
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE MATEMÁTICA APLICADA



**NUMERICAL ANALYSIS OF SECOND ORDER
LAGRANGE-GALERKIN SCHEMES.
APPLICATION TO OPTION PRICING PROBLEMS**

MARÍA RODRÍGUEZ NOGUEIRAS

2005

Agradecementos

Desexo expresar o meu agradecemento aos profesores *Alfredo Bermúdez de Castro* e *Carlos Vázquez*, directores desta tese doutoral. As súa axuda e apoio foron fundamentais na realización desta memoria.

Alfredo, como científico entusiasmado polo seu traballo, é un referente para min e conseguiu, ao longo destes anos, transmitirme o gusto pola Matemática Aplicada. Os seus consellos sempre resultaron acertados e útiles.

Carlos sempre atopou un oco para as miñas dúbidas. As súas iniciativas, coñecementos e experiencia foron indispensabeis para rematar este traballo. Teño que destacar, ademais, a súa enerxía, coa que un millón de veces conseguiu “cargarme as pilas”.

Síntome afortunada de ter traballado con ambos os dous, por se ter complementado nesta dirección.

Agradézolle ao laboratorio CEREMADE de Paris IX a súa acollida. Quero destacar a paciencia, dispoñibilidade e axuda do profesor *Jérôme Busca*. Tamén quero agradecer aos membros do grupo DMTC HSBC-CCF as atencións prestadas durante a miña estada, en particular, ao profesor *H. Berestycki* polas interesantes discusións.

Os “chicos” de fluído-estrutura merecen, cando menos, unhas liñas: graciñas a *Pablo*, *Rodolfo* e *Luis*. Non podo imaxinar a ninguén mellor que *Pablo*, coa súa paciencia e xeito xeneroso de compartir coñecementos, para dar os primeiros pasos no departamento. Desde aquel primeiro verano (lembras?) non deixou de dar probas de amizade. *Rodolfo* só ten un “defecto”, o moito que traballa. Consigue, no traballo e fóra, sacar o mellor das persoas. Gracias tamén a *Luis* pola súa “chispa”.

Quero expresar a miña admiración por *Ana B.*, que nos foi abrindo camiño nesto das finanzas desde a *City*. Foi un placer traballar con ela. Gracias por confiar en min.

Son moitas horas de traballo e moitos favores e axudas, gracias a tod@s, e gracias a *Rubal*, *Fran*, *Ferrín*, *Elena*, *Rafa M.* ... Quero adicarlle unha lembranza especial ás persoas do departamento que sinto máis preto. A *Dolores*, porque a súa presenza medra en situacións difíciles, e agora xa se me fixo imprescindible sempre. A *Duarte* que compartiu comigo cousas marabillosas(∞), pero tamén soubo escoitarme, coidarme e darme agarimo nos meus desacougos. De *Pili* só podo dicir que me gustaría seguila tendo de compañeira de confidencias, despacho, risas, traballo ... En tódalas situacións e sempre me fas falla!.

Nas miñas estadas no CEREMADE contei co apoio de *Cassio Neri* e *Belén Lerena*, amigos desde aquela. Muito obrigada, “graSia” ...

Síntome orgullosa de pertencer ao grupo de bolseir@s da Facultade de Matemáticas. Nada mellor (é un dicir ;-)) que as pachangas de fútbol, baloncesto, e festas varias con vós para axudarme a desconectar! Graciñas a *Álex, J Ignacio, Bea P., Cris, Tere S., Jero, Marco, Andrés*, etc, etc, etc.

Entre os amigos aos que tamén lle tocou “aguantarme”, lembro especialmente a *Vallejo, Xoán Carlos* e *Elías C.* Gracias tamén a *Teresa Fonseca* polo súa amizade, coa que non pode a distancia.

Non sei por onde empezar a darlle as gracias á miña amiga-da-alma *Bea Hervella*, porque foron tantas cousas (todas!) nestes anos de convivencia . . . Que sorte tiveron de dar contigo dentro dese “mar de fueguitos”!. Obrigadissima.

