Instead of the standard centered difference scheme obtained by using

second order centered difference approximations to the derivatives [35, 59, 71, 93], a scheme using three grid points each at zeroth, first and second time level was applied for telegraphic equation in [68]. Two terms were added in the resulting equation in order to get an unconditionally stable scheme, confirmed by the von Neumann method. Furthermore, the difference scheme has order $\mathcal{O}(\Delta t^2 + \Delta x^4)$.

Let us now describe what has been done for another equation, the sine-Gordon equation. The sine-Gordon equation is also a special case of our equation (1.9), with $P = 0$ and a source term $f(x, t, u)$. A compact finite difference scheme was considered in [21]. After approximating the second order derivative in space by the compact finite difference operator (1.20), the sine-Gordon equation was transformed in an ordinary differential equation of second order. Then Padé approximant (1.16) was used to approximate the time derivatives and we get a three level implicit compact difference scheme with the local truncation error being $\mathcal{O}(\Delta t^2 + \Delta x^4)$. The resulting fully discrete nonlinear finite difference equation was solved by a predictor-corrector scheme. Convergence of the method was obtained by the energy method.

The numerical methods described above are applied in problems which do not assume the presence of a potential field, that is, do not include the first order spatial derivative in equation (1.9). Next, we present some of the numerical schemes developed to solve these problems which contain the more general equation (1.9).

For a one dimensional diffusive problem given by equation (1.9) with $\theta = 1$, a hybrid numerical scheme in [14] was derived: the method consists in using the Laplace transform technique to remove the time dependent terms and a finite volume formulation for the spatial discretization that uses hyperbolic shape functions. For numerical inversion of Laplace transform the authors refer to [46]. In works such as [15, 57, 58] we can observe other applications for the same finite volume formulation presented in [14]. This method was applied to solve many different problems: for instance, in [15] the numerical method was used in order to analyze diffusive problems in

a two-layered composite medium; the thermal wave propagation caused by pulsed surface disturbance in an infinite cylinder and also in a sphere was investigated in [57], where hyperbolic heat conduction problems were solved for pulsed surface heat flux in a finite slab, a solid cylinder and a solid sphere; in [58] the behavior of bio-heat transfer in multi-layer living tissues was studied during magnetic tumor hyperthermia treatment.

A piecewise linearized method can be found in [79] to solve equation (1.9) with $\theta = 1$: this numerical method was presented as an alternative to the finite volume formulation in cases where the P value is large, since the finite volume formulation yields oscillatory solutions in these cases. The results of the method have been compared with those obtained with the finite volume formulation presented in [15].

In this thesis, the finite volume formulation and the piecewise linearized method are described in detail and applied to our model problem considering non-trivial initial conditions and different values of the parameter P, for both parabolic ($\theta = 0$) and hyperbolic ($\theta \neq 0$) equations.

Another hybrid numerical method that combines the Laplace transform, a weighting function scheme and a hyperbolic shape function for solving a time dependent hyperbolic heat conduction equation, with a conservation term, can be found in [16]. In the end, the application of the numerical inversion algorithm for the Laplace transform presented in [46] was used. To investigate the effect of the surface curvature of a solid body on hyperbolic heat conduction, equation (1.9) was considered with $\theta = 0.5$ and for constant values of P and D.

In a more recent paper [17] the same author analyzed the hyperbolic heat conduction problems in the cylindrical coordinate system using a slightly different approach. The method combines again the Laplace transform for the time domain and Green's function for the space domain. The efficiency of the method was analyzed with one, two and three dimensional numerical examples. Study of the heat conduction problems in cylindrical coordinate systems has received considerable interest, because of its wide industrial applicability, such as rocket wall, oil reservoirs and boilers. Furthermore,

problems of hyperbolic heat conduction are considered in situations involving energy sources such as laser and microwave with extremely short duration or very high frequency and very high temperature gradients.

1.3.2 Two dimensional linear hyperbolic equation

As mentioned in the last section, some of the numerical methods applied in one dimensional problem were also extended to the corresponding problem in two dimensions [17, 28, 30]. In this section, we focus our attention in models arising in the context of equations similar to equation (1.10). We also start with a review of the numerical methods for the telegraph equation, obtained from equation (1.10) with no potential field and a reactive term or source term.

One of the difficulties that appears in two dimensions is that the system to be solved becomes larger as the lengths of the variables increase. This causes more memory usage and therefore a big computational effort. We note that, from the application of some finite difference schemes, a sparse linear system arising from the implicit spatial discretization must be solved at each time step. Direct methods, based on Gaussian elimination, are not usually practical since they need excessive memory and computational effort for solving the matrix equations associated with difference schemes in two dimensions [90]. A strategy to overcome the computational inefficiency of an implicit scheme in two dimensions, is to use an alternating direction implicit (ADI) scheme after discretization. The ADI methods, which are based on reducing a multidimensional problem in several space variables to a set of independent one dimensional problems and only requiring to solve systems with tridiagonal matrices, are highly efficient procedures for the solutions of parabolic and hyperbolic multidimensional initial-boundary value problems. The underlying idea of the ADI in two dimensional problems is to split the computations in two steps. In the first step we evaluate the spatial variable x implicitly and variable y explicitly, producing an intermediate solution for time. In the second step an implicit method is applied in the y-direction and

an explicit method in the x-direction. Taking into account some important aspects as the dimension considered, the discretization methods that were first applied to approximate the solution of the differential equation and the splitting techniques used afterwards, triggered the appearance of some ADI methods such as the Peaceman-Rachford scheme, the D'Yakonov scheme, the Douglas-Rachford scheme and the Douglas-Gunn scheme [90]. The ADI method was first proposed by Peaceman and Rachford in 1955 [75] for the implicit solution of heat flow (a parabolic partial differential equation) in two geometric dimensions. It has been popular since then due to the gain of computational cost effectiveness. The ADI methods are used in a great variety of applications: from astrophysical and bioengineering applications to tsunami modeling and Black-Scholes option pricing. Once implicit schemes are necessary the ADI approaches are needed in two and higher dimensional problems.

Some of the ADI methods have been used for solving two dimensional linear hyperbolic equations [22, 31, 36, 50, 60, 69, 89, 93], although very few take in consideration the presence of a potential field.

An implicit and unconditionally stable difference scheme was developed in [69] for the solution of telegraph equation which is second order accurate in time and fourth order accurate in space. The equation was discretized by an explicit scheme, constructed with second order centered and averaging difference approximations for the derivatives in both space and time. A conditionally stable numerical method is obtained. In order to get an unconditionally stable difference scheme, the explicit scheme was rewritten as an implicit one by adding two additional terms of higher order that do not affect the second order accuracy of the method. Then, it was applied an alternating direction implicit (ADI) method and the equation was separated in two steps. This ADI method requires solution of tridiagonal systems, first along the y-direction and then along the x-direction. The unconditionally stability was studied applying the von Neumann method. It can be found in [70] the same technique for telegraph equation with variable coefficients in the diffusive and reactive terms.

To achieve higher spatial accuracy and computational cost effectiveness, there has been a continuous interest in the implementation of high order compact ADI (HOC-ADI) methods. The HOC-ADI scheme retains the high efficiency and tridiagonal algorithm of ADI methods, and at the same time achieves fourth order accuracy in space, preserving the high accuracy of HOC schemes. More recently, HOC-ADI methods have been successfully applied to solve hyperbolic problems [22, 31, 36, 60, 68, 93].

The unconditionally stable implicit difference scheme presented in [68] and already described in the last section, was extended to two and three dimensional telegraphic equations. To solve the two dimensional problem, in order to facilitate the computation, the obtained scheme was rewritten in two step ADI form which only requires the solution of tridiagonal systems.

A three level compact difference scheme which is second order accurate in time and fourth order accurate in space was proposed in [36]. The method is stable and follows the ideas presented in [35] for the one dimensional case, described in Section 1.3.1.

In [60], a three level ADI compact scheme which is second order accurate in time and fourth order accurate in space was formulated for the telegraph equation. The method is unconditionally stable which was proved by a von Neumann analysis. In order to give the comparative results with the high order scheme given in [68], the same problems were tested in the numerical results. In one example the numerical errors are less than half of the ones of [68] and in another example the method has the same accuracy order but it only needs half of the computational time.

A HOC-ADI difference scheme was derived in [93] to solve telegraph equations. A Taylor series expansion was used in second order derivatives in space and a three time level discretization was made for the derivatives in time. The result was a three time level difference scheme which, with the introduction of an auxiliary variable, allowed the authors to obtain an equivalent two level compact difference scheme. Three and two level ADI compact difference schemes were constructed. The method was shown to be unconditionally stable by the energy method and is second order accurate in

time and fourth order accurate in space for both time levels.

Another type of HOC-ADI scheme was employed in [22] for solving the generalized sine-Gordon equation similar to (1.10). First, the compact finite difference operator (1.20) was used for both second order space derivatives and the derivatives in time were approximated by first and second order centered difference operators, respectively. Then, a three time level HOC-ADI difference scheme was developed and its convergence was proved by the energy method. This method is second order accurate in time and fourth order accurate in space.

In the work presented in [50], the computational efficiency of the ADI approach and high order accuracy of the HOC scheme were combined to solve the two dimensional convection diffusion equation. This equation plays an important role in computational hydraulics and fluid dynamics to model convection diffusion of quantities such as mass, heat, energy and vorticity. A fourth order polynomial compact difference formula was used in the approximation of the spatial derivatives and the Crank-Nicolson method was used for time discretization resulting in a method which is second order accurate in time and fourth order accurate in space. It was shown through the von Neumann analysis that the method is unconditionally stable. It seems the present HOC-ADI method provides a more accurate solution than the standard Peaceman-Rachford ADI method.

An exponential high order compact (EHOC) alternating direction implicit (ADI) method was presented in [89] for the solution of unsteady convection diffusion problems in two dimensions. The Crank-Nicolson scheme was then used for the time discretization and an exponential fourth order compact difference formula for the steady-state one dimensional convection diffusion problem was used for the spatial discretization. The method is second order accurate in time and fourth order accurate in space. The unconditionally character of the method was verified by a von Neumann analysis. The main difference between the HOC-ADI and the EHOC-ADI schemes is that, for the spatial approximation, the first uses a polynomial compact difference discretization while the second one uses an exponential compact difference

discretization. In each step of the ADI method the EHOC scheme produces a strictly diagonally dominant tridiagonal matrix equation, which can be inverted by simple tridiagonal Gaussian decomposition with a considerable saving in computation time. Numerical experiments were performed to illustrate the performance of the method proposed and to compare it mostly with the HOC-ADI method proposed in [50].

These HOC-ADI and EHOC-ADI methods have fourth order accuracy in space, but only second order accuracy in time. To improve the accuracy in temporal dimension and raise computational efficiency, it was used a Richardson extrapolation. The idea of Richardson extrapolation is to use combinations of numerical approximations obtained previously by the same numerical method with different grid parameters [31, 32, 33, 55].

Motivated by the work of [22], a three level HOC-ADI difference method with a Richardson extrapolation algorithm was used to solve a non-linear wave equation in [31], which includes a nonlinear forcing term $f(u, x, y, t)$. Varying the parameters and terms involved, the equation may represents a telegraph equation, a damped sine-Gordon equation or even a Klein-Gordon equation. First, a three level HOC-ADI difference scheme was derived. The use of energy method showed the conditionally convergence of the numerical solution with accuracy order of $\mathcal{O}(\Delta t^2 + \Delta x^4 + \Delta y^4)$. Then, a Richardson extrapolation algorithm based on three time grid parameters was designed and combined with HOC-ADI method to achieve numerical solution of fourth order accuracy in both time and space. The same combination was used in [33], although the spatial discretizations are different from those in [22]. In fact, two auxiliary functions and a constant parameter were introduced and a Taylor series expansion was performed. An approximate factorization of finite difference operators was carried out to obtain a family of three level compact ADI schemes. The application of energy method proved that this numerical method can attain fourth order accuracy in both time and space.

The computational cost that comes from the application of Richardson extrapolation algorithms is reduced since the resulting high order accuracy methods allow the use of much larger time steps in the computation. The

main advantage is that they preserve the stability of lower order methods used initially.

The differential quadrature method developed in [76] and described in Section 1.3.1 was also applied in the two dimensional problem. In this case, to overcome the problem of the big computational effort required in two dimensions, the following procedure is done: the discretized system is reduced by removing the entries in the coefficient matrix, which correspond to known initial and Dirichlet boundary conditions, and the right hand side of the reduced system is modified taking into account the removed known entries. All required time level values are present in the solution.

1.4 Original contributions

The study of hyperbolic diffusive equations have been appearing in many other works such as [3, 43, 52, 56, 61, 78, 83, 84, 94, 96]. They deserved our attention for their innovation in theoretical analysis and/or interesting applications. However, in this literature review we have mainly described numerical methods for equations more similar to equations (1.9) and (1.10) and that have more closely inspired our original work.

Our original contributions concerns the development of numerical methods, in one and two dimensions, to solve equations (1.9) and (1.10), respectively. The particular features of those contributions can be summarized as the following:

(i) Development of an implicit numerical method taking into account a potential field, as well as its convergence analysis, implemented in Chapter 2 for the one dimensional case.

(ii) Convergence of an inverse Laplace transform algorithm and the convergence of methods presented in Chapter 3. The introduction of a symmetric periodic potential field, which is a practical application of equation (1.9). Some of this work is published in papers:

C. Neves, A. Araújo, E. Sousa, *Numerical approximation of a transport equation with a time-dependent dispersion flux*, AIP Conference Proceedings 1048: 403–406, 2008,

A. Araújo, C. Neves, E. Sousa, *A Laplace transform piecewise linearized method for a second order hyperbolic equation*, AIP Conference Proceedings, 1479: 2187-2190, 2012

and

A. Araújo, A. K. Das, C. Neves, E. Sousa, *Numerical solution for a non-Fickian diffusion in a periodic potential*, Communications in Computational Physics, 13(2): 502–525, 2013.

(iii) Implementation of an ADI numerical method to solve the hyperbolic two dimensional equation (1.10). The presence of functions P and Q makes the proof of stability by the energy method more challenging. The numerical method, its convergence analysis and some numerical results contained in Chapter 4 are submitted for publication:

A. Araújo, C. Neves, E. Sousa, *An alternating direction implicit method for a two-dimensional hyperbolic diffusion equation*, submitted for publication, 2013.

Chapter 2

The Crank-Nicolson method

The main purpose of this chapter is to find approximate solutions of an initial boundary value problem for equation (1.9). We start to implement an implicit finite difference method which is a Crank-Nicolson method [19] with first order discretization in time and second order discretizations in space. The second order accuracy of the method as well as its stability are proved. Finally, in Section 2.3 we illustrate the performance of the numerical method with some numerical results.

2.1 Numerical method

We consider the problem

$$\theta\frac{\partial^2 u}{\partial t^2}(x,t)+\frac{\partial u}{\partial t}(x,t)=-\frac{\partial}{\partial x}\left(P(x)u(x,t)\right)+D\frac{\partial^2 u}{\partial x^2}(x,t), \tag{2.1}$$

with initial conditions

$$u(x,0)=u_0(x),\quad \theta\frac{\partial u}{\partial t}(x,0)=u_1(x),\qquad x\in[a,b], \tag{2.2}$$

and Dirichlet boundary conditions

$$u(a,t)=f(t),\qquad u(b,t)=g(t),\qquad t>0. \tag{2.3}$$

A finite difference scheme based on the Crank-Nicolson method is developed. Due to the second derivative in time, direct discretization of (2.1) leads to a

finite difference scheme that is three-level in time. To avoid a three-level discretization scheme we introduce an auxiliary function, following the idea in [98]:

$$w = \theta \frac{\partial u}{\partial t} + u \tag{2.4}$$

and change the differential equation (2.1) into

$$\frac{\partial w}{\partial t} = -\frac{\partial}{\partial x}(Pu) + D\frac{\partial^2 u}{\partial x^2} = -P'u - P\frac{\partial u}{\partial x} + D\frac{\partial^2 u}{\partial x^2}, \tag{2.5}$$

where P' denotes the derivative of $P(x)$. We consider the mesh points in $\Omega = [a, b]$ given by

$$x_i = a + i\Delta x, \quad i = 0, \ldots, N,$$

with $\Delta x = (b - a)/N$, where N is a positive integer. For $0 \leq t \leq T_f$, let $t_n = n\Delta t$, with Δt being the time increment and $n\Delta t \leq T_f$. We denote the approximate solutions to $u(x_i, t_n)$ and $w(x_i, t_n)$ by U_i^n and W_i^n, respectively, $P(x_i)$ by P_i and $P'(x_i)$ by P_i'. The discretization of equations (2.4) and (2.5) is made using the Crank-Nicolson method:

$$W_i^{n+1} + W_i^n = U_i^{n+1} + U_i^n + \frac{2\theta}{\Delta t}\left(U_i^{n+1} - U_i^n\right) \tag{2.6}$$

and

$$\begin{aligned} \frac{W_i^{n+1} - W_i^n}{\Delta t} = & - \frac{1}{2}P_i'(U_i^{n+1} + U_i^n) - \frac{P_i}{2}\left[\frac{U_{i+1}^{n+1} - U_{i-1}^{n+1}}{2\Delta x} + \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x}\right] \\ & + \frac{D}{2}\left[\frac{U_{i-1}^{n+1} - 2U_i^{n+1} + U_{i+1}^{n+1}}{\Delta x^2} + \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{\Delta x^2}\right]. \end{aligned} \tag{2.7}$$

To write the scheme (2.6)-(2.7) in matrix form we solve equation (2.6) for W_i^{n+1} and get

$$W_i^{n+1} = \left(1 + \frac{2\theta}{\Delta t}\right) U_i^{n+1} + \left(1 - \frac{2\theta}{\Delta t}\right) U_i^n - W_i^n. \tag{2.8}$$

Substituting (2.8) into (2.7) gives

$$\frac{1}{\Delta t}\left[\left(1+\frac{2\theta}{\Delta t}\right)U_i^{n+1}+\left(1-\frac{2\theta}{\Delta t}\right)U_i^n-2W_i^n\right]$$
$$=-\frac{P_i'}{2}(U_i^{n+1}+U_i^n)-\frac{P_i}{4\Delta x}\left(U_{i+1}^{n+1}-U_{i-1}^{n+1}+U_{i+1}^n-U_{i-1}^n\right)$$
$$+\frac{D}{2\Delta x^2}\left[\left(U_{i-1}^{n+1}-2U_i^{n+1}+U_{i+1}^{n+1}\right)+\left(U_{i-1}^n-2U_i^n+U_{i+1}^n\right)\right].$$

After simplification we have

$$\left(-\frac{1}{4\Delta x}P_i-\frac{D}{2\Delta x^2}\right)U_{i-1}^{n+1}+\left[\frac{1}{\Delta t}\left(1+\frac{2\theta}{\Delta t}\right)+\frac{P_i'}{2}+\frac{D}{\Delta x^2}\right]U_i^{n+1}$$
$$+\left(\frac{1}{4\Delta x}P_i-\frac{D}{2\Delta x^2}\right)U_{i+1}^{n+1}$$
$$=\left(\frac{1}{4\Delta x}P_i+\frac{D}{2\Delta x^2}\right)U_{i-1}^n+\left[-\frac{1}{\Delta t}\left(1-\frac{2\theta}{\Delta t}\right)-\frac{P_i'}{2}-\frac{D}{\Delta x^2}\right]U_i^n$$
$$+\left(-\frac{1}{4\Delta x}P_i+\frac{D}{2\Delta x^2}\right)U_{i+1}^n+\frac{2}{\Delta t}W_i^n \quad i=1,\dots,N-1, \tag{2.9}$$

which will be used to compute U_i^{n+1}. After that, U_i^{n+1} is substituted into (2.8) to compute W_i^{n+1}.

From (2.8) and (2.9) we obtain the system

$$\begin{cases} AU^{n+1}=BU^n+\dfrac{2}{\Delta t}W^n+d \\ W^{n+1}=\left(1+\dfrac{2\theta}{\Delta t}\right)U^{n+1}+\left(1-\dfrac{2\theta}{\Delta t}\right)U^n-W^n \end{cases}, \tag{2.10}$$

where A and B are band matrixes of size $(N-1)\times(N-1)$ with bandwidth three,

$$U^{n+1}=\left[U_1^{n+1},\dots,U_{N-1}^{n+1}\right]^T, \qquad U^n=\left[U_1^n,\dots,U_{N-1}^n\right]^T,$$
$$W^{n+1}=\left[W_1^{n+1},\dots,W_{N-1}^{n+1}\right]^T, \qquad W^n=\left[W_1^n,\dots,W_{N-1}^n\right]^T,$$

and d contains boundary conditions. Therefore,

$$A = \begin{bmatrix} A_{21} & A_{31} & & & & \\ A_{12} & A_{22} & A_{32} & & & \\ & \ddots & \ddots & \ddots & & \\ & & A_{1N-2} & A_{2N-2} & A_{3N-2} & \\ & & & A_{1N-1} & A_{2N-1} & \end{bmatrix},$$

$$B = \begin{bmatrix} B_{21} & B_{31} & & & & \\ B_{12} & B_{22} & B_{32} & & & \\ & \ddots & \ddots & \ddots & & \\ & & B_{1N-2} & B_{2N-2} & B_{3N-2} & \\ & & & B_{1N-1} & B_{2N-1} & \end{bmatrix},$$

$$d = \left[-A_{11}U_0^{n+1} + B_{11}U_0^n, 0, \ldots, 0, -A_{3N-1}U_N^{n+1} + B_{3N-1}U_N^n\right]^T,$$

where

$$A_{1i} = -\frac{1}{4\Delta x}P_i - \frac{D}{2\Delta x^2}, \qquad B_{1i} = \frac{1}{4\Delta x}P_i + \frac{D}{2\Delta x^2},$$

$$A_{2i} = \frac{1}{\Delta t}\left(1 + \frac{2\theta}{\Delta t}\right) + \frac{P_i'}{2} + \frac{D}{\Delta x^2}, \qquad B_{2i} = -\frac{1}{\Delta t}\left(1 - \frac{2\theta}{\Delta t}\right) - \frac{P_i'}{2} - \frac{D}{\Delta x^2},$$

$$A_{3i} = \frac{1}{4\Delta x}P_i - \frac{D}{2\Delta x^2}, \qquad B_{3i} = -\frac{1}{4\Delta x}P_i + \frac{D}{2\Delta x^2},$$

for $i = 1, \ldots, N-1$.

In order to simplify the notation of this numerical method, we define the following difference operators. The first order forward and the backward difference operators are given by

$$\delta_x^+ U_i^n = \frac{U_{i+1}^n - U_i^n}{\Delta x} \quad \text{and} \quad \delta_x^- U_i^n = \frac{U_i^n - U_{i-1}^n}{\Delta x}. \tag{2.11}$$

The first order centered difference operator is defined by

$$\delta_x U_i^n = \frac{1}{2}[\delta_x^+ + \delta_x^-]U_i^n = \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x} \tag{2.12}$$

and the second order centered difference operator is defined by

$$\delta_x^2 U_i^n = \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{\Delta x^2}. \tag{2.13}$$

Using the discretization operators defined in (2.12)-(2.13) and by denoting the set of discretization points $U^n = \{U_i^n\}$, $PU^n = \{P_i U_i^n\}$, $P'U^n = \{P_i' U_i^n\}$ and $W^n = \{W_i^n\}$, the numerical method (2.6)-(2.7) can be written in the form

$$W^{n+1} + W^n = U^{n+1} + U^n + \frac{2\theta}{\Delta t}\left(U^{n+1} - U^n\right) \tag{2.14}$$

and

$$W^{n+1} - W^n = -\frac{\Delta t}{2} P'(U^{n+1} + U^n) - \frac{\Delta t}{2} P\delta_x\left(U^{n+1} + U^n\right) + \frac{D\Delta t}{2}\delta_x^2\left(U^{n+1} + U^n\right). \tag{2.15}$$

2.2 Consistency and stability analysis

This section is concerned with the conditions that must be satisfied to ensure the convergence of the numerical method. We start to discuss the consistency, which measures how well the difference equation (2.9) approximates the partial differential equation (2.5) and then, we analyze the stability of the difference scheme.

In the next proposition we prove the second order accuracy of the Crank-Nicolson method.

Proposition 2.2.1. *For the Crank-Nicolson discretization (2.8)-(2.9) we have, for a sufficiently smooth* u*,*

$$\begin{aligned} &A_{1i}u_{i-1}^{n+1} + A_{2i}u_i^{n+1} + A_{3i}u_{i+1}^{n+1} - B_{1i}u_{i-1}^n - B_{2i}u_i^n - B_{3i}u_{i+1}^n - \frac{2}{\Delta t}w_i^n \\ &= \left(\theta\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} + P_i' u + P_i\frac{\partial u}{\partial x} - D\frac{\partial^2 u}{\partial x^2}\right)_i^{n+1/2} + \mathcal{O}(\Delta x^2 + \Delta t^2). \end{aligned} \tag{2.16}$$

Proof: Let us substitute the exact solution $u(x,t)$ in the numerical method (2.9), that is,

$$A_{1i}u_{i-1}^{n+1} + A_{2i}u_i^{n+1} + A_{3i}u_{i+1}^{n+1} - B_{1i}u_{i-1}^n - B_{2i}u_i^n - B_{3i}u_{i+1}^n - \frac{2}{\Delta t}w_i^n = 0.$$

In order to estimate the size of the truncation error, we expand the functions $u_{i-1}^{n+1}, u_i^{n+1}, u_{i+1}^{n+1}, u_{i-1}^n, u_i^n$ and u_{i+1}^n into a Taylor series around the point $\left(x_i, t_{n+1/2}\right)$. Therefore, we have

$$
\begin{aligned}
&A_{1i}u_{i-1}^{n+1} + A_{2i}u_i^{n+1} + A_{3i}u_{i+1}^{n+1} - B_{1i}u_{i-1}^{n} - B_{2i}u_i^{n} - B_{3i}u_{i+1}^{n} \\
&\quad = \left[\left(\frac{2}{\Delta t} + P_i'\right)u + \frac{2\theta}{\Delta t}\frac{\partial u}{\partial t} + \left(\frac{\Delta t}{4} + \frac{\Delta t^2}{8}P_i'\right)\frac{\partial^2 u}{\partial t^2} + \frac{\theta\Delta t}{12}\frac{\partial^3 u}{\partial t^3}\right]_i^{n+1/2} \\
&\qquad + \left[P_i\frac{\partial u}{\partial x} + \frac{\Delta t^2}{8}P_i\frac{\partial^3 u}{\partial t^2\partial x} - D\frac{\partial^2 u}{\partial x^2} + \frac{\Delta x^2}{6}P_i\frac{\partial^3 u}{\partial x^3}\right]_i^{n+1/2} + \cdots . \qquad (2.17)
\end{aligned}
$$

On the other hand, expanding the equality (2.8)

$$w_i^{n+1} = \left(1 + \frac{2\theta}{\Delta t}\right)u_i^{n+1} + \left(1 - \frac{2\theta}{\Delta t}\right)u_i^n - w_i^n$$

around the point $(x_i, t_{n+1/2})$ we reach to

$$
\begin{aligned}
&\left(w + \frac{\Delta t}{2}\frac{\partial w}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 w}{\partial t^2} + \frac{\Delta t^3}{48}\frac{\partial^3 w}{\partial t^3}\right)_i^{n+1/2} \\
&\quad = \left(1 + \frac{2\theta}{\Delta t}\right)\left(u + \frac{\Delta t}{2}\frac{\partial u}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 u}{\partial t^2} + \frac{\Delta t^3}{48}\frac{\partial^3 u}{\partial t^3}\right)_i^{n+1/2} \\
&\qquad + \left(1 - \frac{2\theta}{\Delta t}\right)\left(u - \frac{\Delta t}{2}\frac{\partial u}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 u}{\partial t^2} - \frac{\Delta t^3}{48}\frac{\partial^3 u}{\partial t^3}\right)_i^{n+1/2} \\
&\qquad - \left(w - \frac{\Delta t}{2}\frac{\partial w}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 w}{\partial t^2} - \frac{\Delta t^3}{48}\frac{\partial^3 w}{\partial t^3}\right)_i^{n+1/2} + \cdots ,
\end{aligned}
$$

that is,

$$w_i^{n+1/2} = \left(u + \theta\frac{\partial u}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 u}{\partial t^2} + \theta\frac{\Delta t^2}{24}\frac{\partial^3 u}{\partial t^3} - \frac{\Delta t^2}{8}\frac{\partial^2 w}{\partial t^2} + \cdots\right)_i^{n+1/2} \qquad (2.18)$$

Since

$$w_i^n = \left(w - \frac{\Delta t}{2}\frac{\partial w}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 w}{\partial t^2} - \frac{\Delta t^3}{48}\frac{\partial^3 w}{\partial t^3} + \cdots\right)_i^{n+1/2},$$

substituting $w_i^{n+1/2}$ from (2.18) gives

$$w_i^n = \left(u + \theta\frac{\partial u}{\partial t} + \frac{\Delta t^2}{8}\frac{\partial^2 u}{\partial t^2} + \theta\frac{\Delta t^2}{24}\frac{\partial^3 u}{\partial t^3} - \frac{\Delta t}{2}\frac{\partial w}{\partial t} - \frac{\Delta t^3}{48}\frac{\partial^3 w}{\partial t^3}\right)_i^{n+1/2} + \cdots .$$

From (2.4) we can write

$$\left(\frac{\partial w}{\partial t}\right)_i^{n+1/2} = \left(\frac{\partial u}{\partial t} + \theta\frac{\partial^2 u}{\partial t^2}\right)_i^{n+1/2} \quad \text{and} \quad \left(\frac{\partial^3 w}{\partial t^3}\right)_i^{n+1/2} = \left(\frac{\partial^3 u}{\partial t^3} + \theta\frac{\partial^4 u}{\partial t^4}\right)_i^{n+1/2}.$$

Therefore,

$$w_i^n = \left(u + \left(\theta - \frac{\Delta t}{2}\right)\frac{\partial u}{\partial t} + \left(\frac{\Delta t^2}{8} - \theta\frac{\Delta t}{2}\right)\frac{\partial^2 u}{\partial t^2}\right)_i^{n+1/2} + \left(\theta\frac{\Delta t^2}{24} - \frac{\Delta t^3}{48}\right)\left(\frac{\partial^3 u}{\partial t^3}\right)_i^{n+1/2} + \cdots . \tag{2.19}$$

Finally, rearranging the terms in (2.17) and (2.19),

$$\begin{aligned} &A_{1i}u_{i-1}^{n+1} + A_{2i}u_i^{n+1} + A_{3i}u_{i+1}^{n+1} - B_{1i}u_{i-1}^n - B_{2i}u_i^n - B_{3i}u_{i+1}^n - \frac{2}{\Delta t}w_i^n \\ &= \left(\theta\tfrac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} + P_i' u + P_i\tfrac{\partial u}{\partial x} - D\tfrac{\partial^2 u}{\partial x^2}\right)_i^{n+1/2} \\ &\quad + \left(\frac{\Delta t^2}{8}P_i'\frac{\partial^2 u}{\partial t^2} + \frac{\Delta t^2}{24}\frac{\partial^3 u}{\partial t^3} + \frac{\Delta t^2}{8}P_i\frac{\partial^3 u}{\partial t^2 \partial x} + \frac{\Delta x^2}{6}P_i\frac{\partial^3 u}{\partial x^3}\right)_i^{n+1/2} + \cdots \\ &= \left(\theta\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} + P_i' u + P_i\frac{\partial u}{\partial x} - D\frac{\partial^2 u}{\partial x^2}\right)_i^{n+1/2} + \mathcal{O}(\Delta x^2 + \Delta t^2). \end{aligned}$$

■

According to this result, we can conclude that the truncation error is $\mathcal{O}(\Delta x^2 + \Delta t^2)$ which confirms the difference scheme is consistent and second order accurate.

Now, to prove the stability of the method we use the discrete energy method [53, 98]. Let us start to define the set of discrete values with homogeneous boundary conditions. Assume that

$$\mathcal{G} = \{U | U = \{U_i\}, U_0 = U_N = 0\}.$$

For $U, V \in \mathcal{G}$, we define the inner product and norm respectively as

$$(U, V) = \Delta x \sum_{i=1}^{N-1} U_i V_i, \quad \|U\|^2 = (U, U) = \Delta x \sum_{i=1}^{N-1} U_i^2. \tag{2.20}$$

We also define the following inner products that involve the first order discretization operators of $U, V \in \mathcal{G}$:

$$\left(\delta_x^+ U, \delta_x^+ V\right)_* = \Delta x \sum_{i=0}^{N-1} \delta_x^+ U_i \delta_x^+ V_i \quad \text{and} \quad \|\delta_x^+ U\|_*^2 = \left(\delta_x^+ U, \delta_x^+ U\right)_* .$$

Next, we introduce some lemmas to prove the main result.

Lemma 2.2.1. *For any $W \in \mathcal{G}$,*

$$\|\delta_x W\| \leq \|\delta_x^+ W\|_*.$$

Proof: We have

$$\|\delta_x W\|^2 = \Delta x \sum_{i=1}^{N-1} (\delta_x W_i)^2 = \Delta x \sum_{i=1}^{N-1} (\frac{1}{2}(\delta_x^+ W_i + \delta_x^- W_i))^2.$$

Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and then shifting the index i in both summations yields

$$\begin{aligned}
\|\delta_x W\|^2 &\leq \frac{1}{2}\Delta x \sum_{i=1}^{N-1} \left(\frac{W_{i+1} - W_i}{\Delta x}\right)^2 + \frac{1}{2}\Delta x \sum_{i=1}^{N-1} \left(\frac{W_i - W_{i-1}}{\Delta x}\right)^2 \\
&= \frac{1}{2}\Delta x \sum_{i=0}^{N-1} \left(\frac{W_{i+1} - W_i}{\Delta x}\right)^2 - \frac{1}{2}\Delta x \left(\frac{W_1 - W_0}{\Delta x}\right)^2 \\
&+ \frac{1}{2}\Delta x \sum_{i=0}^{N-2} \left(\frac{W_{i+1} - W_i}{\Delta x}\right)^2 \\
&\leq \frac{1}{2}\Delta x \sum_{i=0}^{N-1} \left(\frac{W_{i+1} - W_i}{\Delta x}\right)^2 + \frac{1}{2}\Delta x \sum_{i=0}^{N-1} \left(\frac{W_{i+1} - W_i}{\Delta x}\right)^2 \\
&- \frac{1}{2}\Delta x \left(\frac{W_N - W_{N-1}}{\Delta x}\right)^2 \\
&\leq \Delta x \sum_{i=0}^{N-1} \left(\frac{W_{i+1} - W_i}{\Delta x}\right)^2 \\
&= \Delta x \sum_{i=0}^{N-1} (\delta_x^+ W_i)^2 = \|\delta_x^+ W\|_*^2.
\end{aligned}$$

■

The following lemma is the well known property of summation by parts [53, 98].

Lemma 2.2.2. *For any $U, V \in \mathcal{G}$,*

$$\left(\delta_x^2 U, V\right) = -\left(\delta_x^+ U, \delta_x^+ V\right)_*.$$

Proof: We have

$$\begin{aligned}
(\delta_x^2 U, V) &= \Delta x \sum_{i=1}^{N-1} \frac{U_{i+1} - 2U_i + U_{i-1}}{\Delta x^2} V_i \\
&= \Delta x \sum_{i=1}^{N-1} \frac{U_{i+1} - U_i}{\Delta x^2} V_i - \Delta x \sum_{i=1}^{N-1} \frac{U_i - U_{i-1}}{\Delta x^2} V_i \\
&= \Delta x \sum_{i=0}^{N-1} \frac{U_{i+1} - U_i}{\Delta x^2} V_i - \Delta x \sum_{i=0}^{N-1} \frac{U_{i+1} - U_i}{\Delta x^2} V_{i+1} \\
&= -\Delta x \sum_{i=0}^{N-1} \frac{U_{i+1} - U_i}{\Delta x} \frac{V_{i+1} - V_i}{\Delta x} \\
&= -(\delta_x^+ U, \delta_x^+ V)_*.
\end{aligned}$$

In the third line we shifted the indices in both summations and use the fact that $V_0 = V_N = 0$.

■

Let us suppose that $P(x)$ has non-negative derivative $P'(x)$. We define

$$\|U^n\|_{P'}^2 = \Delta x \sum_{i=1}^{N-1} P_i' \, (U_i^n)^2 \quad \text{and} \quad \|P\|^2 = \sum_{i=1}^{N-1} (P_i)^2.$$

Theorem 2.2.1. *Suppose that $\{U_i^n, W_i^n\}$ and $\{V_i^n, Y_i^n\}$ are solutions of the finite difference scheme (2.14)-(2.15) which satisfy the boundary conditions (2.3), and have different initial values $\{U_i^0, W_i^0\}$ and $\{V_i^0, Y_i^0\}$ respectively. Let $\omega_i^n = W_i^n - Y_i^n$, $\epsilon_i^n = U_i^n - V_i^n$. For $\Delta t \le \frac{2D}{\|P\|^2}$, then $\{\omega_i^n, \epsilon_i^n\}$ satisfy*

$$\begin{aligned}
&\|\omega^{n+1}\|^2 + \theta \|\epsilon^{n+1}\|_{P'}^2 + D\theta \|\delta_x^+ \epsilon^{n+1}\|_*^2 \\
&\quad \le (1 + C\Delta t) \left(\|\omega^n\|^2 + \theta \|\epsilon^n\|_{P'}^2 + D\theta \|\delta_x^+ \epsilon^n\|_*^2 \right),
\end{aligned} \tag{2.21}$$

where C denotes a constant independent of $\Delta x, \Delta t$.

Proof: For $\omega^n = \{\omega_i^n\}$ and $\epsilon^n = \{\epsilon_i^n\}$, from (2.15) we have

$$\omega^{n+1} - \omega^n = -\frac{\Delta t}{2} P'(\epsilon^{n+1} + \epsilon^n) - \frac{\Delta t}{2} P \delta_x \left(\epsilon^{n+1} + \epsilon^n \right) + \frac{D\Delta t}{2} \delta_x^2 \left(\epsilon^{n+1} + \epsilon^n \right). \tag{2.22}$$

Multiplying both sides of (2.22) by $\omega^{n+1}+\omega^n$ with respect to the inner product (2.20) we obtain

$$\begin{aligned}(\omega^{n+1}-\omega^n,\omega^{n+1}+\omega^n) = &- \frac{\Delta t}{2}(P'(\epsilon^{n+1}+\epsilon^n),\omega^{n+1}+\omega^n)\\ &- \frac{\Delta t}{2}(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n)\\ &+ \frac{D\Delta t}{2}(\delta_x^2\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n). \quad (2.23)\end{aligned}$$

Considering (2.14) and the norm definitions we have

$$\begin{aligned}&\left(\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right),\delta_x^+\left(\omega^{n+1}+\omega^n\right)\right)_*\\ &= \left(\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right),\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right)\right)_* + \frac{2\theta}{\Delta t}\left(\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right),\delta_x^+\left(\epsilon^{n+1}-\epsilon^n\right)\right)_*\\ &= \|\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right)\|_*^2 + \frac{2\theta}{\Delta t}\left(\|\delta_x^+\epsilon^{n+1}\|_*^2 - \|\delta_x^+\epsilon^n\|_*^2\right). \quad (2.24)\end{aligned}$$

Using summation by parts we can rewrite (2.23) as

$$\begin{aligned}&\|\omega^{n+1}\|^2 - \|\omega^n\|^2\\ &= -\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\omega^{n+1}+\omega^n\right) - \frac{\Delta t}{2}\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right)\\ &\quad -\frac{D\Delta t}{2}\left(\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right),\delta_x^+\left(\omega^{n+1}+\omega^n\right)\right)_*\end{aligned}$$

which gives, from (2.24)

$$\begin{aligned}&\|\omega^{n+1}\|^2 - \|\omega^n\|^2\\ &= -\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\omega^{n+1}+\omega^n\right) - \frac{\Delta t}{2}\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right)\\ &\quad -\frac{D\Delta t}{2}\|\delta_x^+\left(\epsilon^{n+1}+\epsilon^n\right)\|_*^2 - D\theta\left(\|\delta_x^+\epsilon^{n+1}\|_*^2 - \|\delta_x^+\epsilon^n\|_*^2\right). \quad (2.25)\end{aligned}$$

Let us now discuss the terms with P. We first consider the term

$$-\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\omega^{n+1}+\omega^n\right)$$

which is, by (2.14),

$$\begin{aligned}&-\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\epsilon^{n+1}+\epsilon^n\right) - \theta\left(P'(\epsilon^{n+1}+\epsilon^n),\epsilon^{n+1}-\epsilon^n\right)\\ &= -\frac{\Delta t\Delta x}{2}\sum_{i=1}^{N-1}P_i'\left(\epsilon_i^{n+1}+\epsilon_i^n\right)^2 - \theta\Delta x\sum_{i=1}^{N-1}P_i'\left(\epsilon_i^{n+1}\right)^2 + \theta\Delta x\sum_{i=1}^{N-1}P_i'\left(\epsilon_i^n\right)^2\\ &\le -\theta\Delta x\sum_{i=1}^{N-1}P_i'\left(\epsilon_i^{n+1}\right)^2 + \theta\Delta x\sum_{i=1}^{N-1}P_i'\left(\epsilon_i^n\right)^2,\end{aligned}$$

with the assumption that $P'(x)$ is non-negative. Then,

$$-\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\omega^{n+1}+\omega^n\right) \leq -\theta\|\epsilon^{n+1}\|^2_{P'}+\theta\|\epsilon^n\|^2_{P'}. \tag{2.26}$$

Let us know consider the term

$$-\frac{\Delta t}{2}\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right).$$

Using the Cauchy-Schwarz inequality, Lemma 2.2.1 and also the inequality $ab \leq \eta a^2 + b^2/4\eta$, for $\eta > 0$, we have

$$\begin{aligned}
&-\frac{\Delta t}{2}\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right)\\
&\leq \frac{\Delta t}{2}\|P\delta_x(\epsilon^{n+1}+\epsilon^n)\|\|\omega^{n+1}+\omega^n\|\\
&\leq \frac{\Delta t}{2}\eta\|P\delta_x(\epsilon^{n+1}+\epsilon^n)\|^2+\frac{\Delta t}{2}\frac{1}{4\eta}\|\omega^{n+1}+\omega^n\|^2\\
&\leq \frac{\Delta t}{2}\eta\|P\|^2\|\delta_x(\epsilon^{n+1}+\epsilon^n)\|^2+\frac{\Delta t}{8\eta}\|\omega^{n+1}+\omega^n\|^2\\
&\leq \frac{\Delta t}{2}\eta\|P\|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|^2_*+\frac{\Delta t}{8\eta}\|\omega^{n+1}+\omega^n\|^2.
\end{aligned}$$

Since $(a+b)^2 \leq 2a^2+2b^2$, we can conclude that

$$\begin{aligned}
&-\frac{\Delta t}{2}\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right)\\
&\leq \frac{\Delta t}{2}\eta\|P\|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|^2_*+\frac{\Delta t}{4\eta}\left(\|\omega^{n+1}\|^2+\|\omega^n\|^2\right).
\end{aligned} \tag{2.27}$$

From (2.25) and the inequalities (2.26)-(2.27), we obtain

$$\begin{aligned}
&\|\omega^{n+1}\|^2\left(1-\frac{\Delta t}{4\eta}\right)+\theta\|\epsilon^{n+1}\|^2_{P'}+D\theta\|\delta_x^+\epsilon^{n+1}\|^2_*\\
&\leq \|\omega^n\|^2\left(1+\frac{\Delta t}{4\eta}\right)+\theta\|\epsilon^n\|^2_{P'}+D\theta\|\delta_x^+\epsilon^n\|^2_*\\
&\quad+\left(\frac{\Delta t}{2}\eta\|P\|^2-D\frac{\Delta t}{2}\right)\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|^2_*.
\end{aligned}$$

Let us choose $\eta \leq \frac{D}{\|P\|^2}$. Then, $\frac{\Delta t}{2}\eta\|P\|^2 - D\frac{\Delta t}{2} \leq 0$ and we can drop the last term of the previous inequality. This means that we can write

$$\begin{aligned}
&\|\omega^{n+1}\|^2\left(1-\frac{\Delta t}{4\eta}\right)+\theta\|\epsilon^{n+1}\|^2_{P'}+D\theta\|\delta_x^+\epsilon^{n+1}\|^2_*\\
&\leq \|\omega^n\|^2\left(1+\frac{\Delta t}{4\eta}\right)+\theta\|\epsilon^n\|^2_{P'}+D\theta\|\delta_x^+\epsilon^n\|^2_*.
\end{aligned}$$

Let $\Delta t \leq 2\eta$. Therefore, $1 - \frac{\Delta t}{4\eta} > 0$ and it follows

$$\left(1 - \frac{\Delta t}{4\eta}\right)\left(\|\omega^{n+1}\|^2 + \theta\|\epsilon^{n+1}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n+1}\|_*^2\right)$$
$$\leq \left(1 + \frac{\Delta t}{4\eta}\right)\left(\|\omega^{n}\|^2 + \theta\|\epsilon^{n}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n}\|_*^2\right)$$

which implies

$$\|\omega^{n+1}\|^2 + \theta\|\epsilon^{n+1}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n+1}\|_*^2 \leq \frac{1 + \frac{\Delta t}{4\eta}}{1 - \frac{\Delta t}{4\eta}}\left(\|\omega^{n}\|^2 + \theta\|\epsilon^{n}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n}\|_*^2\right).$$

Finally, by noting that

$$\begin{aligned}\frac{1 + \frac{\Delta t}{4\eta}}{1 - \frac{\Delta t}{4\eta}} &= 1 + \frac{\frac{\Delta t}{2\eta}}{1 - \frac{\Delta t}{4\eta}} = 1 + \frac{\frac{1}{2\eta}}{1 - \frac{\Delta t}{4\eta}}\Delta t = 1 + \frac{1/2}{\eta - \frac{\Delta t}{4}}\Delta t \\ &\leq 1 + \frac{1/2}{\eta - \frac{\eta}{2}}\Delta t = 1 + \frac{1}{\eta}\Delta t\end{aligned}$$

and defining $C = \frac{1}{\eta}$ we obtain

$$\|\omega^{n+1}\|^2 + \theta\|\epsilon^{n+1}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n+1}\|_*^2$$
$$\leq (1 + C\Delta t)\left(\|\omega^{n}\|^2 + \theta\|\epsilon^{n}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n}\|_*^2\right).$$

■

From the previous theorem we get the following result.

Corollary 2.2.1. *Suppose that* $\{U_i^n, W_i^n\}$ *and* $\{V_i^n, Y_i^n\}$ *are solutions of the finite difference scheme (2.14)-(2.15) which satisfy the boundary conditions (2.3), and have different initial values* $\{U_i^0, W_i^0\}$ *and* $\{V_i^0, Y_i^0\}$ *respectively. Let* $\omega_i^n = W_i^n - Y_i^n$, $\epsilon_i^n = U_i^n - V_i^n$. *For* $\Delta t \leq \frac{2D}{\|P\|^2}$, *then* $\{\omega_i^n, \epsilon_i^n\}$ *satisfy*

$$\|\omega^{n}\|^2 + \theta\|\epsilon^{n}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n}\|_*^2 \leq K\left(\|\omega^{0}\|^2 + \theta\|\epsilon^{0}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{0}\|_*^2\right), \quad (2.28)$$

where K *denotes a constant independent of* $\Delta x, \Delta t$.

Proof: The result follows from Theorem 2.2.1 by making recursion with respect to n. In fact, for a constant C independent of Δx and Δt, we have

$$\begin{aligned}
&\|\omega^n\|^2 + \theta\|\epsilon^n\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^n\|_*^2 \\
&\quad \leq (1 + C\Delta t)\left(\|\omega^{n-1}\|^2 + \theta\|\epsilon^{n-1}\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^{n-1}\|_*^2\right) \\
&\quad \leq (1 + C\Delta t)^n\left(\|\omega^0\|^2 + \theta\|\epsilon^0\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^0\|_*^2\right) \\
&\quad \leq \mathrm{e}^{Cn\Delta t}\left(\|\omega^0\|^2 + \theta\|\epsilon^0\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^0\|_*^2\right) \\
&\quad \leq \mathrm{e}^{CT_f}\left(\|\omega^0\|^2 + \theta\|\epsilon^0\|_{P'}^2 + D\theta\|\delta_x^+\epsilon^0\|_*^2\right),
\end{aligned}$$

that proves the result with $K = \mathrm{e}^{CT_f}$.

■

Thus, we can conclude that the difference scheme is stable.

Remark 2.2.1. *For $P = 0$ it follows that the difference scheme (2.14)-(2.15) is unconditionally stable.*

Therefore, since we already proved that the Crank-Nicolson scheme is consistent and stable, we can conclude that it is a convergent numerical method by the Lax equivalence theorem.

2.3 Numerical results

In this section we present numerical results to test the performance of the Crank-Nicolson method. We compare the numerical results with exact solutions and we also illustrate the behavior of some solutions. We present two problems for which we are able to determine the exact solution in order to compute the errors and the convergence rate. In three other problems we show how the solution behaves, for parabolic and hyperbolic cases with P constant and P non-constant. In this last case, we highlight an application involving a periodic potential.

Let

$$\epsilon_i = u_i - U_i, \qquad \omega_i = w_i - W_i, \tag{2.29}$$

where u is the exact solution, w is defined by (2.4) and U and W are the approximate solutions, respectively. To measure the error we consider the

maximum norm, or ℓ_∞ norm,

$$\|\epsilon\|_\infty = \max|u_i - U_i| \quad , \quad \|\omega\|_\infty = \max|w_i - W_i|, \tag{2.30}$$

for $1 \leq i \leq N-1$, and the energy norm, or $\ell_{2,\Delta x}$ norm,

$$\|\epsilon\| = \left(\Delta x \sum_{i=1}^{N-1} \epsilon_i^2\right)^{1/2} \quad , \quad \|\omega\| = \left(\Delta x \sum_{i=1}^{N-1} \omega_i^2\right)^{1/2}. \tag{2.31}$$

We consider the $\ell_{2,\Delta x}$ norm to measure the errors because it has been used in the theoretical results of the last section. It is a variation of the Euclidean norm ℓ_2 and it is commonly used to measure the errors defined in (2.29) since, as Δx approaches zero, the ℓ_2 norm of these functions goes to infinity [90]. The maximum norm ℓ_∞ measures the maximum of the error over the interval $[a, b]$ and is also a natural choice whenever convergence is discussed.

Example 2.3.1. We start with the parabolic problem

$$\frac{\partial u}{\partial t}(x,t) = -P\frac{\partial u}{\partial x}(x,t) + \frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in]-\infty, \infty[\,, t > 0,$$

which initial condition is $u(x,0) = \mathrm{e}^{-x^2}$ and the boundary conditions are

$$\lim_{x\to-\infty} u(x,t) = 0, \qquad \lim_{x\to+\infty} u(x,t) = 0.$$

The analytical solution is given by

$$u\,(x,t) = \frac{1}{\sqrt{1+4t}} e^{-\frac{(x-Pt)^2}{1+4t}}.$$

In Table 2.1 we present the errors defined by norms ℓ_∞ and $\ell_{2,\Delta x}$, at the instant of time $t = 1$. We consider $\Delta t = \Delta x$ and observe the convergence rate is second order as expected. The norm ℓ_∞ provides slightly smaller errors.

Example 2.3.2. We consider now a more general problem with $\theta \neq 0$*:*

$$\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -P\frac{\partial u}{\partial x}(x,t) + \frac{\partial^2 u}{\partial x^2}(x,t), \ x \in (0,1), \ t > 0,$$

with initial conditions

$$u(x,0) = e^{Px/2}\sinh(x\sqrt{(2+P^2)/2}),$$

Δx	Error $\|\epsilon\|_\infty$	Rate	Error $\|\epsilon\|$	Rate
20/128	0.1438×10^{-2}		0.2138×10^{-2}	
20/256	0.3591×10^{-3}	2.0	0.5306×10^{-3}	2.0
20/512	0.8502×10^{-4}	2.1	0.1274×10^{-3}	2.1
20/1024	0.2071×10^{-4}	2.0	0.3125×10^{-4}	2.0
20/2048	0.5177×10^{-5}	2.0	0.7813×10^{-5}	2.0

Table 2.1: Errors and rates obtained for Example 2.3.1 for $\theta = 0$, $t = 1$, $-10 \leq x \leq 10$, $P = 1$ and $\Delta t = \Delta x$, computed with the norms ℓ_∞ and $\ell_{2,\Delta x}$.

$$\frac{\partial u}{\partial t}(x,0) = -\frac{1+\sqrt{5+P^2}}{2} e^{Px/2} \sinh(x\sqrt{(2+P^2)/2}),$$

and boundary conditions

$$u(0,t) = 0, \quad u(1,t) = e^{-(1+\sqrt{5+P^2})t/2} e^{P/2} \sinh(\sqrt{(2+P^2)/2}).$$

The exact solution is given by

$$u(x,t) = e^{-(1+\sqrt{5+P^2})t/2} e^{Px/2} \sinh(x\sqrt{(2+P^2)/2}).$$

The errors and convergence rate are presented in Table 2.2 and Table 2.3 for $P = 1$, $t = 1$ and different space steps. In this case, the norms defined by $\ell_{2,\Delta x}$ show slightly smaller errors. We observe the method provides second order accurate solutions when different norms are considered, as predicted by Proposition 2.2.1.

Δx	Error $\|\epsilon\|_\infty$	Rate	Error $\|\omega\|_\infty$	Rate
1/128	0.3781×10^{-5}		0.2966×10^{-4}	
1/256	0.9433×10^{-6}	2.0	0.7533×10^{-5}	2.0
1/512	0.2357×10^{-6}	2.0	0.1904×10^{-5}	2.0
1/1024	0.5890×10^{-7}	2.0	0.4800×10^{-6}	2.0
1/2048	0.1477×10^{-7}	2.0	0.1204×10^{-6}	2.0

Table 2.2: Errors and rates obtained for Example 2.3.2 for $\theta = 1$, $P = 1$, $t = 1$, $0 \leq x \leq 1$ and $\Delta t = \Delta x$, computed with the norm ℓ_∞.

Example 2.3.3. To observe the behavior of the solution performed by the numerical method we first consider the problem

$$\theta \frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -P\frac{\partial u}{\partial x}(x,t) + \frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in \mathbb{R}, t > 0,$$

Δx	Error $\|\epsilon\|$	Rate	Error $\|\omega\|$	Rate
$1/128$	0.2765×10^{-5}		0.1596×10^{-4}	
$1/256$	0.6914×10^{-6}	2.0	0.4005×10^{-5}	2.0
$1/512$	0.1729×10^{-6}	2.0	0.1004×10^{-5}	2.0
$1/1024$	0.4321×10^{-7}	2.0	0.2513×10^{-6}	2.0
$1/2048$	0.1083×10^{-7}	2.0	0.6287×10^{-7}	2.0

Table 2.3: Errors and rates obtained for Example 2.3.2 for $\theta = 1$, $P = 1$, $t = 1$, $0 \leq x \leq 1$ and $\Delta t = \Delta x$, computed with the norm $\ell_{2,\Delta x}$.

for P constant, with initial conditions

$$u(x,0) = \frac{1}{2}e^{-x^2}, \qquad \theta\frac{\partial u}{\partial t}(x,0) = xe^{-x^2}$$

and boundary conditions

$$\lim_{x\to-\infty} u(x,t) = 0, \qquad \lim_{x\to+\infty} u(x,t) = 0.$$

In Figure 2.1 we present the parabolic case, $\theta = 0$, and in Figure 2.2 we consider the hyperbolic case with $\theta \neq 0$. We can observe how the solution changes with the direction of P and also the evolution of the solution as we travel in time. The solution $u(x,t)$ moves to the left and to the right, depending on the sign of P. It also dissipates as t increases.

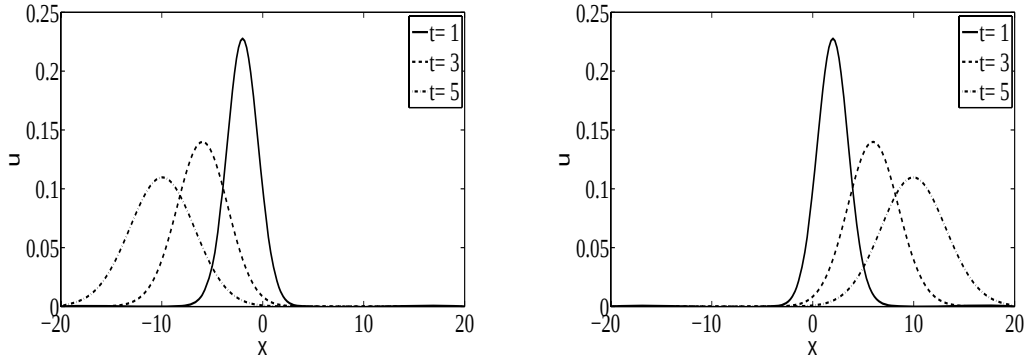

Figure 2.1: Approximate solution obtained for Example 2.3.3 for $\theta = 0$ and $\Delta t = \Delta x = 0.2$. Left: $P = -2$. Right: $P = 2$.

Example 2.3.4. Let us consider now the problem with P non-constant

$$\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}(P(x)u(x,t)) + \frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in \mathbb{R}, t > 0,$$

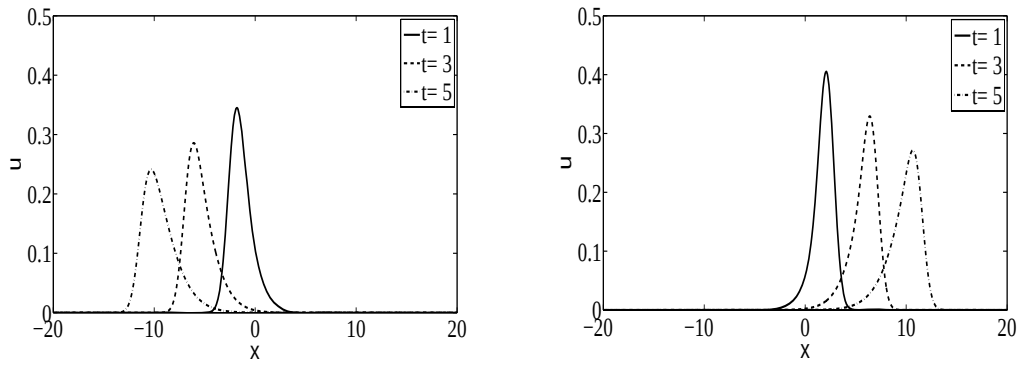

Figure 2.2: Approximate solution obtained for Example 2.3.3 for $\theta = 0.2$, $\Delta x = 0.2$, $\Delta t = 0.002$. Left: $P = -2$. Right: $P = 2$.

with

$$V(x;\alpha) = \frac{1}{J_0(\mathbf{i}\alpha)} e^{\alpha\cos(x)} - 1$$

and

$$P(x) = -\frac{dV}{dx} = \alpha\sin(x)\frac{1}{J_0(\mathbf{i}\alpha)} e^{\alpha\cos(x)},$$

where i is the imaginary unit and J_0 is the Bessel function of the first kind of zero order given by the series [95]

$$J_0(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m!)^2}\left(\frac{x}{2}\right)^{2m}.$$

We consider the initial conditions

$$u(x,0) = \frac{1}{L\sqrt{\pi}} e^{-x^2/L^2}, \qquad \frac{\partial u}{\partial t}(x,0) = 0$$

and the boundary conditions

$$\lim_{x\to-\infty} u(x,t) = 0, \qquad \lim_{x\to+\infty} u(x,t) = 0.$$

To see how the potential field $V(x)$ affects the solution and also the performance of the numerical method, we consider the same problem for $\alpha = 1$ and for $\alpha = 16$. To simulate the results for $\alpha = 1$ we can use the Crank-Nicolson method, as shown in Figure 2.3, but to deal with the type of solution that comes out for $\alpha = 16$ this numerical method can not be used since it presents oscillations, as we can see in Figure 2.4. The oscillations

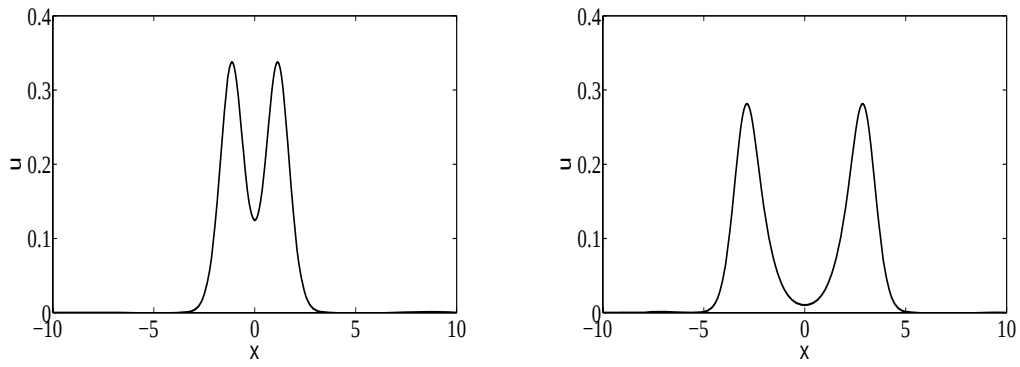

Figure 2.3: Approximate solution obtained for Example 2.3.4 with $\alpha = 1$, $\Delta t = \Delta x = 0.06$. Left: $t = 1$. Right: $t = 3$.

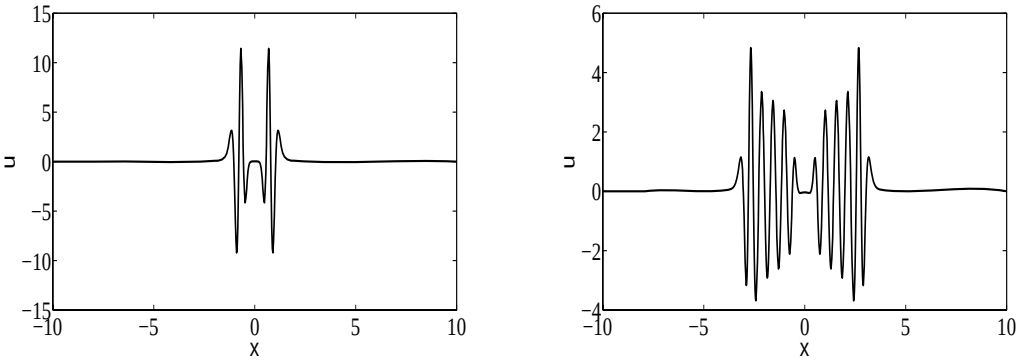

Figure 2.4: Approximate solution obtained for Example 2.3.4 with $\alpha = 16$, $\Delta t = \Delta x = 0.06$. Left: $t = 1$. Right: $t = 3$.

remain even for small values of the time step. With the need to solve this problem, since it is very interesting in the physical point of view [4], a new numerical strategy is required. To also avoid discretization in time, allowing performances for long times, we introduce in the next chapter some alternative numerical methods. All of them are constructed based on the Laplace transform which means that, in the end, an inversion has to be made to obtain the approximate solution.

Chapter 3

Laplace transform numerical methods

Taking in consideration that we are interested in the long time behavior of solutions of equation (2.1), the idea of using the Laplace transform in time arises naturally. Application of the Laplace transform is suitable in many problems but implies its inversion to recover the final approximate solution in time. To this end, we describe a numerical inverse Laplace transform algorithm based on continued fractions [45]. The spatial discretization is implemented with three different schemes. First, we present a scheme that uses centered difference approximations after the application of the Laplace transform; we call it the Laplace transform finite difference method. However, there are problems where the Laplace transform finite difference scheme presents oscillations, namely in the presence of discontinuous initial conditions. Some of these oscillations disappear when we use a very small space step but, in some cases, they still remain in the numerical solution. Therefore, to overcome this disadvantage, we present another method for the spatial discretization that uses a finite volume formulation [14]. Despite the good results obtained with this method, it also presents oscillations when one need to deal with large values of the $|P|$ parameter. Hence, a third spatial discretization is used to suppress this handicap: a piecewise linearized method [79]. Numerical tests are presented to compare the

performance of these three different spatial discretizations. The rate of convergence and the computational cost of methods based on the Laplace transform are compared with the Crank-Nicolson method. Finally, the end of the chapter describes a problem that contains a symmetric periodic potential field.

3.1 The Laplace transform inversion

The Laplace transform has been used in several works, such as [14]-[17], to remove the time dependent terms and obtain an ordinary differential equation in space variable. Using this technique and combining it with an appropriate spatial discretization method has some advantages. First, we can compute the approximate solution for long times accurately and quickly and we do not need to do computations in the time domain using time iterations. Secondly, it also avoids undesirable numerical oscillations that are related with the bad choices of time steps. Any iterative numerical method would take too long to compute the solution for similar times, due to the increased computational effort for discretizing in time, even when we consider an unconditionally implicit numerical method which will allow large time steps. To solve problem (2.1)-(2.3) we apply also this technique that can be separated in three steps. First, we apply the Laplace transform to the partial differential equation and boundary conditions, in order to remove the time dependent terms, yielding an ordinary differential equation in the space variable that depends on the Laplace parameter. Secondly, we solve the ordinary differential equation by an appropriate discretization in space. Depending on the purposes and the specifications of problem (2.1)-(2.3), we use different numerical methods for the spatial discretization. At last, a numerical inverse Laplace transform algorithm is used to obtain the final approximate solution in time and space.

The Laplace transform of the real valued function $u(x,t)$ is defined by

$$\widetilde{u}\,(x,s) = \int_0^{+\infty} \mathrm{e}^{-st} u\,(x,t)\,dt, \tag{3.1}$$

where s is a complex variable. If we apply the Laplace transform (3.1) to equation (2.1) we obtain

$$\int_0^{+\infty} \left(\theta \frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) \right) \mathrm{e}^{-st}\, dt = (\theta s^2 + s)\widetilde{u}(x,s) - (1+\theta s)u_0(x) - u_1(x)$$

and

$$\int_0^{+\infty} \left(-\frac{\partial}{\partial x}\left(P(x)u(x,t)\right) + D\frac{\partial^2 u}{\partial x^2}(x,t) \right) \mathrm{e}^{-st}\, dt = -\frac{d}{dx}(P(x)\widetilde{u}(x,s)) + D\frac{d^2\widetilde{u}}{dx^2}(x,s).$$

Thus, we obtain the ordinary differential equation

$$\frac{d^2\widetilde{u}}{dx^2}(x,s) - \lambda_s^2 \widetilde{u}(x,s) - \frac{d}{dx}\left(\frac{P(x)}{D}\widetilde{u}(x,s) \right) = -\frac{u_0(x)}{D}(1+\theta s) - \frac{u_1(x)}{D}, \quad (3.2)$$

where $\lambda_s = \left((\theta s^2 + s)/D\right)^{1/2}$. In the particular case where P is a constant, we do not need a spatial discretization and can directly apply the inverse Laplace transform algorithm. In fact, if P is a constant the exact solution of (3.2) can be written in the form

$$\widetilde{u}(x,s) = A\mathrm{e}^{\nu_s^+ x} + B\mathrm{e}^{\nu_s^- x} + \tilde{u}_p(x,s)$$

for

$$\nu_s^{\pm} = \frac{P}{2D} \pm \sqrt{\left(\frac{P}{2D}\right)^2 + \lambda_s^2} \quad (3.3)$$

and $\tilde{u}_p(x,s)$ is a particular solution. The constants A, B are to be determined from the following boundary conditions, derived from (2.3), $\widetilde{u}(a,s) = \widetilde{f}(s)$ and $\widetilde{u}(b,s) = \widetilde{g}(s)$. The procedure for inverting $\widetilde{u}(x,s)$ is described in detail in the next section. In some cases we need a spatial discretization to solve (3.2). Figure 3.1 is a schematic explanation that clarifies when we need to discretize in space in order to obtain the solution $\widetilde{u}(x,s)$. The numerical methods used for that purpose are described in Section 3.2.

3.1.1 The inverse Laplace transform algorithm

The principal difficulty in using Laplace transforms is to find their inverses. Although there exist extensive tables of Laplace transforms and

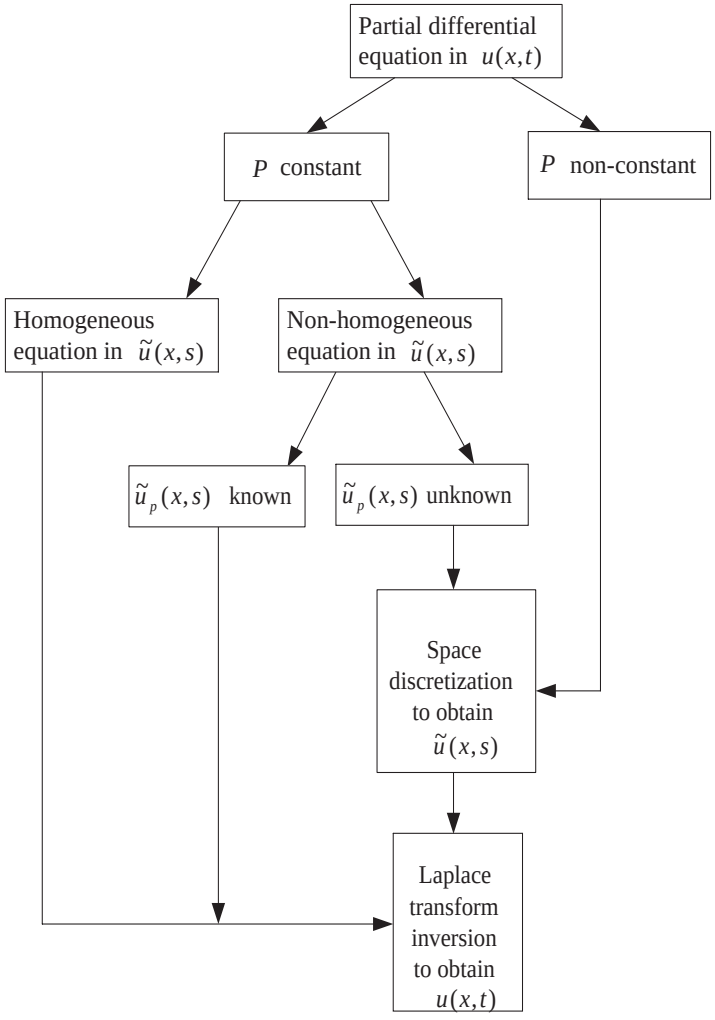

Figure 3.1: Schematic explanation.

their inverses, it is highly desirable to have numerical methods to obtain an approximation of the inversion solution. Based on a different approach from some conventional inverse transform algorithms [20, 46], a numerical inverse Laplace transform algorithm was developed in [2]. This algorithm serves the purpose of the present study and, therefore, a computer code has been written following it to invert the solution numerically into the time domain. Based on the Fourier series expansion and the continued fractional approaches, this algorithm is described in [4].

Let $\widetilde{U}_i(s)$, $i = 0, \ldots, N$, represent the approximation of $\widetilde{u}(x_i, s)$ in the Laplace transform domain. In this section, we determine an approximate solution $U_i(t)$ from $\widetilde{U}_i(s)$ by using a Laplace inversion numerical method. For the sake of clarity we omit the index i, denoting $\widetilde{U}_i(s)$ by $\widetilde{U}(s)$. An exact inverse Laplace transform of $\widetilde{U}\,(s)$ into $\overline{U}\,(t)$ is given through the Bromwich integral [63]

$$\overline{U}\,(t) = \frac{1}{2\pi\mathbf{i}} \int_{\beta-\mathbf{i}\infty}^{\beta+\mathbf{i}\infty} \mathrm{e}^{st}\widetilde{U}\,(s)\, ds, \tag{3.4}$$

where $\mathbf{i}$ is the imaginary unit and β is such that the contour of integration is to the right hand side of any singularity of $\widetilde{U}\,(s)$. If we consider $s = \beta + \mathbf{i}\omega$ [1, 4, 63] and $ds = \mathbf{i}d\omega$ we can write

$$\begin{aligned}
\overline{U}\,(t) &= \frac{1}{2\pi\mathbf{i}} \int_{-\infty}^{+\infty} \mathrm{e}^{\beta t+\mathbf{i}\omega t}\widetilde{U}\,(s)\,\mathbf{i}\,d\omega = \frac{1}{2\pi}\mathrm{e}^{\beta t} \int_{-\infty}^{+\infty} \mathrm{e}^{\mathbf{i}\omega t}\widetilde{U}\,(s)\, d\omega \\
&= \frac{1}{2\pi}\mathrm{e}^{\beta t} \int_{-\infty}^{+\infty} (\cos\omega t + \mathbf{i}\sin\omega t)\,\widetilde{U}\,(s)\, d\omega \\
&= \frac{1}{2\pi}\mathrm{e}^{\beta t} \int_{-\infty}^{+\infty} \mathrm{Re}\left\{\widetilde{U}\,(s)\right\}\cos\omega t - \mathrm{Im}\left\{\widetilde{U}\,(s)\right\}\sin\omega t\, d\omega \\
&\quad + \frac{1}{2\pi}\mathrm{e}^{\beta t}\mathbf{i} \int_{-\infty}^{+\infty} \mathrm{Im}\left\{\widetilde{U}\,(s)\right\}\cos\omega t + \mathrm{Re}\left\{\widetilde{U}\,(s)\right\}\sin\omega t\, d\omega.
\end{aligned} \tag{3.5}$$

Since

$$\begin{aligned}
\widetilde{U}\,(s) &= \int_0^{+\infty} \mathrm{e}^{-st}\overline{U}\,(t)\, dt = \int_0^{+\infty} \mathrm{e}^{-(\beta+\mathbf{i}\omega)t}\overline{U}\,(t)\, dt \\
&= \int_0^{+\infty} \mathrm{e}^{-\beta t}\overline{U}\,(t)\,(\cos\omega t - \mathbf{i}\sin\omega t)\, dt,
\end{aligned}$$

we have

$$\mathrm{Re}\left\{\widetilde{U}\,(s)\right\} = \int_0^{+\infty} \mathrm{e}^{-\beta t}\overline{U}\,(t)\cos\omega t\, dt \tag{3.6}$$

and

$$\mathrm{Im}\left\{\widetilde{U}\left(s\right)\right\} = -\int_0^{+\infty} \mathrm{e}^{-\beta t}\overline{U}\left(t\right)\sin\omega t\, dt. \tag{3.7}$$

From (3.6) and (3.7) the second integral of (3.5) vanishes and we get

$$\overline{U}\left(t\right) = \frac{1}{2\pi}\mathrm{e}^{\beta t}\int_{-\infty}^{+\infty} \mathrm{Re}\left\{\widetilde{U}\left(s\right)\right\}\cos\omega t - \mathrm{Im}\left\{\widetilde{U}\left(s\right)\right\}\sin\omega t\, d\omega. \tag{3.8}$$

It is easy to see from (3.6) that $\mathrm{Re}\left\{\widetilde{U}\left(s\right)\right\}$ is an even function. Therefore, $\mathrm{Re}\left\{\widetilde{U}\left(s\right)\right\}\cos\omega t$ is also an even function. Furthermore, from (3.7) we see that the function $\mathrm{Im}\left\{\widetilde{U}\left(s\right)\right\}$ is odd and the product $\mathrm{Im}\left\{\widetilde{U}\left(s\right)\right\}\sin\omega t$ is an even function. This means that the function of the integral (3.8) is even and we can write

$$\begin{aligned}
\overline{U}\left(t\right) &= \frac{1}{\pi}\mathrm{e}^{\beta t}\int_0^{+\infty} \mathrm{Re}\left\{\widetilde{U}\left(s\right)\right\}\cos\omega t - \mathrm{Im}\left\{\widetilde{U}\left(s\right)\right\}\sin\omega t\, d\omega \\
&= \frac{1}{\pi}\mathrm{e}^{\beta t}\int_0^{+\infty} \mathrm{Re}\left\{\widetilde{U}\left(s\right)\left(\cos\omega t + \mathbf{i}\sin\omega t\right)\right\}\, d\omega.
\end{aligned}$$

Finally,

$$\overline{U}\left(t\right) = \frac{1}{\pi}\mathrm{e}^{\beta t}\int_0^{+\infty} \mathrm{Re}\left\{\widetilde{U}\left(s\right)\mathrm{e}^{\mathbf{i}\omega t}\right\}\, d\omega.$$

The integral is now evaluated through the trapezoidal rule, with step size π/T, and we obtain

$$\overline{U}(t) = \frac{1}{T}\mathrm{e}^{\beta t}\left[\frac{\widetilde{U}\left(\beta\right)}{2} + \sum_{k=1}^{\infty}\mathrm{Re}\left\{\widetilde{U}\left(\beta + \frac{\mathbf{i}k\pi}{T}\right)\mathrm{e}^{\frac{\mathbf{i}k\pi t}{T}}\right\}\right] - E_T, \tag{3.9}$$

for $0 < t < 2T$ and where E_T is the discretization error. It is known that the infinite series in this equation converges very slowly. To accelerate the convergence, we apply the quotient-difference algorithm, proposed in [2], and also used in [4, 73], to calculate the series obtained in (3.9) by the rational approximation in the form of a continued fraction. With the purpose of applying this scheme to (3.9), set

$$\sum_{k=0}^{\infty}\widetilde{U}_k z^k = \frac{\widetilde{U}\left(\beta\right)}{2} + \sum_{k=1}^{\infty}\widetilde{U}\left(\beta + \frac{\mathbf{i}k\pi}{T}\right)\mathrm{e}^{\frac{\mathbf{i}k\pi t}{T}}, \tag{3.10}$$

where $\widetilde{U}_0 = \frac{\widetilde{U}(\beta)}{2}$, $\widetilde{U}_k = \widetilde{U}\left(\beta + \frac{\mathbf{i}k\pi}{T}\right), k = 1, 2, \cdots$, and $z = \mathrm{e}^{\mathbf{i}\pi t/T}$.

Under some conditions we can always associate a continued fraction to a given power series [45]. We denote $v\left(z\right)$ the continued fraction

$$v\left(z\right) = d_0/\left(1 + d_1 z/\left(1 + d_2 z/\left(1 + \cdots\right)\right)\right) \tag{3.11}$$

associated to the power series in (3.10) and write

$$v\left(z\right) = \frac{\widetilde{U}\left(\beta\right)}{2} + \sum_{k=1}^{\infty} \widetilde{U}\left(\beta + \frac{\mathbf{i}k\pi}{T}\right) \mathrm{e}^{\frac{\mathbf{i}k\pi t}{T}}. \tag{3.12}$$

The coefficients d_p's of (3.11) are obtained by recurrence relations from the coefficients $\widetilde{U}\left(\beta + \frac{\mathbf{i}k\pi}{T}\right)$, that is, let

$$e_0^{(k)} = 0, \qquad q_1^{(k)} = \widetilde{U}_{k+1}/\widetilde{U}_k, \qquad k = 0, 1, \ldots.$$

From the recurrence relations

$$e_p^{(k)} + q_p^{(k)} = e_{p-1}^{(k+1)} + q_p^{(k+1)}, \qquad k = 0, 1, \ldots, \qquad p = 1, 2, \ldots,$$

$$q_{p+1}^{(k)} e_p^{(k)} = q_p^{(k+1)} e_p^{(k+1)}, \qquad k = 0, 1, \ldots, \qquad p = 1, 2, \ldots,$$

we obtain the coefficients d_p's,

$$d_0 = \widetilde{U}_0, \qquad d_{2p-1} = -q_p^{(0)}, \qquad d_{2p} = -e_p^{(0)}, \qquad p = 1, 2, \ldots.$$

This way, the algorithm computes the coefficients of the continued fraction needed for the inversion process. Let the M-th partial fraction $v(z, M)$ be

$$v(z, M) = d_0/\left(1 + d_1 z/\left(1 + \cdots + d_M z\right)\right).$$

We obtain $v(z, M) = A_M/B_M$ using the following recurrence relations

$$A_p = A_{p-1} + d_p z A_{p-2}$$

$$B_p = B_{p-1} + d_p z B_{p-2}, \qquad p = 1, 2, \ldots,$$

with $A_{-1} = 0$, $B_{-1} = 1$, $A_0 = d_0$, $B_0 = 1$. Therefore

$$v\left(z\right) = v\left(z, M\right) + E_F^M,$$

where E_F^M is the truncation error. Then

$$\overline{U}\,(t) = \frac{1}{T}\mathrm{e}^{\beta t}\mathrm{Re}\left\{v\,(z,M) + E_F^M\right\} - E_T.$$

In the next section an error analysis is made to see how we can control these errors. The approximation for $\overline{U}\,(t)$ is denoted by $U(t)$ and given by

$$U\,(t) = \frac{1}{T}\mathrm{e}^{\beta t}\mathrm{Re}\left\{v\,(z,M)\right\}.$$

3.1.2 Convergence of the inverse Laplace transform algorithm

We discuss in this section some convergence aspects of the inverse Laplace transform algorithm. Let $\widetilde{U}_i(s),\ i = 0,\ldots,N$ represent the approximation of $\widetilde{u}\,(x_i,s)$ in the Laplace transform domain and denote by $\widetilde{E}_S$ the error that is associated to the spatial discretization,

$$\widetilde{u}(x_i,s) = \widetilde{U}_i(s) + \widetilde{E}_S(x_i,s). \tag{3.13}$$

The Laplace inverse transform of $\widetilde{U}_i(s)$, as described in the previous section, is the solution

$$\overline{U}_i\,(t) = \frac{1}{T}\mathrm{e}^{\beta t}\mathrm{Re}\left\{v\,(z,M_i) + E_F^M(x_i,t)\right\} - E_T(x_i,t), \tag{3.14}$$

where E_T is the error associated with the trapezoidal approximation and E_F^M is the truncation error associated to the continued fraction. Note that for each x_i the algorithm chooses an M_i and therefore for each x_i we have a different value of the approximation of the continued fraction, $v\,(z,M_i)$. Therefore, from (3.13)-(3.14) we have

$$u(x_i,t) = \frac{1}{T}\mathrm{e}^{\beta t}\mathrm{Re}\left\{v\,(z,M_i) + E_F^M(x_i,t)\right\} - E_T(x_i,t) + E_S(x_i,t),$$

where $E_S(x_i,t)$ is the inverse Laplace transform of the error $\widetilde{E}_S(x_i,s)$.

Leaving for now the spatial discretization error, we have two errors which controll the discretization obtained from the Laplace transform inversion. The first error, E_T, that comes from the integral approximation using the trapezoidal rule, according to Crump [20], is

$$E_T = \sum_{n=1}^{\infty}\mathrm{e}^{-2n\beta T}u(x_i,2nT+t).$$

Assume that our function is bounded such as $|u(x_i,t)| \leq \mathrm{e}^{\sigma t}$, for all x_i. In this case the Laplace transform $\widetilde{u}(s)$ is defined for $\mathrm{Re}(s) > \sigma$ which means that β on (3.4) must be $\beta > \sigma$. Thus, the error can be bounded by

$$E_T \leq \mathrm{e}^{\sigma t} \sum_{n=1}^{\infty} \mathrm{e}^{-2nT(\beta-\sigma)} = \frac{\mathrm{e}^{\sigma t}}{\mathrm{e}^{2T(\beta-\sigma)}-1}, \quad 0 < t < 2T.$$

It follows that by choosing β sufficiently larger than σ, we can make E_T as small as desired. For practical purposes and in order to choose a convenient β we replace the previous inequality by

$$E_T \leq \mathrm{e}^{\sigma t - 2T(\beta-\sigma)}, \quad 0 < t < 2T.$$

If we want to have the bound $E_T \leq b_T$ then by applying the logarithm in both sides of the previous inequality we have

$$\beta \geq \sigma \frac{2T+t}{2T} - \frac{1}{2T} \ln\left(b_T\right).$$

Assuming $\sigma \geq 0$ we can write

$$\beta \geq \sigma - \frac{\ln\left(b_T\right)}{2T}.$$

We will consider $\sigma = 0$. In practice the trapezoidal error E_T is controlled by the parameter β we choose.

The second error, E_F^M, comes from the approximation of the continued fraction given by (3.12). This error is controlled by imposing a tolerance TOL such as

$$\left|v\left(z,M\right) - v\left(z,M-1\right)\right| < TOL,$$

in order to get the approximation $U_i(t)$ given by

$$U_i(t) = \frac{1}{T}\mathrm{e}^{\beta t}\mathrm{Re}\{v\left(z,M_i\right)\},$$

where M_i changes according to which x_i we are considering.

In order to understand better how to control the trapezoidal error with the parameter β and how the tolerance TOL affects the error, we present a test example which is an analytically exactly solvable model.

Example 3.1.1. We assume P constant and $\theta = 0$*:*

$$\frac{\partial u}{\partial t}(x,t) = -P\frac{\partial u}{\partial x}(x,t) + D\frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in]0,\infty[\,, t > 0. \tag{3.15}$$

The initial condition is $u(x,0) = 0$ and the boundary conditions are

$$u(0,t) = u_0, \qquad \lim_{x\to+\infty} u(x,t) = 0. \tag{3.16}$$

We choose this test example for two reasons: firstly, equation (3.15) can be analytically exactly solved by first applying the time Laplace transform and then through the inverse Laplace transform. Secondly, this example is chosen to analyze the convergence aspects of the Laplace inversion algorithm without spatial discretization. If we apply the Laplace transform to this problem we obtain the ordinary differential equation

$$\frac{d^2\widetilde{u}}{dx^2}(x,s) - \frac{P}{D}\frac{d\widetilde{u}}{dx}(x,s) - \frac{s}{D}\widetilde{u}(x,s) = 0,$$

where s is a complex variable. Since P is constant, by using the boundary conditions we obtain the solution of this equation as

$$\widetilde{u}(x,s) = u_0\frac{1}{s}\mathrm{e}^{\left[P/2D - \sqrt{(P/2D)^2 + s/D}\right]x}. \tag{3.17}$$

The analytical solution is given by [2]

$$u\,(x,t) = \frac{u_0}{2}\left(\operatorname{erfc}\left[\frac{x-Pt}{2\sqrt{Dt}}\right] + \mathrm{e}^{Px/D}\operatorname{erfc}\left[\frac{x+Pt}{2\sqrt{Dt}}\right]\right), \tag{3.18}$$

where $\mathrm{erfc(x)}$ is the error function complement [18], $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, being $\operatorname{erf}(x)$ the error function defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-\rho^2}\,d\rho.$$

In Figures 3.2 and 3.3, for $u_0 = 1$, $P = 2$, $t = 1$ and $0 \le x \le 12$, we plot the following errors,

$$E_F = \max_{1\le i\le N-1} |v(z,M_i) - v(z,M_i - 1)| \quad \text{and} \quad E_G = \max_{1\le i\le N-1} |u(x_i,t) - U_i(t)|.$$

We choose the interval $0 \le x \le 12$ in order to avoid the influence of the right numerical boundary condition in the numerical computations, that in this case is $u(12,t) = 0$.