Nas últimas liñas van ser para a miña familia sen a que non tería superado os momentos máis duros. Os meus irmáns, *Ra, Fon, Sana*, que sempre estiveron “aí”, entendendo as miñas loucuras matemáticas. Non teño verbas para expresar admiración que sinto pola *miña nai*, a máis forte entre nós . . . Gracias por coidarme, por esperar sempre o mellor de min e por facerme libre. Quérote moito *****!!!

Santiago de Compostela, primavera de 2005

Este traballo foi financiado polo *Ministerio de Educación, Cultura e Deporte* a través das súas bolsas *FPU*, e foi parcialmente realizado dentro dun proxecto financiado pola *Fundación Caixa Galicia*.

À memória do
meu pai.

Contents

Introduction	1
1 Mathematical models for pricing financial derivatives	7
1.1 Introduction	7
1.2 Financial derivatives: definitions and examples	7
1.3 Black-Scholes models for path dependent option pricing	9
1.3.1 No arbitrage opportunity	10
1.3.2 Stochastic model for the underlying asset	11
1.3.3 Black-Scholes and European path dependent options. Eurasian options	12
1.3.4 Black-Scholes and American path dependent options. Amerasian options	16
1.4 Fixed strike Asian options with continuous arithmetic averaging	17
1.4.1 Some option price properties from the no arbitrage assumption	18
1.4.2 Black-Scholes model for fixed-strike Eurasian options	22
1.4.3 Black-Scholes model for fixed-strike Amerasian options	24
1.4.4 One dimensional model before averaging	24
2 Functional framework for option pricing	27
2.1 Introduction	27
2.2 General problems	28
2.2.1 Unbounded domains	28
2.2.2 Truncated domains	29
2.3 Path-dependent options	32
2.3.1 Unbounded domain	32
2.3.2 Truncated domain	32
2.4 Fixed-strike Eurasian options	33
2.4.1 Existence of solution in the unbounded domain	33
2.4.2 Put-call parity relation and other properties	36

2.4.3	Truncated domain: boundary conditions and weak formulation	37
2.5	Fixed strike Amerasian options	40
2.5.1	Unbounded domain	40
2.5.2	Truncated domain: boundary conditions and weak formulation	41
3	Semi-Lagrangian time discretization	45
3.1	Introduction	45
3.2	Cauchy problem and notations	47
3.3	Characteristic curves	48
3.3.1	Approximate characteristic lines	54
3.4	Weak formulation	58
3.5	Time discretization: characteristics method	59
3.5.1	Approximation of the material derivative	60
3.5.2	Theta semi-Lagrangian scheme	61
3.5.3	Two-step semi-Lagrangian scheme	64
3.6	Analysis of the Crank-Nicholson semi-Lagrangian scheme	65
3.6.1	Stability of the semidiscretized scheme	66
3.6.2	Another stability result of the semidiscretized scheme	75
3.6.3	Error estimate of the semidiscretized scheme	79
3.7	Classical semi-Lagrangian method	91
4	Lagrange-Galerkin approximation	95
4.1	Introduction	95
4.2	Space discretization: finite element method	96
4.3	Analysis of the Crank-Nicholson Lagrange-Galerkin method	97
4.3.1	Stability for the fully discretized scheme	97
4.3.2	Error estimate of the fully discretized scheme	97
4.4	Classical Lagrange-Galerkin scheme	105
4.5	Finite element spaces and quadrature formulas	107
4.5.1	Study of the one dimensional problem	111
4.5.2	Analysis in m dimensions	115
4.6	Numerical results	116
5	Numerical solution of Asian options pricing problems	131
5.1	Introduction	131
5.2	Fixed-strike Eurasian options	133
5.2.1	Crank-Nicholson semi-Lagrangian time discretization	134