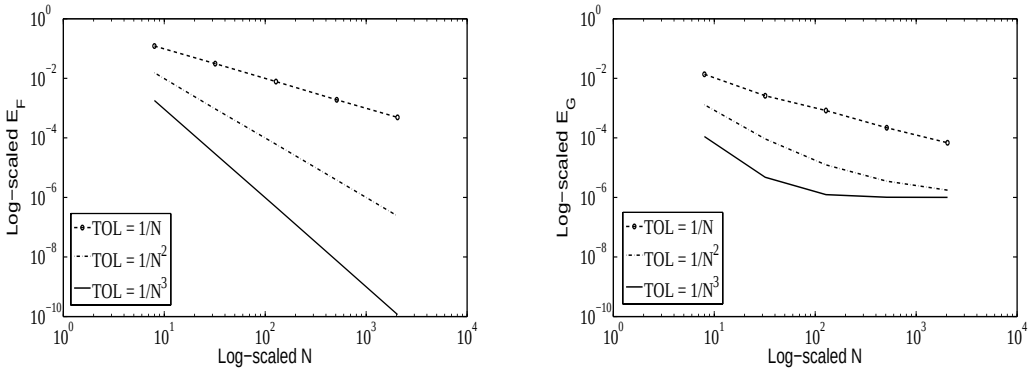

Figure 3.2: Error E_F and E_G for Example 3.1.1 for $u_0 = 1$, $P = 2$, $t = 1$, $0 \leq x \leq 12$ and $\beta = -\ln(10^{-6})/2T$ with $T = 20$ and different values of TOL. The global error is controlled by the parameter β.

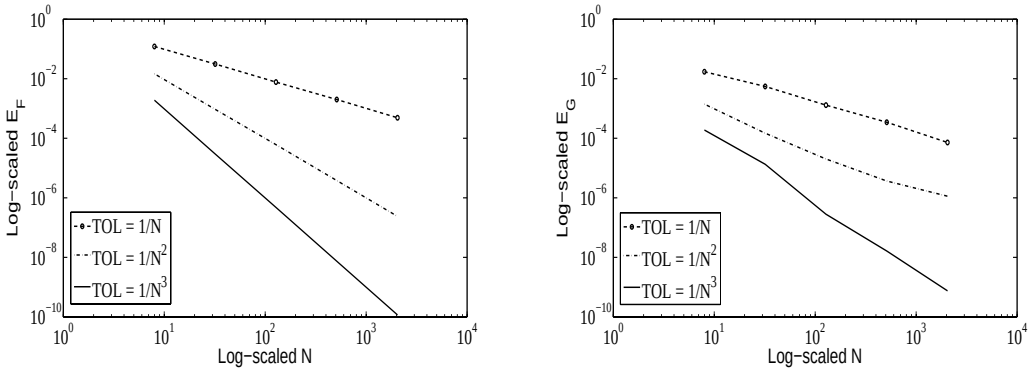

Figure 3.3: Error E_F and E_G for Example 3.1.1 for $u_0 = 1$, $P = 2$, $t = 1$, $0 \leq x \leq 12$ and $\beta = -\ln(10^{-10})/2T$ with $T = 20$ and different values of TOL. The parameter β is chosen such that the global error is not affected.

The error E_F is related with the error E_F^M since we can control E_F^M by controlling E_F with the tolerance TOL. Figure 3.2 and Figure 3.3 illustrate how the parameter β, defined by $\beta = -\ln(10^{-6})/2T$ in Figure 3.2 and $\beta = -\ln(10^{-10})/2T$ in Figure 3.3, affects the global convergence. Note that in Figure 3.2 the precision does not go further than 10^{-6}. The global error of Figure 3.2 and Figure 3.3 is not affected by the spatial error E_S since we apply the Laplace inversion algorithm directly in (3.17).

3.2 Spatial discretization methods

In this section we will deal with the time derivatives through time Laplace transform. We use three different approaches for the spatial discretization described in Sections 3.2.1, 3.2.2 and 3.2.3. After spatial discretization we obtain the linear system

$$K(s)\,\widetilde{U}(s) = \widetilde{b}(s), \tag{3.19}$$

where $K(s) = [K_{i,j}(s)]$ is a band matrix of size $(N-1)\times(N-1)$, the unknown vector is $\widetilde{U}(s) = [\widetilde{U}_1(s), \ldots, \widetilde{U}_{N-1}(s)]^T$ and $\widetilde{b}(s)$ contains source terms and boundary conditions. In this study we obtain a matrix K with bandwidth three. We provide the three spatial discretizations by giving the entries of the matrix K and the vector $\widetilde{b}$. At last, the algorithm described in Section 3.1.1 is used to perform the Laplace inversion and get the final approximate solution $u(x,t)$.

The errors that come from the numerical inversion of Laplace transform were already described in Section 3.1.2. We will prove further on that the spatial discretization error $\widetilde{E}_S(x_i,s)$, defined in (3.13), has a truncation error of second order for the three spatial discretizations.

3.2.1 A Laplace transform finite difference method

The Laplace transform finite difference method (Laplace-FD) consists of applying first the Laplace transform to remove the time dependent terms from equation (2.1) and boundary conditions (2.3). We obtain an ordinary differential equation which is discretized using a finite difference method. We consider centered differences to approximate the first derivative and the second derivative of equation (3.2). For a fixed s, the finite difference scheme is given by

$$\begin{aligned}&\frac{\widetilde{U}_{i-1}(s) - 2\widetilde{U}_i(s) + \widetilde{U}_{i+1}(s)}{\Delta x^2} - \lambda_s^2\widetilde{U}_i(s) - \frac{1}{D}\frac{P_{i+1}\widetilde{U}_{i+1}(s) - P_{i-1}\widetilde{U}_{i-1}(s)}{2\Delta x}\\ &\quad = -\frac{1}{D}\left(u_0(x_i)(1+\theta s) + u_1(x_i)\right),\end{aligned} \tag{3.20}$$

for $i = 1, \ldots, N-1$, where $P_i = P(x_i)$.

The matrix $K(s)$ has entries of the form

$$\begin{aligned} K_{i,i-1}(s) &= \frac{1}{\Delta x^2} + \frac{P_{i-1}}{2D\Delta x}, \\ K_{i,i}(s) &= -\frac{2}{\Delta x^2} - \lambda_s^2, \\ K_{i,i+1}(s) &= \frac{1}{\Delta x^2} - \frac{P_{i+1}}{2D\Delta x}, \end{aligned} \tag{3.21}$$

and $\widetilde{b}\,(s)$ contains boundary conditions being represented by

$$\widetilde{b}\,(s) = \frac{1}{D}\begin{bmatrix} -u_0(x_1)(1+\theta s) - u_1(x_1) \\ -u_0(x_2)(1+\theta s) - u_1(x_2) \\ \vdots \\ -u_0(x_{N-2})(1+\theta s) - u_1(x_{N-2}) \\ -u_0(x_{N-1})(1+\theta s) - u_1(x_{N-1}) \end{bmatrix} - \begin{bmatrix} K_{1,0}(s)\widetilde{U}_0(s) \\ 0 \\ \vdots \\ 0 \\ K_{N-1,N}(s)\widetilde{U}_N(s) \end{bmatrix}.$$

We now prove the Laplace-FD method is second order accurate for the spatial discretization and, for simplicity, we assume $D = 1$.

Proposition 3.2.1. *For the finite difference discretization we have*

$$\begin{aligned} &K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\ &\quad = \frac{d^2\widetilde{u}_i}{dx^2}(s) - \lambda_s^2\widetilde{u}_i(s) - \frac{d}{dx}(P\widetilde{u})_i(s) + \mathcal{O}(\Delta x^2), \end{aligned} \tag{3.22}$$

where the $K's$ are defined by (3.21).

Proof: Let us substitute the exact solution $\widetilde{u}(x,s)$ in the numerical method (3.20),

$$K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s) + u_0(x_i)(1+\theta s) + u_1(x_i) = 0.$$

Now, for a fixed s, we do Taylor expansions of the functions $\widetilde{u}_{i-1}$, $\widetilde{u}_{i+1}$, P_{i-1} and P_{i+1}, around the point x_i. We have

$$K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s)$$

$$= \frac{d^2\widetilde{u}_i}{dx^2}(s) - \lambda_s^2\widetilde{u}(s) - \frac{d}{dx}(P\widetilde{u})_i(s)$$

$$+ \left[-\frac{1}{6}P_i'''\widetilde{u}_i(s) - \frac{1}{2}P_i''\frac{d\widetilde{u}_i}{dx}(s) - \frac{1}{2}P_i'\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{1}{6}P_i\frac{d^3\widetilde{u}_i}{dx^3}(s) + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\right]\Delta x^2$$

$$+\mathcal{O}(\Delta x^3),$$

where P', P'' and P''' denote the derivatives of P.

■

From this result we can conclude the Laplace-FD method is second order accurate in space since the truncation error is of order $\mathcal{O}(\Delta x^2)$.

3.2.2 A Laplace transform finite volume method

Although the Laplace transform finite difference method described in the previous section has second order accuracy, it presents numerical oscillations for some initial conditions as we will see in some numerical tests. Therefore, we introduce in this section a finite volume method, presented in [14], for the spatial discretization to suppress those oscillations from the vicinity of sharp discontinuities. We will apply it to our model problem (2.1)-(2.3) considering non-trivial initial conditions and different values of the parameter P, for both parabolic ($\theta = 0$) and hyperbolic ($\theta \neq 0$) equations. The combination of Laplace transform with the finite volumes generates the Laplace transform finite volume method (Laplace-FV).

The discretization consists of using the finite volume formulation by integrating in x the ordinary differential equation (3.2) in the i-th control volume $[x_i - \Delta x/2, x_i + \Delta x/2]$,

$$\int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} \left[\frac{d^2\widetilde{u}\,(x,s)}{dx^2} - \lambda_s^2\widetilde{u}\,(x,s) - \frac{d}{dx}\left(\frac{P(x)}{D}\widetilde{u}\,(x,s)\right)\right]dx$$

$$= -\frac{1}{D}\int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} ((1+\theta s)u_0(x) + u_1(x))dx. \tag{3.23}$$

We compute the integral on the right hand side by the midpoint rule, that is,

$$\int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} ((1+\theta s)u_0(x)+u_1(x))dx \simeq \Delta x\left[(1+\theta s)u_0(x_i)+u_1(x_i)\right].$$

We can write the integral on the left hand side as

$$\left[\frac{d}{dx}\widetilde{U}(x,s)\right]_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} - \lambda_s^2\left[\int_{x_i-\frac{\Delta x}{2}}^{x_i}\widetilde{U}(x,s)\,dx + \int_{x_i}^{x_i+\frac{\Delta x}{2}}\widetilde{U}(x,s)\,dx\right]$$

$$-\frac{P(x_i+\Delta x/2)}{D}\widetilde{U}(x_i+\Delta x/2,s) + \frac{P(x_i-\Delta x/2)}{D}\widetilde{U}(x_i-\Delta x/2,s). \quad (3.24)$$

As suggested in [14], for $x\in[x_i,x_{i+1}]$, $i=0,\ldots,N-1$, we approximate $\widetilde{U}(x,s)$ by the following combination of hyperbolic functions,

$$\widetilde{U}(x,s) = \frac{\sinh(\lambda_s(x_{i+1}-x))}{\sinh(\lambda_s\Delta x)}\widetilde{U}_i(s) + \frac{\sinh(\lambda_s(x-x_i))}{\sinh(\lambda_s\Delta x)}\widetilde{U}_{i+1}(s),$$

where $\widetilde{U}_i(s)$, $i=0,\ldots,N$, represents the approximation of $\widetilde{U}(x_i,s)$ in the Laplace transform domain. Substituting this approximation in (3.24) yields

$$\begin{aligned}
&\frac{\lambda_s}{\sinh(\lambda_s\Delta x)}\left[\widetilde{U}_{i-1}(s) - 2\cosh(\lambda_s\Delta x)\widetilde{U}_i(s) + \widetilde{U}_{i+1}(s)\right]\\
-\;&\frac{P(x_i+\Delta x/2)}{D}\frac{\sinh(\lambda_s\Delta x/2)}{\sinh(\lambda_s\Delta x)}\left(\widetilde{U}_i(s)+\widetilde{U}_{i+1}(s)\right)\\
+\;&\frac{P(x_i-\Delta x/2)}{D}\frac{\sinh(\lambda_s\Delta x/2)}{\sinh(\lambda_s\Delta x)}\left(\widetilde{U}_{i-1}(s)+\widetilde{U}_i(s)\right).
\end{aligned}$$

Finally, the evaluation of (3.23) produces the following discretized equations, for $i=1,\ldots,N-1$,

$$\begin{aligned}
&K_{i,i-1}(s)\widetilde{U}_{i-1}(s) + K_{i,i}(s)\widetilde{U}_i(s) + K_{i,i+1}(s)\widetilde{U}_{i+1}(s)\\
&\qquad = -\frac{\sinh(\lambda_s\Delta x)}{D\lambda_s}\Delta x\left[(1+\theta s)u_0(x_i)+u_1(x_i)\right], \qquad (3.25)
\end{aligned}$$

for

$$\begin{aligned} K_{i,i-1}(s) &= 1 + P_{i-1/2}\frac{\sinh(\lambda_s \Delta x/2)}{D\lambda_s}, \\ K_{i,i}(s) &= -2\cosh(\lambda_s \Delta x) - \left(P_{i+1/2} - P_{i-1/2}\right)\frac{\sinh(\lambda_s \Delta x/2)}{D\lambda_s}, \\ K_{i,i+1}(s) &= 1 - P_{i+1/2}\frac{\sinh(\lambda_s \Delta x/2)}{D\lambda_s}, \end{aligned} \tag{3.26}$$

where $P_{i\pm 1/2} = P(x_i \pm \Delta x/2)$. The vector that contains boundary terms is given by

$$\widetilde{b}(s) = R_{x,s}\begin{bmatrix} (1+\theta s)u_0(x_1) + u_1(x_1) \\ (1+\theta s)u_0(x_2) + u_1(x_2) \\ \vdots \\ (1+\theta s)u_0(x_{N-2}) + u_1(x_{N-2}) \\ (1+\theta s)u_0(x_{N-1}) + u_1(x_{N-1}) \end{bmatrix} - \begin{bmatrix} K_{1,0}(s)\widetilde{U}_0(s) \\ 0 \\ \vdots \\ 0 \\ K_{N-1,N}(s)\widetilde{U}_N(s) \end{bmatrix},$$

where $R_{x,s} = -\frac{\Delta x \sinh(\lambda_s \Delta x)}{D\lambda_s}$. Thus, equation (3.25) can be written in the matrix form (3.19) where matrix K and vector $\widetilde{b}$ are defined by the entries given above.

We will assume the diffusion coefficient $D = 1$ to show the Laplace-FV method has a local truncation error of second order for the spatial discretization.

Proposition 3.2.2. *For the finite volume discretization (3.25) we have*

$$\begin{aligned} &K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\ &\quad + \frac{\sinh(\lambda_s \Delta x)}{\lambda_s}\Delta x\left[(1+\theta s)u_0(x_i) + u_1(x_i)\right] \\ &= \frac{d^2\widetilde{u}_i}{dx^2}(s) - \lambda_s^2\widetilde{u}_i(s) - \frac{d}{dx}(P\widetilde{u})_i(s) + u_0(x_i)(1+\theta s) + u_1(x_i) + \mathcal{O}(\Delta x^2), \end{aligned} \tag{3.27}$$

where the $K's$ are defined by (3.26).

Proof: Let us substitute the exact solution $\widetilde{u}(x,s)$ in the numerical method (3.25), that is,

$$K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s)$$

$$+\frac{\sinh(\lambda_s \Delta x)}{\lambda_s}\Delta x\left[(1+\theta s)u_0(x_i)+u_1(x_i)\right]=0. \tag{3.28}$$

Making Taylor expansions of the hyperbolic functions,

$$\sinh\left(\lambda_s \Delta x/2\right)=\frac{1}{2}\left(\lambda_s\Delta x+\frac{\lambda_s^3\Delta x^3}{24}+\mathcal{O}(\Delta x^5)\right)$$

and

$$\cosh\left(\lambda_s \Delta x\right)=1+\frac{\lambda_s^2\Delta x^2}{2}+\frac{\lambda_s^4\Delta x^4}{24}+\mathcal{O}(\Delta x^5),$$

then the first three terms in the equation (3.28) turn into

$$\begin{aligned}
&\widetilde{u}_{i-1}(s)\left(1+\tfrac{P_{i-1/2}}{2\lambda_s}\left(\lambda_s\Delta x+\tfrac{\lambda_s^3\Delta x^3}{24}+\mathcal{O}(\Delta x^5)\right)\right)\\
&+\widetilde{u}_i(s)\left(-2-\lambda_s^2\Delta x^2-\tfrac{\lambda_s^4\Delta x^4}{12}\right)\\
&-\widetilde{u}_i(s)\tfrac{(P_{i+1/2}-P_{i-1/2})}{2\lambda_s}\left(\lambda_s\Delta x+\tfrac{\lambda_s^3\Delta x^3}{24}+\mathcal{O}(\Delta x^5)\right)\\
&+\widetilde{u}_{i+1}(s)\left(1-\frac{P_{i+1/2}}{2\lambda_s}\left(\lambda_s\Delta x+\frac{\lambda_s^3\Delta x^3}{24}+\mathcal{O}(\Delta x^5)\right)\right).
\end{aligned} \tag{3.29}$$

For a fixed s, we do also Taylor expansions of the functions $\widetilde{u}_{i-1}(s)$, $\widetilde{u}_{i+1}(s)$, $P_{i-1/2}$ and $P_{i+1/2}$ around the point x_i obtaining

$$P_{i+1/2}=P\left(x_i+\Delta x/2\right)=P_i+\frac{\Delta x}{2}P_i'+\frac{\Delta x^2}{8}P_i''+\frac{\Delta x^3}{48}P_i'''+\frac{\Delta x^4}{16\times 24}P_i''''+\mathcal{O}(\Delta x^5),$$

$$P_{i-1/2}=P\left(x_i-\Delta x/2\right)=P_i-\frac{\Delta x}{2}P_i'+\frac{\Delta x^2}{8}P_i''-\frac{\Delta x^3}{48}P_i'''+\frac{\Delta x^4}{16\times 24}P_i''''+\mathcal{O}(\Delta x^5),$$

$$\widetilde{u}_{i-1}(s)=\widetilde{u}_i(s)-\Delta x\frac{d\widetilde{u}_i}{dx}(s)+\frac{\Delta x^2}{2}\frac{d^2\widetilde{u}_i}{dx^2}(s)-\frac{\Delta x^3}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s)+\frac{\Delta x^4}{24}\frac{d^4\widetilde{u}_i}{dx^4}(s)+\mathcal{O}(\Delta x^5),$$

$$\widetilde{u}_{i+1}(s)=\widetilde{u}_i(s)+\Delta x\frac{d\widetilde{u}_i}{dx}(s)+\frac{\Delta x^2}{2}\frac{d^2\widetilde{u}_i}{dx^2}(s)+\frac{\Delta x^3}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s)+\frac{\Delta x^4}{24}\frac{d^4\widetilde{u}_i}{dx^4}(s)+\mathcal{O}(\Delta x^5),$$

where P', P'', P''' and P'''' denote the derivatives of P. Using these equalities in (3.29) we obtain, after some algebraic manipulations

$$\begin{aligned}
&\widetilde{u}_i(s)\left(-\lambda_s^2\Delta x^2-\frac{\lambda_s^4}{12}\Delta x^4-P_i'\Delta x^2-\frac{\lambda_s^2}{24}P_i'\Delta x^4-\frac{1}{24}P_i'''\Delta x^4\right)\\
&\quad+\frac{d\widetilde{u}_i}{dx}(s)\left(-P_i\Delta x^2-\frac{\lambda_s^2}{24}P_i\Delta x^4-\frac{1}{8}P_i''\Delta x^4\right)+\frac{d^2\widetilde{u}_i}{dx^2}(s)\left(\Delta x^2-\frac{1}{4}P_i'\Delta x^4\right)\\
&\quad-\frac{1}{6}P_i\Delta x^4\frac{d^3\widetilde{u}_i}{dx^3}(s)+\frac{1}{12}\Delta x^4\frac{d^4\widetilde{u}_i}{dx^4}(s)+\mathcal{O}(\Delta x^5)
\end{aligned}$$

which is the same as

$$\left(\frac{d^2\widetilde{u}_i}{dx^2}(s) - \lambda_s^2 - P_i\frac{d\widetilde{u}_i}{dx}(s) - P_i'\widetilde{u}_i(s)\right)\Delta x^2$$
$$+\left(\left(-\frac{\lambda_s^4}{12} - \frac{\lambda_s^2}{24}P_i' - \frac{P_i'''}{24}\right)\widetilde{u}_i(s) - \left(\frac{\lambda_s^2}{24}P_i + \frac{P_i''}{8}\right)\frac{d\widetilde{u}_i}{dx}(s)\right)\Delta x^4$$
$$+\left(-\frac{P_i'}{4}\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{P_i}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s) + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\right)\Delta x^4 + \mathcal{O}(\Delta x^5). \tag{3.30}$$

The last term of equation (3.28), using Taylor's formula of the hyperbolic function, becomes

$$\frac{\sinh(\lambda_s\Delta x)}{\lambda_s}\Delta x\left[(1+\theta s)u_0(x_i) + u_1(x_i)\right]$$
$$= \frac{\Delta x}{\lambda_s}(\lambda_s\Delta x + \frac{\lambda_s^3\Delta x^3}{6} + \mathcal{O}(\Delta x^5))[(1+\theta s)u_0(x_i) + u_1(x_i)]$$
$$= [(1+\theta s)u_0(x_i) + u_1(x_i)]\Delta x^2 + \frac{\lambda_s^2}{6}[(1+\theta s)u_0(x_i) + u_1(x_i)]\Delta x^4$$
$$+\mathcal{O}(\Delta x^5). \tag{3.31}$$

From (3.30) and (3.31), and after dividing by Δx^2, we obtain

$$K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s)$$
$$+\frac{\sinh(\lambda_s\Delta x)}{\lambda_s}\Delta x\left[(1+\theta s)u_0(x_i) + u_1(x_i)\right]$$
$$= \frac{d^2\widetilde{u}_i}{dx^2}(s) - \lambda_s^2\widetilde{u}_i(s) - \frac{d}{dx}(P\widetilde{u}_i)(s) + u_0(x_i)(1+\theta s) + u_1(x_i)$$
$$+\left(\left(-\frac{\lambda_s^4}{12} - \frac{\lambda_s^2}{24}P_i' - \frac{P_i'''}{24}\right)\widetilde{u}_i(s) - \left(\frac{\lambda_s^2}{24}P_i + \frac{P_i''}{8}\right)\frac{d\widetilde{u}_i}{dx}(s)\right)\Delta x^2$$
$$+\left(-\frac{P_i'}{4}\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{P_i}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s) + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\right)\Delta x^2$$
$$+\frac{\lambda_s^2}{6}[(1+\theta s)u_0(x_i) + u_1(x_i)]\Delta x^2 + \mathcal{O}(\Delta x^3).$$

Finally, since

$$\frac{d^2\widetilde{u}_i}{dx^2}(s) - (\lambda_s^2 + P_i')\widetilde{u}_i(s) - P_i\frac{d\widetilde{u}_i}{dx}(s) = -(1+\theta s)u_0(x_i) - u_1(x_i),$$

then

$$
\begin{aligned}
&K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\
&\quad + \frac{\sinh(\lambda_s \Delta x)}{\lambda_s}\Delta x\left[(1+\theta s)u_0(x_i) + u_1(x_i)\right] \\
&= \frac{d^2\widetilde{u}_i}{dx^2}(s) - \lambda_s^2\widetilde{u}_i(s) - \frac{d}{dx}(P\widetilde{u})_i(s) + u_0(x_i)(1+\theta s) + u_1(x_i) \\
&\quad + \left(\left(\frac{\lambda_s^4}{12} + \frac{\lambda_s^2}{8}P_i' - \frac{P_i'''}{24}\right)\widetilde{u}_i(s) + \left(\frac{\lambda_s^2}{8}P_i - \frac{P_i''}{8}\right)\frac{d\widetilde{u}_i}{dx}(s)\right)\Delta x^2 \\
&\quad + \left(-\left(\frac{P_i'}{4} + \frac{\lambda_s^2}{6}\right)\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{P_i}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s) + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\right)\Delta x^2 + \mathcal{O}(\Delta x^3),
\end{aligned}
$$

which confirms the second order accuracy in space of Laplace-FV method.

■

3.2.3 A Laplace transform piecewise linearized method

In this section we present the Laplace transform piecewise linearized method (Laplace-PL) which is obtained from the application of the Laplace transform in conjunction with the piecewise linearized method used for the spatial discretization. This method was first proposed in [79]. It was also applied to our model problem in [5], where we present some numerical tests to show the convergence and efficiency of the numerical method. In fact, we shall see this method performs better than the Laplace transform finite volume method, since the latter presents oscillations for large values of the parameter $|P|$.

Following the suggestions applied in [79] for a slightly different problem, we can rewrite the ordinary differential equation (3.2) in the form

$$
\frac{d^2\widetilde{u}}{dx^2}(x,s) - \frac{P}{D}\frac{d\widetilde{u}}{dx}(x,s) - \left(\frac{P'}{D} + \lambda_s^2\right)\widetilde{u}(x,s) = -\frac{u_0(x)}{D}(1+\theta s) - \frac{u_1(x)}{D}, \quad (3.32)
$$

where P' represents the x derivative of P. In each interval $[x_{i-1}, x_{i+1}]$, for $i = 1, \ldots, N-1$, the equation can be approximated by

$$
\frac{d^2\widetilde{U}}{dx^2}(x,s) - \frac{P_i}{D}\frac{d\widetilde{U}}{dx}(x,s) - \left(\frac{P_i'}{D} + \lambda_s^2\right)\widetilde{U}(x,s) = -\frac{u_0(x_i)}{D}(1+\theta s) - \frac{u_1(x_i)}{D}, \quad (3.33)
$$

where $\widetilde{U}(x,s)$ is an approximation of $\widetilde{u}(x,s)$, $P_i = P(x_i)$ and $P_i^{'} = P'(x_i)$. This equation is obtained from (3.32) by freezing the coefficients at the mid-point

of the interval $x_{i-1} \leq x \leq x_{i+1}$. The solution of (3.33) in $[x_{i-1}, x_{i+1}]$ is

$$\widetilde{U}(x,s) = A_i \mathrm{e}^{\nu_{s,i}^{+}(x-x_i)} + B_i \mathrm{e}^{\nu_{s,i}^{-}(x-x_i)} + \widetilde{U}_p\left(x_i, s\right), \tag{3.34}$$

with $\nu_{s,i}^{\pm} = \frac{P_i}{2D} \pm \sqrt{\left(\frac{P_i}{2D}\right)^2 + (\frac{P_i'}{D} + \lambda_s^2)}$ and $\widetilde{U}_p\left(x_i, s\right)$ is a particular solution given by

$$\widetilde{U}_p\left(x_i, s\right) = \frac{u_0(x_i)(1+\theta s) + u_1(x_i)}{P_i' + s(1+\theta s)}.$$

The values A_i and B_i can be directly determined from the solution (3.34) as

$$\begin{aligned}
B_i &= \widetilde{U}_i(s) - A_i - \widetilde{U}_{p_i}, & (3.35)\\
\widetilde{U}_{i-1}(s) &= A_i \mathrm{e}^{-\nu_{s,i}^{+}\Delta x} + (\widetilde{U}_i(s) - A_i - \widetilde{U}_{p_i})\mathrm{e}^{-\nu_{s,i}^{-}\Delta x} + \widetilde{U}_{p_i}, & (3.36)\\
\widetilde{U}_{i+1}(s) &= A_i \mathrm{e}^{\nu_{s,i}^{+}\Delta x} + (\widetilde{U}_i(s) - A_i - \widetilde{U}_{p_i})\mathrm{e}^{\nu_{s,i}^{-}\Delta x} + \widetilde{U}_{p_i}, & (3.37)
\end{aligned}$$

where $\widetilde{U}_i(s)$, $i = 0, \ldots, N$, represents the approximation of $\widetilde{U}(x_i, s)$ in the Laplace transform domain and $\widetilde{U}_{p_i}$ denotes $\widetilde{U}_p(x_i, s)$. From (3.36)-(3.37), we have

$$\begin{aligned}
A_i &= \frac{\widetilde{U}_{i-1}(s) - \widetilde{U}_i(s)\mathrm{e}^{-\nu_{s,i}^{-}\Delta x} + \widetilde{U}_{p_i}\mathrm{e}^{-\nu_{s,i}^{-}\Delta x} - \widetilde{U}_{p_i}}{\mathrm{e}^{-\nu_{s,i}^{+}\Delta x} - \mathrm{e}^{-\nu_{s,i}^{-}\Delta x}}\\
&= \frac{[-\widetilde{U}_{i-1}(s) + \widetilde{U}_i(s)\mathrm{e}^{-\nu_{s,i}^{-}\Delta x} - \widetilde{U}_{p_i}\mathrm{e}^{-\nu_{s,i}^{-}\Delta x} + \widetilde{U}_{p_i}]\mathrm{e}^{(\nu_{s,i}^{+}+\nu_{s,i}^{-})\Delta x}}{\mathrm{e}^{\nu_{s,i}^{+}\Delta x} - \mathrm{e}^{\nu_{s,i}^{-}\Delta x}}
\end{aligned}$$

and

$$A_i = \frac{\widetilde{U}_{i+1}(s) - \widetilde{U}_i(s)\mathrm{e}^{\nu_{s,i}^{-}\Delta x} + \widetilde{U}_{p_i}\mathrm{e}^{\nu_{s,i}^{-}\Delta x} - \widetilde{U}_{p_i}}{\mathrm{e}^{\nu_{s,i}^{+}\Delta x} - \mathrm{e}^{\nu_{s,i}^{-}\Delta x}}.$$

By equaling the values of A_i, we obtain the following three-point finite difference equations, for $i = 1, \ldots, N-1$,

$$\begin{aligned}
&K_{i,i-1}(s)\widetilde{U}_{i-1}(s) + K_{i,i}(s)\widetilde{U}_i(s) + K_{i,i+1}(s)\widetilde{U}_{i+1}(s)\\
&\quad = \widetilde{U}_{p_i}\left(K_{i,i-1}(s) + K_{i,i}(s) + K_{i,i+1}(s)\right), \qquad (3.38)
\end{aligned}$$

where

$$\begin{aligned}
K_{i,i-1}(s) &= \mathrm{e}^{(\nu_{s,i}^{+}+\nu_{s,i}^{-})\Delta x},\\
K_{i,i}(s) &= -\mathrm{e}^{\nu_{s,i}^{+}\Delta x} - \mathrm{e}^{\nu_{s,i}^{-}\Delta x},\\
K_{i,i+1}(s) &= 1. \qquad (3.39)
\end{aligned}$$

The vector $\widetilde{b}(s)$ containing the boundary conditions and source terms is represented by

$$\begin{bmatrix} \widetilde{U}_{p_1}(K_{1,0}+K_{1,1}+K_{1,2}) \\ \widetilde{U}_{p_2}(K_{2,1}+K_{2,2}+K_{2,3}) \\ \vdots \\ \widetilde{U}_{p_{N-2}}(K_{N-2,N-3}+K_{N-2,N-2}+K_{N-2,N-1}) \\ \widetilde{U}_{p_{N-1}}(K_{N-1,N-2}+K_{N-1,N-1}+K_{N-1,N}) \end{bmatrix} - \begin{bmatrix} K_{1,0}\widetilde{U}_0(s) \\ 0 \\ \vdots \\ 0 \\ K_{N-1,N}\widetilde{U}_N(s) \end{bmatrix},$$

where $K_{i,i-1}=K_{i,i-1}(s)$, $K_{i,i}=K_{i,i}(s)$ and $K_{i,i+1}=K_{i,i+1}(s)$, $i=1,\ldots,N-1$. Once again, we can write equation (3.38) in the matrix form (3.19). The next step is to determine an approximate solution $U(x_i,t)$ from $\widetilde{U}(x_i,s)$ by using the Laplace inversion numerical method described in Section 3.1.1.

At last, we will confirm the second order accuracy of the Laplace-PL method, for the spatial discretization, when $D=1$.

Proposition 3.2.3. *For the piecewise linearized method we have*

$$\begin{aligned} &K_{i,i-1}(s)\widetilde{u}_{i-1}(s)+K_{i,i}(s)\widetilde{u}_i(s)+K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\ &\quad -\widetilde{u}_{p_i}\left(K_{i,i-1}(s)+K_{i,i}(s)+K_{i,i+1}(s)\right) \\ &\quad =\frac{d^2\widetilde{u}_i}{dx^2}(s)-P_i\frac{d\widetilde{u}_i}{dx}(s)-\left(P_i'+\lambda_s^2\right)\widetilde{u}_i(s)+u_0(x_i)(1+\theta s)+u_1(x_i) \\ &\quad +\mathcal{O}(\Delta x^2), \end{aligned} \tag{3.40}$$

where the $K's$ are defined by (3.39).

Proof: We first substitute the exact solution $\widetilde{u}(x,s)$ in the numerical method (3.38), that is,

$$\begin{aligned} &K_{i,i-1}(s)\widetilde{u}_{i-1}(s)+K_{i,i}(s)\widetilde{u}_i(s)+K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\ &\quad -\widetilde{u}_{p_i}\left(K_{i,i-1}(s)+K_{i,i}(s)+K_{i,i+1}(s)\right)=0. \end{aligned} \tag{3.41}$$

Let $R^2=\frac{P_i^2}{4}+(P_i'+\lambda_s^2)$. Using the definition of the $K's$ coefficients we have

$$\begin{aligned} &K_{i,i-1}(s)\widetilde{u}_{i-1}(s)+K_{i,i}(s)\widetilde{u}_i(s)+K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\ &\quad =\mathrm{e}^{P_i\Delta x}\widetilde{u}_{i-1}(s)-\left(\mathrm{e}^{(P_i/2+R)\Delta x}+\mathrm{e}^{(P_i/2-R)\Delta x}\right)\widetilde{u}_i(s)+\widetilde{u}_{i+1}(s) \end{aligned}$$

$$= \mathrm{e}^{(P_i/2)\Delta x}\left(\mathrm{e}^{(P_i/2)\Delta x}\widetilde{u}_{i-1}(s) - (\mathrm{e}^{R\Delta x} + \mathrm{e}^{-R\Delta x})\widetilde{u}_i(s) + \mathrm{e}^{-(P_i/2)\Delta x}\widetilde{u}_{i+1}(s)\right) \tag{3.42}$$

and

$$\begin{aligned}
&\widetilde{u}_{p_i}\left(K_{i,i-1}(s) + K_{i,i}(s) + K_{i,i+1}(s)\right)\\
&= \widetilde{u}_{p_i}\left(\mathrm{e}^{P_i\Delta x} - (\mathrm{e}^{(P_i/2+R)\Delta x} + \mathrm{e}^{(P_i/2-R)\Delta x}) + 1\right)\\
&= \mathrm{e}^{(P_i/2)\Delta x}\widetilde{u}_{p_i}\left(\mathrm{e}^{(P_i/2)\Delta x} - (\mathrm{e}^{R\Delta x} + \mathrm{e}^{-R\Delta x}) + \mathrm{e}^{(-P_i/2)\Delta x}\right).
\end{aligned} \tag{3.43}$$

For convenience, we multiply equation (3.41) by $\mathrm{e}^{-(P_i/2)\Delta x}$. Next, by doing Taylor expansions of the exponential functions in (3.42) and (3.43) results in

$$\begin{aligned}
&\mathrm{e}^{-(P_i/2)\Delta x}(K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s))\\
&= \widetilde{u}_{i-1}(s)\left(1 + \frac{P_i}{2}\Delta x + \frac{P_i^2}{8}\Delta x^2 + \frac{P_i^3}{48}\Delta x^3 + \frac{P_i^4}{4!16}\Delta x^4 + \mathcal{O}(\Delta x^5)\right)\\
&\quad -\widetilde{u}_i(s)\left(2 + R^2\Delta x^2 + \frac{R^4}{12}\Delta x^4 + \mathcal{O}(\Delta x^5)\right)\\
&\quad +\widetilde{u}_{i+1}(s)\left(1 - \frac{P_i}{2}\Delta x + \frac{P_i^2}{8}\Delta x^2 - \frac{P_i^3}{48}\Delta x^3 + \frac{P_i^4}{4!16}\Delta x^4 + \mathcal{O}(\Delta x^5)\right)
\end{aligned} \tag{3.44}$$

and

$$\begin{aligned}
&\mathrm{e}^{-(P_i/2)\Delta x}\widetilde{u}_{p_i}\left(K_{i,i-1}(s) + K_{i,i}(s) + K_{i,i+1}(s)\right)\\
&= \widetilde{u}_{p_i}\left(\frac{P_i^2}{4}\Delta x^2 + \frac{P_i^4}{4!8}\Delta x^4 - R^2\Delta x^2 - \frac{R^4}{12}\Delta x^4 + \mathcal{O}(\Delta x^5)\right)\\
&= \widetilde{u}_{p_i}\left(-(P_i' + \lambda_s^2)\Delta x^2 - \frac{(P_i' + \lambda_s^2)^2}{12}\Delta x^4 - \frac{P_i^2}{24}(P_i' + \lambda_s^2)\Delta x^4 + \mathcal{O}(\Delta x^5)\right),
\end{aligned}$$

as $R^2 = \frac{P_i^2}{4} + (P_i' + \lambda_s^2)$. Furthermore, since

$$\widetilde{u}_{p_i} = \frac{u_0(x_i)(1+\theta s) + u_1(x_i)}{P_i' + \lambda_s^2}$$

we can simplify the last expression and get

$$\begin{aligned}
&\mathrm{e}^{-(P_i/2)\Delta x}\widetilde{u}_{p_i}\left(K_{i,i-1}(s) + K_{i,i}(s) + K_{i,i+1}(s)\right)\\
&\quad = (-u_0(x_i)(1+\theta s) - u_1(x_i))\Delta x^2\\
&\qquad +(-u_0(x_i)(1+\theta s) - u_1(x_i))\left(\frac{P_i^2}{24} + \frac{(P_i' + \lambda_s^2)}{12}\right)\Delta x^4 + \mathcal{O}(\Delta x^5).
\end{aligned} \tag{3.45}$$

Next, we do Taylor expansions of the functions $\widetilde{u}_{i-1}(s)$ and $\widetilde{u}_{i+1}(s)$ around the point x_i. After some algebraic manipulations we can write (3.44) in the form

$$\begin{aligned}
&\mathrm{e}^{-(P_i/2)\Delta x}(K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s)) \\
&\quad = \left(\frac{d^2\widetilde{u}_i}{dx^2}(s) - P_i\frac{d\widetilde{u}_i}{dx}(s) - (P_i' + \lambda_s^2)\widetilde{u}_i(s)\right)\Delta x^2 \\
&\qquad + \left(\left(-\frac{1}{12}(P_i' + \lambda_s^2)^2 - \frac{P_i^2}{24}(P_i' + \lambda_s^2)\right)\widetilde{u}_i(s) - \frac{P_i^3}{24}\frac{d\widetilde{u}_i}{dx}(s)\right)\Delta x^4 \\
&\qquad + \left(\frac{P_i^2}{8}\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{P_i}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s) + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\right)\Delta x^4 + \mathcal{O}(\Delta x^5). \qquad (3.46)
\end{aligned}$$

After dividing (3.45) and (3.46) by Δx^2, we obtain

$$\begin{aligned}
&(K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\
&\qquad -\widetilde{u}_{p_i}\left(K_{i,i-1}(s) + K_{i,i}(s) + K_{i,i+1}(s)\right) \\
&\quad = \frac{d^2\widetilde{u}_i}{dx^2}(s) - P_i\frac{d\widetilde{u}_i}{dx}(s) - (P_i' + \lambda_s^2)\widetilde{u}_i(s) + u_0(x_i)(1+\theta s) + u_1(x_i) \\
&\qquad + \left(\left(-\frac{1}{12}(P_i' + \lambda_s^2)^2 - \frac{P_i^2}{24}(P_i' + \lambda_s^2)\right)\widetilde{u}_i(s) - \frac{P_i^3}{24}\frac{d\widetilde{u}_i}{dx}(s)\right)\Delta x^2 \\
&\qquad + \left(\frac{P_i^2}{8}\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{P_i}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s) + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\right)\Delta x^2 \\
&\qquad + (u_0(x_i)(1+\theta s)u_1(x_i))\left(\frac{P_i^2}{24} + \frac{(P_i' + \lambda_s^2)}{12}\right)\Delta x^2.
\end{aligned}$$

Finally, since

$$\frac{d^2\widetilde{u}_i}{dx^2}(s) - (\lambda_s^2 + P_i')\widetilde{u}_i(s) - P_i\frac{d\widetilde{u}_i}{dx}(s) = -(1+\theta s)u_0(x_i) - u_1(x_i),$$

we obtain the desired result

$$\begin{aligned}
&K_{i,i-1}(s)\widetilde{u}_{i-1}(s) + K_{i,i}(s)\widetilde{u}_i(s) + K_{i,i+1}(s)\widetilde{u}_{i+1}(s) \\
&\qquad -\widetilde{u}_{p_i}\left(K_{i,i-1}(s) + K_{i,i}(s) + K_{i,i+1}(s)\right) \\
&\quad = \frac{d^2\widetilde{u}_i}{dx^2}(s) - P_i\frac{d\widetilde{u}_i}{dx}(s) - (P_i' + \lambda_s^2)\widetilde{u}_i(s) + u_0(x_i)(1+\theta s) + u_1(x_i) \\
&\qquad + \left(\frac{P_i}{12}(P_i' + \lambda_s^2)\frac{d\widetilde{u}_i}{dx}(s) + \frac{1}{12}(P_i^2 - P_i' - \lambda_s^2)\frac{d^2\widetilde{u}_i}{dx^2}(s) - \frac{P_i}{6}\frac{d^3\widetilde{u}_i}{dx^3}(s)\right)\Delta x^2 \\
&\qquad + \frac{1}{12}\frac{d^4\widetilde{u}_i}{dx^4}(s)\Delta x^2.
\end{aligned}$$

■

We can conclude that the truncation error has order $\mathcal{O}(\Delta x^2)$, that is, the Laplace-PL method is consistent and second order accurate in space.

3.2.4 Some remarks on stability

In order to extend the study of the convergence of methods, we will now discuss some properties of matrix $K(s)$ of the system (3.19), to draw some conclusions about the stability of the numerical methods. By denoting $\widetilde{E}_i = \widetilde{E}_S(x_i, s)$, $i = 1, \ldots, N-1$ we have

$$\mathcal{L}_\Delta \widetilde{E}_i = \mathcal{T}_\Delta(x_i, s),$$

that is,

$$K(s)\widetilde{E}(s) = \mathcal{T}_\Delta(s),$$

with $\mathcal{T}_\Delta(x_i, s)$ the truncation error for the Laplace-FD method. If, for some constant $c > 0$, $\|K^{-1}(s)\|_\infty \leq c$ then we would have $|\widetilde{E}_i| \leq c\|\mathcal{T}_\Delta\|_\infty$. Regardless of the method, the matrix $K(s)$ contains the parameter λ_s^2, given by

$$\lambda_s^2 = (\theta s^2 + s)/D = (\theta\beta^2 + \beta - \theta\omega^2 + \mathbf{i}\omega(2\theta\beta + 1))/D, \;\; \omega = \frac{k\pi}{T}, \;\; k = 1, \ldots, M,$$

with $s = \beta + \mathbf{i}\omega$, M defines the set of values in the Laplace domain given by $M = \max_i M_i$, where M_i is the iteration for each x_i and T defines the stepsize of the trapezoidal rule applied in (3.9). For $\theta\omega^2 > \theta\beta^2 + \beta$ the complex λ_s^2 has negative real part. Hence, it is not easy to prove analytically the inverse of $K(s)$ is bounded, since the matrix $K(s)$ is not an M-matrix [91, 92]. Despite this, we will analyze $K(s)$ to get some conclusions. For this study we consider the example

Example 3.2.1. Problem (2.1)-(2.3) for

$$P(x) = \alpha \sin(x) \frac{1}{J_0(\mathbf{i}\alpha)} e^{\alpha \cos x},$$

with $-15 \leq x \leq 15$, $t = 1$, $\theta = 1$, $D = 1$, with initial conditions

$$u_0(x) = \frac{1}{L\sqrt{\pi}} e^{-x^2/L^2}, \qquad u_1(x) = 0,$$

and boundary conditions $f(t) = g(t) = 0$.

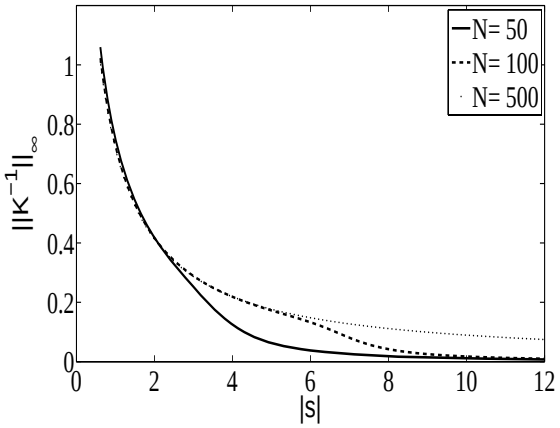

Figure 3.4: Infinity norm for matrix $K^{-1}(s)$ for $T = 30$ and different values of N. We have considered Example 3.2.1 for $P(x)$ with $\alpha = 1$.

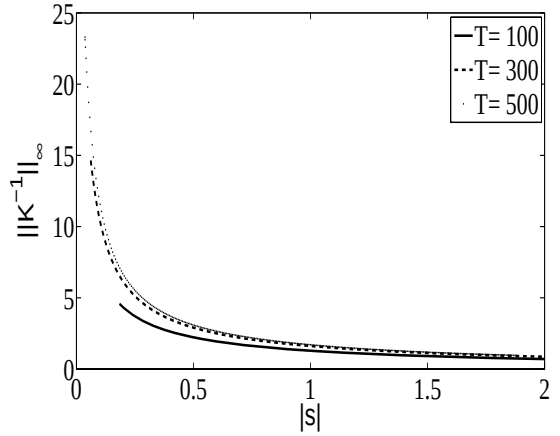

Figure 3.5: Infinity norm for matrix $K^{-1}(s)$ for $N = 250$ and different values of T. We have considered Example 3.2.1 for $P(x)$ with $\alpha = 1$.

It is easy to see numerically that for a fixed T, as we refine the space step, the value $\|K^{-1}(s)\|_\infty$ does not change significantly as illustrated in Figure 3.4. We also notice that $\|K^{-1}(s)\|_\infty$ is larger for values of $|s|$ close to zero, indicating that the convergence can be lower for these values. The same features are also evident when we consider different values of T, as can be observed in Figure 3.5.

Regarding the accuracy of the numerical methods, additionally to the

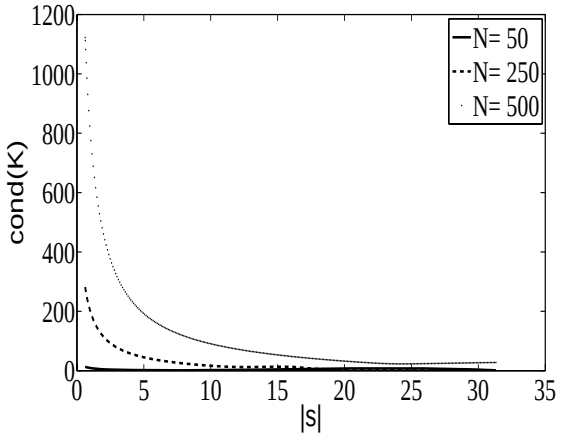

Figure 3.6: Condition number for the matrix $K(s)$ for $T = 30$ and different values of N. We have considered Example 3.2.1 for $P(x)$ with $\alpha = 1$.

truncation errors, let us look at the condition number of the matrix $K(s)$, cond(K), that determines how accurately we can solve the system (3.19). The condition number of the matrix $K(s)$ is affected by the values of N and T. Although the matrix K is different for the three spatial discretizations, the numerical tests performed exhibited the same behavior (same curve) of the condition number for all the three methods. Therefore, we only present the results for the Laplace-FD method. We can infer from Figures 3.6 and 3.7 that the condition number of the matrix $K(s)$ increases if we increase N or T and decays with $|s|$.

Usually one must always expect to loose $\log_{10}(\text{cond}(K))$ digits of precision in computing the solution, except under very special circumstances. Since we work with double precision numbers, about 16 decimal digits of accuracy, caution is advised when the condition number is much greater than $1/\sqrt{10^{-16}}$, which in general does not happen for our problem. The condition number of order 10^6 is reached for very large values of T and N, such as, both larger than 10^5. We plotted the results for $\alpha = 1$, although for different values of α we have similar results. The same conclusions are valid for problems with different functions P and different initial and boundary conditions. We also notice that the Laplace-FV and Laplace-PL methods perform similarly.

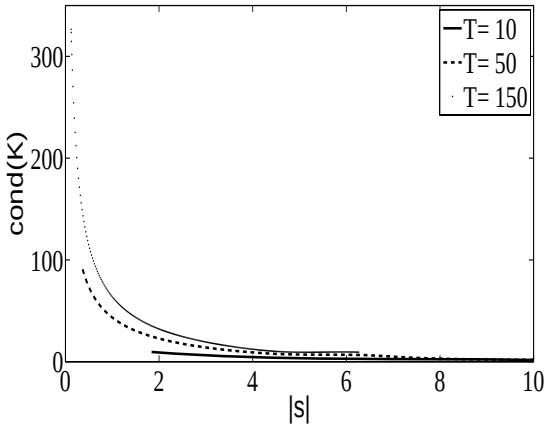

Figure 3.7: Condition number for the matrix $K(s)$ for $N = 250$ and different values of T. We have considered Example 3.2.1 for $P(x)$ with $\alpha = 1$.

Additionally we observe that we have a similar phenomenon to the so called pollution effect [8] observed for the Helmholtz equation and for high wavenumbers, where the discretization space step has to be sufficiently refined to avoid numerical dispersion. Also in this context, it is observed that if we have a complex number as a coefficient in the equation, which is our case with λ_s, the imaginary part acts as an absorption parameter, which seems to allow us to better control the solution by decreasing the solution magnitude [40]. Following what is reported in literature [8, 9, 74], a useful rule observed for an adjustment of the space step is to force some relation between T and the Δx. For our problem a similar condition is

$$\omega \Delta x \leq \frac{2\pi}{10}.$$

This leads to $(M/T)\Delta x \leq 2/10$.

3.2.5 Convergence of the numerical methods

We have seen in previous sections three different approaches for the spatial discretization after the application of the Laplace transform: the Laplace transform finite difference (Laplace-FD), the Laplace transform finite volume (Laplace-FV) and the Laplace transform piecewise linearized

Δx	Laplace-FD	Rate	Laplace-FV	Rate	Laplace-PL	Rate
10/128	0.2549×10^{-3}		0.1815×10^{-3}		0.6307×10^{-4}	
10/256	0.6601×10^{-4}	2.0	0.4238×10^{-4}	2.0	0.7612×10^{-5}	3.1
10/512	0.1615×10^{-4}	2.0	0.1093×10^{-4}	2.0	0.9513×10^{-6}	3.0
10/1024	0.4063×10^{-5}	2.0	0.2681×10^{-5}	2.0	0.1392×10^{-6}	2.8
10/2048	0.1018×10^{-5}	2.0	0.6670×10^{-6}	2.0	0.1733×10^{-7}	3.0

Table 3.1: Global error E_{G1} for $\theta = 0$, $P = 2$, $t = 1$, $0 \leq x \leq 10$, $TOL = 1/N^3$, $T = 3$, $\beta = -\ln(10^{-16})/2T$, computed with the norm ℓ_∞.

Δx	Laplace-FD	Rate	Laplace-FV	Rate	Laplace-PL	Rate
10/128	0.7229×10^{-3}		0.7199×10^{-3}		0.3034×10^{-4}	
10/256	0.1800×10^{-3}	2.0	0.1792×10^{-3}	2.0	0.3575×10^{-5}	3.1
10/512	0.4529×10^{-4}	2.0	0.4511×10^{-4}	2.0	0.7431×10^{-6}	2.3
10/1024	0.1128×10^{-4}	2.0	0.1123×10^{-4}	2.0	0.5582×10^{-7}	3.7
10/2048	0.2824×10^{-5}	2.0	0.2817×10^{-5}	2.0	0.9191×10^{-8}	2.6

Table 3.2: Global error E_{G1} for $\theta = 0$, $P = -2$, $t = 1$, $0 \leq x \leq 10$, $TOL = 1/N^3$, $T = 3$, $\beta = -\ln(10^{-16})/2T$, computed with the norm ℓ_∞.

(Laplace-PL) methods. Even though the second order accuracy of the spatial discretization was proved theoretically for the three schemes, we illustrate this property with some numerical tests.

In order to compare the approximate solution $U_i(t) = U_i$ with the exact solution $u(x_i, t) = u_i$, $i = 1, \ldots, N-1$, we consider two problems. The first problem has $\theta = 0$ and P is constant, that is, the problem (3.15)-(3.16) for $u_0 = 1$, from Example 3.1.1, with exact solution (3.18). To have information about the spatial discretization errors we define the global errors as

$$E_{G1} = \|u - U\|_\infty = \max_{1 \leq i \leq N-1} |u_i - U_i|, \tag{3.47}$$

and

$$E_{G2} = \|u - U\| = \left(\Delta x \sum_{i=1}^{N-1} |u_i - U_i|^2 \right)^{1/2}, \tag{3.48}$$

where these two norms, ℓ_∞ and $\ell_{2,\Delta x}$, are the same defined in Section 2.3. We show the results in Table 3.1, Table 3.2 and Figure 3.8 for the three schemes, computed with the norm ℓ_∞ (3.47).

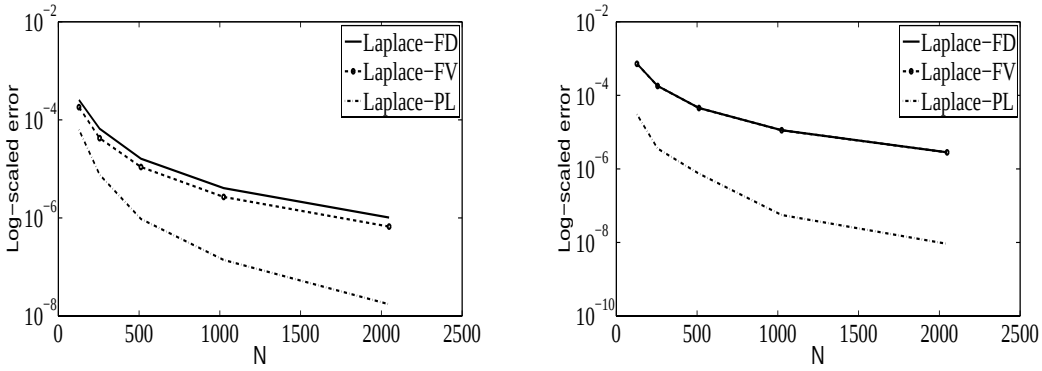

Figure 3.8: Global error E_{G1}. Left: $P = 2$ (Table 3.1). Right: $P = -2$ (Table 3.2).

Δx	Laplace-FD	Rate	Laplace-FV	Rate	Laplace-PL	Rate
$10/128$	0.4319×10^{-3}		0.2314×10^{-3}		0.4466×10^{-4}	
$10/256$	0.1070×10^{-3}	2.0	0.5659×10^{-4}	2.0	0.5036×10^{-5}	3.1
$10/512$	0.2669×10^{-4}	2.0	0.1411×10^{-4}	2.0	0.8055×10^{-6}	2.6
$10/1024$	0.6686×10^{-5}	2.0	0.3553×10^{-5}	2.0	0.1031×10^{-6}	3.0
$10/2048$	0.1670×10^{-5}	2.0	0.8800×10^{-6}	2.0	0.1578×10^{-7}	2.7

Table 3.3: Global error E_{G2} for $\theta = 0$, $P = 2$, $t = 1$, $0 \leq x \leq 10$, $TOL = 1/N^3$, $T = 3$, $\beta = -\ln(10^{-16})/2T$, computed with the norm $\ell_{2,\Delta x}$.

We observe the Laplace-PL method has a smaller error than the other two schemes. From Table 3.1 ($P = 2$) and Table 3.2 ($P = -2$), it is evident that the Laplace-PL has a significant higher convergence rate in comparison with the Laplace-FD and the Laplace-FV methods. For $P = 2$ the Laplace-FV method is slightly more accurate than the Laplace-FD method, although both methods perform similarly for $P = -2$. We obtain the same conclusions for E_{G2} defined by (3.48). However, the contrast is even greater between the Laplace-PL method and the other two methods when $P = 2$, as illustrated in Table 3.3 and Figure 3.9. In fact, the Laplace-PL method has smaller errors when we use the norm $\ell_{2,\Delta x}$ while the Laplace-FD and the Laplace-FV methods have better global errors when defined by the norm ℓ_∞. For $P = -2$ the errors are smaller when the norm $\ell_{2,\Delta x}$ is applied in all the methods as shown in Table 3.4.

We also notice that for large values of P positive the Laplace-FD method

Δx	Laplace-FD	Rate	Laplace-FV	Rate	Laplace-PL	Rate
10/128	0.6663×10^{-3}		0.6682×10^{-3}		0.2718×10^{-4}	
10/256	0.1673×10^{-3}	2.0	0.1678×10^{-3}	2.0	0.3456×10^{-5}	3.0
10/512	0.4170×10^{-4}	2.0	0.4181×10^{-4}	2.0	0.4828×10^{-6}	2.8
10/1024	0.1043×10^{-4}	2.0	0.1046×10^{-4}	2.0	0.4586×10^{-7}	3.4
10/2048	0.2606×10^{-5}	2.0	0.2616×10^{-5}	2.0	0.9324×10^{-8}	2.3

Table 3.4: Global error E_{G2} for $\theta = 0$, $P = -2$, $t = 1$, $0 \leq x \leq 10$, $TOL = 1/N^3$, $T = 3$, $\beta = -\ln(10^{-16})/2T$, computed with the norm $\ell_{2,\Delta x}$.

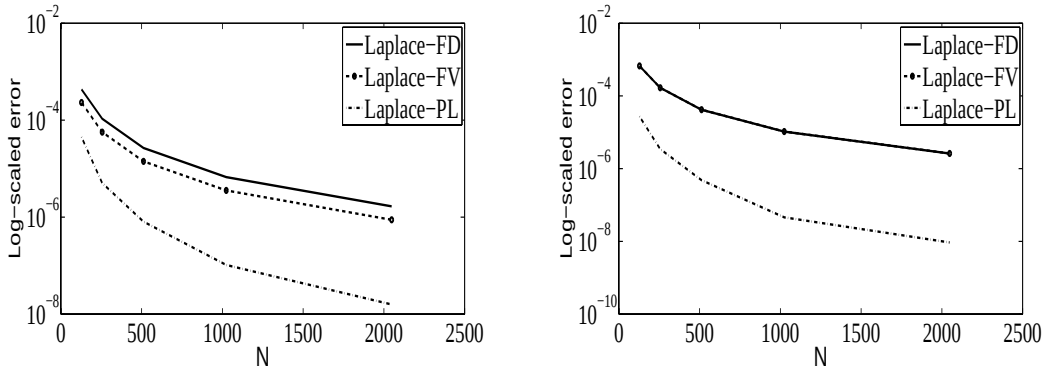

Figure 3.9: Global error E_{G2}. Left: $P = 2$ (Table 3.3). Right: $P = -2$ (Table 3.4).

has a smaller global error than the Laplace-FV method, has shown in Figure 3.10.

For our experiments the Laplace-PL method is more accurate, although theoretically we have proved the truncation errors of the numerical methods have the same order.

Example 3.2.2. Another example is considered for a problem with $\theta = 1$ and $P = 0$:

$$\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in]0, 2\pi[\,, t > 0. \tag{3.49}$$

The initial conditions are

$$u(x,0) = \sin(\frac{x}{2}), \qquad \frac{\partial u}{\partial t}(x,0) = -\frac{1}{2}\sin(\frac{x}{2}), \tag{3.50}$$

and the boundary conditions are

$$u(0,t) = 0, \qquad u(2\pi,t) = 0. \tag{3.51}$$

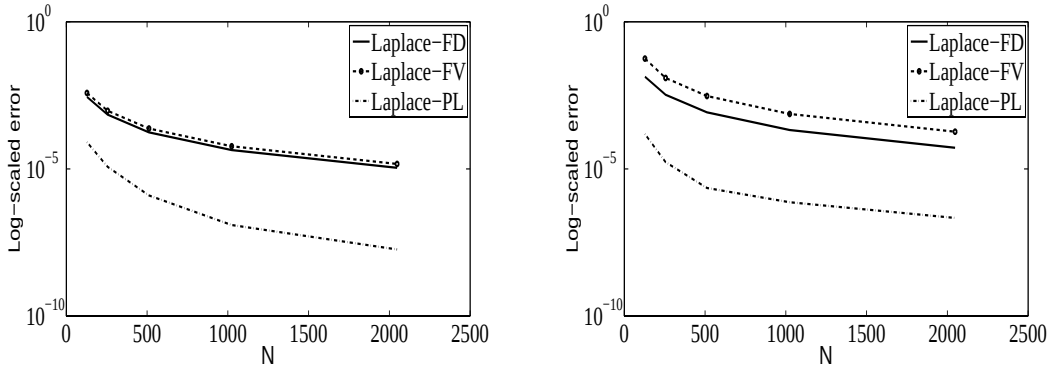

Figure 3.10: Global error E_{G1} for $t = 1$, $TOL = 1/N^3$, $\beta = -\ln(10^{-16})/2T$, $T = 3$, computed with the norm ℓ_∞. Left: $P = 5$, $0 \leq x \leq 20$. Right: $P = 15$, $0 \leq x \leq 25$.

Δx	Laplace-FD	Rate	Laplace-FV	Rate	Laplace-PL	Rate
$2\pi/128$	0.4860×10^{-5}		0.1215×10^{-5}		0.3151×10^{-4}	
$2\pi/256$	0.8596×10^{-6}	2.5	0.1733×10^{-6}	2.8	0.7555×10^{-5}	2.1
$2\pi/512$	0.2338×10^{-6}	1.9	0.8236×10^{-7}	1.1	0.1892×10^{-5}	2.0
$2\pi/1024$	0.6272×10^{-7}	1.9	0.1070×10^{-7}	2.9	0.4847×10^{-6}	2.0
$2\pi/2048$	0.1813×10^{-7}	1.8	0.2094×10^{-8}	2.4	0.1220×10^{-6}	2.0

Table 3.5: Global error E_{G1} for Example 3.2.2 for $\theta = 1$, $P = 0$, $t = 1$, $0 \leq x \leq 2\pi$, $TOL = 1/N^3$, $T = 5$, $\beta = -\ln(10^{-16})/2T$, computed with the norm ℓ_∞.

The analytical solution is easily obtained as

$$u\,(x,t) = e^{-\frac{t}{2}} \sin(\frac{x}{2}). \tag{3.52}$$

We note that, for this problem, the conclusions obtained when comparing the three methods are the same whether we use the norm ℓ_∞ or $\ell_{2,\Delta x}$. We present only the norm ℓ_∞, since it provides smaller errors for all the methods. Since the exact solution is very smooth, it can be seen in Table 3.5 and Figure 3.11 that the Laplace-FD and the Laplace-FV methods perform better in this particular case. In fact, the Laplace-FV method has a smaller error than the other two schemes.

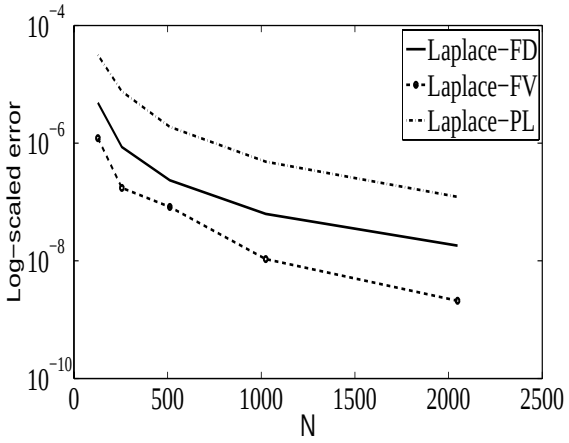

Figure 3.11: Global error E_{G1} for Example 3.2.2 for $P = 0$ (Table 3.5).

3.2.6 Behavior of the solution and comparison of performance

It is our purpose to show in this section the different performances of the numerical methods based on the Laplace transform, considering several numerical tests, to analyze different aspects of the spatial discretizations. Taking into account the initial and boundary conditions, different values of the parameters θ and P and the value of the space step, in some specific situations it will be clear what method should or not should be used to obtain the numerical solution of our problem. Numerical tests are presented to focus their advantages and disadvantages. In all of our examples we consider $D = 1$.

First, we choose a problem with a discontinuous initial condition to see how the numerical methods handle discontinuities. In the second and third examples we consider problems with non-zero initial conditions leading to a non-homogeneous ordinary differential equation for $\widetilde{u}(x, s)$, obtained from the application of the Laplace transform.

Example 3.2.3. Let us consider the problem

$$\theta\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}(P(x)u(x,t)) + D\frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in]0,\infty[\,, t > 0, \tag{3.53}$$

with initial conditions

$$u(x,0) = 0, \qquad \frac{\partial u}{\partial t}(x,0) = 0$$

and boundary conditions

$$u(0,t) = 1, \qquad \lim_{x \to +\infty} u(x,t) = 0.$$

For this case we have a homogeneous differential equation in $\widetilde{u}(x,s)$ to solve, that is,

$$\frac{d^2\widetilde{u}}{dx^2}(x,s) - \lambda_s^2 \widetilde{u}(x,s) - \frac{d}{dx}\left(\frac{P(x)}{D}\widetilde{u}(x,s)\right) = 0. \tag{3.54}$$

In what follows, we consider different values of P. First, we assume P constant and secondly, we consider P non-constant. These examples show how the numerical methods perform in the presence of discontinuities.

We display in Figure 3.12 and Figure 3.13 the results for P constant and for $\theta = 0$ and $\theta = 1$. For P constant, the solution of the homogeneous ordinary differential equation (3.54) is given by

$$\widetilde{u}\,(x,s) = \frac{1}{s}\mathrm{e}^{\nu_s^- x},$$

where ν_s^- is defined in (3.3).

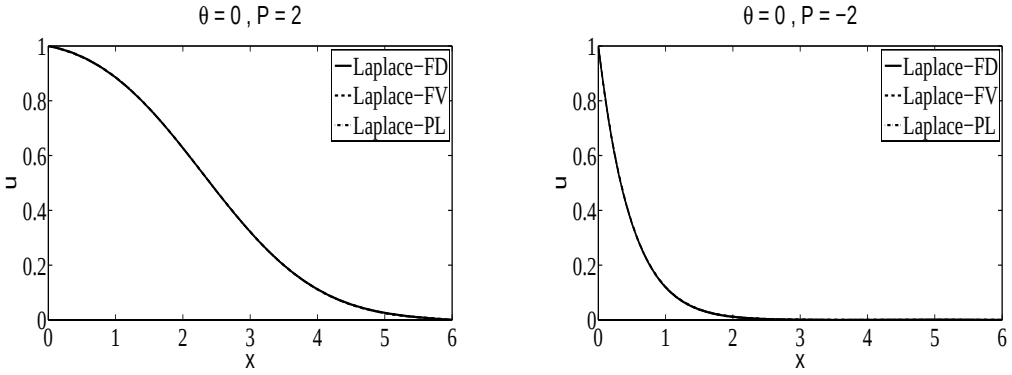

Figure 3.12: Approximate solution of Example 3.2.3 using different space discretizations, for different values of P, $\theta = 0$, $t = 1$ and $\Delta x = 0.1$.

In Figure 3.12 we consider the parabolic case, $\theta = 0$, and in Figure 3.13 we consider the hyperbolic case, $\theta = 1$. The Laplace-FD formulation, the Laplace-FV method and the Laplace-PL method are compared.

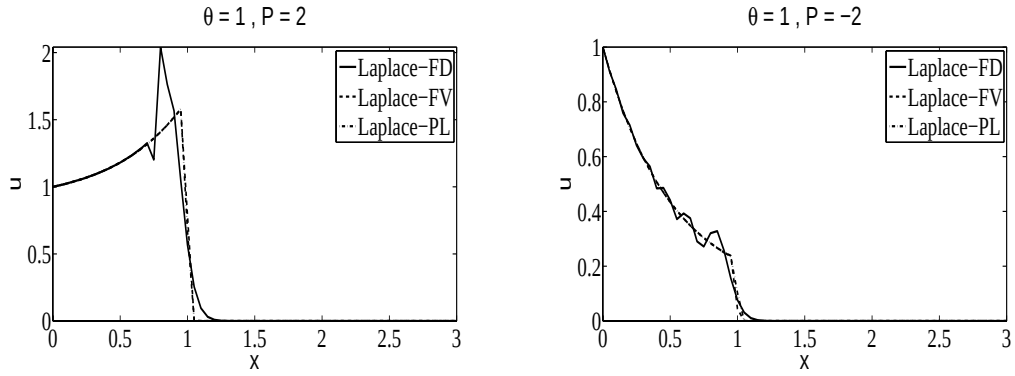

Figure 3.13: Approximate solution of Example 3.2.3 using different space discretizations, for different values of P, $\theta = 1$, $t = 1$ and $\Delta x = 0.05$.

For the parabolic case, $\theta = 0$, all the three methods perform similarly. The Laplace-FD method performs worse than the other two methods in the hyperbolic case, $\theta = 1$, since oscillations are not avoided for small space steps near the discontinuity.

However, these oscillations are easily removed for smaller values of the space step, as can be seen in Figure 3.14.

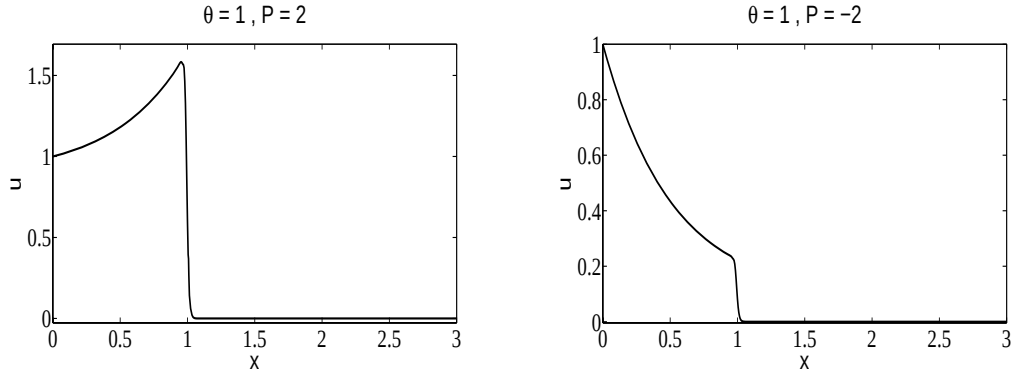

Figure 3.14: Approximate solution of Example 3.2.3 using the Laplace-FD discretization, for $\theta = 1$, $t = 1$ and $\Delta x = 0.005$. The oscillations are removed for small values of Δx.

Since for $\theta \neq 0$ the equation is hyperbolic, the discontinuity of the initial condition at $x = 0$ is transported along the characteristics. The characteristic

equation associated to equation (3.53) is

$$\theta\left(\frac{dx}{dt}\right)^2 - D = 0.$$

The characteristic curves are defined by $x = t\sqrt{\frac{D}{\theta}} + \xi$ and the waves are transmitted with finite velocity $v = \sqrt{D/\theta}$. At $\xi = 0$ we have $x = t\sqrt{\frac{D}{\theta}}$. For $\theta = 1$ and $D = 1$, the discontinuity at $t = 1$ appears at $x = 1$ as shown in Figure 3.13. The jump discontinuity at a specific time is the same for different values of P. When θ increases, the hyperbolic effects start to appear in a lower point x, that is, the velocity decreases with θ as the jump discontinuity peak increases.

Additionally, if we change the values of P the behavior of the Laplace-FV method is quite different from the Laplace-PL method and this one performs better. The Laplace-FV method in certain situations oscillates as shown in Figure 3.15. We can observe two examples for $\theta = 1$, with $P = -20$ and $P = -1000$. For $P = -20$ we have a space step of $\Delta x = 0.2$ and for $P = -1000$ we consider a smaller space step, $\Delta x = 0.02$. Therefore, for large values of $|P|$, the oscillations are not avoided as we refine the mesh.

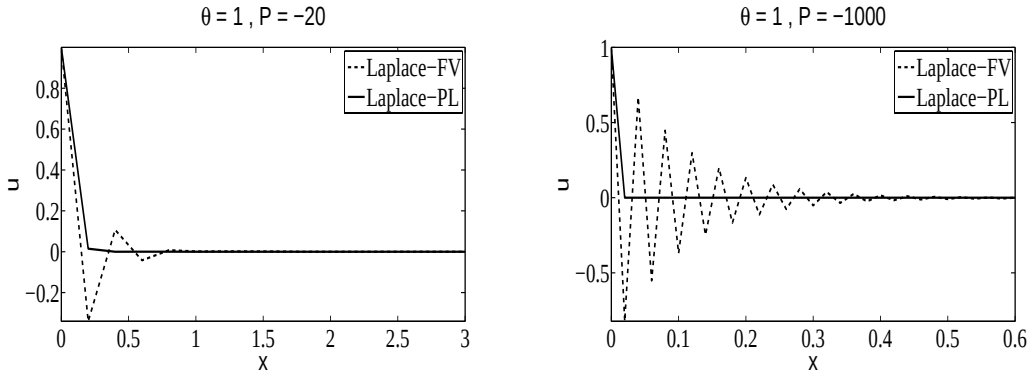

Figure 3.15: Approximate solution of Example 3.2.3 using the Laplace-FV and the Laplace-PL methods for $\theta = 1$ at $t = 1$. Left: $\Delta x = 0.2$. Right: $\Delta x = 0.02$.

Next, we consider the case for non-constant P. In Figure 3.16 we show the behavior of the solution for $P(x) = -2x$ and $P(x) = 2\mathrm{e}^{-x}$, for $\theta = 0.5$. The Laplace-FV method has oscillations for $P(x) = -2x$, while the Laplace-FD

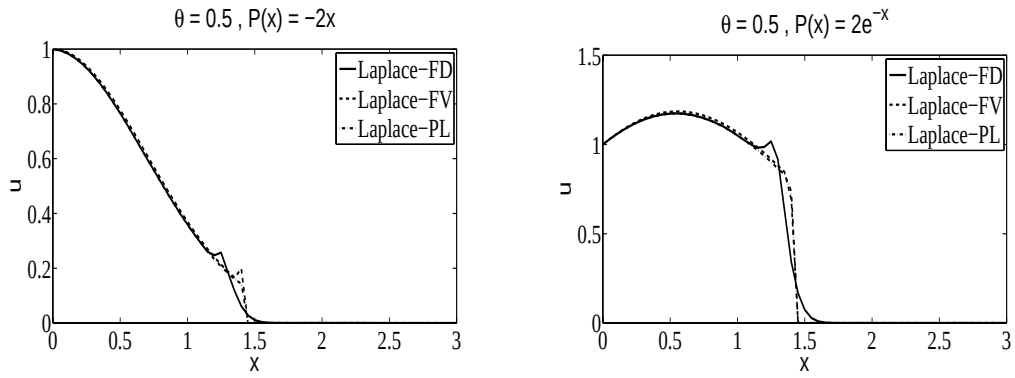

Figure 3.16: Approximate solution of Example 3.2.3 for $\theta = 0.5$, $\Delta x = 0.05$ at $t = 1$. Left figure: $P(x) = -2x$. Right figure: $P(x) = 2\mathrm{e}^{-x}$.

method shows oscillations in both cases for $\Delta x = 0.05$. These oscillations can be removed if we consider smaller values of the space step. For this reason, the next numerical tests are performed using only the Laplace-PL method, since we want to focus on the dependence of the solution taking account the variation of the parameters θ, P and t.

Figure 3.17 illustrates the behavior of the solution as we change the parameter θ. When θ increases, the hyperbolic effects start to appear in a lower point x, that is, the velocity decreases with θ as the jump discontinuity peak increases.

The effect of the parameter P in the solution is shown in Figure 3.18. The peak increases with $|P|$ for positive P and decreases for negative P. Once again, the location of the jump discontinuity is the same for different values of P.

In Figures 3.19 and 3.20 we can observe the evolution of the solution u as we travel in time. As stated before, when $\theta \neq 0$, the jump discontinuity does not depend on P and quickly dissipates with time. For constant P the hyperbolic solution tends to the parabolic solution as t increases.

In the next two problems we consider non-zero initial conditions. Thus, application of the Laplace transform yields a non-homogeneous differential equation in $\widetilde{u}(x, s)$.

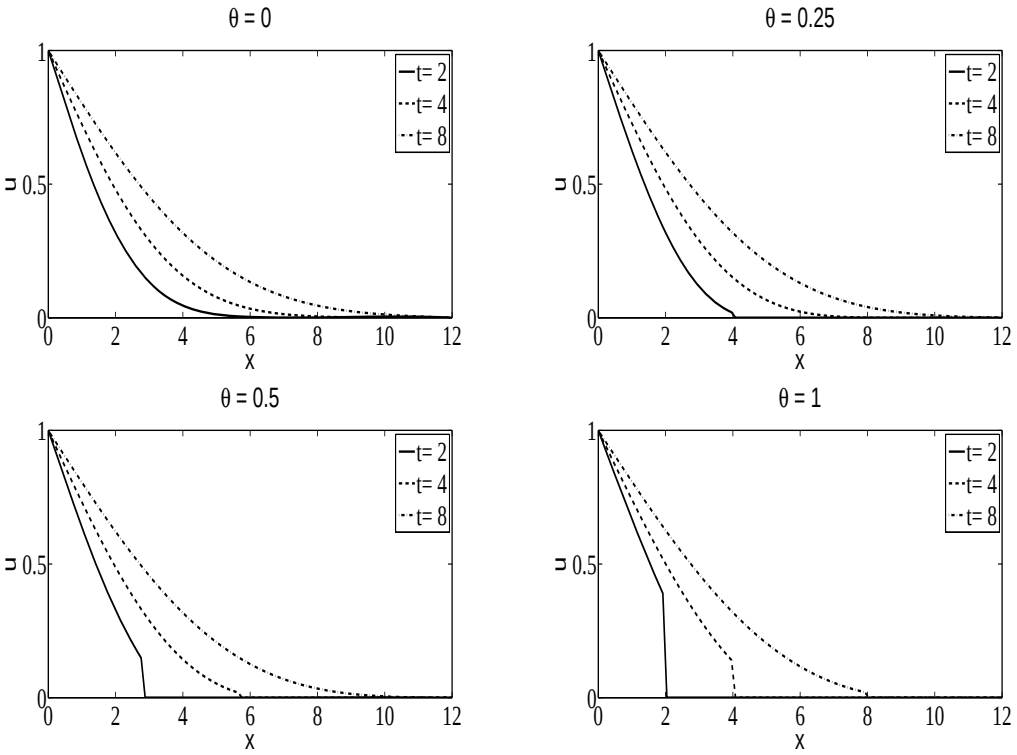

Figure 3.17: Approximate solution of Example 3.2.3 for $P = 0$ and $\Delta x = 0.12$.

Example 3.2.4. Let us consider the first problem given by

$$\theta\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}(P(x)u(x,t)) + D\frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in]0,\infty[\,, t > 0,$$

for non-constant P, with non-zero initial conditions

$$u(x,0) = e^{-x}, \qquad \theta\frac{\partial u}{\partial t}(x,0) = -e^{-x}$$

and boundary conditions

$$u(0,t) = 1, \qquad \lim_{x\to+\infty} u(x,t) = 0.$$

For this case we have a non-homogeneous differential equation in $\widetilde{u}(x,s)$ to solve, that is

$$\frac{d^2\widetilde{u}}{dx^2}(x,s) - \frac{P}{D}\frac{d\widetilde{u}}{dx}(x,s) - \left(\frac{P'}{D} + \lambda_s^2\right)\widetilde{u}(x,s) = -\frac{\mathrm{e}^{-x}}{D}\theta s. \tag{3.55}$$

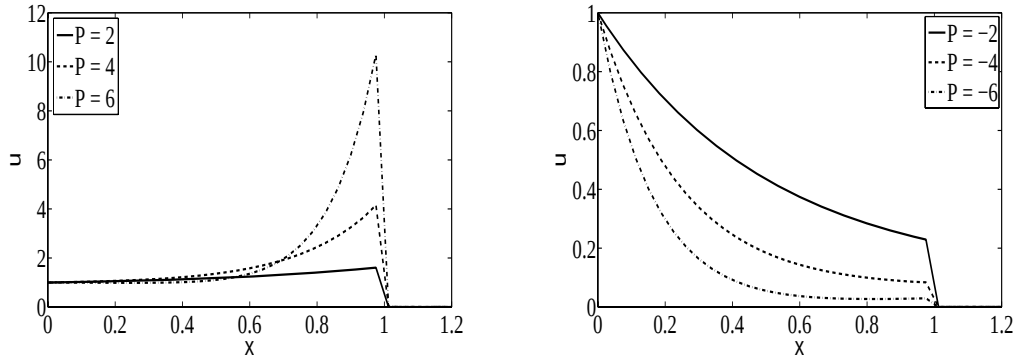

Figure 3.18: Approximate solution of Example 3.2.3 for $\theta = 1$, $\Delta x = 0.0375$ at $t = 1$.

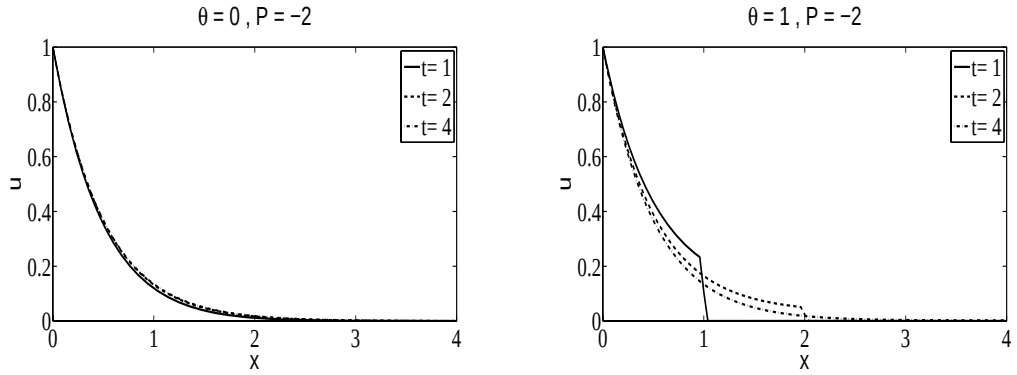

Figure 3.19: Approximate solution of Example 3.2.3 for P constant and $\Delta x = 0.04$.

We are able to compute a particular solution for this equation only for P constant, which is

$$\widetilde{u}_p(x,s) = -\frac{\theta s}{D + P - s(1+\theta s)} \mathrm{e}^{-x}.$$

Applying the boundary conditions we obtain the solution in the Laplace transform domain as

$$\widetilde{u}(x,s) = \left(\frac{1}{s} - c\right) \mathrm{e}^{\nu_s^- x} + c\mathrm{e}^{-x},$$

with $c = -\frac{\theta s}{D+P-s(1+\theta s)}$. In Figure 3.21, we display the three numerical methods for $\theta = 0.25$, $P = 2\mathrm{e}^{-x}$ at $t = 1$. The behavior of the solution is similar for all the methods.

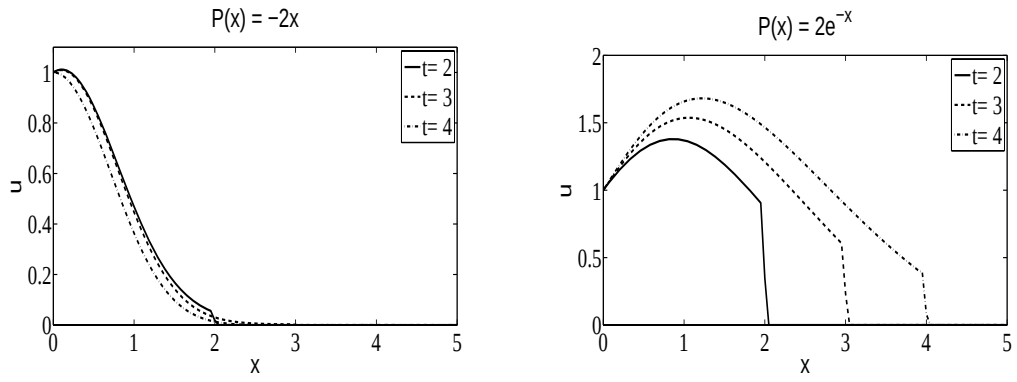

Figure 3.20: Approximate solution of Example 3.2.3 for $\theta = 1$, $\Delta x = 0.05$ and P non-constant.

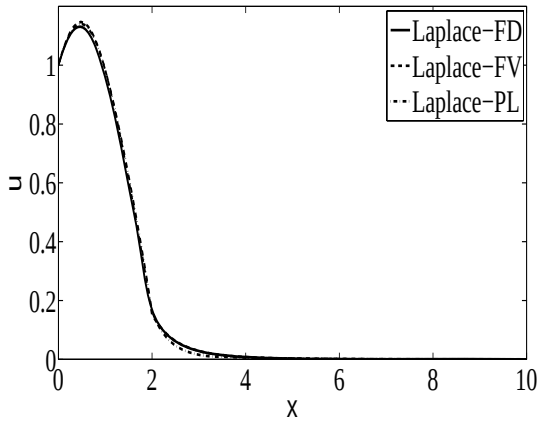

Figure 3.21: Approximate solution of Example 3.2.4 for $\theta = 0.25$, $P = 2\mathrm{e}^{-x}$ and $\Delta x = 0.08$ at $t = 1$.

Our last problem has also non-zero initial conditions, but in this case a particular solution of the non-homogeneous ordinary differential equation is not available.