5.2.2	Galerkin projection, quadrature formulas and computational issues	140
5.2.3	Numerical results	141
5.3	Fixed-strike Amerasian options	143
5.3.1	The Bermúdez-Moreno iterative algorithm (BM)	144
5.3.2	Augmented Lagrangian Active set method (ALAS)	146
5.3.3	Numerical results	148
5.4	Qualitative results and the Greeks	149
A	Background in Stochastic Calculus	163
B	Black-Scholes pricing framework for vanilla options	171
B.1	European vanilla options	171
B.2	Black-Scholes and American options	176
B.3	Black-Scholes and m-factor models	178
C	Summary of the thesis in Galician	179
	Bibliography	194

Introduction

In this work we study the numerical solution of linear convection-diffusion-reaction partial differential equations by a second order Lagrange-Galerkin method. This study has been motivated by an application in financial mathematics: the pricing of derivatives by using partial differential equations (pdes). We focus on *two-factor models*, i.e., models to price financial products whose value depends on two random variables. As a representative example we deal with *path-dependent Asian options* of both *European* and *American* style.

Linear convection-diffusion-reaction equations arise as mathematical models in fluid mechanics, meteorology, semiconductor devices, finance, ... Very often, the convective term is distinctly greater than the diffusive one, giving rise to the convection-dominated problems. For instance, this is the case of Navier-Stokes equations with high a Reynolds number and convection-diffusion equations with a high Peclet number. This (dominant) convective term leads to computational difficulties in the numerical solution of such equations (see, for instance, [81, 89]).

In a Lagrangian coordinate system the effect of convection is not present. Thus, inspired in the solution of hyperbolic problems, numerical methods which use finite differences approximations of the material derivative along the characteristic trajectories for time discretization have been considered in the literature (see, for instance, [45]). In problems with significant convection, the solution changes much less rapidly in the characteristic direction than in the time direction, so a Lagrangian time discretization allows for the use of large time steps while maintaining stability and accuracy. Moreover, it leads to symmetric systems. In contrast to them, we have Eulerian methods which develop time discretization along the time direction. They cannot accurately simulate all the wave interactions that take place if the information propagates more than one cell per time step, because of stability reasons, in the case of explicit methods, or because of accuracy reasons, in the case of implicit methods (see the review paper [47]).

The Lagrangian time discretization we consider in this work traces backward the position at time t_n of the particles which will reach the points of a fixed spatial mesh at time t_{n+1} . This method, proposed in the early 1980s [45, 88], is known as the *(modified) method of characteristics*, or *semi-Lagrangian method*. It has been combined with space discretizations by finite differences [45], finite elements [88, 16, 21, 82, 103, 102, 91], spectral finite elements [104, 4], discontinuous finite elements [6, 5, 7], and so on.

In this work we are interested in the combination of the above semi-Lagrangian method with finite elements, also called the *characteristics finite element method* or the *Lagrange-Galerkin method*, which has been applied to a wide range of pde problems. For instance, this method applied to the Navier-Stokes equations was first analyzed in [88], in which unconditional stability and error estimates were obtained. These were improved in [102] under some stability conditions.

The numerical solution of the convection-diffusion equation with a time independent “velocity” field is addressed in [45, 44], and in [88, 103, 14] for a time dependent “velocity field”.

Unconditional stability, independent of the diffusion coefficient, has been obtained in these works. Moreover, if Δt denotes the time step, h the spatial step, and k the degree of the finite element space, error estimates of the form $O(h^k + \Delta t)$ in $l^\infty(L^2(\mathbb{R}^m))$ -norm are stated in [103] (m denotes the dimension of the spatial domain). In [88] estimates of the form $O(h^k + \Delta t + h^{k+1}/\Delta t)$ in $l^\infty(L^2(\Omega))$ -norm are shown under the assumption that the normal velocity component vanishes on the boundary of the spatial domain Ω . All of these estimates involve constants which depend on solution norms. More recently, for linear finite elements and for a velocity field vanishing on the boundary, a convergence of order $O(h^2 + \min(h, h^2/\Delta t) + \Delta t)$ in $l^\infty(L^2(\Omega))$ -norm is shown in [14], where the constants in the estimate only depend on the data.