Example 3.2.5. We consider the problem

$$\theta \frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -P\frac{\partial u}{\partial x}(x,t) + D\frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in \mathbb{R}, t > 0,$$

for P constant, with initial conditions

$$u(x,0) = \frac{1}{2}e^{-x^2}, \qquad \theta \frac{\partial u}{\partial t}(x,0) = xe^{-x^2}$$

and boundary conditions

$$\lim_{x\to-\infty} u(x,t) = 0, \qquad \lim_{x\to+\infty} u(x,t) = 0.$$

The corresponding non-homogeneous differential equation in $\widetilde{u}(x,s)$ is

$$\frac{d^2\widetilde{u}}{dx^2}(x,s) - \lambda_s^2\widetilde{u}(x,s) - \frac{P}{D}\frac{d\widetilde{u}}{dx}(x,s) = -\frac{\mathrm{e}^{-x^2}}{2D}(1+\theta s) - \frac{x\mathrm{e}^{-x^2}}{D}. \tag{3.56}$$

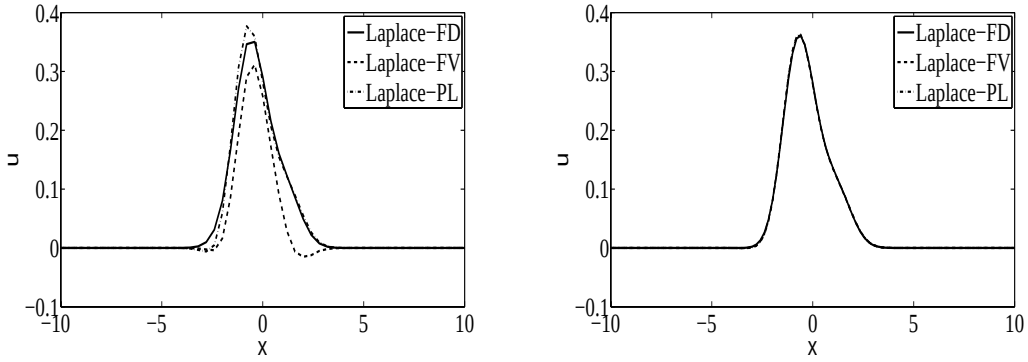

Figure 3.22: Approximate solution of Example 3.2.5 for $\theta = 0.5$, $P = -2$ at $t = 1$. Left figure: $\Delta x = 0.4$. Right figure: $\Delta x = 0.2$.

In this example the solution is smooth, with no discontinuities, and we observe in Figure 3.22 that for large space steps, such as $\Delta x = 0.4$, the Laplace-PL method performs more successfully. As we refine the mesh, by doing $\Delta x = 0.2$, the Laplace-FD method and the Laplace-FV method become similar to the Laplace-PL method.

3.3 Crank-Nicolson *vs* Laplace transform methods

The computational efficiency of a numerical method gives the relation between the computational cost of the method and the precision of the results obtained by the method.

In what follows, two test problems are considered in order to compare the performance of the Crank-Nicolson method with the Laplace transform finite difference method, described in Sections 2.1 and 3.2.1, respectively. We want to highlight the advantages of using the Laplace transform technique

in some specific problems instead of iterative methods in time. As mentioned before in Section 3.1, numerical methods that use the Laplace transform in time have the advantage of not iterating in time. This means that it is the same to compute the solution for short times or long times, while iterative methods in time, including methods such as the Crank-Nicolson, usually takes too long to compute the solution. We will show that the convergence order of both numerical methods is second order as evidenced by previous theoretical analysis. However, the computational efficiency is higher by applying the Laplace transform finite difference scheme. The computational cost of the algorithm [2] for the Laplace inversion process is $\mathcal{O}\left(\sum_{i=1}^{N} M_i^2\right)$, where M_i is the number of iterations performed by the algorithm for each x_i, $i = 1, \ldots, N$ with N the length of the domain in x. On the other hand, the computational cost for solving the linear system of equations (3.19) resulting from the finite difference discretization is approximately $\mathcal{O}(\overline{M}N)$, where we define $\overline{M}$ as the average of $M_i, i = 1, \ldots, N$. Therefore, the total cost of the Laplace transform finite difference method (Laplace-FD) is approximately

$$\mathcal{O}\left(\sum_{i=1}^{N} M_i^2 + \overline{M}N\right) \approx \mathcal{O}\left((\overline{M}^2 + \overline{M})N\right).$$

We will use only the Laplace-FD method although we obtain similar results for the Laplace-FV and Laplace-PL methods.

For each step of the Crank-Nicolson (CN) method, we need to solve a linear system with a tridiagonal matrix. This can be done efficiently in $\mathcal{O}(N)$ operations. If we consider $0 \leq t \leq T_f$ and $L = \frac{T_f}{\Delta t}$ the number of timesteps, then the computational cost of CN method is $\mathcal{O}(NL)$. Thus, since we expect $\mathcal{O}(\overline{M})$ to be smaller than $\mathcal{O}(N)$, it is expectable the computational cost of Laplace-FD scheme to be less than the computational cost of CN method. In order to measure the gain in efficiency, we consider the variable Gain given by

$$\text{Gain} = \frac{\#\text{ operations CN} - \#\text{ operations Laplace-FD}}{\max\{\#\text{ operations CN}\ ,\ \#\text{ operations Laplace-FD}\}}.$$

We remark that, in this section, the conclusions obtained are the same

Δx	CN	Rate	Laplace-FD	Rate	M	$\overline{M}$	Gain
10/128	0.2567×10^{-3}		0.2549×10^{-3}		14	9	29.7%
10/256	0.6427×10^{-4}	2.0	0.6601×10^{-4}	2.0	16	11	48.4%
10/512	0.1608×10^{-4}	2.0	0.1615×10^{-4}	2.0	18	13	64.5%
10/1024	0.4019×10^{-5}	2.0	0.4063×10^{-5}	2.0	20	15	76.6%
10/2048	0.1005×10^{-5}	2.0	0.1018×10^{-5}	2.0	22	17	85.1%

Table 3.6: Global error E_G of Example 3.1.1 for $P = 2$, $t = 1$, $0 \leq x \leq 10$, $TOL = 1/N^3$, $T = 3$, $\beta = -\ln(10^{-16})/2T$, $\Delta t = \Delta x/10$, computed with the norm ℓ_∞.

whether we use the norm ℓ_∞ or norm $\ell_{2,\Delta x}$. We present only the norm ℓ_∞, $E_G = \max |u_i - U_i|$, $i = 1, \ldots, N-1$, since it provides more precise results for both methods.

First, we consider the problem (3.15)-(3.16) from Example 3.1.1 which has the exact solution (3.18). In Table 3.6 and Figure 3.23 we show the rate of convergence and the global error for $u_0 = 1$, $P = 2$, $t = 1$ and different values of the space step. For the Laplace-FD method we also present the number of iterations $M = \max_i M_i$ and $\overline{M}$. As we can see, the Laplace-FD method is more efficient than the CN method. Furthermore, the advantage of the Laplace-FD method in the computational cost becomes clear as N increases, that is, as Δx decreases. In Table 3.7 and Figure 3.24 we increase the time to $t = 20$ and for the CN method we consider the same timestep Δt considered previously in Table 3.6. As expected, the efficiency of Laplace-FD method is more evident as shown by the values of the variable Gain. We also note that the computational cost of the inverse Laplace transform algorithm is reduced since the values of M and $\overline{M}$ decreased in comparison with Table 3.6. In Table 3.8 we also consider $t = 20$ but we increase the time step Δt. The Laplace-FD method is now slightly more accurate and is still with less computational effort than the CN method as shown in Figure 3.25.

For the second problem we consider an hyperbolic equation with $\theta = 1$ and $P = 0$: the problem (3.49)-(3.51) from Example 3.2.2 already discussed in Section 3.2.5. The rate of convergence and the global error are present in Table 3.9 for $t = 1$ and different space steps. Note that the solution of

Δx	CN	Rate	Laplace-FD	Rate	M	$\overline{M}$	Gain
70/128	0.3116×10^{-2}		0.3210×10^{-2}		12	8	97.2%
70/256	0.7788×10^{-3}	2.0	0.8620×10^{-3}	1.9	13	10	97.9%
70/512	0.1949×10^{-3}	2.0	0.1886×10^{-3}	2.2	14	12	98.5%
70/1024	0.4871×10^{-4}	2.0	0.4675×10^{-4}	2.0	16	13	99.1%
70/2048	0.1218×10^{-4}	2.0	0.1229×10^{-4}	1.9	17	15	99.4%

Table 3.7: Global error E_G of Example 3.1.1 for $P = 2$, $t = 20$, $0 \leq x \leq 70$, $TOL = 1/N^3$, $T = 30$, $\beta = -\ln(10^{-16})/2T$, $\Delta t = \Delta x/70$, computed with the norm ℓ_∞.

Δx	CN	Rate	Laplace-FD	Rate	M	$\overline{M}$	Gain
70/128	0.3589×10^{-2}		0.3210×10^{-2}		12	8	43.8%
70/256	0.8991×10^{-3}	2.0	0.8620×10^{-3}	1.9	13	10	57.0%
70/512	0.2249×10^{-3}	2.0	0.1886×10^{-3}	2.2	14	12	69.5%
70/1024	0.5622×10^{-4}	2.0	0.4675×10^{-4}	2.0	16	13	82.2%
70/2048	0.1406×10^{-4}	2.0	0.1229×10^{-4}	1.9	17	15	88.3%

Table 3.8: Global error E_G of Example 3.1.1 for $P = 2$, $t = 20$, $0 \leq x \leq 70$, $TOL = 1/N^3$, $T = 30$, $\beta = -\ln(10^{-16})/2T$, $\Delta t = 20\Delta x/70$, computed with the norm ℓ_∞.

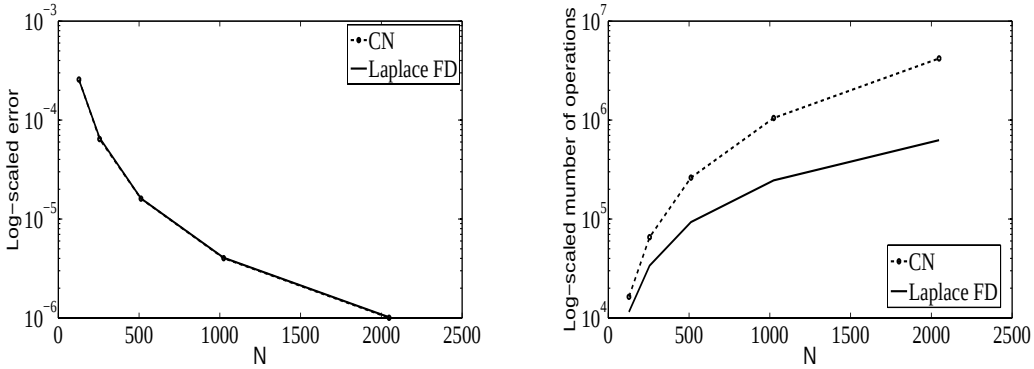

Figure 3.23: CN method *vs* Laplace-FD for $t = 1$. Left: global error. Right: total cost. (Table 3.6)

this problem is smoother than the solution of the previous problem which has an initial discontinuity in the corner $(x, t) = (0, 0)$. In this case the CN method behaves better than previously. However, the Laplace-FD method is still more efficient. In fact, for $t = 1$ the variable Gain gives an advantage to the CN method in some space steps, but then the superiority of the Laplace-

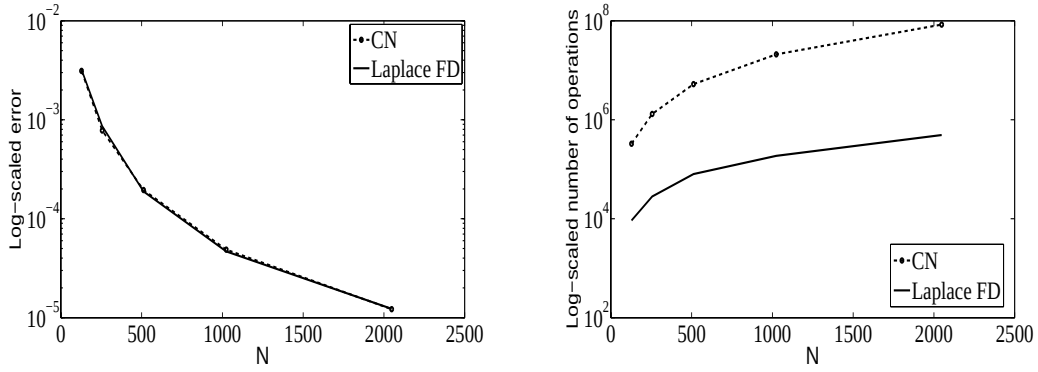

Figure 3.24: CN method *vs* Laplace-FD for $t = 20$. Left: global error. Right: total cost. (Table 3.7)

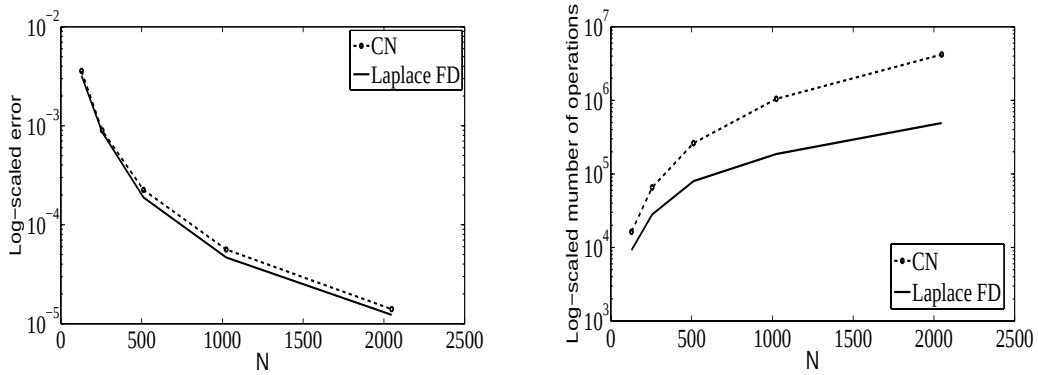

Figure 3.25: CN method *vs* Laplace-FD for $t = 20$. Left: global error. Right: total cost. (Table 3.8)

FD method becomes evident in Table 3.10 and Figure 3.27 where we increase the time to $t = 3$. Our conclusion is that the CN method performs well for short times.

We observe from the previous results the second order convergence of the numerical methods as predicted by the theoretical analysis for the main problem.

3.4 Numerical solution for a periodic potential

Based on the results published in [4], we now present the numerical solutions of a diffusion equation that involves the symmetric and periodic

Δx	CN	Rate	Laplace-FD	Rate	M	$\overline{M}$	Gain
$2\pi/128$	0.3420×10^{-5}		0.4860×10^{-5}		19	18	-62.6%
$2\pi/256$	0.8551×10^{-6}	2.0	0.8596×10^{-6}	2.5	21	20	-39.1%
$2\pi/512$	0.2138×10^{-6}	2.0	0.2338×10^{-6}	1.9	26	24	-14.7%
$2\pi/1024$	0.5357×10^{-7}	2.0	0.6272×10^{-7}	1.9	31	28	20.7%
$2\pi/2048$	0.1339×10^{-7}	2.0	0.1813×10^{-7}	1.8	37	33	45.2%

Table 3.9: Global error E_G of Example 3.2.2 for $t = 1$, $T = 5$, $0 \leq x \leq 2\pi$, $TOL = 1/N^3$, $\beta = -\ln(10^{-16})/2T$, $\Delta t = \Delta x/(2\pi)$, computed with the norm ℓ_∞.

Δx	CN	Rate	Laplace-FD	Rate	M	$\overline{M}$	Gain
$2\pi/128$	0.1218×10^{-4}		0.1045×10^{-4}		16	15	37.5%
$2\pi/256$	0.3044×10^{-5}	2.0	0.3368×10^{-5}	1.6	18	17	60.2%
$2\pi/512$	0.7610×10^{-6}	2.0	0.7792×10^{-6}	2.1	20	19	75.3%
$2\pi/1024$	0.1907×10^{-6}	2.0	0.1892×10^{-6}	2.0	23	22	83.5%
$2\pi/2048$	0.4767×10^{-7}	2.0	0.5016×10^{-7}	1.9	29	26	88.6%

Table 3.10: Global error E_G of Example 3.2.2 for $t = 3$, $T = 10$, $0 \leq x \leq 2\pi$, $TOL = 1/N^3$, $\beta = -\ln(10^{-16})/2T$, $\Delta t = \Delta x/(2\pi)$, computed with the norm ℓ_∞.

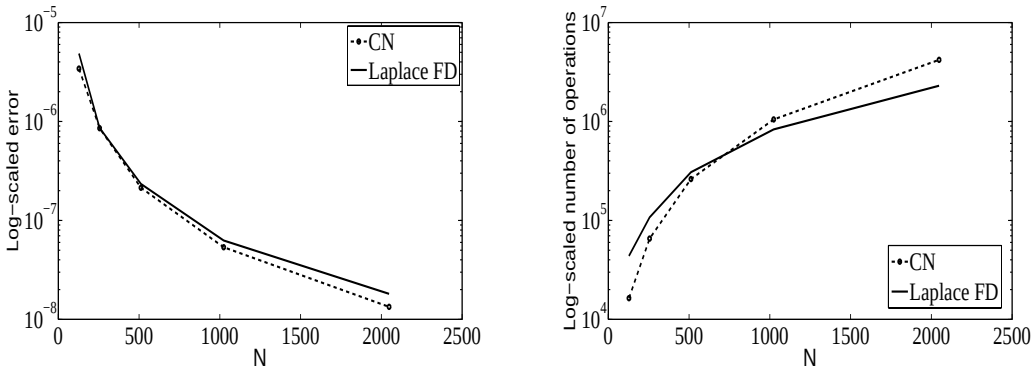

Figure 3.26: CN method *vs* Laplace-FD for $t = 1$. Left: global error. Right: total cost. (Table 3.9)

potential field, as previously studied in [12] and [62], given by

$$V(x;\alpha) = \frac{1}{J_0(\mathbf{i}\alpha)} e^{\alpha \cos(x)} - 1,$$

where $\mathbf{i}$ is the imaginary unit and $J_0(\mathbf{i}\alpha)$ is the Bessel function of the first kind of zero order. The parameter α controls the shape and height of the

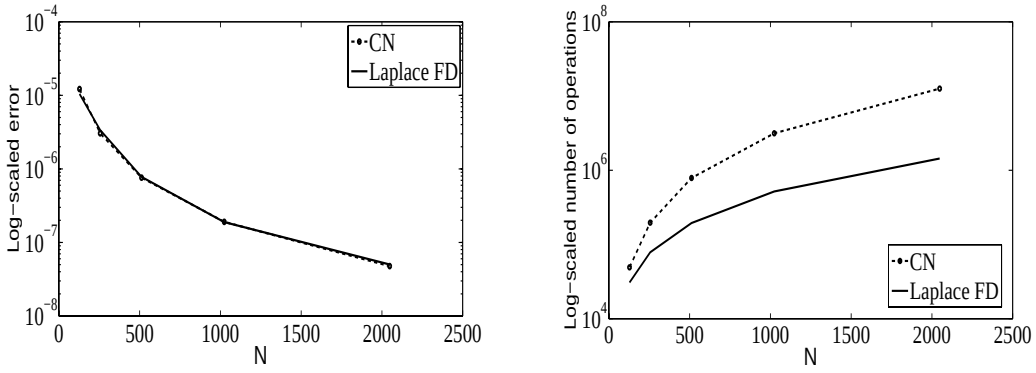

Figure 3.27: CN method *vs* Laplace-FD for $t = 3$. Left: global error. Right: total cost. (Table 3.10)

potential barrier which affects the diffusion behavior.

Example 3.4.1. We consider the following equation

$$\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}(P(x)u(x,t)) + \frac{\partial^2 u}{\partial x^2}(x,t),$$

with $P(x) = -\frac{dV}{dx}$. The initial conditions are

$$u\,(x,0) = \frac{1}{L\sqrt{\pi}} e^{-x^2/L^2}, \qquad \frac{\partial u}{\partial t}\,(x,0) = 0,$$

and the boundary conditions are given by

$$\lim_{x\to-\infty} u(x,t) = 0 \quad \text{and} \quad \lim_{x\to+\infty} u(x,t) = 0.$$

We illustrate the flexible form of $P(x)$ in Figure 3.28 for two values of the parameter α, $\alpha = 1$ and $\alpha = 16$. We observe that $P(x)$ for $\alpha = 1$ changes between -1 and 1, whereas for $\alpha = 16$ changes between -20 and 20 and the change is not smooth. As already seen in Section 2.3 the CN method does not give good results for $\alpha = 16$. The Laplace-FD can deal very well with both cases.

For $\alpha = 1$ the behavior of the solution can be observed in Figures 3.29 and 3.30 as we increase time from $t = 1$ until $t = 500$. The peak starts to split into two and then into several waves forming, that goes to the right and left. The domain where the function is not zero becomes larger as we travel in time. For that reason the computational domain increases considerably which

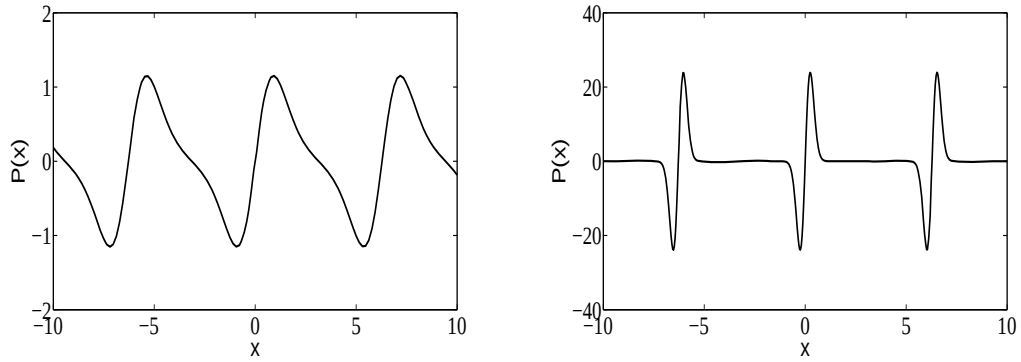

Figure 3.28: $P(x)$ from Example 3.4.1. Left: $\alpha = 1$. Right: $\alpha = 16$.

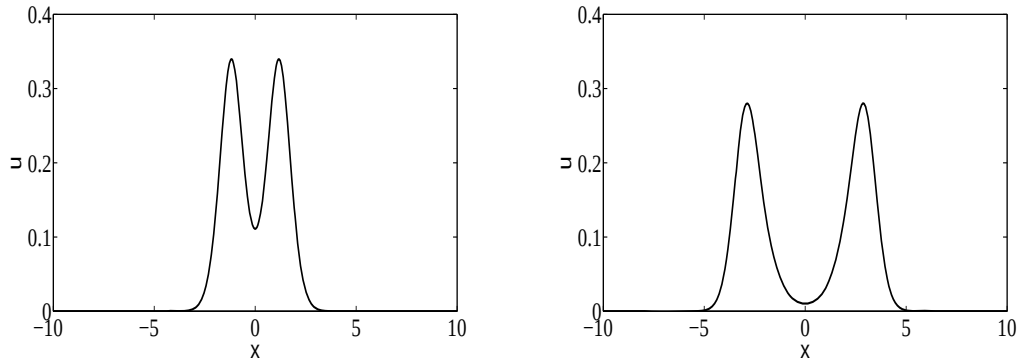

Figure 3.29: Approximate solution $u(x,t)$ of Example 3.4.1 for $\alpha = 1$. Left: curve for instant of time $t = 1$. Right: curve for instant of time $t = 3$.

requires more computational effort regarding the discretization in space. For an iterative method where we need to consider a discretization in time, it would require more computational effort for long times. This is the reason why the Laplace-FD is applied in this example.

For $\alpha = 16$ the evolution of the solution is considered in Figure 3.31 in the first instants of time. We observe the solution presents very steep gradients and the Laplace-FD is able to give accurate solutions. In Figure 3.32 the behavior of the solution is presented for long times. Since the method uses the Laplace transform technique, it is able to give very quickly and accurate solutions for very large times. Once again, this is evident in Figure 3.33, where $t = 5000$ and $t = 10000$.

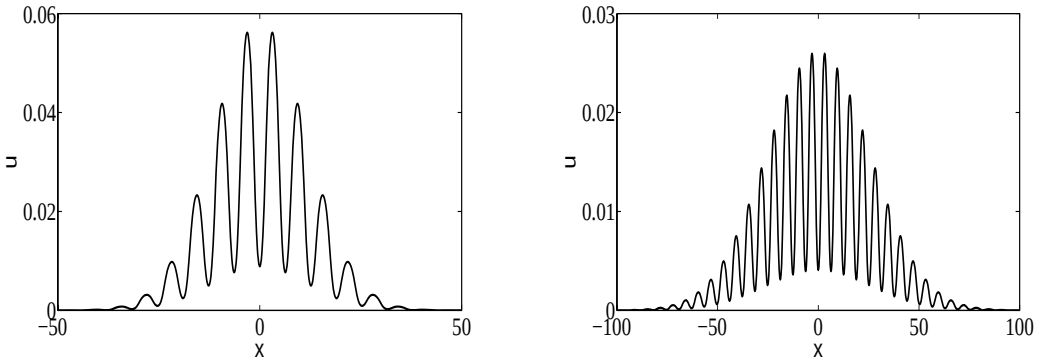

Figure 3.30: Approximate solution $u(x,t)$ of Example 3.4.1 for $\alpha = 1$. Left: curve for instant of time $t = 100$. Right: curve for instant of time $t = 500$.

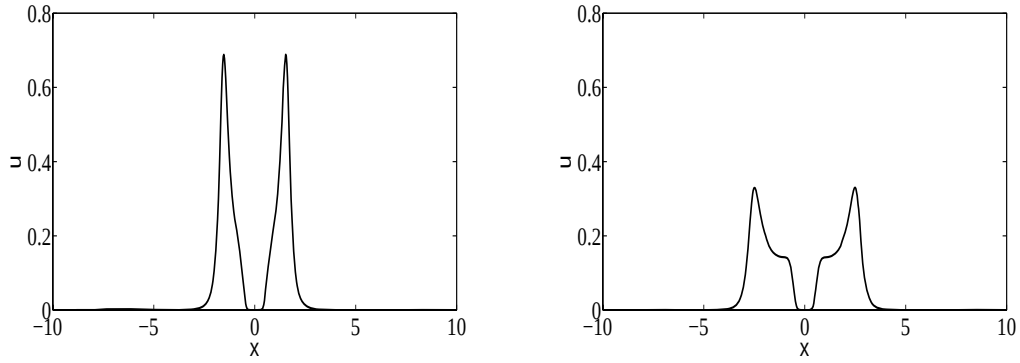

Figure 3.31: Approximate solution $u(x,t)$ of Example 3.4.1 for $\alpha = 16$. Left: curve for instant of time $t = 1$. Right: curve for instant of time $t = 2$.

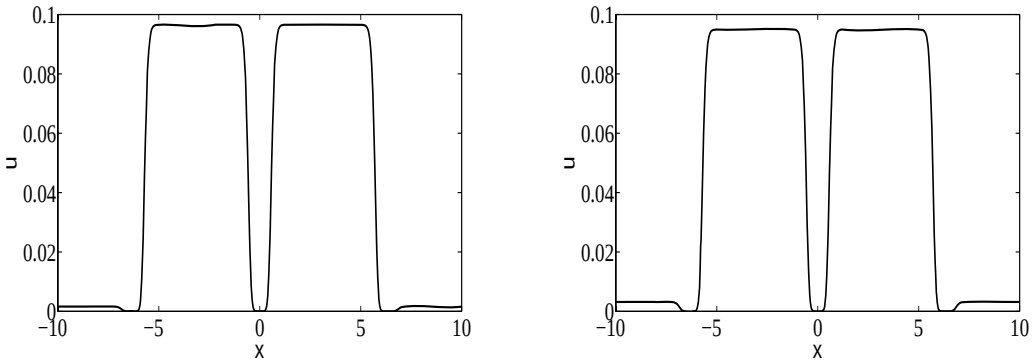

Figure 3.32: Approximate solution $u(x,t)$ of Example 3.4.1 for $\alpha = 16$. Left: curve for instant of time $t = 500$. Right: curve for instant of time $t = 1000$.

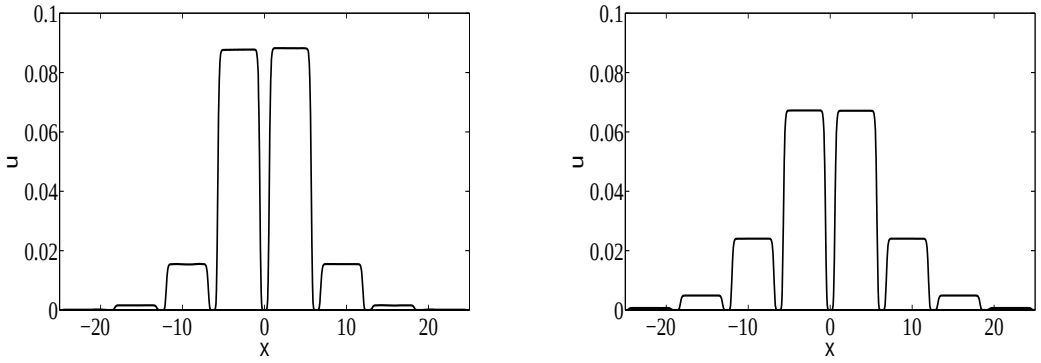

Figure 3.33: Approximate solution $u(x,t)$ of Example 3.4.1 for $\alpha = 16$. Left: curve for instant of time $t = 5000$. Right: curve for instant of time $t = 10000$.

Chapter 4

A two dimensional hyperbolic diffusion equation

In this chapter we extend to two dimensional problems the numerical methods presented in the previous chapters. The generalization of the Laplace-PL method is not straightforward because the solution (3.34) is not easily attainable in two dimensions. Hereby, we turn our attention to the CN, Laplace-FD and Laplace-FV methods described in Section 4.1. However, the methods obtained are computational inefficient for a large number of the discretization points of the spatial variables. The need to find a computationally efficient method led us to develop an alternating direction implicit (ADI) method based on Crank-Nicolson scheme. As already explained in Section 1.3.2, this technique has been used to solve problems in two dimensions although we did not find in the literature a numerical method suitable for our equation (1.10). Deduction, theoretical analysis and performance of this numerical method described in Section 4.2 are all new contributions.

4.1 Extension of the numerical methods to two dimensions

The numerical methods described for problems in one dimension are extended in this section to solve the two dimensional hyperbolic diffusion

equation (1.10), defined in a rectangular domain $\Omega \subset \mathbb{R}^2$,

$$\begin{aligned}\theta\frac{\partial^2 u}{\partial t^2}(x,y,t)+\frac{\partial u}{\partial t}(x,y,t) = & - \frac{\partial}{\partial x}(P(x,y)u(x,y,t))-\frac{\partial}{\partial y}(Q(x,y)u(x,y,t)) \\ & + D\frac{\partial^2 u}{\partial x^2}(x,y,t)+D\frac{\partial^2 u}{\partial y^2}(x,y,t), \end{aligned} \tag{4.1}$$

$(x,y)\in\Omega$, $t>0$. We consider the initial conditions given by

$$u(x,y,0) = u_0(x,y), \qquad (x,y)\in\Omega, \tag{4.2}$$

$$\theta\frac{\partial u}{\partial t}(x,y,0) = u_1(x,y), \qquad (x,y)\in\Omega, \tag{4.3}$$

and Dirichlet boundary conditions

$$u(x,y,t)=f(x,y,t), \qquad (x,y)\in\partial\Omega,\ t>0. \tag{4.4}$$

4.1.1 The Crank-Nicolson method

Similarly to what was done in the one dimensional case, Section 2.1, we introduce the auxiliary function

$$w=\theta\frac{\partial u}{\partial t}+u \tag{4.5}$$

and change (4.1) into

$$\begin{aligned}\frac{\partial w}{\partial t} &= -\frac{\partial}{\partial x}(Pu)-\frac{\partial}{\partial y}(Qu)+D\frac{\partial^2 u}{\partial x^2}+D\frac{\partial^2 u}{\partial y^2} \\ &= -P_x u - P\frac{\partial u}{\partial x}-Q_y u - Q\frac{\partial u}{\partial y}+D\frac{\partial^2 u}{\partial x^2}+D\frac{\partial^2 u}{\partial y^2}, \end{aligned} \tag{4.6}$$

where P_x denotes the derivative of $P(x,y)$ in the variable x and Q_y denotes the derivative of $Q(x,y)$ in the variable y. We consider the mesh points in $\Omega=[a,b]\times[c,d]$ given by $x_i=a+i\Delta x$, $i=0,\ldots,N_x$, $y_j=c+j\Delta y$, $j=0,\ldots,N_y$, with $\Delta x=(b-a)/N_x$ and $\Delta y=(d-c)/N_y$, where N_x and N_y are positive integers. For $0\le t\le T_f$, let $t_n=n\Delta t$, with Δt being the time increment and $n\Delta t\le T_f$. We denote the approximate solutions to $u(x_i,y_j,t_n)$ and $w(x_i,y_j,t_n)$ by $U_{i,j}^n$ and $W_{i,j}^n$ respectively. We also denote $P(x_i,y_j)$ by $P_{i,j}$, $P_x(x_i,y_j)$ by $(P_x)_{i,j}$, $Q(x_i,y_j)$ by $Q_{i,j}$ and $Q_y(x_i,y_j)$ by $(Q_y)_{i,j}$. We discretize equations (4.5) and (4.6) using the Crank-Nicolson method:

$$W_{i,j}^{n+1}+W_{i,j}^n = U_{i,j}^{n+1}+U_{i,j}^n+\frac{2\theta}{\Delta t}\left(U_{i,j}^{n+1}-U_{i,j}^n\right) \tag{4.7}$$

and

$$\begin{aligned}
&\frac{W_{i,j}^{n+1}-W_{i,j}^{n}}{\Delta t}\\
&\quad=-\frac{(P_x)_{i,j}}{2}(U_{i,j}^{n+1}+U_{i,j}^{n})-\frac{P_{i,j}}{2}\left[\frac{U_{i+1,j}^{n+1}-U_{i-1,j}^{n+1}}{2\Delta x}+\frac{U_{i+1,j}^{n}-U_{i-1,j}^{n}}{2\Delta x}\right]\\
&\quad-\frac{(Q_y)_{i,j}}{2}(U_{i,j}^{n+1}+U_{i,j}^{n})-\frac{Q_{i,j}}{2}\left[\frac{U_{i,j+1}^{n+1}-U_{i,j-1}^{n+1}}{2\Delta y}+\frac{U_{i,j+1}^{n}-U_{i,j-1}^{n}}{2\Delta y}\right]\\
&\quad+\frac{D}{2}\left[\frac{U_{i-1,j}^{n+1}-2U_{i,j}^{n+1}+U_{i+1,j}^{n+1}}{\Delta x^2}+\frac{U_{i-1,j}^{n}-2U_{i,j}^{n}+U_{i+1,j}^{n}}{\Delta x^2}\right]\\
&\quad+\frac{D}{2}\left[\frac{U_{i,j-1}^{n+1}-2U_{i,j}^{n+1}+U_{i,j+1}^{n+1}}{\Delta y^2}+\frac{U_{i,j-1}^{n}-2U_{i,j}^{n}+U_{i,j+1}^{n}}{\Delta y^2}\right].
\end{aligned}\tag{4.8}$$

To write the scheme (4.7)-(4.8) in matrix form we solve equation (4.7) for $W_{i,j}^{n+1}$ and get

$$W_{i,j}^{n+1}=\left(1+\frac{2\theta}{\Delta t}\right)U_{i,j}^{n+1}+\left(1-\frac{2\theta}{\Delta t}\right)U_{i,j}^{n}-W_{i,j}^{n}.\tag{4.9}$$

Substituting (4.9) into (4.8) gives

$$\begin{aligned}
&\frac{1}{\Delta t}\left[\left(1+\frac{2\theta}{\Delta t}\right)U_{i,j}^{n+1}+\left(1-\frac{2\theta}{\Delta t}\right)U_{i,j}^{n}-2W_{i,j}^{n}\right]\\
&\quad=-\frac{(P_x)_{i,j}}{2}(U_{i,j}^{n+1}+U_{i,j}^{n})-\frac{P_{i,j}}{4\Delta x}\left(U_{i+1,j}^{n+1}-U_{i-1,j}^{n+1}+U_{i+1,j}^{n}-U_{i-1,j}^{n}\right)\\
&\quad-\frac{(Q_y)_{i,j}}{2}(U_{i,j}^{n+1}+U_{i,j}^{n})-\frac{Q_{i,j}}{4\Delta y}\left(U_{i,j+1}^{n+1}-U_{i,j-1}^{n+1}+U_{i,j+1}^{n}-U_{i,j-1}^{n}\right)\\
&\quad+\frac{D}{2\Delta x^2}\left[\left(U_{i-1,j}^{n+1}-2U_{i,j}^{n+1}+U_{i+1,j}^{n+1}\right)+\left(U_{i-1,j}^{n}-2U_{i,j}^{n}+U_{i+1,j}^{n}\right)\right]\\
&\quad+\frac{D}{2\Delta y^2}\left[\left(U_{i,j-1}^{n+1}-2U_{i,j}^{n+1}+U_{i,j+1}^{n+1}\right)+\left(U_{i,j-1}^{n}-2U_{i,j}^{n}+U_{i,j+1}^{n}\right)\right].
\end{aligned}$$

After simplification we can write

$$\begin{aligned}
&A_{i,j}^{1}U_{i-1,j}^{n+1}+A_{i,j}^{2}U_{i,j}^{n+1}+A_{i,j}^{3}U_{i+1,j}^{n+1}+A_{i,j}^{4}U_{i,j-1}^{n+1}+A_{i,j}^{5}U_{i,j+1}^{n+1}\\
&\quad=B_{i,j}^{1}U_{i-1,j}^{n}+B_{i,j}^{2}U_{i,j}^{n}+B_{i,j}^{3}U_{i+1,j}^{n}+B_{i,j}^{4}U_{i,j-1}^{n}+B_{i,j}^{5}U_{i,j+1}^{n}\\
&\qquad+\frac{2}{\Delta t}W_{i,j}^{n},
\end{aligned}\tag{4.10}$$

where

$$A^1_{i,j} = -\frac{P_{i,j}}{4\Delta x} - \frac{D}{2\Delta x^2}, \qquad B^1_{i,j} = \frac{P_{i,j}}{4\Delta x} + \frac{D}{2\Delta x^2},$$

$$A^3_{i,j} = \frac{P_{i,j}}{4\Delta x} - \frac{D}{2\Delta x^2}, \qquad B^3_{i,j} = -\frac{P_{i,j}}{4\Delta x} + \frac{D}{2\Delta x^2},$$

$$A^4_{i,j} = -\frac{Q_{i,j}}{4\Delta y} - \frac{D}{2\Delta y^2}, \qquad B^4_{i,j} = \frac{Q_{i,j}}{4\Delta y} + \frac{D}{2\Delta y^2},$$

$$A^5_{i,j} = \frac{Q_{i,j}}{4\Delta y} - \frac{D}{2\Delta y^2}, \qquad B^5_{i,j} = -\frac{Q_{i,j}}{4\Delta y} + \frac{D}{2\Delta y^2},$$

$$A^2_{i,j} = \frac{1}{\Delta t}\left(1 + \frac{2\theta}{\Delta t}\right) + \frac{(P_x)_{i,j}}{2} + \frac{(Q_y)_{i,j}}{2} + \frac{D}{\Delta x^2} + \frac{D}{\Delta y^2},$$

$$B^2_{i,j} = -\frac{1}{\Delta t}\left(1 - \frac{2\theta}{\Delta t}\right) - \frac{(P_x)_{i,j}}{2} - \frac{(Q_y)_{i,j}}{2} - \frac{D}{\Delta x^2} - \frac{D}{\Delta y^2},$$

for $i = 1, \ldots, N_x - 1, j = 1, \ldots, N_y - 1$. Equation (4.10) will be used to compute $U^{n+1}_{i,j}$. After that, $U^{n+1}_{i,j}$ is substituted into (4.9) to compute $W^{n+1}_{i,j}$.

From (4.9) and (4.10) we obtain the system

$$\begin{cases} AU^{n+1} = BU^n + \dfrac{2}{\Delta t}W^n + d \\ W^{n+1} = \left(1 + \dfrac{2\theta}{\Delta t}\right)U^{n+1} + \left(1 - \dfrac{2\theta}{\Delta t}\right)U^n - W^n \end{cases}, \tag{4.11}$$

where A and B are band matrixes of size $((N_x-1)\times(N_y-1))^2$ with bandwidth five, U^{n+1}, U^n, W^{n+1} and W^n are column matrixes of size $(N_x-1)\times(N_y-1)\times 1$ and d is the vector that contains appropriate initial and boundary conditions.

We note that the convergence of the difference scheme is obtained in the same way as we did for the one dimensional case in Section 2.1. Thus, this is an implicit numerical method of order $\mathcal{O}(\Delta t^2 + \Delta x^2 + \Delta y^2)$.

4.1.2 A Laplace transform finite difference method

We now consider the extension of the Laplace transform finite difference method to equation (4.1). Similarly to what was done in one dimension, we apply the Laplace transform to remove the time dependent terms and obtain

the equation

$$\frac{\partial^2\widetilde{u}}{\partial x^2}+\frac{\partial^2\widetilde{u}}{\partial y^2}-\lambda_s^2\widetilde{u}-\frac{\partial}{\partial x}\left(\frac{P}{D}\widetilde{u}\right)-\frac{\partial}{\partial y}\left(\frac{Q}{D}\widetilde{u}\right)=-\frac{u_0(x,y)}{D}(1+\theta s)-\frac{u_1(x,y)}{D}, \quad (4.12)$$

with $\widetilde{u}(x,y,s)$ the Laplace transform of $u(x,y,t)$ and $\lambda_s^2=(\theta s^2+s)/D$. For a fixed s, the finite difference scheme is given by

$$\begin{aligned}&\frac{\widetilde{U}_{i-1,j}(s)-2\widetilde{U}_{i,j}(s)+\widetilde{U}_{i+1,j}(s)}{\Delta x^2}+\frac{\widetilde{U}_{i,j-1}(s)-2\widetilde{U}_{i,j}(s)+\widetilde{U}_{i,j+1}(s)}{\Delta y^2}-\lambda_s^2\widetilde{U}_{i,j}(s)\\&\quad-\frac{P_{i+1,j}\widetilde{U}_{i+1,j}(s)-P_{i-1,j}\widetilde{U}_{i-1,j}(s)}{2D\Delta x}-\frac{Q_{i,j+1}\widetilde{U}_{i,j+1}(s)-Q_{i,j-1}\widetilde{U}_{i,j-1}(s)}{2D\Delta y}\\&\quad=-\frac{1}{D}\left(u_0(x_i,y_j)(1+\theta s)+u_1(x_i,y_j)\right), \quad (4.13)\end{aligned}$$

for $i=1,\ldots,N_x-1$, $j=1,\ldots,N_y-1$, $P_{i,j}=P(x_i,y_j)$ and $Q_{i,j}=Q(x_i,y_j)$. This discretized equation can be written in the form

$$\begin{aligned}&K_W\widetilde{U}_{i-1,j}+K_O\widetilde{U}_{i,j}+K_E\widetilde{U}_{i+1,j}+K_S\widetilde{U}_{i,j-1}+K_N\widetilde{U}_{i,j+1}\\&\quad=-\frac{1}{D}\left(u_0(x_i,y_j)(1+\theta s)+u_1(x_i,y_j)\right),\end{aligned}$$

where the coefficients are given by

$$\begin{aligned}K_W&=\frac{1}{\Delta x^2}+\frac{P_{i-1,j}}{2D\Delta x}, &K_O&=-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}-\lambda_s^2,\\K_E&=\frac{1}{\Delta x^2}-\frac{P_{i+1,j}}{2D\Delta x}, &K_S&=\frac{1}{\Delta y^2}+\frac{Q_{i,j-1}}{2D\Delta y},\\K_N&=\frac{1}{\Delta y^2}-\frac{Q_{i,j+1}}{2D\Delta y}.\end{aligned}$$

We obtain the linear system

$$K(s)\widetilde{U}(s)=\widetilde{b}(s), \quad (4.14)$$

where $K(s)$ is a band matrix of size $((N_x-1)\times(N_y-1))^2$, $\widetilde{U}(s)$ a vector of size $(N_x-1)\times(N_y-1)\times 1$ and $\widetilde{b}(s)$ a vector that contains the appropriate initial and boundary conditions.