Keeping in mind a second order accuracy in time by using semi-Lagrangian methods, we have found two different approaches in the literature. The first consists of approximating the material derivative with a second order three point scheme, and implicitly evaluating the rest of the equation. This method, that we call a *two-step Lagrange-Galerkin method*, has been proposed and analyzed for one-dimensional convection-diffusion equations in [46], and for the incompressible Navier-Stokes equations in [31]. The second approach consists of using a Crank-Nicholson Lagrangian discretization. More precisely, the material derivative is approximated by a two-point formula which is second order in an intermediate “characteristics-time point” and approximating the rest of the equation at this point. Although this idea was already suggested in [88, 91], the proposed upwind for the other terms failed to preserve second order accuracy [99]. The correct second order characteristics method for solving a constant coefficient convection-diffusion equation with Dirichlet boundary conditions is proposed and analyzed in [99]. For a divergence-free velocity field vanishing on the boundary and a smooth enough solution, stability and $O(\Delta t^2 + h^k)$ error estimates in $l^\infty(L^2(\Omega))$ -norm are stated.

In the present work we propose an extension of this latter characteristics finite element method, which we call the *Crank-Nicholson Lagrange-Galerkin method* or the *Crank-Nicholson characteristics method*: we allow for a (possibly degenerated) variable coefficient diffusive term, instead of the more classical Laplacian one; non divergence-free velocity fields and nonzero reaction functions. Furthermore, general mixed Dirichlet-Robin boundary conditions are considered. We remark also that the mathematical formalism of continuum mechanics (see for instance [58]) is used to express the results and notations related to the characteristic lines. In fact, the scheme is introduced and motivated after a correct weak formulation of the problem in terms of the characteristic lines. This formulation could help if we want to apply a similar scheme with other conditions or to other equations.

Stability results are given, and an error estimate of $O(\Delta t^2 + h^k)$ is obtained. A summary of these results can be found in [24, 25].

All of the properties of the characteristics finite element method (both the classical first order method and the second order ones) previously explained are established under the assumption that the inner products in the Galerkin formulation are calculated exactly. Since this is rarely possible in practice, numerical quadrature must be used instead. In some cases, this produces the loss of unconditional stability and adds some terms to the final error estimates (see [82, 103, 94, 52, 105, 25]).

There are some papers in the literature which study the effect of numerical integration in the classical first order Lagrange-Galerkin with piecewise linear finite elements. In particular, conditional instability is shown in [82] for a wide class of quadrature formulas when applied to the linear convection equation. This work was extended to the linear convection-diffusion equation in [103] and to a wider class of quadrature formulas in [94]. For both convection and convection-diffusion equations, Gauss-Lobatto quadrature formulas lead to the most stable

schemes. However, only the trapezoidal rule (or two-point Gauss-Lobatto rule) preserves unconditional stability. A Fourier based analysis of the (second order) two-step Lagrange-Galerkin method with linear finite elements is addressed in [52]. The author concludes that the two-step method seems to be more unstable than the classical first order one. With respect to the (second order) Crank-Nicholson Lagrange-Galerkin method, in [105] the authors experimentally show that it is more robust than the first order scheme with respect to numerical integration errors and produces better numerical results.

In this work we also study the effect of numerical quadrature for the (first order) classical method and the (second order) Crank-Nicholson Lagrange-Galerkin one using Lagrange finite elements of degrees $k = 1, 2$, on “triangular” and “quadrangular” meshes. Adequate quadrature formulas are proposed, and, a theoretical Fourier analysis is developed for some of them.

We have written a computer FORTRAN code for the above methods. Numerical results illustrate and complete our analysis.

Our motivation to study accurate and stable numerical solutions of convection-dominated problems is that equations of this sort model the price evolution of some of the financial contracts we are interested in. In fact, the numerical solution of *two-factor* pricing problems constitutes the second part of this work.

The trading of financial derivatives on organized markets goes back to the early 1970s, and to the mid-1980s for non-organized ones. Over the last three decades, this market has spectacularly grown up, and the financial contracts have become more and more complex. Likewise, the research on financial derivatives has also dramatically increased since the initial works by Black and Scholes [29] and Merton [77] in 1973. Within their framework for mathematical modelling of option pricing, (degenerate) second order parabolic final value problems arise (see, for instance, [111, 110, 71]). These problems eventually involve inequality constraints related to early exercise features. The dimension of the spatial domain depends on the particular financial product. A product with m spatial variables is termed an *m-factor product*.

In another approach to option pricing, the value of an option can be expressed as the expected present value of its payoff, i.e., a probabilistic representation of the solution of pricing problems exists. This is considered in [15]:

“Le lien fondamental (entre des équations aux dérivées partielles du second ordre et des problèmes unilatéraux d’une part, et du contrôle stochastique et des problèmes de temps d’arrêt optimal d’autre part) réside dans l’interprétation des solutions de certaines équations aux dérivées partielles. Cette interprétation est une extension de la méthode des caractéristiques, qui permet d’exprimer la solution d’une équation hyperbolique du premier ordre linéaire, de manière explicite comme une fonctionnelle définie sur les trajectoires caractéristiques. Pour les équations paraboliques ou elliptiques, un phénomène similaire se produit, mais les trajectoires caractéristiques sont alors des processus stochastiques.”

There are some particular examples of options, like the standard *vanilla* ones, for which a closed form solution is known. However, in many cases, numerical methods have to be employed in option valuation. We can distinguish methods focusing on estimating the conditional expectation of the final payoff from methods aiming at solving the Cauchy problem numerically.

An important example of the first group is the Monte Carlo method (see [98] for a textbook on the subject and [32] for an early application to finance). This basically consists of simulating

many realizations of the underlying price path, computing the corresponding option values and calculating the average payoff over the realizations. It is easy to implement and to understand. Its main drawbacks are that it is slow compared to standard methods for pdes having up to three factors and that it is not easily applicable when there are inequality constraints.

With respect to standard numerical methods for solving parabolic equations, those more oftenly extended in finance involve Eulerian discretizations of time derivatives combined with finite differences for spatial discretization and projected relaxation methods to deal with unilateral constraints (see, for instance, [111]). Most of them are applied to one-factor problems. However, in the last two decades more sophisticated methods from fluid mechanics have been used in computational finance. Take, for instance, finite volume methods [116] and finite element methods [90, 76] for spatial discretization; characteristics methods for time discretization [107, 10, 23, 43]. For handling the nonlinearities of obstacle type we can find, for example, relaxation methods combined with multigrid methods in [37], Uzawa's algorithm in [107], and implicit penalty with variable time step in [51].

The methodology we propose lies in the second group of methods, i.e., we study the numerical solution of partial differential equations of two-factor pricing problems. We use classical and higher order Lagrange-Galerkin methods for time-space discretization (see [99, 24, 25]) and two different iterative algorithms based on the mixed formulation, introduced in [22] and [69] respectively, to discretize the unilateral obstacle problem. The developed regularization does not introduce any further source of error as penalty methods do.

The combination of classical Lagrange-Galerkin methods with the iterative algorithm proposed in [22] was successfully used to price two-factor products, such as convertible bonds and Asian options [23].

As an example of application, we consider the Asian options pricing problem. These options are *path-dependent* financial derivatives whose *payoffs* (i.e., their values at the end of the contract) somehow depend on an averaging price of another financial product called the *underlying asset* during a period of time.

Following Black-Scholes and Merton's methodology, the value of an Asian option of *European* type solves a two-dimensional linear parabolic convection-diffusion-reaction equation strongly degenerated, as there is no diffusion in one of the spatial directions. The first formulation of this mathematical model is addressed in [60]. Furthermore, when the *American feature* is considered, the pricing problem becomes nonlinear. It is formulated as a linear complementarity problem in [42]. In both cases, the parabolic differential operator is strongly convection-dominated, so the use of Lagrange-Galerkin methods is justified. The smoothness of solution [13] allows us to apply higher order schemes successfully.