It follows directly, by doing Taylor expansions, that this finite difference scheme has accuracy order in space given by $\mathcal{O}(\Delta x^2+\Delta y^2)$.

4.1.3 A Laplace transform finite volume method

The Laplace transform finite volume method presented in Section 3.2.2 is also applied in two dimensions for the spatial discretization of equation (4.12). We consider the control volume Ω, see Figure 4.1, where

$$\Omega = [x_i - \Delta x/2, x_i + \Delta x/2] \times [y_j - \Delta y/2, y_j + \Delta y/2]\,,$$

$i = 1, \ldots, N_x - 1$, $j = 1, \ldots, N_y - 1$. Note that the point O represents (x_i, y_j).

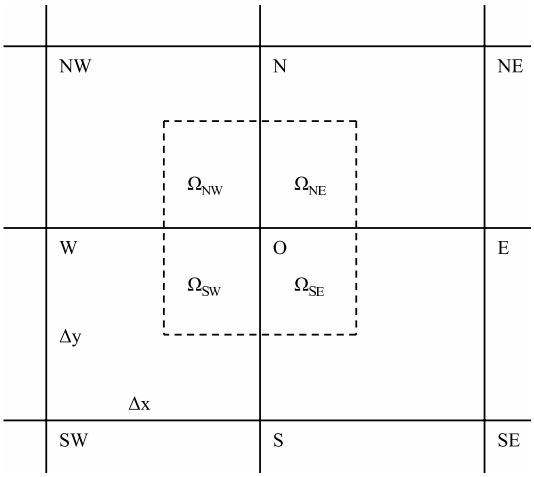

Figure 4.1: Control volume Ω.

We integrate the ordinary differential equation (4.12) within the control volume Ω, that is,

$$\begin{aligned}\int_\Omega \frac{\partial^2 \widetilde{u}}{\partial x^2} + \frac{\partial^2 \widetilde{u}}{\partial y^2} - \lambda_s^2 \widetilde{u} - \frac{\partial}{\partial x}\left(\frac{P}{D}\widetilde{u}\right) - \frac{\partial}{\partial y}\left(\frac{Q}{D}\widetilde{u}\right) dxdy \\ = -\frac{1}{D}\int_\Omega (1+\theta s)u_0(x,y) + u_1(x,y)dxdy. \qquad (4.15)\end{aligned}$$

The domain Ω is subdivided in four rectangular elements as shown in Figure 4.1. To derive the discretization, we approximate $\widetilde{u}(x, y, s)$ in terms of the nodal points and the shape functions in each element. Next, four shape functions are chosen in a similar way to what was done for the one dimensional case, as explained in [13]. For the element Ω_{NE}, and assuming

O represents the point (x_i, y_j), the shape functions are given by

$$\begin{aligned}
N_O(x,y,s) &= \frac{1}{\sinh(\mu\Delta x)\sinh(\mu\Delta y)}\sinh(\mu(x_{i+1}-x))\sinh(\mu(y_{j+1}-y)),\\
N_E(x,y,s) &= \frac{1}{\sinh(\mu\Delta x)\sinh(\mu\Delta y)}\sinh(\mu(x-x_i))\sinh(\mu(y_{j+1}-y)),\\
N_N(x,y,s) &= \frac{1}{\sinh(\mu\Delta x)\sinh(\mu\Delta y)}\sinh(\mu(x_{i+1}-x))\sinh(\mu(y-y_j)),\\
N_{NE}(x,y,s) &= \frac{1}{\sinh(\mu\Delta x)\sinh(\mu\Delta y)}\sinh(\mu(x-x_i))\sinh(\mu(y-y_j)),
\end{aligned}$$

where $\mu = \lambda_s/\sqrt{2}$. For this element the solution is then approximated by

$$\begin{aligned}
\widetilde{U}(x,y,s) &= N_O(x,y,s)\widetilde{U}_{i,j} + N_E(x,y,s)\widetilde{U}_{i+1,j} + N_N(x,y,s)\widetilde{U}_{i,j+1}\\
&\quad + N_{NE}(x,y,s)\widetilde{U}_{i+1,j+1}.
\end{aligned}$$

For the other three elements $\widetilde{U}(x,y,s)$ can be represented in a similar way.

We compute the integral on the right hand side of equation (4.15) by the midpoint rule and obtain

$$\begin{aligned}
&\frac{1}{D}\int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}}\int_{y_j-\frac{\Delta y}{2}}^{y_j+\frac{\Delta y}{2}}((1+\theta s)u_0(x,y)+u_1(x,y))dxdy\\
&\quad\simeq \frac{\Delta x\Delta y}{D}\left[(1+\theta s)u_0(x_i,y_j)+u_1(x_i,y_j)\right].
\end{aligned}$$

In order to ensure the viability of the calculation of the first integral in (4.15), we consider $P(x,y) = P(x)$ and $Q(x,y) = Q(y)$. After the integration of the first member of equation (4.15), the complete discretized equation that corresponds to node O is obtained by the contribution of all the four elements and is given by

$$\begin{aligned}
&K_O\widetilde{U}_{i,j} + K_E\widetilde{U}_{i+1,j} + K_W\widetilde{U}_{i-1,j} + K_N\widetilde{U}_{i,j+1} + K_S\widetilde{U}_{i,j-1} + K_{NE}\widetilde{U}_{i+1,j+1}\\
&\quad + K_{NW}\widetilde{U}_{i-1,j+1} + K_{SE}\widetilde{U}_{i+1,j-1} + K_{SW}\widetilde{U}_{i-1,j-1}\\
&= -\frac{\Delta x\Delta y}{D}\sinh(\mu\Delta x)\sinh(\mu\Delta y)\left((1+\theta s)u_0(x_i,y_j)+u_1(x_i,y_j)\right),
\end{aligned}$$

where the coefficients are given by

$$
\begin{aligned}
K_O &= 4[\cosh(\mu\Delta x)\cosh(\mu\Delta y/2) + \cosh(\mu\Delta y)\cosh(\mu\Delta x/2)] \\
&\quad -8\cosh(\mu\Delta x)\cosh(\mu\Delta y) \\
&\quad +\frac{2}{\mu}(P_{i+1/2} - P_{i-1/2})\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y) - \cosh(\mu\Delta y/2)) \\
&\quad +\frac{2}{\mu}(Q_{j+1/2} - Q_{j-1/2})\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x) - \cosh(\mu\Delta x/2)),
\end{aligned}
$$

$$
\begin{aligned}
K_E &= 2[2\cosh(\mu\Delta y) - \cosh(\mu\Delta y/2) - \cosh(\mu\Delta x/2)\cosh(\mu\Delta y)] \\
&\quad +\frac{2}{\mu}P_{i+1/2}\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y) - \cosh(\mu\Delta y/2)) \\
&\quad +\frac{1}{\mu}(Q_{j+1/2} - Q_{j-1/2})\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x/2) - 1),
\end{aligned}
$$

$$
\begin{aligned}
K_W &= 2[2\cosh(\mu\Delta y) - \cosh(\mu\Delta y/2) - \cosh(\mu\Delta x/2)\cosh(\mu\Delta y)] \\
&\quad -\frac{2}{\mu}P_{i-1/2}\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y) - \cosh(\mu\Delta y/2)) \\
&\quad +\frac{1}{\mu}(Q_{j+1/2} - Q_{j-1/2})\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x/2) - 1),
\end{aligned}
$$

$$
\begin{aligned}
K_N &= 2[2\cosh(\mu\Delta x) - \cosh(\mu\Delta x/2) - \cosh(\mu\Delta x)\cosh(\mu\Delta y/2)] \\
&\quad +\frac{1}{\mu}(P_{i+1/2} - P_{i-1/2})\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y/2) - 1) \\
&\quad +\frac{2}{\mu}Q_{j+1/2}\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x) - \cosh(\mu\Delta x/2)),
\end{aligned}
$$

$$
\begin{aligned}
K_S &= 2[2\cosh(\mu\Delta x) - \cosh(\mu\Delta x/2) - \cosh(\mu\Delta x)\cosh(\mu\Delta y/2)] \\
&\quad +\frac{1}{\mu}(P_{i+1/2} - P_{i-1/2})\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y/2) - 1) \\
&\quad -\frac{2}{\mu}Q_{j-1/2}\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x) - \cosh(\mu\Delta x/2)),
\end{aligned}
$$

$$
\begin{aligned}
K_{NE} &= [\cosh(\mu\Delta x/2) + \cosh(\mu\Delta y/2) - 2] \\
&\quad +\frac{1}{\mu}P_{i+1/2}\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y/2) - 1) \\
&\quad +\frac{1}{\mu}Q_{j+1/2}\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x/2) - 1),
\end{aligned}
$$

$$
\begin{aligned}
K_{NW} &= [\cosh(\mu\Delta x/2) + \cosh(\mu\Delta y/2) - 2] \\
&\quad -\frac{1}{\mu}P_{i-1/2}\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y/2) - 1) \\
&\quad +\frac{1}{\mu}Q_{j+1/2}\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x/2) - 1),
\end{aligned}
$$

$$
\begin{aligned}
K_{SE} &= [\cosh(\mu\Delta x/2) + \cosh(\mu\Delta y/2) - 2] \\
&\quad +\frac{1}{\mu}P_{i+1/2}\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y/2) - 1) \\
&\quad -\frac{1}{\mu}Q_{j-1/2}\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x/2) - 1),
\end{aligned}
$$

$$
\begin{aligned}
K_{SW} &= [\cosh(\mu\Delta x/2) + \cosh(\mu\Delta y/2) - 2] \\
&\quad -\frac{1}{\mu}P_{i-1/2}\sinh(\mu\Delta x/2)(\cosh(\mu\Delta y/2) - 1) \\
&\quad -\frac{1}{\mu}Q_{j-1/2}\sinh(\mu\Delta y/2)(\cosh(\mu\Delta x/2) - 1).
\end{aligned}
$$

As for the Laplace transform finite difference method, we also obtain the linear system (4.14)

$$K(s)\widetilde{U}(s) = \widetilde{b}(s).$$

The consistency of the numerical method follows by doing Taylor series expansions as we did in Section 3.2.2. Using convenient and practical algebraic manipulations we can show that this finite volume difference scheme has accuracy order $\mathcal{O}(\Delta x^2 + \Delta y^2)$.

4.1.4 Comparison of performance

Three different numerical methods were developed to get an approximate solution of equation (4.1) in two dimensions: the Crank-Nicolson method (CN-2D), the Laplace transform finite difference method (Laplace-FD-2D) and the Laplace transform finite volume method (Laplace-FV-2D). In the next numerical example, for which we are able to determine the analytical solution, we compare their performance. At first we compute the errors and the convergence rate and then we show how the solution behaves.

$\Delta x = \Delta y$	CN-2D	Rate	Laplace-FD-2D	Rate	Laplace-FV-2D	Rate
$\sqrt{8}\pi/40$	0.2611×10^{-3}		0.3700×10^{-2}		0.3700×10^{-2}	
$\sqrt{8}\pi/70$	0.8510×10^{-4}	2.0	0.1200×10^{-2}	2.0	0.1200×10^{-2}	2.0
$\sqrt{8}\pi/100$	0.4331×10^{-4}	1.9	0.5984×10^{-3}	2.0	0.5984×10^{-3}	2.0
$\sqrt{8}\pi/130$	0.2542×10^{-4}	2.0	0.3542×10^{-3}	2.0	0.3542×10^{-3}	2.0
$\sqrt{8}\pi/160$	0.1705×10^{-4}	1.9	0.2338×10^{-3}	2.0	0.2338×10^{-3}	2.0

Table 4.1: Global error E_{G1} of Example 4.1.1 for $t = 1$, $0 \leq x, y \leq \sqrt{8}\pi$, $\Delta t = \Delta x$, $TOL = 1/N^3$, $T = 20$, $\beta = -\ln(10^{-16})/2T$, computed with the norm ℓ_∞.

Example 4.1.1. We consider equation (4.1) with $P = Q = 0$ and $\theta = D = 1$:

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2},\ (x, y) \in (0, b) \times (0, b),\ t > 0, \tag{4.16}$$

with initial conditions

$$\begin{aligned} u(x, y, 0) &= \sin\left(\frac{\pi}{b}x\right) \sin\left(\frac{\pi}{b}y\right), \\ \frac{\partial u}{\partial t}(x, y, 0) &= -\frac{a}{2} \sin\left(\frac{\pi}{b}x\right) \sin\left(\frac{\pi}{b}y\right), \end{aligned} \tag{4.17}$$

and boundary conditions

$$u(0, y, t) = u(x, 0, t) = 0, \quad u(b, y, t) = u(x, b, t) = 0. \tag{4.18}$$

The exact solution is given by

$$u(x, y, t) = e^{-at/2} \sin\left(\frac{\pi}{b}x\right) \sin\left(\frac{\pi}{b}y\right), \tag{4.19}$$

where constants a and b satisfy the relation

$$b^2 = \frac{8\pi^2}{2a - a^2}.$$

We will consider this problem with $a = 1$ and $b = \sqrt{8}\pi$. In Table 4.1 we present the global error, defined by

$$E_{G1} = \|u - U\|_\infty = \max_{1 \leq i \leq N_x - 1, 1 \leq j \leq N_y - 1} |u(x_i, y_j, t) - U(x_i, y_j, t)|, \tag{4.20}$$

and the convergence rate for the three methods.

$\Delta x = \Delta y$	CN-2D	Rate	Laplace-FD-2D	Rate	Laplace-FV-2D	Rate
$\sqrt{8}\pi/40$	0.1200×10^{-2}		0.8473×10^{-3}		0.1100×10^{-2}	
$\sqrt{8}\pi/70$	0.3781×10^{-3}	2.1	0.1650×10^{-3}	2.9	0.2742×10^{-3}	2.5
$\sqrt{8}\pi/100$	0.1924×10^{-3}	1.9	0.5996×10^{-4}	2.8	0.1229×10^{-3}	2.2
$\sqrt{8}\pi/130$	0.1129×10^{-3}	2.0	0.2924×10^{-4}	2.7	0.6990×10^{-4}	2.2
$\sqrt{8}\pi/160$	0.7576×10^{-4}	1.9	0.1690×10^{-4}	2.6	0.4519×10^{-4}	2.1

Table 4.2: Global error E_{G2} of Example 4.1.1 for $t = 1$, $0 \leq x, y \leq \sqrt{8}\pi$, $\Delta t = \Delta x$, $TOL = 1/N^3$, $T = 20$, $\beta = -\ln(10^{-16})/2T$, computed with the norm $\ell_{2,\Delta}$.

All methods have a truncation error of second order and the CN-2D presents a smaller error. However, let us now consider the error defined by the norm $\ell_{2,\Delta}$:

$$E_{G2} = \|u - U\| = \left(\Delta x \Delta y \sum_{i=1}^{N_x - 1} \sum_{j=1}^{N_y - 1} |u(x_i, y_j, t) - U(x_i, y_j, t)|^2 \right)^{1/2}. \quad (4.21)$$

In Table 4.2 we can observe that the Laplace-FD-2D method has a significant higher convergence rate. The same happens with the Laplace-FV-2D method.

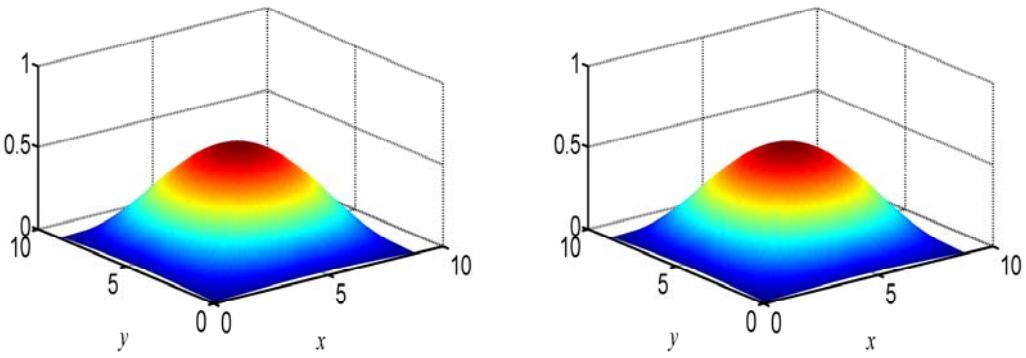

Figure 4.2: Results for Example 4.1.1. Left: exact solution $u(x, y, t)$. Right: approximate solution by CN-2D for $\Delta t = \Delta x = \Delta y = \sqrt{8}\pi/40$.

In Figure 4.2 and Figure 4.3 we show how the solution behaves for this problem: the solution u and the numerical solutions match very well.

Despite the good results obtained from these methods, they proved to be computationally inefficient: the resolution of system (4.14), used for the

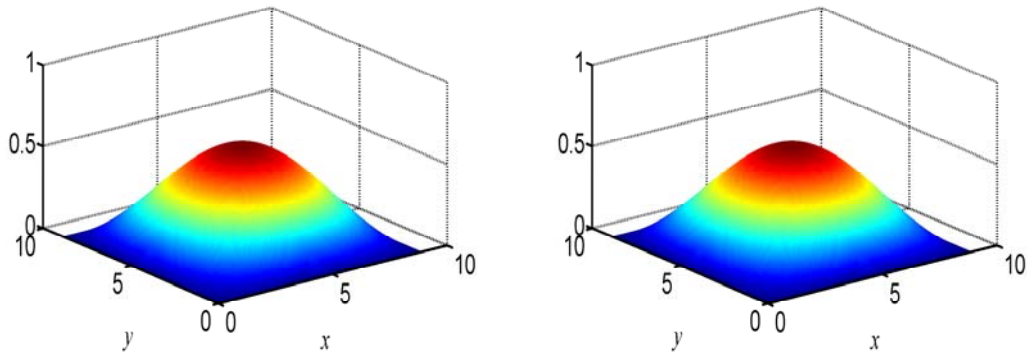

Figure 4.3: Approximate solution of Example 4.1.1 for $\Delta t = \Delta x = \Delta y = \sqrt{8}\pi/40$. Left: Laplace-FD-2D. Right: Laplace-FV-2D.

Laplace-FD-2D and Laplace-FV-2D methods, and system (4.11) used for the CN-2D method, require excessive memory and computational effort. In fact, the three schemes lead to a system that has matrices with five diagonals. For large values of N_x and N_y, lengths of the domain in the x and in the y directions respectively, it quickly fills the computer memory. For example, in the numerical test given above the methods Laplace-FD-2D and Laplace-FV-2D do not deal with $N_x, N_y > 200$. For the CN-2D method this drawback is even more evident since the system (4.11) needs to be solved at each time step; the method does not support $N_x, N_y > 160$. Therefore a new strategy is developed in the next section to overcome this difficulty.

4.2 An alternating direction implicit method

The numerical method proposed consists first of deriving a scheme based on the Crank-Nicolson method giving rise to a sparse linear system that need to be solved at each time step. Then, to overcome the computational inefficiency of this implicit scheme in two dimensions, after discretization we apply an alternating direction implicit (ADI) method. This approach allows the reduction of our two dimensional problem into two problems in one dimension which only require to solve systems with tridiagonal matrices.

The main idea of the ADI method is to split the computations in two

steps. In the first step, we apply an implicit method in the x-direction and an explicit method in the y-direction, producing an intermediate solution for time. In the second step, we apply an implicit method in the y-direction and an explicit method in the x-direction. This is described bellow.

For convenience of numerical analysis, we introduce the dimensionless parameters

$$x^* = \frac{x}{\sqrt{D\theta}}, \quad y^* = \frac{y}{\sqrt{D\theta}}, \quad t^* = \frac{t}{\theta}, \quad \theta \in]0, 1],$$
$$P^* = \sqrt{\frac{D}{\theta}}P, \quad Q^* = \sqrt{\frac{D}{\theta}}Q$$

and the dimensionless form of equation (4.1) can be written as, omitting asterisks for notational simplicity,

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2}(x,y,t) + \frac{\partial u}{\partial t}(x,y,t) = & - \frac{\partial}{\partial x}(P(x,y)u(x,y,t)) - \frac{\partial}{\partial y}(Q(x,y)u(x,y,t)) \\ & + \frac{\partial^2 u}{\partial x^2}(x,y,t) + \frac{\partial^2 u}{\partial y^2}(x,y,t), \end{aligned} \tag{4.22}$$

where $(x, y) \in \Omega$, $t > 0$ and $\Omega \subset \mathbb{R}^2$ is a rectangular domain. We consider the initial conditions given by

$$u(x,y,0) = u_0(x,y), \qquad (x,y) \in \Omega, \tag{4.23}$$
$$\frac{\partial u}{\partial t}(x,y,0) = u_1(x,y), \qquad (x,y) \in \Omega, \tag{4.24}$$

and Dirichlet boundary condition

$$u(x,y,t) = f(x,y,t), \qquad (x,y) \in \partial\Omega, \ t > 0. \tag{4.25}$$

Next, we describe some difference operators so that we can easily handle the discretized equations (4.7)-(4.8). We define the first order forward and backward difference operators as

$$\delta_x^+ U_{i,j}^n = \frac{U_{i+1,j}^n - U_{i,j}^n}{\Delta x}, \qquad \delta_y^+ U_{i,j}^n = \frac{U_{i,j+1}^n - U_{i,j}^n}{\Delta y} \tag{4.26}$$

and

$$\delta_x^- U_{i,j}^n = \frac{U_{i,j}^n - U_{i-1,j}^n}{\Delta x}, \qquad \delta_y^- U_{i,j}^n = \frac{U_{i,j}^n - U_{i,j-1}^n}{\Delta y}. \tag{4.27}$$

The first order centered difference operators are given by

$$\delta_x U_{i,j}^n = \frac{1}{2}[\delta_x^+ + \delta_x^-]U_{i,j}^n = \frac{U_{i+1,j}^n - U_{i-1,j}^n}{2\Delta x}, \tag{4.28}$$

$$\delta_y U_{i,j}^n = \frac{1}{2}[\delta_y^+ + \delta_y^-]U_{i,j}^n = \frac{U_{i,j+1}^n - U_{i,j-1}^n}{2\Delta y}, \tag{4.29}$$

and the second order centered difference operators are defined by

$$\delta_x^2 U_{i,j}^n = \frac{U_{i-1,j}^n - 2U_{i,j}^n + U_{i+1,j}^n}{\Delta x^2}, \quad \delta_y^2 U_{i,j}^n = \frac{U_{i,j-1}^n - 2U_{i,j}^n + U_{i,j+1}^n}{\Delta y^2}. \tag{4.30}$$

Considering the new variables, we use the discretization operators defined in (4.28)-(4.30) and denote the set of discretization points $U^n = \{U_{i,j}^n\}$, $PU^n = \{P_{i,j}U_{i,j}^n\}$, $QU^n = \{Q_{i,j}U_{i,j}^n\}$, $P_xU^n = \{(P_x)_{i,j}U_{i,j}^n\}$, $Q_yU^n = \{(Q_y)_{i,j}U_{i,j}^n\}$ and $W^n = \{W_{i,j}^n\}$. The numerical method can be written in the form

$$W^{n+1} + W^n = U^{n+1} + U^n + \frac{2}{\Delta t}\left(U^{n+1} - U^n\right) \tag{4.31}$$

and

$$\begin{aligned} W^{n+1} - W^n = & - \frac{\Delta t}{2}P_x(U^{n+1} + U^n) - \frac{P\Delta t}{2}\delta_x\left(U^{n+1} + U^n\right) \\ & - \frac{\Delta t}{2}Q_y(U^{n+1} + U^n) - \frac{Q\Delta t}{2}\delta_y\left(U^{n+1} + U^n\right) \\ & + \frac{\Delta t}{2}\delta_x^2\left(U^{n+1} + U^n\right) + \frac{\Delta t}{2}\delta_y^2\left(U^{n+1} + U^n\right). \end{aligned} \tag{4.32}$$

Let us define the operators

$$L_P = \frac{\Delta t}{2}(P_x + P\delta_x - \delta_x^2) \qquad \text{and} \qquad L_Q = \frac{\Delta t}{2}(Q_y + Q\delta_y - \delta_y^2) \tag{4.33}$$

and consider the numerical method

$$\begin{aligned} & \left(\left(1 + \frac{2}{\Delta t}\right) + L_P\right)\left(1 + \frac{\Delta t}{\Delta t + 2}L_Q\right)U^{n+1} \\ = & \left(\left(1 + \frac{2}{\Delta t}\right) - L_P\right)\left(1 - \frac{\Delta t}{\Delta t + 2}L_Q\right)U^n + 2(W^n - U^n). \end{aligned} \tag{4.34}$$

Proposition 4.2.1. *The numerical method (4.31) and (4.34) is a discretization of the equation (4.22) and, for a sufficiently smooth u, the numerical method is $\mathcal{O}(\Delta t^2 + \Delta^2)$ accurate, where $\Delta^2 = \Delta x^2 + \Delta y^2$.*

Proof: If we replace the exact solution in the numerical method (4.31)-(4.32) then, after expanding functions u^{n+1} and w^{n+1} into a Taylor series around the point $(x_i, y_j, t_{n+1/2})$,

$$w^{n+1} + w^n = u^{n+1} + u^n + \frac{2}{\Delta t}\left(u^{n+1} - u^n\right) + \mathcal{O}\left(\Delta t^2\right) \tag{4.35}$$

and

$$\begin{aligned} w^{n+1} - w^n = & - \frac{\Delta t}{2} P_x (u^{n+1} + u^n) - \frac{P\Delta t}{2}\delta_x \left(u^{n+1} + u^n\right) \\ & - \frac{\Delta t}{2} Q_y (u^{n+1} + u^n) - \frac{Q\Delta t}{2}\delta_y \left(u^{n+1} + u^n\right) \\ & + \frac{\Delta t}{2}\left(\delta_x^2 + \delta_y^2\right)\left(u^{n+1} + u^n\right) + \mathcal{O}\left(\Delta t^2 + \Delta^2 + \Delta t \Delta^2\right) \end{aligned} \tag{4.36}$$

and the CN-2D method is $\mathcal{O}\left(\Delta t^2 + \Delta^2\right)$ accurate. Replacing (4.35) in (4.36) yields

$$\begin{aligned} \left(1 + \frac{2}{\Delta t}\right) u^{n+1} + \left(1 - \frac{2}{\Delta t}\right) u^n = & - \frac{\Delta t}{2}\left(P_x + Q_y\right)\left(u^{n+1} + u^n\right) \\ & - \frac{\Delta t}{2}\left(P\delta_x + Q\delta_y\right)\left(u^{n+1} + u^n\right) \\ & + \frac{\Delta t}{2}\left(\delta_x^2 + \delta_y^2\right)\left(u^{n+1} + u^n\right) + 2w^n \\ & + \mathcal{O}\left(\Delta t^2 + \Delta^2 + \Delta t \Delta^2\right). \end{aligned}$$

Using the operators defined in (4.33), the previous equation can be written as

$$\begin{aligned} \left(\left(1 + \frac{2}{\Delta t}\right) + L_P + L_Q\right) u^{n+1} = & \left(\left(1 + \frac{2}{\Delta t}\right) - L_P - L_Q\right) u^n \\ & + 2(w^n - u^n) + \mathcal{O}\left(\Delta t^2 + \Delta^2 + \Delta t \Delta^2\right). \end{aligned} \tag{4.37}$$

The accuracy of the numerical method is $\mathcal{O}\left(\Delta t^2 + \Delta^2\right)$, and therefore we can add to it any term of the same or higher order without changing the accuracy of the scheme. With this in mind, let us consider the term

$$\frac{\Delta t}{\Delta t + 2} L_P L_Q (u^{n+1} - u^n). \tag{4.38}$$

Since

$$
\begin{aligned}
L_P L_Q(u^{n+1}-u^n) &= \frac{\Delta t^2}{4}\left(P_x + P\delta x - \delta x^2\right)\left(Q_y + Q\delta y - \delta y^2\right)\left(u^{n+1}-u^n\right) \\
&= \frac{\Delta t^2}{4}\left(P_x + P\frac{\partial}{\partial x} - \frac{\partial^2}{\partial x^2} + \mathcal{O}(\Delta x^2)\right)\left(Q_y + Q\frac{\partial}{\partial y} - \frac{\partial^2}{\partial y^2} + \mathcal{O}(\Delta y^2)\right) \\
&\quad \times \left(\Delta t\left(\frac{\partial u}{\partial t}\right)^{n+1/2} + \mathcal{O}(\Delta t^3)\right) \\
&= \frac{\Delta t^3}{4}\left(P_x + P\frac{\partial}{\partial x} - \frac{\partial^2}{\partial x^2}\right)\left(Q_y + Q\frac{\partial}{\partial y} - \frac{\partial^2}{\partial y^2}\right)\frac{\partial u^{n+1/2}}{\partial t} \\
&\quad + \mathcal{O}\left(\Delta t^3\Delta^2 + \Delta t^5\right),
\end{aligned}
$$

the term (4.38) is $\mathcal{O}\left(\Delta t^3\Delta^2 + \Delta t^5\right)$ and can be added to (4.37) without any change in the accuracy order. Then, we obtain

$$
\begin{aligned}
&\left(\left(1+\frac{2}{\Delta t}\right) + L_P\right)\left(1 + \frac{\Delta t}{\Delta t+2}L_Q\right)u^{n+1} \\
&= \left(\left(1+\frac{2}{\Delta t}\right) - L_P\right)\left(1 - \frac{\Delta t}{\Delta t+2}L_Q\right)u^{n} \\
&\qquad + 2(w^n - u^n) + \mathcal{O}\left(\Delta t^2 + \Delta^2 + \Delta t\Delta^2 + \Delta t^3\Delta^2\right). \qquad (4.39)
\end{aligned}
$$

Therefore, we conclude the numerical method (4.34) is $\mathcal{O}(\Delta t^2 + \Delta^2)$ accurate. ■

To implement the numerical method (4.34) we can split the equation in two, following the Peaceman-Rachford strategy [75], by introducing a new intermediate variable $\widehat{U}$, which represents a solution computed for an intermediate time. Thus we obtain a type of Peaceman-Rachford ADI,

$$
\left(\left(1+\frac{2}{\Delta t}\right) + L_P\right)\widehat{U} = \left(1 - \frac{\Delta t}{\Delta t+2}L_Q\right)U^n + \frac{\Delta t}{\Delta t+2}(W^n - U^n), \qquad (4.40)
$$

$$
\left(1 + \frac{\Delta t}{\Delta t+2}L_Q\right)U^{n+1} = \left(\left(1+\frac{2}{\Delta t}\right) - L_P\right)\widehat{U} + \frac{\Delta t}{\Delta t+2}(W^n - U^n). \qquad (4.41)
$$

If we apply the operator $\left(\left(1+\frac{2}{\Delta t}\right) + L_P\right)$ to both sides of (4.41), it can be seen that (4.40) and (4.41) are equivalent to (4.34). To implement the difference scheme (4.40)-(4.41) we need to solve two systems. The resolution

of (4.40) gives

$$\begin{aligned}&\left(1+\frac{2}{\Delta t}+\frac{\Delta t}{2}P_x+\frac{\Delta t}{2}P\delta_x-\frac{\Delta t}{2}\delta_x^2\right)\widehat{U}\\ =\;&\left(1-\frac{\Delta t^2}{2(\Delta t+2)}(Q_y+Q\delta_y-\delta_y^2)\right)U^n+\frac{\Delta t}{\Delta t+2}(W^n-U^n).\end{aligned}$$

After developing the difference operators we obtain the equation

$$\begin{aligned}&\left(1+\frac{2}{\Delta t}+\frac{\Delta t}{2}(P_x)_{i,j}\right)\widehat{U}_{i,j}+\frac{\Delta t P_{i,j}}{2}\frac{\widehat{U}_{i+1,j}-\widehat{U}_{i-1,j}}{2\Delta x}\\ &\quad-\frac{\Delta t}{2}\frac{\widehat{U}_{i-1,j}-2\widehat{U}_{i,j}+\widehat{U}_{i+1,j}}{\Delta x^2}\\ &=\left(1-\frac{\Delta t^2(Q_y)_{i,j}}{2(\Delta t+2)}\right)U_{i,j}^n-\frac{\Delta t^2 Q_{i,j}}{2(\Delta t+2)}\frac{U_{i,j+1}^n-U_{i,j-1}^n}{2\Delta y}\\ &\quad+\frac{\Delta t^2}{2(\Delta t+2)}\frac{U_{i,j-1}^n-2U_{i,j}^n+U_{i,j+1}^n}{\Delta y^2}+\frac{\Delta t}{\Delta t+2}\left(W_{i,j}^n-U_{i,j}^n\right).\end{aligned}$$

Dividing the equation by Δt results in

$$\begin{aligned}&A_{i,j}^1\widehat{U}_{i-1,j}+A_{i,j}^2\widehat{U}_{i,j}+A_{i,j}^3\widehat{U}_{i+1,j}\\ &\quad=B_{i,j}^1U_{i,j-1}^n+B_{i,j}^2U_{i,j}^n+B_{i,j}^3U_{i,j+1}^n+\frac{1}{\Delta t+2}\left(W_{i,j}^n-U_{i,j}^n\right),\end{aligned}\tag{4.42}$$

for $i=1,\ldots,N_x-1, j=1,\ldots,N_y-1$, where

$A_{i,j}^1=-\frac{P_{i,j}}{4\Delta x}-\frac{1}{2\Delta x^2}$, $A_{i,j}^2=\frac{1}{\Delta t}+\frac{2}{\Delta t^2}+\frac{(P_x)_{i,j}}{2}+\frac{1}{\Delta x^2}$,

$A_{i,j}^3=\frac{P_{i,j}}{4\Delta x}-\frac{1}{2\Delta x^2}$, $B_{i,j}^1=\frac{\Delta t}{4(\Delta t+2)\Delta y}Q_{i,j}+\frac{\Delta t}{2(\Delta t+2)\Delta y^2}$,

$B_{i,j}^2=\frac{1}{\Delta t}-\frac{\Delta t}{2(\Delta t+2)}(Q_y)_{i,j}-\frac{\Delta t}{(\Delta t+2)\Delta y^2}$,

$B_{i,j}^3=-\frac{\Delta t}{4(\Delta t+2)\Delta y}Q_{i,j}+\frac{\Delta t}{2(\Delta t+2)\Delta y^2}$.

The matrix form of (4.42) is

$$
\begin{bmatrix}
A^2_{1,j} & A^3_{1,j} & & & \\
A^1_{2,j} & A^2_{2,j} & A^3_{2,j} & & \\
 & \ddots & \ddots & \ddots & \\
 & & A^1_{N_x-2,j} & A^2_{N_x-2,j} & A^3_{N_x-2,j} \\
 & & & A^1_{N_x-1,j} & A^2_{N_x-1,j}
\end{bmatrix}
\begin{bmatrix}
\widehat{U}_{1,j} \\ \widehat{U}_{2,j} \\ \vdots \\ \widehat{U}_{N_x-2,j} \\ \widehat{U}_{N_x-1,j}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
B^1_{1,j}U^n_{1,j-1} + B^2_{1,j}U^n_{1,j} + B^3_{1,j}U^n_{1,j+1} \\
B^1_{2,j}U^n_{2,j-1} + B^2_{2,j}U^n_{2,j} + B^3_{2,j}U^n_{2,j+1} \\
\vdots \\
B^1_{N_x-2,j}U^n_{N_x-2,j-1} + B^2_{N_x-2,j}U^n_{N_x-2,j} + B^3_{N_x-2,j}U^n_{N_x-2,j+1} \\
B^1_{N_x-1,j}U^n_{N_x-1,j-1} + B^2_{N_x-1,j}U^n_{N_x-1,j} + B^3_{N_x-1,j}U^n_{N_x-1,j+1}
\end{bmatrix}
$$

$$
+ \begin{bmatrix}
\frac{1}{\Delta t+2}\left(W^n_{1,j} - U^n_{1,j}\right) - A^1_{1,j}\widehat{U}_{0,j} \\
\frac{1}{\Delta t+2}\left(W^n_{2,j} - U^n_{2,j}\right) \\
\vdots \\
\frac{1}{\Delta t+2}\left(W^n_{N_x-2,j} - U^n_{N_x-2,j}\right) \\
\frac{1}{\Delta t+2}\left(W^n_{N_x-1,j} - U^n_{N_x-1,j}\right) - A^3_{N_x-1,j}\widehat{U}_{N_x,j}
\end{bmatrix}.
$$

For a fixed j, we need to solve a tridiagonal system of dimension $(N_x - 1)$. This implies the resolution of $(N_y - 1)$ tridiagonal systems. In a similar way, the resolution of (4.41) gives

$$
\left(1 + \frac{\Delta t^2}{2(\Delta t + 2)}(Q_y + Q\delta_y - \delta_y^2)\right) U^{n+1}
$$

$$
= \left(1 + \frac{2}{\Delta t} - \frac{\Delta t}{2}P_x - \frac{\Delta t}{2}P\delta_x + \frac{\Delta t}{2}\delta_x^2\right)\widehat{U} + \frac{\Delta t}{\Delta t + 2}(W^n - U^n).
$$

Again, we develop the operators and divide by Δt to get

$$
C^1_{i,j}U^{n+1}_{i,j-1} + C^2_{i,j}U^{n+1}_{i,j} + C^3_{i,j}U^{n+1}_{i,j+1}
$$

$$
= D^1_{i,j}\widehat{U}_{i-1,j} + D^2_{i,j}\widehat{U}_{i,j} + D^3_{i,j}\widehat{U}_{i+1,j} + \frac{1}{\Delta t + 2}\left(W^n_{i,j} - U^n_{i,j}\right), \quad (4.43)
$$

for $i = 1, \ldots, N_x - 1, j = 1, \ldots, N_y - 1$, where

$$C^1_{i,j} = -\frac{\Delta t}{4(\Delta t + 2)\Delta y} Q_{i,j} - \frac{\Delta t}{2(\Delta t + 2)\Delta y^2},$$

$$C^2_{i,j} = \frac{1}{\Delta t} + \frac{\Delta t}{2(\Delta t + 2)} (Q_y)_{i,j} + \frac{\Delta t}{(\Delta t + 2)\Delta y^2},$$

$$C^3_{i,j} = \frac{\Delta t}{4(\Delta t + 2)\Delta y} Q_{i,j} - \frac{\Delta t}{2(\Delta t + 2)\Delta y^2}, \quad D^1_{i,j} = \frac{P_{i,j}}{4\Delta x} + \frac{1}{2\Delta x^2},$$

$$D^2_{i,j} = \frac{1}{\Delta t} + \frac{2}{\Delta t^2} - \frac{(P_x)_{i,j}}{2} - \frac{1}{\Delta x^2}, \quad D^3_{i,j} = -\frac{P_{i,j}}{4\Delta x} + \frac{1}{2\Delta x^2}.$$

The matrix form of (4.43) is

$$\begin{bmatrix} C^2_{i,1} & C^3_{i,1} & & & & \\ C^1_{i,2} & C^2_{i,2} & C^3_{i,2} & & & \\ & \ddots & \ddots & \ddots & & \\ & & C^1_{i,N_y-2} & C^2_{i,N_y-2} & C^3_{i,N_y-2} & \\ & & & C^1_{i,N_y-1} & C^2_{i,N_y-1} \end{bmatrix} \begin{bmatrix} U^{n+1}_{i,1} \\ U^{n+1}_{i,2} \\ \vdots \\ U^{n+1}_{i,N_y-2} \\ U^{n+1}_{i,N_y-1} \end{bmatrix}$$

$$= \begin{bmatrix} D^1_{i,1}\widehat{U}_{i-1,1} + D^2_{i,1}\widehat{U}_{i,1} + D^3_{i,1}\widehat{U}_{i+1,1} \\ D^1_{i,2}\widehat{U}_{i-1,2} + D^2_{i,2}\widehat{U}_{i,2} + D^3_{i,2}\widehat{U}_{i+1,2} \\ \vdots \\ D^1_{i,N_y-2}\widehat{U}_{i-1,N_y-2} + D^2_{i,N_y-2}\widehat{U}_{i,N_y-2} + D^3_{i,N_y-2}\widehat{U}_{i+1,N_y-2} \\ D^1_{i,N_y-1}\widehat{U}_{i-1,N_y-1} + D^2_{i,N_y-1}\widehat{U}_{i,N_y-1} + D^3_{i,N_y-1}\widehat{U}_{i+1,N_y-1} \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{1}{\Delta t+2}\left(W^n_{i,1} - U^n_{i,1}\right) - C^1_{i,1}U^{n+1}_{i,0} \\ \frac{1}{\Delta t+2}\left(W^n_{i,2} - U^n_{i,2}\right) \\ \vdots \\ \frac{1}{\Delta t+2}\left(W^n_{i,N_y-2} - U^n_{i,N_y-2}\right) \\ \frac{1}{\Delta t+2}\left(W^n_{i,N_y-1} - U^n_{i,N_y-1}\right) - C^3_{i,N_y-1}U^{n+1}_{i,N_y} \end{bmatrix}$$

For a fixed i, we need to solve a tridiagonal system of dimension $(N_y - 1)$. This implies the resolution of $(N_x - 1)$ tridiagonal systems.

We have seen the difference scheme (4.34) has a truncation error of order $\mathcal{O}\left(\Delta t^2 + \Delta^2\right)$. Next, we focus our attention in its stability studied in the following section.

4.2.1 Stability analysis

We prove the stability of finite difference scheme (4.34) for two distinct cases: in the first we consider P and Q constants and the second one has $P(x,y) = P(x) \neq 0$ and $Q(x,y) = 0$. This last case appears in the study of hyperbolic diffusion equations in two dimensions containing a first order spatial derivative, namely in hyperbolic heat conduction problems due to their wide industrial applicability [3, 17]. The discrete energy method is used for both results. To this end, let us define the set of discrete values with homogeneous boundary conditions.

Assume that

$$\mathcal{G} = \{U |\, U = \{U_{i,j}\},\ \text{and}\ U_{0,j} = U_{N_x,j} = U_{i,0} = U_{i,N_y} = 0\}.$$

For $U, V \in \mathcal{G}$, we define the inner product and norm respectively as

$$(U,V) = \Delta x \Delta y \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y-1} U_{i,j} V_{i,j}, \qquad \|U\|^2 = (U,U)\,. \tag{4.44}$$

We also define the following inner products that involve the first and second order discretization operators of $U, V \in \mathcal{G}$:

$$\left(\delta_x^+ U, \delta_x^+ V\right)_{*x} = \Delta x \Delta y \sum_{i=0}^{N_x-1} \sum_{j=1}^{N_y-1} \delta_x^+ U_{i,j} \delta_x^+ V_{i,j}, \qquad \|\delta_x^+ U\|_{*x}^2 = \left(\delta_x^+ U, \delta_x^+ U\right)_{*x},$$

$$\left(\delta_y^+ U, \delta_y^+ V\right)_{*y} = \Delta x \Delta y \sum_{i=1}^{N_x-1} \sum_{j=0}^{N_y-1} \delta_y^+ U_{i,j} \delta_y^+ V_{i,j}, \qquad \|\delta_y^+ U\|_{*y}^2 = \left(\delta_y^+ U, \delta_y^+ U\right)_{*y},$$

$$\left(\delta_x^+ \delta_y^+ U, \delta_x^+ \delta_y^+ V\right)_{*} = \Delta x \Delta y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \delta_x^+ \delta_y^+ U_{i,j} \delta_x^+ \delta_y^+ V_{i,j},$$

and

$$\|\delta_x^+\delta_y^+U\|_*^2 = \left(\delta_x^+\delta_y^+U, \delta_x^+\delta_y^+U\right)_*.$$

Next, we introduce some lemmas that will be useful to prove the main theorems.