In [26, 27] higher order Lagrange-Galerkin methods for Eurasian options are used. They have been extended to the general constrained case in [28].

The outline of the present thesis is as follows:

In *Chapter 1* we first introduce the main financial concepts and notations required for the mathematical modelling of option pricing in the Black-Scholes and Merton framework. Then, the mathematical model for Asian options is given and some properties are discussed mainly by using financial arguments.

In *Chapter 2* we formulate general pricing problems in a suitable functional framework. Keeping in mind their numerical solution, the truncation of the (unbounded) spatial domain is explained. Then, we consider the particular Eurasian options pricing problem for which existence

and regularity of solution results are given within the framework proposed in [13]. For both Eurasian and Amerasian options, the problem in bounded domain is posed.

In *Chapter 3* we propose a second order characteristics method for time discretization of linear convection-diffusion-reaction equations in a bounded domain. Throughout the entire chapter, the mathematical formalism of continuum mechanics is used. Firstly, we give some results related to the characteristic lines associated with a velocity field. Then, after formulating an appropriate variational formulation of the problem in terms of the characteristic lines, we introduce the second order characteristics scheme. Next, stability and error estimates are established.

In *Chapter 4* a fully discretized scheme, which results from combining the second order semidiscretization from the previous chapter and finite elements, is analyzed. After that, a study of the influence of the numerical quadrature is developed for quadrangular and triangular Lagrange finite elements. Finally, numerical results for some academic tests are included.

In *Chapter 5* we first discuss the numerical discretization of Eurasian options pricing problems by using a slightly different version of Lagrange-Galerkin methods introduced in Chapters 3 and 4. Moreover, we compare two algorithms based on the augmented Lagrange formulation for the unilateral obstacle problem appearing in the Amerasian pricing problems. Numerical results for real life options, compared with other results appearing in the literature, show the performance of the proposed methods.

In *Appendix A* we include some definitions and results from Stochastic Calculus, and, in *Appendix B* the mathematical modelling of a more basic option pricing model, the vanilla options pricing problem, is discussed. Both of the appendices could aid the reader's understanding of Chapter 1.

Chapter 1

Mathematical models for pricing financial derivatives

1.1 Introduction

In the present chapter we first introduce some definitions and notations from financial pricing theory, focusing our study in the assumptions and techniques proposed by Black-Scholes [29] and Merton [77]. We point out that professors M.S. Scholes and R.C. Merton were awarded the Nobel Prize in Economic Sciences in 1997 for “a new method to determine the value of derivatives”.

Within this framework, the mathematical model for pricing path-dependent options, and, as a particular case, for fixed-strike Asian options, is deduced. Finally, this latter model is deeply studied, and some properties satisfied by the option values are given. They concern, for instance, the put-call parity relation, some boundary conditions, the regions where it is optimal to hold the Amerasian options, etc.

The present chapter can be completed with Appendices A and B. Firstly, since the modelling procedure requires stochastic calculus tools, we have included in Appendix A mostly of the definitions and results needed. Secondly, similar modelling techniques to those developed in this chapter are recalled in Appendix B for pricing the more basic model of standard vanilla options.

1.2 Financial derivatives: definitions and examples

A *derivative product*, *financial derivative* or *contingent claim* is a financial contract whose value depends on one or several risk factors, such as the price of a bond, commodity, currency, share, etc; a yield or rate of interest; an index of prices or yields; weather data, such as inches of rain or heating-degree-days; insurance data, such as claims paid for a disastrous earthquake or flood, etc.

The risk factors are the primary or more basic products for which the financial derivative is the secondary product. In this memory we will only consider primary products that are prices of financial products, called *underlying assets*. In this case, the risk factor is the market risk or the risk of loss from being on the wrong side of a bet about a market movement.