Lemma 4.2.1. *For any $W \in \mathcal{G}$,*

$$\|\delta_x W\| \leq \|\delta_x^+ W\|_{*x}, \qquad \|\delta_y W\| \leq \|\delta_y^+ W\|_{*y}, \qquad \|\delta_x\delta_y W\| \leq \|\delta_x^+\delta_y^+ W\|_*.$$

Proof: The first two inequalities are obtained in a similar way to inequality in Lemma 2.2.1. We only prove the third inequality. We have

$$\begin{aligned}\|\delta_x\delta_y W\|^2 &= \Delta x\Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}(\delta_x\delta_y W_{i,j})^2 \\ &= \Delta x\Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}\left(\delta_x\left(\frac{1}{2}\delta_y^+ W_{i,j} + \frac{1}{2}\delta_y^- W_{i,j}\right)\right)^2.\end{aligned}$$

Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ we obtain

$$\begin{aligned}\|\delta_x\delta_y W\|^2 &\leq \frac{1}{2}\Delta x\Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}\left(\delta_x\left(\frac{W_{i,j+1}-W_{i,j}}{\Delta y}\right)\right)^2 \\ &+ \frac{1}{2}\Delta x\Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}\left(\delta_x\left(\frac{W_{i,j}-W_{i,j-1}}{\Delta y}\right)\right)^2.\end{aligned}$$

The first inequality $\|\delta_x W\| \leq \|\delta_x^+ W\|_{*x}$ leads to

$$\begin{aligned}\|\delta_x\delta_y W\|^2 &\leq \frac{1}{2}\Delta x\Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1}\left(\delta_x^+\left(\frac{W_{i,j+1}-W_{i,j}}{\Delta y}\right)\right)^2 \\ &+ \frac{1}{2}\Delta x\Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1}\left(\delta_x^+\left(\frac{W_{i,j}-W_{i,j-1}}{\Delta y}\right)\right)^2.\end{aligned}$$

Now, we shift the index j in the second summation and get

$$\begin{aligned}
\|\delta_x\delta_y W\|^2 &\leq \frac{1}{2}\Delta x\Delta y\sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1}\left(\delta_x^+\left(\frac{W_{i,j+1}-W_{i,j}}{\Delta y}\right)\right)^2\\
&+ \frac{1}{2}\Delta x\Delta y\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-2}\left(\delta_x^+\left(\frac{W_{i,j+1}-W_{i,j}}{\Delta y}\right)\right)^2\\
&= \frac{1}{2}\Delta x\Delta y\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}\left(\delta_x^+\left(\frac{W_{i,j+1}-W_{i,j}}{\Delta y}\right)\right)^2\\
&- \frac{1}{2}\Delta x\Delta y\sum_{i=0}^{N_x-1}\left(\delta_x^+\left(\frac{W_{i,1}-W_{i,0}}{\Delta y}\right)\right)^2\\
&+ \frac{1}{2}\Delta x\Delta y\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}\left(\delta_x^+\left(\frac{W_{i,j+1}-W_{i,j}}{\Delta y}\right)\right)^2\\
&- \frac{1}{2}\Delta x\Delta y\sum_{i=0}^{N_x-1}\left(\delta_x^+\left(\frac{W_{i,N_y}-W_{i,N_y-1}}{\Delta y}\right)\right)^2\\
&\leq \Delta x\Delta y\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}\left(\delta_x^+\delta_y^+W_{i,j}\right)^2\\
&= \|\delta_x^+\delta_y^+W\|_*^2.
\end{aligned}$$

■

The following lemma is the well known property of summation by parts [53, 98]. The proof is similar to the one in Lemma 2.2.2 and we do not include it.

Lemma 4.2.2. *For any $U, V \in \mathcal{G}$,*

$$\left(\delta_x^2U, V\right) = -\left(\delta_x^+U, \delta_x^+V\right)_{*x}, \qquad \left(\delta_y^2U, V\right) = -\left(\delta_y^+U, \delta_y^+V\right)_{*y}.$$

The next lemma can be seen, for instance in [22, 31, 86].

Lemma 4.2.3. *For any $U \in \mathcal{G}$, the following inequalities hold*

$$\|\delta_x^+U\|_{*x}^2 \leq \frac{4}{\Delta x^2}\|U\|^2, \qquad \|\delta_y^+U\|_{*y}^2 \leq \frac{4}{\Delta y^2}\|U\|^2,$$

$$\|\delta_x^+\delta_y^+U\|_*^2 \leq \frac{4}{\Delta y^2}\|\delta_x^+U\|_{*x}^2, \qquad \|\delta_x^+\delta_y^+U\|_*^2 \leq \frac{4}{\Delta x^2}\|\delta_y^+U\|_{*y}^2.$$

Proof: Let us start to prove the first inequality. We have

$$\begin{aligned}
\|\delta_x^+ U\|_{*x}^2 &= \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1} \left(\delta_x^+ U_{i,j}\right)^2 = \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1} \left(\frac{U_{i+1,j}-U_{i,j}}{\Delta x}\right)^2 \\
&\leq \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1} \frac{2\left(U_{i+1,j}\right)^2 + 2\left(U_{i,j}\right)^2}{\Delta x^2} \\
&= \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1} \frac{2}{\Delta x^2}\left(U_{i+1,j}\right)^2 + \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1} \frac{2}{\Delta x^2}\left(U_{i,j}\right)^2 .
\end{aligned}$$

Shifting the index i in the first summation and using the conditions $U_{0,j} = U_{N_x,j} = 0$ we obtain

$$\begin{aligned}
\|\delta_x^+ U\|_{*x}^2 &\leq \Delta x \Delta y \sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1} \frac{2\left(U_{i,j}\right)^2}{\Delta x^2} + \Delta x \Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} \frac{2\left(U_{i,j}\right)^2}{\Delta x^2} \\
&+ \Delta x \Delta y \sum_{j=1}^{N_y-1} \frac{2\left(U_{0,j}\right)^2}{\Delta x^2} \\
&= \Delta x \Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} \frac{2\left(U_{i,j}\right)^2}{\Delta x^2} + \Delta x \Delta y \sum_{j=1}^{N_y-1} \frac{2\left(U_{N_x,j}\right)^2}{\Delta x^2} \\
&+ \Delta x \Delta y \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} \frac{2\left(U_{i,j}\right)^2}{\Delta x^2} \\
&= \frac{4}{\Delta x^2}\|U\|^2.
\end{aligned}$$

Let us now consider the norm $\|\delta_x^+ \delta_y^+ U\|_*^2$. We have

$$\begin{aligned}
\|\delta_x^+ \delta_y^+ U\|_*^2 &= \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \left(\delta_x^+ \delta_y^+ U_{i,j}\right)^2 \\
&= \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \left(\delta_x^+ \left(\frac{U_{i,j+1}-U_{i,j}}{\Delta y}\right)\right)^2 \\
&= \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \frac{\left(\delta_x^+ U_{i,j+1} - \delta_x^+ U_{i,j}\right)^2}{\Delta y^2} \\
&\leq \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \frac{2\left(\delta_x^+ U_{i,j+1}\right)^2 + 2\left(\delta_x^+ U_{i,j}\right)^2}{\Delta y^2} \\
&= \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \frac{2\left(\delta_x^+ U_{i,j+1}\right)^2}{\Delta y^2} + \Delta x \Delta y \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \frac{2\left(\delta_x^+ U_{i,j}\right)^2}{\Delta y^2}.
\end{aligned}$$

Shifting the index j in the first summation and using the conditions $U_{i,0} = U_{i,N_y} = 0$ we obtain

$$\begin{aligned}
\|\delta_x^+\delta_y^+U\|_*^2 &\leq \Delta x\Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y}\frac{2\left(\delta_x^+U_{i,j}\right)^2}{\Delta y^2} + \Delta x\Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1}\frac{2\left(\delta_x^+U_{i,j}\right)^2}{\Delta y^2} \\
&+ \Delta x\Delta y \sum_{i=0}^{N_x-1}\frac{2\left(\delta_x^+U_{i,0}\right)^2}{\Delta y^2} \\
&= \Delta x\Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1}\frac{2\left(\delta_x^+U_{i,j}\right)^2}{\Delta y^2} + \Delta x\Delta y \sum_{i=0}^{N_x-1}\frac{2\left(\delta_x^+U_{i,N_y}\right)^2}{\Delta y^2} \\
&+ \Delta x\Delta y \sum_{i=0}^{N_x-1}\sum_{j=1}^{N_y-1}\frac{2\left(\delta_x^+U_{i,j}\right)^2}{\Delta y^2} \\
&= \frac{4}{\Delta y^2}\|\delta_x^+U\|_{*x}^2.
\end{aligned}$$

Similarly we obtain $\|\delta_y^+U\|_{*y}^2 \leq \frac{4}{\Delta y^2}\|U\|^2$, $\|\delta_x^+\delta_y^+U\|_*^2 \leq \frac{4}{\Delta x^2}\|\delta_y^+U\|_{*y}^2$.

■

Before proving the main result, note that the ADI method (4.34) can be written in the form

$$\begin{aligned}
&\left(1+\frac{2}{\Delta t}\right)U^{n+1} + \left(1-\frac{2}{\Delta t}\right)U^n - 2W^n + \frac{\Delta t}{\Delta t+2}L_PL_Q\left(U^{n+1}-U^n\right) \\
=\ & -\left(L_P+L_Q\right)\left(U^{n+1}+U^n\right). \qquad (4.45)
\end{aligned}$$

Taking into account (4.31) we have

$$W^{n+1} - W^n + \frac{\Delta t}{\Delta t+2}L_PL_Q\left(U^{n+1}-U^n\right) = -\left(L_P+L_Q\right)\left(U^{n+1}+U^n\right). \quad (4.46)$$

Theorem 4.2.1. *Suppose that $\{U_{i,j}^n, W_{i,j}^n\}$ and $\{V_{i,j}^n, Y_{i,j}^n\}$ are solutions of the finite difference scheme (4.31) and (4.46) with constants P and Q, which satisfy the boundary condition (4.25), and have different initial values $\{U_{i,j}^0, W_{i,j}^0\}$ and $\{V_{i,j}^0, Y_{i,j}^0\}$ respectively. Let $\omega_{i,j}^n = W_{i,j}^n - Y_{i,j}^n$, $\epsilon_{i,j}^n = U_{i,j}^n - V_{i,j}^n$. For $\Delta t \leq 1$, such that, $\Delta t \leq c_p\Delta x$, $\Delta t \leq c_q\Delta y$, with constants c_p, c_q, then $\{\omega_{i,j}^n, \epsilon_{i,j}^n\}$ satisfy*

$$\|\omega^{n+1}\|^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \leq (1+C\Delta t)\left(\|\omega^n\|^2 + \|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2\right), \quad (4.47)$$

where C denotes a constant independent of $\Delta x, \Delta y, \Delta t$.

Proof: For $\omega^n = \{\omega_{i,j}^n\}$ and $\epsilon^n = \{\epsilon_{i,j}^n\}$, from (4.46) we have

$$\omega^{n+1} - \omega^n + \frac{\Delta t}{\Delta t + 2} L_P L_Q \left(\epsilon^{n+1} - \epsilon^n\right) = -\left(L_P + L_Q\right)\left(\epsilon^{n+1} + \epsilon^n\right). \tag{4.48}$$

Multiplying both sides of (4.48) by $\omega^{n+1}+\omega^n$ with respect to the inner product (4.44) we obtain, using (4.33) with P and Q constants,

$$\begin{aligned}
&\|\omega^{n+1}\|^2 - \|\omega^n\|^2 \\
&\quad + \frac{\Delta t^3}{4(\Delta t + 2)} \left(\left(PQ\delta_x\delta_y - P\delta_x\delta_y^2 - Q\delta_x^2\delta_y + \delta_x^2\delta_y^2\right)\left(\epsilon^{n+1} - \epsilon^n\right), \omega^{n+1} + \omega^n\right) \\
&\quad + \frac{\Delta t}{2}\left(\left(P\delta_x - \delta_x^2 + Q\delta_y - \delta_y^2\right)\left(\epsilon^{n+1} + \epsilon^n\right), \omega^{n+1} + \omega^n\right) = 0.
\end{aligned} \tag{4.49}$$

By (4.31) and summation by parts we have

$$\begin{aligned}
&\left(\delta_x^2\delta_y^2\left(\epsilon^{n+1} - \epsilon^n\right), \omega^{n+1} + \omega^n\right) \\
= \;& \left(\delta_x^+\delta_y^+\left(\epsilon^{n+1} - \epsilon^n\right), \delta_x^+\delta_y^+\left(\epsilon^{n+1} + \epsilon^n\right)\right) + \frac{2}{\Delta t}\left(\delta_x^+\delta_y^+\left(\epsilon^{n+1} - \epsilon^n\right), \delta_x^+\delta_y^+\left(\epsilon^{n+1} - \epsilon^n\right)\right) \\
= \;& \|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2 - \|\delta_x^+\delta_y^+\epsilon^n\|_*^2 + \frac{2}{\Delta t}\|\delta_x^+\delta_y^+(\epsilon^{n+1} - \epsilon^n)\|_*^2,
\end{aligned} \tag{4.50}$$

$$\left(\delta_x^2\left(\epsilon^{n+1} + \epsilon^n\right), \omega^{n+1} + \omega^n\right) = -\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2 - \frac{2}{\Delta t}\left(\|\delta_x^+\epsilon^{n+1}\|_{*x}^2 - \|\delta_x^+\epsilon^n\|_{*x}^2\right) \tag{4.51}$$

and

$$\left(\delta_y^2\left(\epsilon^{n+1} + \epsilon^n\right), \omega^{n+1} + \omega^n\right) = -\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2 - \frac{2}{\Delta t}\left(\|\delta_y^+\epsilon^{n+1}\|_{*y}^2 - \|\delta_y^+\epsilon^n\|_{*y}^2\right). \tag{4.52}$$

We can rewrite (4.49) as

$$\begin{aligned}
&\|\omega^{n+1}\|^2 + \frac{\Delta t^3}{4(\Delta t + 2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \\
&\quad = \|\omega^n\|^2 + \frac{\Delta t^3}{4(\Delta t + 2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2 + \|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2 \\
&\qquad - \frac{\Delta t}{2}\left(\|\delta_x^+(\epsilon^{n+1} + \epsilon^n)\|_{*x}^2 + \|\delta_y^+(\epsilon^{n+1} + \epsilon^n)\|_{*y}^2\right) \\
&\qquad - \frac{\Delta t^2}{2(\Delta t + 2)}\|\delta_x^+\delta_y^+(\epsilon^{n+1} - \epsilon^n)\|_*^2 \\
&\qquad - \frac{\Delta t}{2}\left(\left(P\delta_x + Q\delta_y\right)\left(\epsilon^{n+1} + \epsilon^n\right), \omega^{n+1} + \omega^n\right)
\end{aligned}$$

$$-\frac{\Delta t^3}{4(\Delta t+2)}\left(\left(PQ\delta_x\delta_y - P\delta_x\delta_y^2 - Q\delta_x^2\delta_y\right)\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right). \quad (4.53)$$

Let us now discuss the terms with P or Q. We first consider the terms

$$-\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right) \quad \text{and} \quad -\left(Q\delta_y\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right).$$

Using the Cauchy-Schwarz inequality, Lemma 4.2.1 and the inequality

$$ab \leq \eta a^2 + b^2/4\eta, \quad (4.54)$$

for $\eta > 0$, we have

$$\begin{aligned} -\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right) &\leq \|P\delta_x(\epsilon^{n+1}+\epsilon^n)\|\|\omega^{n+1}+\omega^n\| \\ &\leq \||P|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}\|\omega^{n+1}+\omega^n\| \\ &\leq \eta_1|P|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2 \\ &\quad +\frac{1}{4\eta_1}\|\omega^{n+1}+\omega^n\|^2. \end{aligned}$$

Using the inequality

$$(a+b)^2 \leq 2a^2+2b^2, \quad (4.55)$$

we conclude that

$$\begin{aligned} -\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right) &\leq \eta_1|P|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2 \\ &\quad +\frac{1}{2\eta_1}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \quad (4.56) \end{aligned}$$

Similarly

$$\begin{aligned} -\left(Q\delta_y\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right) &\leq \eta_2|Q|^2\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2 \\ &\quad +\frac{1}{2\eta_2}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \quad (4.57) \end{aligned}$$

Let us now consider the term

$$-\left(PQ\delta_x\delta_y\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right).$$

Using the Cauchy-Schwarz inequality, Lemma 4.2.1 and the inequalities (4.54)-(4.55), we have

$$\begin{aligned} -\left(PQ\delta_x\delta_y\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) &\leq \eta_3|PQ|^2\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2 \\ &\quad +\frac{1}{2\eta_3}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \quad (4.58) \end{aligned}$$

Finally, we consider the terms

$$\left(P\delta_x\delta_y^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) \quad \text{and} \quad \left(Q\delta_y\delta_x^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right).$$

Using summation by parts, Lemma 4.2.1 and the inequalities (4.54)-(4.55), we obtain

$$\begin{aligned}
\left(P\delta_x\delta_y^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) &= -\left(P\delta_x\delta_y^+\left(\epsilon^{n+1}-\epsilon^n\right),\delta_y^+\left(\omega^{n+1}+\omega^n\right)\right)_{*y} \\
&\leq \|P\delta_x\delta_y^+\left(\epsilon^{n+1}-\epsilon^n\right)\|_{*y}\|\delta_y^+\left(\omega^{n+1}+\omega^n\right)\|_{*y} \\
&\leq \||P|\delta_x^+\delta_y^+\left(\epsilon^{n+1}-\epsilon^n\right)\|_{*}\|\delta_y^+\left(\omega^{n+1}+\omega^n\right)\|_{*y} \\
&\leq \eta_4|P|^2\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2 \\
&\quad +\frac{1}{4\eta_4}\|\delta_y^+(\omega^{n+1}+\omega^n)\|_{*y}^2.
\end{aligned}$$

Using Lemma 4.2.3 and the inequalities (4.54)-(4.55), we can conclude that

$$\begin{aligned}
\left(P\delta_x\delta_y^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) &\leq \eta_4|P|^2\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2 \\
&\quad +\frac{2}{\eta_4\Delta y^2}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \qquad (4.59)
\end{aligned}$$

Similarly

$$\begin{aligned}
\left(Q\delta_y\delta_x^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) &\leq \eta_5|Q|^2\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2 \\
&\quad +\frac{2}{\eta_5\Delta x^2}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \qquad (4.60)
\end{aligned}$$

From (4.53) and the inequalities (4.56)-(4.60), we obtain

$$\begin{aligned}
&\|\omega^{n+1}\|^2+\tfrac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2+\|\delta_x^+\epsilon^{n+1}\|_{*x}^2+\|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \\
&\leq \|\omega^n\|^2+\tfrac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2+\|\delta_x^+\epsilon^n\|_{*x}^2+\|\delta_y^+\epsilon^n\|_{*y}^2 \\
&-\tfrac{\Delta t}{2}\left(\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2+\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2\right) \\
&-\tfrac{\Delta t^2}{2(\Delta t+2)}\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2 \\
&+\tfrac{\Delta t}{2}\Big(\eta_1|P|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2+\eta_2|Q|^2\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2 \\
&\qquad +\tfrac{1}{2}\left(\eta_1^{-1}+\eta_2{}^{-1}\right)\left(\|\omega^{n+1}\|^2+\|\omega^n\|^2\right)\Big) \\
&+\tfrac{\Delta t^3}{4(\Delta t+2)}\Big(\left(|PQ|^2\eta_3+|P|^2\eta_4+|Q|^2\eta_5\right)\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2 \\
&\qquad +\tfrac{1}{2}\left(\eta_3^{-1}+4\eta_4{}^{-1}\Delta y^{-2}+4\eta_5{}^{-1}\Delta x^{-2}\right)\left(\|\omega^{n+1}\|^2+\|\omega^n\|^2\right)\Big). \qquad (4.61)
\end{aligned}$$

Reorganizing the terms and using the conditions $\Delta t \leq c_p \Delta x$ and $\Delta t \leq c_q \Delta y$, we obtain

$$\begin{aligned}
&A\|\omega^{n+1}\|^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \\
&\leq B\|\omega^n\|^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2 + \|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2 \\
&\quad + \frac{\Delta t}{2}\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2(\eta_1|P|^2-1) + \frac{\Delta t}{2}\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2(\eta_2|Q|^2-1) \\
&\quad + \frac{\Delta t^2}{4(\Delta t+2)}\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2\left(\Delta t\left(|PQ|^2\eta_3 + |P|^2\eta_4 + |Q|^2\eta_5\right) - 2\right),
\end{aligned} \tag{4.62}$$

where A and B are given by

$$A = 1 - \frac{\Delta t}{2}\left(\frac{\eta_1^{-1}+\eta_2^{-1}}{2} + \frac{c_p^2\eta_5^{-1} + c_q^2\eta_4^{-1}}{\Delta t+2} + \frac{\Delta t^2\eta_3^{-1}}{4(\Delta t+2)}\right) \tag{4.63}$$

and

$$B = 1 + \frac{\Delta t}{2}\left(\frac{\eta_1^{-1}+\eta_2^{-1}}{2} + \frac{c_p^2\eta_5^{-1} + c_q^2\eta_4^{-1}}{\Delta t+2} + \frac{\Delta t^2\eta_3^{-1}}{4(\Delta t+2)}\right). \tag{4.64}$$

Let us choose $\eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5 = \eta$ with

$$\eta \leq \min\left\{\frac{1}{P^2}, \frac{1}{Q^2}, \frac{2}{(PQ)^2+P^2+Q^2}\right\}. \tag{4.65}$$

Then, from (4.62) we have

$$\begin{aligned}
&A\|\omega^{n+1}\|^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \\
&\leq B\|\omega^n\|^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2 + \|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2,
\end{aligned} \tag{4.66}$$

with

$$A = 1 - \frac{\Delta t}{2}\left(\eta^{-1} + \eta^{-1}\frac{c_p^2+c_q^2}{\Delta t+2} + \eta^{-1}\frac{\Delta t^2}{4(\Delta t+2)}\right) \tag{4.67}$$

and

$$B = 1 + \frac{\Delta t}{2}\left(\eta^{-1} + \eta^{-1}\frac{c_p^2+c_q^2}{\Delta t+2} + \eta^{-1}\frac{\Delta t^2}{4(\Delta t+2)}\right). \tag{4.68}$$

Using Lemma 4.2.3 and $\Delta t \leq c_q \Delta y$, from (4.66) we obtain

$$\begin{aligned}
&A\|\omega^{n+1}\|^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \\
&\leq B\|\omega^n\|^2 + \left(1 + \frac{c_q^2\Delta t}{\Delta t+2}\right)\|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2.
\end{aligned} \tag{4.69}$$

If we choose η such that

$$\eta > \frac{1}{2}(1+M), \text{ for } M = \max\left\{\frac{c_p^2+c_q^2}{2}, \frac{4(c_p^2+c_q^2)+1}{12}\right\},$$

we can easily check that $0 < A \leq 1$. Additionally if we choose η such that

$$\eta \leq \frac{1}{c_q^2} + \frac{c_p^2+c_q^2}{2c_q^2}$$

we have $B \geq 1 + c_q^2\Delta t/(\Delta t+2)$. Therefore by choosing η such that

$$\frac{1}{2}(1+M) < \eta \leq \min\left\{\frac{1}{P^2}, \frac{1}{Q^2}, \frac{2}{(PQ)^2+P^2+Q^2}, \frac{2+c_p^2+c_q^2}{2c_q^2}\right\}, \tag{4.70}$$

it follows

$$A\left(\|\omega^{n+1}\|^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2\right) \leq B\left(\|\omega^n\|^2 + \|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2\right). \tag{4.71}$$

Consequently, by noting that

$$\frac{B}{A} = 1 + \Delta t\frac{\eta^{-1} + \eta^{-1}\frac{c_p^2+c_q^2}{\Delta t+2} + \eta^{-1}\frac{\Delta t^2}{4(\Delta t+2)}}{1 - \frac{\Delta t}{2}\left(\eta^{-1} + \eta^{-1}\frac{c_p^2+c_q^2}{\Delta t+2} + \eta^{-1}\frac{\Delta t^2}{4(\Delta t+2)}\right)} \leq 1 + C\Delta t,$$

where C denotes a constant independent of $\Delta x, \Delta y, \Delta t$, we obtain the main result.

■

From the previous theorem we get the following result.

Corollary 4.2.1. *Suppose that $\{U_{i,j}^n, W_{i,j}^n\}$ and $\{V_{i,j}^n, Y_{i,j}^n\}$ are solutions of the finite difference scheme (4.31) and (4.46) which satisfy the boundary condition (4.25), and have different initial values $\{U_{i,j}^0, W_{i,j}^0\}$ and $\{V_{i,j}^0, Y_{i,j}^0\}$ respectively. Let $\omega_{i,j}^n = W_{i,j}^n - Y_{i,j}^n$, $\epsilon_{i,j}^n = U_{i,j}^n - V_{i,j}^n$. For $\Delta t \leq 1$, such that, $\Delta t \leq c_p\Delta x$, $\Delta t \leq c_q\Delta y$, with constants c_p, c_q, then $\{\omega_{i,j}^n, \epsilon_{i,j}^n\}$ satisfy*

$$||\omega^n||^2 + ||\delta_x^+\epsilon^n||_{*x}^2 + ||\delta_y^+\epsilon^n||_{*y}^2 \leq K\left(||\omega^0||^2 + ||\delta_x^+\epsilon^0||_{*x}^2 + ||\delta_y^+\epsilon^0||_{*y}^2\right), \tag{4.72}$$

where K denotes a constant independent of $\Delta x, \Delta y, \Delta t$.

The proof follows from Theorem 4.2.1 by making recursion with respect to n. We can easily obtain a similar proof as the one in Corollary 2.2.1. Thus, we can conclude that the difference scheme is stable.

Remark 4.2.1. *The previous results require that the maximum time step size is directly proportional to the space mesh sizes. Usually optimal results are obtained when time step and space steps are comparable and therefore this is a natural condition. Similar conditions can be seen in literature for ADI numerical methods for hyperbolic problems, that usually do not include the first order derivatives in space [31, 33, 55, 93, 97].*

Remark 4.2.2. *Since P and Q are constants we assume they are less than one in absolute value, that is, less than the diffusion coefficient. If P and Q are larger than the diffusion coefficient, asymptotic analysis of exact solutions shows that the Cauchy problem of equation (4.22) can be unstable [95]. The choice of constants c_p and c_q mentioned in the previous theorem can depend on the values of P and Q as can be concluded by observing the condition (4.70). A practical choice could be to consider $c_p^2 + c_q^2 \leq 2/3$, for all P, Q.*

In the next theorem we establish the stability result in the case where $P(x,y) = P(x) \neq 0$ and $Q(x,y) = 0$. The operators in (4.33) become

$$L_P = \frac{\Delta t}{2}(P' + P\delta_x - \delta_x^2) \qquad \text{and} \qquad L = -\frac{\Delta t}{2}\delta_y^2, \tag{4.73}$$

where $P' = P'(x)$ denotes the derivative of $P(x)$ in the x variable. The finite difference scheme (4.46) turns to

$$W^{n+1} - W^n + \frac{\Delta t}{\Delta t + 2} L_P L \left(U^{n+1} - U^n\right) = -\left(L_P + L\right)\left(U^{n+1} + U^n\right). \tag{4.74}$$

Let us suppose that $P(x)$ has non-negative derivative $P'(x)$ and define

$$\|P\|^2 = \sum_{i=1}^{N-1} (P_i)^2, \qquad \|U^n\|_{P'}^2 = \Delta x \Delta y \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y-1} P_i' \left(U_{i,j}^n\right)^2$$

and

$$\|\delta y^+ U^n\|_{*yP'}^2 = \Delta x \Delta y \sum_{i=1}^{N_x-1} \sum_{j=0}^{N_y-1} P_i' \left(\delta y^+ U_{i,j}^n\right)^2.$$

Theorem 4.2.2. *Suppose that $\{U_{i,j}^n, W_{i,j}^n\}$ and $\{V_{i,j}^n, Y_{i,j}^n\}$ are solutions of the finite difference scheme (4.31) and (4.74) where $P(x,y) = P(x) \neq 0$ has*

non-negative derivative $P'(x)$ *and* $Q(x,y) = 0$*, which satisfy the boundary condition (4.25), and have different initial values* $\{U^0_{i,j}, W^0_{i,j}\}$ *and* $\{V^0_{i,j}, Y^0_{i,j}\}$ *respectively. Let* $\omega^n_{i,j} = W^n_{i,j} - Y^n_{i,j}$, $\epsilon^n_{i,j} = U^n_{i,j} - V^n_{i,j}$*. For* $\Delta t \leq 1$ *such that* $\Delta t \leq c\Delta y$*, with constant* c*, then* $\{\omega^n_{i,j}, \epsilon^n_{i,j}\}$ *satisfy*

$$\begin{aligned} & \|\omega^{n+1}\|^2 + \|\delta_x^+ \epsilon^{n+1}\|^2_{*x} + \|\delta_y^+ \epsilon^{n+1}\|^2_{*y} + \|\epsilon^{n+1}\|^2_{P'} \\ \leq \; & (1 + C_1 \Delta t) \left(\|\omega^n\|^2 + \|\delta_x^+ \epsilon^n\|^2_{*x} + \|\delta_y^+ \epsilon^n\|^2_{*y} + \|\epsilon^n\|^2_{P'} \right), \end{aligned} \tag{4.75}$$

where C_1 *denotes a constant independent of* $\Delta x, \Delta y, \Delta t$*.*

Proof: For $\omega^n = \{\omega^n_{i,j}\}$ and $\epsilon^n = \{\epsilon^n_{i,j}\}$, from (4.74) we have

$$\omega^{n+1} - \omega^n + \frac{\Delta t}{\Delta t + 2} L_P L \left(\epsilon^{n+1} - \epsilon^n \right) = - \left(L_P + L \right) \left(\epsilon^{n+1} + \epsilon^n \right). \tag{4.76}$$

Multiplying both sides of (4.76) by $\omega^{n+1} + \omega^n$ with respect to the inner product (4.44) we obtain, using (4.73),

$$\begin{aligned} & \|\omega^{n+1}\|^2 - \|\omega^n\|^2 \\ & + \frac{\Delta t^3}{4(\Delta t + 2)} \left(\left(-P' \delta_y^2 - P \delta_x \delta_y^2 + \delta_x^2 \delta_y^2 \right) \left(\epsilon^{n+1} - \epsilon^n \right), \omega^{n+1} + \omega^n \right) \\ & + \frac{\Delta t}{2} \left(\left(P' + P \delta_x - \delta_x^2 - \delta_y^2 \right) \left(\epsilon^{n+1} + \epsilon^n \right), \omega^{n+1} + \omega^n \right) = 0. \end{aligned} \tag{4.77}$$

Using the equalities (4.50)-(4.52) we can rewrite (4.77) as

$$\begin{aligned} & \|\omega^{n+1}\|^2 + \frac{\Delta t^3}{4(\Delta t + 2)} \|\delta_x^+ \delta_y^+ \epsilon^{n+1}\|^2_* + \|\delta_x^+ \epsilon^{n+1}\|^2_{*x} + \|\delta_y^+ \epsilon^{n+1}\|^2_{*y} \\ = \; & \|\omega^n\|^2 + \frac{\Delta t^3}{4(\Delta t + 2)} \|\delta_x^+ \delta_y^+ \epsilon^n\|^2_* + \|\delta_x^+ \epsilon^n\|^2_{*x} + \|\delta_y^+ \epsilon^n\|^2_{*y} \\ & - \frac{\Delta t}{2} \left(\|\delta_x^+ (\epsilon^{n+1} + \epsilon^n)\|^2_{*x} + \|\delta_y^+ (\epsilon^{n+1} + \epsilon^n)\|^2_{*y} \right) \\ & - \frac{\Delta t^2}{2(\Delta t + 2)} \|\delta_x^+ \delta_y^+ (\epsilon^{n+1} - \epsilon^n)\|^2_* \\ & - \frac{\Delta t}{2} \left(\left(P' + P \delta_x \right) \left(\epsilon^{n+1} + \epsilon^n \right), \omega^{n+1} + \omega^n \right) \\ & + \frac{\Delta t^3}{4(\Delta t + 2)} \left(\left(P' \delta_y^2 + P \delta_x \delta_y^2 \right) \left(\epsilon^{n+1} - \epsilon^n \right), \omega^{n+1} + \omega^n \right). \end{aligned} \tag{4.78}$$

Let us now discuss the terms with P. With the assumption that P' is non-negative and by (4.31), we obtain

$$
\begin{aligned}
& -\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\omega^{n+1}+\omega^n\right) \\
= & -\frac{\Delta t}{2}\left(P'(\epsilon^{n+1}+\epsilon^n),\epsilon^{n+1}+\epsilon^n\right)-\left(P'(\epsilon^{n+1}+\epsilon^n),\epsilon^{n+1}-\epsilon^n\right) \\
= & -\frac{\Delta t}{2}\Delta x\Delta y\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}P_i'\left(\epsilon_{i,j}^{n+1}+\epsilon_{i,j}^n\right)^2 \\
& -\Delta x\Delta y\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}P_i'\left(\left(\epsilon_{i,j}^{n+1}\right)^2-\left(\epsilon_{i,j}^n\right)^2\right) \\
\le & -\Delta x\Delta y\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}P_i'\left(\epsilon_{i,j}^{n+1}\right)^2+\Delta x\Delta y\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}P_i'\left(\epsilon_{i,j}^n\right)^2 \\
= & -\|\epsilon^{n+1}\|_{P'}^2+\|\epsilon^n\|_{P'}^2. \qquad (4.79)
\end{aligned}
$$

We also have, from (4.56),

$$
\begin{aligned}
-\frac{\Delta t}{2}\left(P\delta_x\left(\epsilon^{n+1}+\epsilon^n\right),\omega^{n+1}+\omega^n\right) \le & \frac{\Delta t}{2}\eta_1\|P\|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2 \\
& +\frac{\Delta t}{4\eta_1}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \qquad (4.80)
\end{aligned}
$$

On the other hand, by (4.31) and summation by parts we have

$$
\begin{aligned}
& \left(P'\delta_y^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) \\
& = -\left(P'\delta_y^+\left(\epsilon^{n+1}-\epsilon^n\right),\delta_y^+\left(\epsilon^{n+1}+\epsilon^n\right)\right)_{*y} \\
& \quad -\frac{2}{\Delta t}\left(P'\delta_y^+\left(\epsilon^{n+1}-\epsilon^n\right),\delta_y^+\left(\epsilon^{n+1}-\epsilon^n\right)\right)_{*y} \\
& = -\Delta x\Delta y\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}P_i'\left(\left(\delta y^+\epsilon_{i,j}^{n+1}\right)^2-\left(\delta y^+\epsilon_{i,j}^n\right)^2\right) \\
& \quad -\frac{2}{\Delta t}\Delta x\Delta y\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1}P_i'\left(\delta y^+\left(\epsilon_{i,j}^{n+1}-\epsilon_{i,j}^n\right)\right)^2 \\
& \le -\|\delta_y^+\epsilon^{n+1}\|_{*yP'}^2+\|\delta_y^+\epsilon^n\|_{*yP'}^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \frac{\Delta t^3}{4(\Delta t+2)}\left(P'\delta_y^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right) \\
& \quad \le -\frac{\Delta t^3}{4(\Delta t+2)}\left(\|\delta_y^+\epsilon^{n+1}\|_{*yP'}^2-\|\delta_y^+\epsilon^n\|_{*yP'}^2\right). \qquad (4.81)
\end{aligned}
$$

Finally, from (4.59) we have

$$\begin{aligned}
&\frac{\Delta t^3}{4(\Delta t+2)}\left(P\delta_x\delta_y^2\left(\epsilon^{n+1}-\epsilon^n\right),\omega^{n+1}+\omega^n\right)\\
\leq\ &\frac{\eta_2\Delta t^3}{4(\Delta t+2)}\|P\|^2\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2\\
+\ &\frac{\Delta t^3}{2(\Delta t+2)\eta_2\Delta y^2}(\|\omega^{n+1}\|^2+\|\omega^n\|^2). \qquad (4.82)
\end{aligned}$$

From (4.78) and the inequalities (4.79)-(4.82), we obtain

$$\begin{aligned}
&\|\omega^{n+1}\|^2+\frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2+\|\delta_x^+\epsilon^{n+1}\|_{*x}^2+\|\delta_y^+\epsilon^{n+1}\|_{*y}^2\\
&\leq\|\omega^n\|^2+\frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2+\|\delta_x^+\epsilon^n\|_{*x}^2+\|\delta_y^+\epsilon^n\|_{*y}^2\\
&-\frac{\Delta t}{2}\left(\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2+\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2\right)\\
&-\frac{\Delta t^2}{2(\Delta t+2)}\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2-\|\epsilon^{n+1}\|_{P'}^2+\|\epsilon^n\|_{P'}^2\\
&+\frac{\Delta t}{2}\Big(\eta_1\|P\|^2\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2+\frac{\eta_1^{-1}}{2}\left(\|\omega^{n+1}\|^2+\|\omega^n\|^2\right)\Big)\\
&-\frac{\Delta t^3}{4(\Delta t+2)}\left(\|\delta_y^+\epsilon^{n+1}\|_{*yP'}^2-\|\delta_y^+\epsilon^n\|_{*yP'}^2\right)\\
&+\frac{\Delta t^3}{4(\Delta t+2)}\Big(\eta_2\|P\|^2\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2+2\eta_2{}^{-1}\Delta y^{-2}\left(\|\omega^{n+1}\|^2+\|\omega^n\|^2\right)\Big). \qquad (4.83)
\end{aligned}$$

Reorganizing the terms and using the condition $\Delta t\leq c\Delta y$, we obtain

$$\begin{aligned}
&A\|\omega^{n+1}\|^2+\tfrac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2+\|\delta_x^+\epsilon^{n+1}\|_{*x}^2+\|\delta_y^+\epsilon^{n+1}\|_{*y}^2\\
&\quad+\|\epsilon^{n+1}\|_{P'}^2+\tfrac{\Delta t^3}{4(\Delta t+2)}\|\delta_y^+\epsilon^{n+1}\|_{*yP'}^2\\
&\leq B\|\omega^n\|^2+\tfrac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2+\|\delta_x^+\epsilon^n\|_{*x}^2+\|\delta_y^+\epsilon^n\|_{*y}^2\\
&\quad+\|\epsilon^n\|_{P'}^2+\tfrac{\Delta t^3}{4(\Delta t+2)}\|\delta_y^+\epsilon^n\|_{*yP'}^2\\
&\quad+\tfrac{\Delta t}{2}\|\delta_x^+(\epsilon^{n+1}+\epsilon^n)\|_{*x}^2(\eta_1\|P\|^2-1)-\tfrac{\Delta t}{2}\|\delta_y^+(\epsilon^{n+1}+\epsilon^n)\|_{*y}^2\\
&\qquad+\frac{\Delta t^2}{4(\Delta t+2)}\|\delta_x^+\delta_y^+(\epsilon^{n+1}-\epsilon^n)\|_*^2\left(\Delta t\eta_2\|P\|^2-2\right), \qquad (4.84)
\end{aligned}$$

where A and B are given by

$$A=1-\frac{\Delta t}{2}\left(\frac{\eta_1^{-1}}{2}+\frac{c^2\eta_2^{-1}}{\Delta t+2}\right) \qquad (4.85)$$

and

$$B = 1 + \frac{\Delta t}{2}\left(\frac{\eta_1^{-1}}{2} + \frac{c^2\eta_2^{-1}}{\Delta t + 2}\right). \tag{4.86}$$

Let us choose $\eta_1 = \eta_2 = \eta$ with

$$\eta \leq \frac{1}{\|P\|^2}. \tag{4.87}$$

Then, from (4.84) we have

$$\begin{aligned} &A\|\omega^{n+1}\|^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^{n+1}\|_*^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 \\ &\quad + \|\epsilon^{n+1}\|_{P'}^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_y^+\epsilon^{n+1}\|_{*yP'}^2 \\ &\leq B\|\omega^n\|^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_x^+\delta_y^+\epsilon^n\|_*^2 + \|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2 \\ &\quad + \|\epsilon^n\|_{P'}^2 + \frac{\Delta t^3}{4(\Delta t+2)}\|\delta_y^+\epsilon^n\|_{*yP'}^2, \end{aligned} \tag{4.88}$$

with

$$A = 1 - \frac{\Delta t}{2}\left(\frac{\eta^{-1}}{2} + \eta^{-1}\frac{c^2}{\Delta t + 2}\right) \tag{4.89}$$

and

$$B = 1 + \frac{\Delta t}{2}\left(\frac{\eta^{-1}}{2} + \eta^{-1}\frac{c^2}{\Delta t + 2}\right). \tag{4.90}$$

Using Lemma 4.2.3. and $\Delta t \leq c\Delta y$, from (4.88) we obtain

$$\begin{aligned} &A\|\omega^{n+1}\|^2 + \|\delta_x^+\epsilon^{n+1}\|_{*x}^2 + \|\delta_y^+\epsilon^{n+1}\|_{*y}^2 + \|\epsilon^{n+1}\|_{P'}^2 \\ &\quad \leq B\|\omega^n\|^2 + \left(1 + \frac{c^2\Delta t}{\Delta t + 2}\right)\|\delta_x^+\epsilon^n\|_{*x}^2 + \|\delta_y^+\epsilon^n\|_{*y}^2 + \left(1 + \frac{c^2\Delta t}{\Delta t + 2}\right)\|\epsilon^n\|_{P'}^2. \end{aligned} \tag{4.91}$$

If we choose η such that

$$\eta > \frac{1}{4}(1 + c^2),$$

we can easily check that $0 < A \leq 1$. Additionally if we choose η such that

$$\eta \leq \frac{1 + c^2}{2c^2}$$

we have $B \geq 1 + c^2\Delta t/(\Delta t + 2)$. Therefore by choosing η such that

$$\frac{1}{4}(1 + c^2) < \eta \leq \min\left\{\frac{1}{\|P\|^2}, \frac{1 + c^2}{2c^2}\right\}, \tag{4.92}$$

it follows

$$A\left(\|\omega^{n+1}\|^2+\|\delta_x^+\epsilon^{n+1}\|_{*x}^2+\|\delta_y^+\epsilon^{n+1}\|_{*y}^2+\|\epsilon^{n+1}\|_{P'}^2\right)$$
$$\leq B\left(\|\omega^{n}\|^2+\|\delta_x^+\epsilon^{n}\|_{*x}^2+\|\delta_y^+\epsilon^{n}\|_{*y}^2+\|\epsilon^{n}\|_{P'}^2\right). \tag{4.93}$$

Consequently, by noting that

$$\frac{B}{A}=1+\Delta t\frac{\frac{\eta^{-1}}{2}+\eta^{-1}\frac{c^2}{\Delta t+2}}{1-\frac{\Delta t}{2}\left(\frac{\eta^{-1}}{2}+\eta^{-1}\frac{c^2}{\Delta t+2}\right)}\leq 1+C_1\Delta t,$$

where C_1 denotes a constant independent of $\Delta x, \Delta y, \Delta t$, we obtain the main result.

■

From the previous theorem we get the following result.

Corollary 4.2.2. *Suppose that $\{U_{i,j}^n, W_{i,j}^n\}$ and $\{V_{i,j}^n, Y_{i,j}^n\}$ are solutions of the finite difference scheme (4.31) and (4.74) which satisfy the boundary condition (4.25), and have different initial values $\{U_{i,j}^0, W_{i,j}^0\}$ and $\{V_{i,j}^0, Y_{i,j}^0\}$ respectively. Let $\omega_{i,j}^n = W_{i,j}^n - Y_{i,j}^n$, $\epsilon_{i,j}^n = U_{i,j}^n - V_{i,j}^n$. For $\Delta t \leq 1$, such that, $\Delta t \leq c\Delta y$, with constant c, then $\{\omega_{i,j}^n, \epsilon_{i,j}^n\}$ satisfy*

$$\|\omega^{n}\|^2+\|\delta_x^+\epsilon^{n}\|_{*x}^2+\|\delta_y^+\epsilon^{n}\|_{*y}^2+\|\epsilon^{n}\|_{P'}^2$$
$$\leq K_1\left(\|\omega^{0}\|^2+\|\delta_x^+\epsilon^{0}\|_{*x}^2+\|\delta_y^+\epsilon^{0}\|_{*y}^2+\|\epsilon^{0}\|_{P'}^2\right)$$

where K_1 denotes a constant independent of $\Delta x, \Delta y, \Delta t$.

The proof follows from Theorem 4.2.2 by making recursion with respect to n or by considering the proof of Corollary 2.2.1. Thus, we can conclude that the difference scheme (4.74) is stable.

Remark 4.2.3. *The choice of constant c mentioned in the previous theorem can depend on the value of P as can be concluded by observing the condition (4.92).*

Remark 4.2.4. *In the case where $P = Q = 0$, by following the same steps of the previous proof, we can easily conclude that the difference scheme (4.46) is not unconditionally stable.*

4.2.2 Numerical results

In this section we present numerical tests which confirm the previous theoretical results obtained for the difference scheme. We compare some numerical results with exact solutions and we also illustrate the behavior of some solutions. Let

$$\epsilon_{i,j} = u_{i,j} - U_{i,j}, \qquad \omega_{i,j} = w_{i,j} - W_{i,j}, \tag{4.94}$$

where u is the exact solution, w is defined by (4.5) and U and W are the approximate solutions, respectively. To measure the error and the rate of convergence we consider the norms defined by

$$\|\epsilon\|_\infty = \max|u_{i,j} - U_{i,j}|, \qquad \|\omega\|_\infty = \max|w_{i,j} - W_{i,j}|, \tag{4.95}$$

for $1 \le i \le N_x - 1, 1 \le j \le N_y - 1$, and

$$\begin{aligned} \|\epsilon\| &= \left(\Delta x \Delta y \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y-1} |u_{i,j} - U_{i,j}|^2 \right)^{1/2} \\ \|\omega\| &= \left(\Delta x \Delta y \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y-1} |w_{i,j} - W_{i,j}|^2 \right)^{1/2}. \end{aligned} \tag{4.96}$$

We present two problems for which we are able to determine the exact solution in order to compute the errors and the convergence rate.

Example 4.2.1. We consider the problem

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} = -P\frac{\partial u}{\partial x} - Q\frac{\partial u}{\partial y} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, \ (x,y) \in (0,1)\times(0,1), \ t > 0,$$

with initial conditions

$$u(x,y,0) = e^{Px/2+Qy/2}\sinh(bx)\sinh(cy),$$

$$\frac{\partial u}{\partial t}(x,y,0) = -\frac{a}{2}e^{Px/2+Qy/2}\sinh(bx)\sinh(cy),$$

and boundary conditions

$$u(0,y,t) = u(x,0,t) = 0,$$

$$u(1, y, t) = e^{-at/2} e^{P/2+Qy/2} \sinh(b) \sinh(cy),$$

$$u(x, 1, t) = e^{-at/2} e^{Px/2+Q/2} \sinh(bx) \sinh(c).$$

The exact solution is given by

$$u(x, y, t) = e^{-at/2} e^{Px/2+Qy/2} \sinh(bx) \sinh(cy),$$

where constants a, b and c satisfy the relation

$$b^2 + c^2 = \frac{a^2 - 2a + P^2 + Q^2}{4}.$$

We consider this problem with $a = 1 + \sqrt{17 + P^2 + Q^2}$, $b = \sqrt{(4 + P^2)/2}$ and $c = \sqrt{(4 + Q^2)/2}$. We show the errors and convergence rates for different P values, such as, in Table 4.3 and Table 4.4 for $P = 0.5, Q = 0.4$, in Table 4.5 and Table 4.6 for $P = 0.5, Q = -0.4$, and in Table 4.7 and Table 4.8 for $P = 0.5, Q = 0$. For all the cases we observe the convergence rate is second order as expected, although the norm $\ell_{2,\Delta}$ provides smaller errors.

$\Delta x = \Delta y$	Error $\|\epsilon\|_\infty$	Rate	Error $\|\omega\|_\infty$	Rate
$1/128$	0.1945×10^{-4}		0.1324×10^{-3}	
$1/256$	0.4837×10^{-5}	2.0	0.4098×10^{-4}	1.7
$1/512$	0.1201×10^{-5}	2.0	0.1205×10^{-4}	1.8
$1/1024$	0.2996×10^{-6}	2.0	0.3424×10^{-5}	1.8
$1/2048$	0.7487×10^{-7}	2.0	0.9439×10^{-6}	1.9

Table 4.3: Errors and rates obtained from Example 4.2.1 with $P = 0.5$, $Q = 0.4$, $t = 1$, $0 \leq x, y \leq 1$ and $\Delta t = \Delta x$, computed with the norm ℓ_∞.

$\Delta x = \Delta y$	Error $\|\epsilon\|$	Rate	Error $\|\omega\|$	Rate
$1/128$	0.8624×10^{-5}		0.4189×10^{-4}	
$1/256$	0.2156×10^{-5}	2.0	0.1059×10^{-4}	2.0
$1/512$	0.5391×10^{-6}	2.0	0.2670×10^{-5}	2.0
$1/1024$	0.1348×10^{-6}	2.0	0.6719×10^{-6}	2.0
$1/2048$	0.3369×10^{-7}	2.0	0.1687×10^{-6}	2.0

Table 4.4: Errors and rates obtained from Example 4.2.1 with $P = 0.5$, $Q = 0.4$, $t = 1$, $0 \leq x, y \leq 1$ and $\Delta t = \Delta x$, computed with the norm $\ell_{2,\Delta}$.

$\Delta x = \Delta y$	Error $\|\epsilon\|_\infty$	Rate	Error $\|\omega\|_\infty$	Rate
$1/128$	0.1729×10^{-4}		0.1294×10^{-3}	
$1/256$	0.4306×10^{-5}	2.0	0.4032×10^{-4}	1.7
$1/512$	0.1071×10^{-5}	2.0	0.1193×10^{-4}	1.8
$1/1024$	0.2666×10^{-6}	2.0	0.3402×10^{-5}	1.8
$1/2048$	0.6666×10^{-7}	2.0	0.9399×10^{-6}	1.9

Table 4.5: Errors and rates obtained from Example 4.2.1 with $P = 0.5$, $Q = -0.4$, $t = 1$, $0 \leq x, y \leq 1$ and $\Delta t = \Delta x$, computed with the norm ℓ_∞.

$\Delta x = \Delta y$	Error $\|\epsilon\|$	Rate	Error $\|\omega\|$	Rate
$1/128$	0.7423×10^{-5}		0.3617×10^{-4}	
$1/256$	0.1856×10^{-5}	2.0	0.9179×10^{-5}	2.0
$1/512$	0.4641×10^{-6}	2.0	0.2319×10^{-5}	2.0
$1/1024$	0.1160×10^{-6}	2.0	0.5843×10^{-6}	2.0
$1/2048$	0.2900×10^{-7}	2.0	0.1468×10^{-6}	2.0

Table 4.6: Errors and rates obtained from Example 4.2.1 with $P = 0.5$, $Q = -0.4$, $t = 1$, $0 \leq x, y \leq 1$ and $\Delta t = \Delta x$, computed with the norm $\ell_{2,\Delta}$.

$\Delta x = \Delta y$	Error $\|\epsilon\|_\infty$	Rate	Error $\|\omega\|_\infty$	Rate
$1/128$	0.1767×10^{-4}		0.1256×10^{-3}	
$1/256$	0.4401×10^{-5}	2.0	0.3899×10^{-4}	1.7
$1/512$	0.1094×10^{-5}	2.0	0.1150×10^{-4}	1.8
$1/1024$	0.2726×10^{-6}	2.0	0.3272×10^{-5}	1.8
$1/2048$	0.6813×10^{-7}	2.0	0.9028×10^{-6}	1.9

Table 4.7: Errors and rates obtained from Example 4.2.1 with $P = 0.5$, $Q = 0$, $t = 1$, $0 \leq x, y \leq 1$ and $\Delta t = \Delta x$, computed with the norm ℓ_∞.

Example 4.2.2. In the second problem we consider equation (4.1) with $\theta = 0$, $D = 1$ and P, Q constants, that is, the parabolic equation

$$\frac{\partial u}{\partial t} = -P\frac{\partial u}{\partial x} - Q\frac{\partial u}{\partial y} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, \quad x, y \in \,]-\infty, \infty[\,, t > 0.$$

The initial condition is $u(x, y, 0) = e^{-\left(x^2+y^2\right)}$ and the boundary conditions are

$$\lim_{x\to-\infty} u(x, y, t) = 0, \qquad \lim_{x\to+\infty} u(x, y, t) = 0,$$
$$\lim_{y\to-\infty} u(x, y, t) = 0, \qquad \lim_{y\to+\infty} u(x, y, t) = 0.$$

$\Delta x = \Delta y$	Error $\|\epsilon\|$	Rate	Error $\|\omega\|$	Rate
$1/128$	0.7741×10^{-5}		0.3732×10^{-4}	
$1/256$	0.1935×10^{-5}	2.0	0.9449×10^{-5}	2.0
$1/512$	0.4839×10^{-6}	2.0	0.2385×10^{-5}	2.0
$1/1024$	0.1210×10^{-6}	2.0	0.6004×10^{-6}	2.0
$1/2048$	0.3024×10^{-7}	2.0	0.1509×10^{-6}	2.0

Table 4.8: Errors and rates obtained from Example 4.2.1 with $P = 0.5$, $Q = 0$, $t = 1$, $0 \leq x, y \leq 1$ and $\Delta t = \Delta x$, computed with the norm $\ell_{2,\Delta}$.

The analytical solution is given by

$$u\left(x, y, t\right) = \frac{1}{\sqrt{1+4t}} \mathrm{e}^{-\frac{(x-Pt)^2+(y-Qt)^2}{1+4t}}. \tag{4.97}$$

We present the errors and convergence rates for different P values, such as, in Table 4.9 for $P = 1, Q = 1$, and in Table 4.10 for $P = 0, Q = 0$. For both cases we observe the convergence rate is second order, with and without the values P, Q. Unlike the previous example, the smaller errors are provided by the norm ℓ_∞.

$\Delta x = \Delta y$	Error $\|\epsilon\|_\infty$	Rate	Error $\|\epsilon\|$	Rate
$20/128$	0.1226×10^{-2}		0.2373×10^{-2}	
$20/256$	0.3078×10^{-3}	2.0	0.5890×10^{-3}	2.0
$20/512$	0.7215×10^{-4}	2.1	0.1405×10^{-3}	2.1
$20/1024$	0.1748×10^{-4}	2.1	0.3434×10^{-4}	2.0
$20/2048$	0.4370×10^{-5}	2.0	0.8585×10^{-5}	2.0

Table 4.9: Errors and rates obtained from Example 4.2.2 for $P = Q = 1$, $t = 1$, $-10 \leq x, y \leq 10$, and $\Delta t = \Delta x$, computed with the norms ℓ_∞ and $\ell_{2,\Delta}$.

$\Delta x = \Delta y$	Error $\|\epsilon\|_\infty$	Rate	Error $\|\epsilon\|$	Rate
$20/128$	0.4641×10^{-3}		0.8902×10^{-3}	
$20/256$	0.1176×10^{-3}	2.0	0.2226×10^{-3}	2.0
$20/512$	0.2618×10^{-4}	2.1	0.5083×10^{-4}	2.1
$20/1024$	0.6174×10^{-5}	2.1	0.1215×10^{-4}	2.1
$20/2048$	0.1544×10^{-5}	2.0	0.3038×10^{-5}	2.0

Table 4.10: Errors and rates obtained from Example 4.2.2 for $P = Q = 0$, $t = 1$, $-10 \leq x, y \leq 10$, and $\Delta t = \Delta x$, computed with the norms ℓ_∞ and $\ell_{2,\Delta}$.

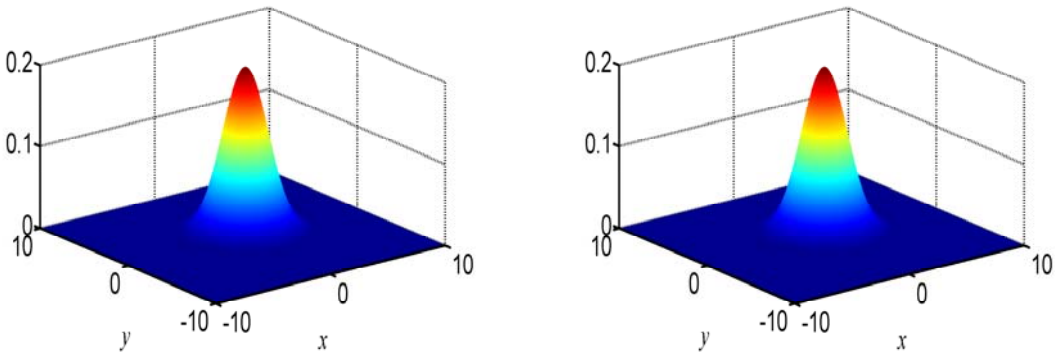

Figure 4.4: Solution $u(x, y, t)$ of Example 4.2.2 versus approximate solution for $P = Q = 1$ at $t = 1$. Left: exact solution $u(x, y, t)$. Right: approximate solution for $\Delta t = \Delta x = \Delta y = 0.02$.

In Figure 4.4 we show how the solution behaves for this problem: the solution u and the numerical solution match very well.

Example 4.2.3. Another example that gives an insight on the physical behavior of the solution is the equation

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(Pu) - \frac{\partial}{\partial y}(Qu) + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, \quad x, y \in \mathbb{R},\ t > 0,$$

with initial conditions

$$u(x, y, 0) = \frac{1}{\sqrt{\pi}} e^{-(x^2+y^2)}, \qquad \frac{\partial u}{\partial t}(x, y, 0) = 0,$$

and boundary conditions

$$\lim_{x\to-\infty} u(x, y, t) = 0, \qquad \lim_{x\to+\infty} u(x, y, t) = 0,$$
$$\lim_{y\to-\infty} u(x, y, t) = 0, \qquad \lim_{y\to+\infty} u(x, y, t) = 0.$$

In Figures 4.5 and 4.6 we display the approximate solutions for P and Q constants and observe how the solution changes with the direction of those values, for $t = 3$.

In Figure 4.7 we consider P non-constant, $P(x) = x/2$, and $Q = 0$. The behavior of the solution can be observed as we travel in time.

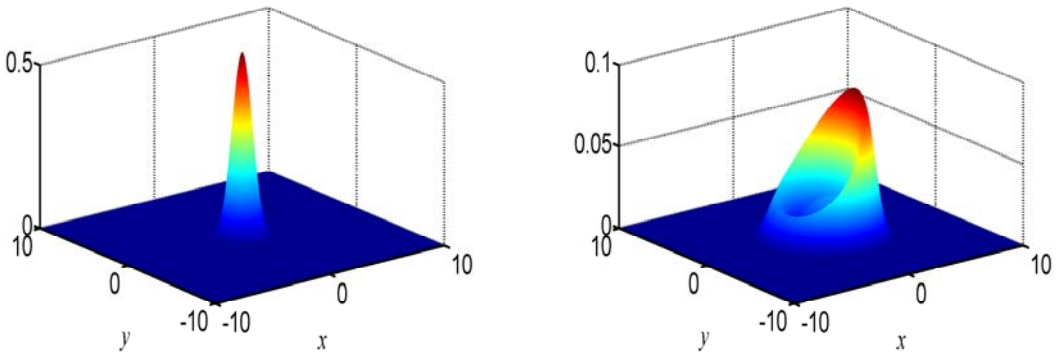

Figure 4.5: Results of Example 4.2.3. Left: initial condition. Right: approximate solution for $P = 0.5$ and $Q = 0$ at $t = 3$. Computed with $\Delta t = \Delta x = \Delta y = 0.02$.

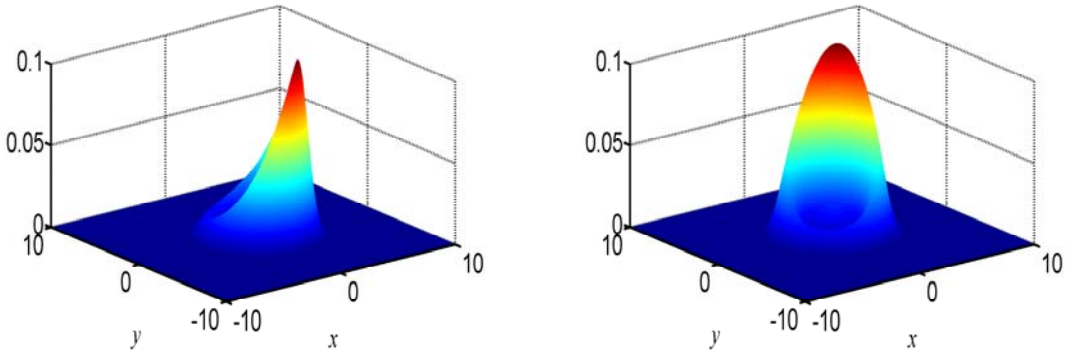

Figure 4.6: Approximate solution of Example 4.2.3 for $t = 3$ computed with $\Delta t = \Delta x = \Delta y = 0.02$. Left: $P = 0.5$, $Q = -0.5$. Right: $P = 0.5$, $Q = 0.5$.

Example 4.2.4. We consider equation (4.22) in the domain $[0, \infty[\times[0, 1]$ with P and Q constants, to observe the behavior of the solution performed with nonzero boundary conditions. The initial conditions are $u_0(x, y) = 0$ and $u_1(x, y) = 0$ and the boundary conditions are

$$u(x, 0, t) = 0, \qquad u(x, 1, t) = 0, \tag{4.98}$$

$$u(0, y, t) = \sin(\pi y), \qquad u(\infty, y, t) = 0. \tag{4.99}$$

Next we present the solution for $t = 1$, $P = 1$ and $Q = 0$. We observe that for $\theta = 0$, Figure 4.8, the solution is smooth. For $\theta = 1$ the solution presents a jump discontinuity at $x = 1$ but the CN-ADI method performs quite well

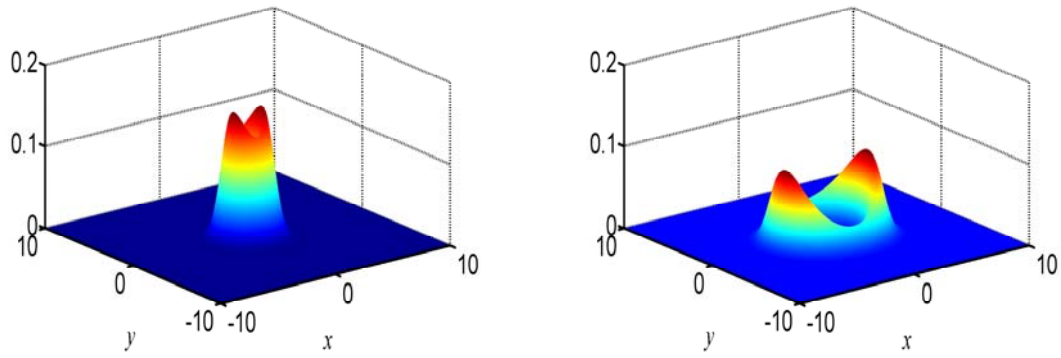

Figure 4.7: Approximate solution of Example 4.2.3 computed with $P(x) = x/2$, $Q = 0$, $\Delta t = \Delta x = \Delta y = 0.02$. Left: $t = 1$. Right: $t = 3$.

without oscillations, although performed with a very small space step.

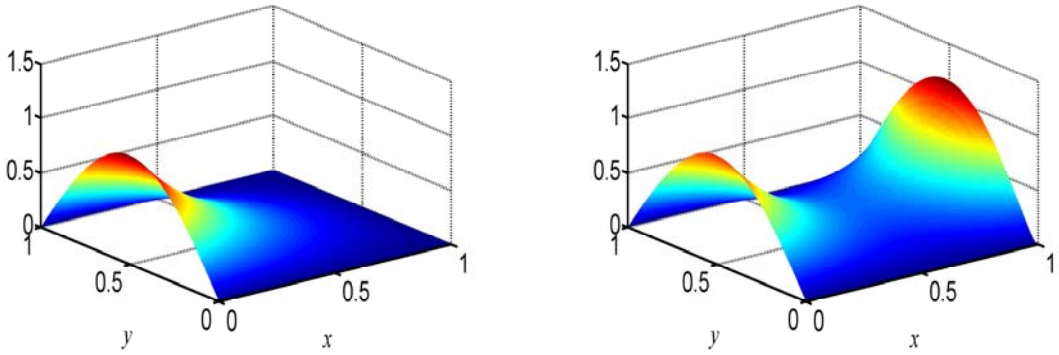

Figure 4.8: Approximate solution of Example 4.2.4 for $P = 1$ and $Q = 0$ at $t = 1$. Computed with $\Delta t = \Delta x = \Delta y = 0.001$. Left: $\theta = 0$. Right: $\theta = 1$.

Chapter 5

Final remarks and perspectives of future research

In this final chapter we draw some conclusions about the subjects addressed in this work. We also leave some new and open questions we intend to consider in future investigation.

5.1 Conclusions

In this work we focused our attention in the study and development of numerical methods to solve a second order hyperbolic diffusion equation in the presence of a potential field. The solution of an initial boundary value problem is under consideration, with Dirichlet boundary conditions, in one and two dimensions.

In one dimension, we have studied a differential equation that takes into account the existence of a potential field and a relaxation time parameter, and we have seen its solution is affected by those values. For the relaxation time parameter $\theta = 0$, the equation is parabolic and therefore the solution is smooth. For $\theta \neq 0$ we have a hyperbolic equation which transports an initial discontinuity at the inflow boundary. Such discontinuity dissipates as we

travel in time.

We considered first a finite difference scheme based on Crank-Nicolson method. We proved its convergence and the numerical results confirmed our theoretical analysis. To avoid integration in time, we introduce numerical methods based on the Laplace transform. These numerical methods consist first of applying the Laplace transform and then the resulting equation is approximated by a proper spatial discretization. Finally, an inverse Laplace transform algorithm using continued fractions, which is accurate and very efficient for time integration, is implemented to obtain the desired numerical solution. For the spatial discretization three different methods have been used: the Laplace-FD, the Laplace-FV and the Laplace-PL methods.

If P is constant and the ordinary differential equation obtained with the Laplace transform is homogeneous, we are able to apply the inverse Laplace algorithm directly. If we have a non-homogeneous equation, we can apply the inverse Laplace algorithm directly only if we know a particular solution, otherwise we must consider a spatial discretization. If P is non-constant, the spatial discretization is mandatory.

Considering the spatial discretization, if we compare the Laplace-FD and the Laplace-FV methods, we have seen that the Laplace-FV method yields better performances near discontinuities. The Laplace-PL method is also accurate and performs quite well near discontinuities, avoiding oscillations where other methods do not. On the whole, the Laplace-PL method has the best performance since it can be applied to every problems considered with good results.

In summary, we can conclude that the methods based on the Laplace transform are a good choice for problems where there is interest to deal with very large times. An example that illustrates this behavior was studied in Section 3.4, where a symmetric periodic potential field was included. Any iterative numerical method would take too long to compute the solution for similar times. Even the unconditionally implicit numerical methods that allow large time steps are not able to give solutions so quickly. This is the case of the CN method, although we have seen its good performance and

efficiency for short times and this is the reason it was used in this work.

Three of the numerical methods applied in one dimension, Chapters 2 and 3, were extended in Chapter 4 in order to solve the analogous hyperbolic diffusion equation in two dimensions: CN-2D, Laplace-FD-2D and Laplace-FV-2D methods. Due to the computational limitations of these schemes, we have derived a second order accurate ADI finite difference method to solve the two dimensional problem. The stability of the method was proved by the discrete energy method. Several numerical results demonstrate the second order accuracy of the method, when compared with some analytical solutions, and are in agreement with the theoretical analysis presented. We highlight the efficiency of this finite difference scheme, which was achieved by the procedure of splitting the resolution of one system in two tridiagonal systems. The good performance of the CN-ADI method is confirmed with the contribution of problems containing a great variety of initial and boundary conditions.

5.2 Future research

Throughout this research we found some issues that may be worth being explored in the future:

(i) As already mentioned in the literature review, there are a significative number of numerical methods that contributed to achieve the fourth order accuracy in space, when solving diffusion problems in one and two dimensions. We have the same purpose for our model problem.

(ii) The computational inefficiency patent in the Laplace-FD-2D and the Laplace-FV-2D methods should be overcome in order to facilitate their application in two dimensions. This would permit us to consider a wide range of initial and boundary conditions, without restricting the length of the spatial variables. A possible approach to accomplish this is to introduce also an ADI method as we did for the Crank-Nicolson method.

Furthermore, our interest in this topic increases when, once again, we focus our attention in the behavior of the solution for long times. Yet, we are aware of the difficulties involved due to the Laplace transform mechanism and the viability of this idea is still in study.

(iii) Another possibility of study is the temperature equation of the Jeffreys type [49], which models heat conduction problems. Although it is not directly related to the equation presented in this thesis, we find it interesting since it contains a mixed derivative term.

(iv) The numerical solution of partial differential equations, when defined on unbounded domains, is sometimes obtained with artificial boundary conditions to limit the area of computation [38]. To minimize possible reflections that occur in these boundaries, incorporation of absorbing boundary conditions (ABCs) have been used to guarantee a realistic, accurate and stable approximation to the solution on the original and unbounded domain. The study and design of (ABCs) can be found in the literature for linear and nonlinear problems, in one and two space dimensions, for instance, in [7, 37, 38, 48, 88]. We want to investigate deeper this issue, since we have to confine to a computational domain when we deal with artificial boundary conditions in unbounded domains. The artificial boundary conditions justify this strategy.

Bibliography

[1] J. Abate, W. Whitt, *Numerical inversion of Laplace transforms of probability distributions*, ORSA Journal on Computing, 7(1): 36–43, 1995.

[2] J. Ahn, S. Kang, Y.H. Kwon, *A flexible inverse Laplace transform algorithm and its application*, Computing, 71(2): 115–131, 2003.

[3] Y. M. Ali, L. C. Zhang, *Relativistic moving heat source*, International Journal of Heat and Mass Transfer, 48: 2741–2758, 2005.

[4] A. Araújo, A. K. Das, C. Neves, E. Sousa, *Numerical solution for a non-Fickian diffusion in a periodic potential*, Communications in Computational Physics, 13(2): 502–525, 2013.

[5] A. Araújo, C. Neves, E. Sousa, *A Laplace transform piecewise linearized method for a second order hyperbolic equation*, AIP Conference Proceedings, 1479: 2187-2190, 2012.

[6] A. Araújo, C. Neves, E. Sousa, *An alternating direction implicit method for a two-dimensional hyperbolic diffusion equation*, submitted for publication, 2013.

[7] A. Aydemir, *Comparison of absorbing boundary conditions: A correlation of reflection coefficients and determination of the optimum grid interval for migration*, Journal of Applied Geophysics, 67: 143–149, 2009.

[8] I. M. Babuska, S. A. Sauter, *Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?*, SIAM Review, 42(3): 451–484, 2000.

[9] G. Bao, G. W. Wei, S. Zhao, *Numerical solution of the Helmholtz equation with high wavenumbers*, International Journal for Numerical Methods in Engineering, 59: 389–408, 2004.

[10] G. Barbero, J. R. Macdonald, *Transport process of ions in insulating media in the hyperbolic diffusion regime*, Physical Review E, 81(5): 051503, 2010.

[11] H. Bradt, *Astrophysics Processes: The Physics of Astronomical Phenomena*, Cambridge University Press, 2008.

[12] A. C. Branka, A. K. Das, D. M. Heyes, *Overdamped Brownian motion in periodic symmetric potentials*, Journal of Chemical Physics, 113(22): 9911–9919, 2000.

[13] H.-T. Chen, J.-Y. Lin, *Analysis of two-dimensional hyperbolic heat conduction problems*, International Journal of Heat and Mass Transfer, 37(1): 153–164, 1994.

[14] H.-T. Chen, K.-C. Liu, *Numerical analysis of a non-Fickian diffusion problems in a potential field*, Numerical Heat Transfer, Part B 40: 265–282, 2001.

[15] H.-T. Chen, K.-C. Liu, *Analysis of non-Fickian diffusion problems in a composite medium*, Computer Physics Communications, 150: 31–42, 2003.

[16] T.-M. Chen, *Numerical solution of hyperbolic heat conduction in thin surface layers*, International Journal of Heat and Mass Transfer, 50: 4424–4429, 2007.

[17] T.-M. Chen, *Numerical solution of hyperbolic heat conduction problems in the cylindrical coordinate system by the hybrid Green's function method*, International Journal of Heat and Mass Transfer, 53: 1319–1325, 2010.

[18] J. Crank, *The Mathematics of Diffusion*, Second Edition, Oxford University Press, 1975.

[19] J. Crank, 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: 50–67, 1947.

[20] K. Crump, *Numerical inversion of Laplace transforms using a Fourier series approximation*, Journal of the Association for Computing Machinery, 23(1): 89–96, 1976.

[21] M. Cui, *Fourth-order compact scheme for the one-dimensional sine-Gordon equation*, Numerical Methods of Partial Differential Equations, 25: 685–711, 2009.

[22] M. Cui, *High order compact Alternating Direction Implicit method for the generalized sine-Gordon equation*, Journal of Computational and Applied Mathematics, 235: 837–849, 2010.

[23] A. K. Das, *A non-fickian diffusion equation*, Journal of Applied Physics, 70: 1355–1358, 1991.

[24] A. K. Das, *Some non-fickian diffusion equations: theory and applications*, Defect and Diffusion Forum, 162–163: 97–118, 1998.

[25] M. Dehghan, A. Ghesmati, *Solution of the second-order one-dimensional hyperbolic telegraph equation by using the dual reciprocity boundary integral equation (DRBIE) method*, Engineering Analysis with Boundary Elements, 34: 51–59, 2010.

[26] M. Dehghan, A. Ghesmati, *Combination of meshless local weak and strong (MLWS) forms to solve the two dimensional hyperbolic telegraph equation*, Engineering Analysis with Boundary Elements, 34: 324–336, 2010.

[27] M. Dehghan, M. Lakestani, *The use of Chebyshev cardinal functions for solution of the second-order one-dimensional telegraph equation*, Numerical Methods for Partial Differential Equations, 25: 931–938, 2009.

[28] M. Dehghan, A. Mohebbi, *High order implicit collocation method for the solution of two-dimensional linear hyperbolic equation*, Numerical Methods for Partial Differential Equations, 25: 232–243, 2009.

[29] M. Dehghan, A. Shokri, *A numerical method for solving the hyperbolic telegraph equation*, Numerical Methods for Partial Differential Equations, 24: 1080-1093, 2008.

[30] M. Dehghan, A. Shokri, *A meshless method for numerical solution of a linear hyperbolic equation with variable coefficients in two space dimensions*, Numerical Methods for Partial Differential Equations, 25: 494-506, 2009.

[31] D. Deng, C. Zhang, *A new fourth-order numerical algorithm for a class of nonlinear wave equations*, Applied Numerical Mathematics, 62: 1864–1879, 2012.

[32] D. Deng, C. Zhang, *A new fourth-order numerical algorithm for a class of three-dimensional nonlinear evolution equations*, Numerical Methods for Partial Differential Equations, 29: 102–130, 2013.

[33] D. Deng, C. Zhang, *A family of new fourth-order solvers for a nonlinear damped wave equation*, Computer Physics Communications, 184: 86–101, 2013.

[34] W. Deng, C. Li, *Finite difference methods and their physical constraints for the fractional Klein-Kramers equation*, Numerical Methods for Partial Differential Equations, 27: 1561–1583, 2011.

[35] H. Ding, Y. Zhang, *A new unconditionally stable compact difference scheme of $\mathcal{O}(\tau^2 + h^4)$ for the 1D linear hyperbolic equation*, Applied Mathematics and Computation, 207: 236-241, 2009.

[36] H. Ding, Y. Zhang, *A new fourth-order compact finite difference scheme for the two-dimensional second-order hyperbolic equation*, Journal of Computational and Applied Mathematics, 230: 626-632, 2009.

[37] M. Ehrhardt, *Absorbing boundary conditions for hyperbolic systems*, Numerical Mathematics: Theory, Methods and Applications, 3(3): 295–337, 2010.

[38] B. Engquist, A. Majda, *Absorbing boundary conditions for the numerical simulation of waves*, Mathematics of Computation, 31(139): 629–651, 1977.

[39] D. Evans, H. Bulut, *The numerical solution of the telegraph equation by the alternating group explicit (AGE) method*, International Journal of Computer Mathematics, 80(10): 1289–1297, 2003.

[40] G. Fibichi, B. Ilan, S. Tsynkov, *Backscattering and nonparaxiality arrest collapse of damped nonlinear waves*, SIAM Journal of Applied Mathematics, 63(5): 1718–1736, 2003.

[41] J. Fort, V. Méndez, *Wavefronts in time-delayed reaction-diffusion systems. Theory and comparison to experiment*, Reports on Progress in Physics, 65: 895–954, 2002.

[42] F. Gao, C. Chi, *Unconditionally stable difference schemes for a one-space dimensional linear hyperbolic equation*, Applied Mathematics and Computation, 187: 1272-1276, 2007.

[43] H. Goméz, I. Colominas, F. Navarrina, M. Casteleiro, *A discontinuous Galerkin method for a hyperbolic model for convection-diffusion problems in CFD*, 71: 1342–1364, 2007.

[44] S. M. Hassanizadeh, *On the transient non-Fickian dispersion theory*, Transport in Porous Media, 23: 107–124, 1996.

[45] P. Henrici, *Applied and Computational Complex Analysis*, vol.2, John Wiley, New York, 1977.

[46] G. Honig, U. Hirdes, *A method for the numerical inversion of Laplace transform*, Journal of Computational and Applied Mathematics, 10: 113–132, 1984.

[47] R. Huang et al, *Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid*, Nature Physics, 7: 576–580, 2011.

[48] J. Jackiewicz, R. A. Renaut, *A note on stability of pseudospectral methods for wave propagation*, Journal of Computational and Applied Mathematics, 143: 127–139, 2002.

[49] D. D. Joseph, L. Preziosi, *Heat waves*, Reviews of Modern Physics, 62: 41–73, 1989.

[50] S. Karaa, J. Zhang, *High order ADI method for solving unsteady convection-diffusion problems*, Journal of Computational Physics, 198: 1–9, 2004.

[51] S. R. Karur, P. A. Ramachandran, *Radial basis function approximation in the dual reciprocity method*, Mathematical and Computer Modelling, 20: 59–70, 1994.

[52] V. Kulish, K. Poletkin, *A generalized relation between the local values of temperature and the corresponding heat flux in a one-dimensional semi-infinite domain with the moving boundary*, International Journal of Heat and Mass Transfer, 55: 6595–6599, 2012.

[53] M. Lees, *Alternating direction and semi-explicit difference methods for parabolic partial differential equations*, Numerische Mathematik, 3: 398–412, 1961.

[54] T. Li, S. Kheifets, D. Medellin, M. G. Raizen, *Measurement of the instantaneous velocity of a Brownian particle*, Science 25, 328(5986): 1673–1675, 2010.

[55] H. Liao, Z. Sun, *Maximum norm error estimates of efficient difference schemes for second-order wave equations*, Journal of Computational and Applied Mathematics, 235: 2217–2233, 2011.

[56] H.-C. Lin, M.-I. Char, W.-J. Chang, *Soret effects on non-Fourier heat and non-Fickian mass diffusion transfer in a slab*, Numerical Heat and Mass Transfer, Part A, 55: 1096–1115, 2009.

[57] K.-C. Liu, H.-T. Chen, *Numerical analysis for the hyperbolic heat conduction problem under a pulsed surface disturbance*, Applied Mathematics and Computation, 159: 887–901, 2004.

[58] K.-C. Liu, H.-T. Chen, *Analysis for the dual-phase-lag bio-heat transfer during magnetic hyperthermia treatment*, International Journal of Heat and Mass Transfer, 52: 1185–1192, 2009.

[59] H.-W. Liu, L.-B. Liu, *An unconditionally stable spline difference scheme of $\mathcal{O}(k^2 + h^4)$ for solving the second-order 1D linear hyperbolic equation*, Mathematical and Computer Moddeling, 49: 1985-1993, 2009.

[60] J. Liu, K. Tang, *A new unconditionally stable ADI compact scheme for the two-space-dimensional linear hyperbolic equation*, International Journal of Computer Mathematics, 87(10): 2259-2267, 2010.

[61] J. E. Macíaz-Díaz, *Sufficient conditions for the preservation of the boundedness in a numerical method for a physical model with transport memory and nonlinear damping*, Computer Physics Communications, 182: 2471–2478, 2011.

[62] M. O. Magnasco, *Forced thermal ratchets*, Physical Review Letters, 71(10): 1477–1481, 1993.

[63] J. E. Marsden, M. J. Hoffman, *Basic Complex Analysis*, W. H. Freeman, 1999.

[64] R. E. Mickens, P. M. Jordan, *A positivity-preserving nonstandard finite difference scheme for the damped wave equation*, Numerical Methods for Partial Differential Equations, 20(5): 639–649, 2004.

[65] R. E. Mickens, P. M. Jordan, *A new positivity-preserving nonstandard finite difference scheme for the DWE*, Numerical Methods for Partial Differential Equations, 21: 976–985, 2005.

[66] R. K. Mohanty, *An unconditionally stable difference scheme for the one-space-dimensional linear hyperbolic equation*, Applied Mathematics Letters, 17: 101–105, 2004.

[67] R. K. Mohanty, *An unconditionally stable finite difference formula for a linear second order one space dimensional hyperbolic equation with variable coefficients*, Applied Mathematics and Computation, 165: 229–236, 2005.

[68] R. K. Mohanty, *New unconditionally stable difference schemes for the solution of multi-dimensional telegraphic equations*, International Journal of Computer Mathematics, 86(12): 2061–2071, 2009.

[69] R. K. Mohanty, M. K. Jain, *An unconditionally stable alternating direction implicit scheme for the two space dimensional linear hyperbolic equation*, Numerical Methods for Partial Differential Equations, 17: 684–688, 2001.

[70] R. K. Mohanty, *An operator splitting method for an unconditionally stable difference scheme for a linear hyperbolic equation with variable coefficients in two space dimensions*, Applied Mathematics and Computation, 152: 799–806, 2004.

[71] A. Mohebbi, M. Dehghan, *High order compact solution of the one-space-dimensional linear hyperbolic equation*, Numerical Methods for Partial Differential Equations, 24: 1222–1235, 2008.

[72] K. W. Morton, D. F. Mayers, *Numerical Solution of Partial Differential Equations*, Cambrigde University Press, 2005.

[73] C. Neves, A. Araújo, E. Sousa, *Numerical approximation of a transport equation with a time-dependent dispersion flux*, AIP Conference Proceedings 1048: 403–406, 2008.

[74] F. S. B. F. Oliveira, K. Anastasiou, *An efficient computational model for water wave propagation in coastal regions*, Applied Ocean Research, 20(5): 263–271, 1998.

[75] D. W. Peaceman, H. H. Rachford Jr., *The numerical solution of parabolic and elliptic differential equations*, Journal of the Society for Industrial and Applied Mathematics, 3: 28–41, 1955.

[76] B. Pekmen, M. Tezer-Sezgin, *Differential quadrature solution of hyperbolic telegraph equation*, Journal of Applied Mathematics, 2012, Article ID 924765, 18 pages, doi:10.1155/2012/924765, 2012.

[77] P. N. Pusey, *Brownian motion goes ballistic*, Science 13, 332(6031): 802–803, 2011.

[78] J. Qin, *The new alternating direction implicit difference methods for the wave equation*, Journal of Computational and Applied Mathematics, 230: 213–223, 2009.

[79] J. I. Ramos, *On the numerical treatment of an ordinary differential equation arising in one-dimensional non-Fickian diffusion problems*, Computer Physics Communications ,170: 231–238, 2005.

[80] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Second Edition, Springer-Verlag, 1989.

[81] V. S. Ryaben' kii, S. V. Tsynkov, *A Theoretical Introduction to Numerical Analysis*, Chapman and Hall/CRC, 2007.

[82] A. Saleh, M. Al-Nimr, *Variational formulation of hyperbolic heat conduction problems aplying Laplace transform technique*, International Communications in Heat and Mass Transfer, 35: 204–214, 2008.

[83] R. Shirmohammadi, A. Moosaie, *Non-Fourier heat conduction in a hollow sphere with periodic surface heat flux*, International Communications in Heat and Mass Transfer, 36: 827–833, 2009.

[84] A. Sluzalec, *Thermal waves propagation in porous material undergoing thermal loading*, International Journal of Heat and Mass Transfer, 46: 1607–1611, 2003.

[85] G. D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd ed., Clarendon Press, Oxford, 1993.

[86] G.A. Sod, *Numerical methods in fluid dynamics: initial and initial boundary value problems*, Cambridge University Press, 1985.

[87] O. D. L. Strack, *A mathematical model for dispersion with a moving front in groundwater*, Water Resources Research, 28(11): 2973–2980, 1992.

[88] J. Szeftel, *Absorbing boundary conditions for nonlinear scalar partial differential equations*, Computer Methods in Applied Mechanics and Engineering, 195: 3760–3775, 2006.

[89] Z. F. Tian, Y. B. Ge, *A fourth-order compact ADI method for solving two-dimensional unsteady convection-diffusion problems*, Journal of Computational and Applied Mathematics, 198: 268–286, 2007.

[90] J. W. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, Springer-Verlag, New York, 1995.

[91] R. S. Varga, *Matrix Iterative Analysis*, Springer, 2000.

[92] K. Wang, C. You, *A note on identifying generalized diagonally dominant matrices*, International Journal of Computer Mathematics, 84(12): 1863–1870, 2007.

[93] S.-S. Xie, S.-C. Yi, T. I. Kwon, *Fourth-order compact difference and alternating direction implicit schemes for telegraph equations*, Computer Physics Communications, 183: 552–569, 2012.

[94] M. Yang, *Two-time level ADI finite volume method for a class of second-order hyperbolic problems*, Applied Mathematics and Computations, 215: 3239–3248, 2010.

[95] E. Zauderer, *Partial differential equations of applied mathematics*, John Wiley & Sons, 1983.

[96] Z. Zhang, *An economical difference scheme for heat transport equation at the microscale*, Numerical Methods for Partial Differential Equations, 20: 855–863, 2004.

[97] Z. Zhang, D. Deng, *A new alternating-direction finite element method for hyperbolic equation*, Numerical Methods for Partial Differential Equations, 23: 1530–1559, 2007.

[98] J. Zhang, J. Zhao, *Unconditionally stable finite difference scheme and iterative solution of 2D microscale heat transport equation*, Journal of Computational Physics, 170: 261–275, 2001.