It is clear that risk comes from uncertainty about future values of the underlying assets, and

thus the derivative products are *forward contracts*, in contraposition to *spot contracts*.

Although derivative products have been bought and sold for generations, it was not until 1973, with the opening of the Chicago Board of Options Exchange (CBOE), that a formal exchange market was organized for that specific purpose. Since then, many other official markets of financial derivatives have appeared. However, many contracts continue to be written in non standardized markets, called *over the counter* (OTC) markets.

Examples of derivative securities are options, futures, convertible bonds, swaps, etc. A *forward contract* (or *future contract* if traded in an exchange) is an agreement between two parts, one of them buys an asset from the counterpart on a certain date in the future for a pre-determined price.

Let us explain what an option is by considering a particular example. An *European vanilla option* is a contract which gives to its *holder* the right, but not the obligation (this is the difference between option and future contract), to buy or sell an underlying asset at a predetermined date for a fixed price. An option to buy is a *call option*, and an option to sell is a *put option*. The amount for which the underlying can be bought or sold is the *strike (price)*, and the date on which the option ceases to exist (or gives the holder any right) is the *expiration date* or *maturity*.

The option is said to be *exercised* when the holder chooses to buy or to sell the asset. He will make the decision depending on whether the deal is favorable to him or not. The counterpart to the holder of the option contract is called the *writer* of the option. Such a contract is not symmetrical: the holder has a right and the writer has an obligation, they are said to be in a *long* and *short* position of the option contract, respectively.

A concept related to the above is the *short selling*, which is the trading practice of borrowing a stock and selling it immediately, buying the stock later and returning it to the borrower. Usually, there are rules in the stock exchanges that restrict the timing and the use of the short selling.

The options have a value called *option value* or *premium*, that has to be paid by the holder to the writer when they both enter into the option contract. The problem of mathematical finance is to find the *fair value* of an option (pricing problem).

Notice that the option contract specifies the value of the option as a function of the underlying asset value at maturity. Considering, for instance, a vanilla call option from a long holder position viewpoint, if $S(T)$ denotes the underlying asset price at maturity T and K the strike price, the option value at maturity is

$$(S(T) - K)_+, \quad (1.1)$$

and for a vanilla put option

$$(K - S(T))_+. \quad (1.2)$$

This function of the underlying asset is called the *payoff function* or *exercise value function*. Notice that the payoff function is always positive.

Options are characterized by the payoff function and the kind of exercise allowed. More precisely, attending to the exercise date we can distinguish three types of options:

- *European* options: they can be exercised only at maturity.
- *American* options: they can be exercised at any time prior to expiry.
- *Bermudan* options: they can be exercised at some predetermined time intervals.

In the case of American (respectively, Bermudan) options, the contract specifies the value the holder get in case of exercise at any time before maturity (respectively, at some predertermines time intervals before maturity). In the case of American vanilla call options (respectively, put options), for instance, the exercise value function at time t is given by

$$(S(t) - K)_+, \quad (\text{respectively, by } (K - S(t))_+),$$

where $S(t)$ denotes the value of the underlying asset at time t . We will mainly use the payoff function referred to exercise at maturity, and exercise value function referred to exercise prior to maturity.

For American (and Bermudan) options another question is formulated in option pricing: what should be the optimal strategy to exercise prior to the expiration date?

Options that have a payoff function different from (1.1) or (1.2) are called *exotic* options. Conceptually, there is no limit in the payoff functions we can imagine, so there is no limit for the types of exotic options. In practice, the “market-makers” are continuously defining new financial products, so the list of exotic options is highly increasing. Some common options are the *path-dependent* options, whose payoffs depend on the history of the asset price, not just on its value at exercise date. Examples of path dependent options are *barrier options*, *lookback options*, *Asian options*, etc. Barrier options are characterized by a value, the barrier, in such a way that they cease to exist (*knock out options*) or start to exists (*knock in options*) if the asset value crosses the barrier. The payoffs of lookback options depend on the maximum/minimum value reached by the underlying asset during a certain period. Asian options are characterized by payoffs depending on average values of the underlying asset over some prescribed period.

The advantage of the above path-dependent options with respect to the vanilla ones is clear: very often it is more difficult to manipulate the maximum, minimum, or average price of an asset over some period of time than to do so on a particular target date. Asian options are interesting, for example, for a company that works with products not very much traded in the markets, as the commodities; traders must buy the product every year at a certain moment, and must sell it during the year regularly. In this case, the underlying asset is the commodity. Asian options are also used in markets of currency exchange by companies that have continuous sales in a currency but must buy the raw materials in different currencies at a fixed date. In this case, the underlying is the rate of currency exchange. In general, Asian options allow investors to protect themselves against losses due to adverse movements in the prices of the assets, without needing to hedge continuously their portfolios. Due to it, its volume of negotiation quickly grew up in OTC markets. There are different types of Asian options depending on the payoff and on the averaging form (see [114], for example).

1.3 Black-Scholes models for option pricing.

Path-dependent options

Eventhough the research in option pricing goes back as far as Bachelier’s thesis in 1900 [3], academic and applied research on financial derivatives has exploded over the last three decades, i.e., after the Black-Scholes [29] and Merton [77] seminal papers in 1973. Simultaneously, the trading of financial derivatives, either in organized exchanges or in OTC markets, has experienced a strong growth.

In this section we firstly discuss some modelling hypotheses and explain how to use them in order to obtain information about options price. Then, we explain how to formulate a deterministic partial differential equation modelling European and American path-dependent options prices, within Black-Scholes and Merton framework.

For the sake of completeness, we have included in Appendix A the basic definitions and results from the stochastic calculus that are used in the modelling process.

Let us first introduce some notation:

- V , option value or premium.
- S , underlying asset value.
- K , strike price.
- T , maturity date.

If t denotes the current time, we have already mentioned that V is equal to the payoff function when $t = T$. Our aim is to find V for $t < T$ and, in the case of an American option, to know whether the holder should exercise it or not.

1.3.1 No arbitrage opportunity

The basic hypothesis underlying the theory of financial derivative pricing is the *no arbitrage hypothesis*, which is a kind of financial equivalent to conservation laws in continuum mechanics. It claims that there are never opportunities to make a profit without risk.

In a proper financial market such kind of arbitrage opportunity can not occur because alert traders (*arbitrageurs*) should notice the opportunity and trade in the right direction until it is removed.

Firstly, we have to assume the existence of *risk-free* investments that give a guaranteed return with no chance of default and, moreover, that the *interest rate* associated to this risk-free investment is a known function of time, $r = r(t)$.

Let B be the price of a riskless investment. Its evolution in time is given by the ordinary differential equation (ODE)

$$\frac{dB}{dt}(t) = r(t)B(t), \quad (1.3)$$

where we consider a continuously compounded interest rate. Thus, if the initial investment date is t_0 , then the value of B at any time $t > t_0$ is

$$B(t) = B(t_0)e^{\int_{t_0}^t r(\tau)d\tau}, \quad (1.4)$$

which is, for known r , a deterministic quantity.

Thus the greatest return that one can make on a risk-free portfolio of assets is the same as the return of the equivalent amount of cash placed in a risk-free bank.

It is possible to establish some bounds for options value and some relations between options price and underlying assets price without having specified any hypotheses about the underlying asset movement. The only assumption is that there are no arbitrage opportunities and that investors prefer wealth to less.

As an example, let us give some properties satisfied by standard vanilla options, which are directly deduced by the no arbitrage argument (see [71]). The European call option value (respectively, European put option value) at time $t > t_0$ is denoted by $V_c^{Eu}(t)$ (respectively, by $V_p^{Eu}(t)$); and the American call option value (respectively, American put option value) at time t is denoted by $V_c^{Am}(t)$ (respectively, by $V_p^{Am}(t)$).

- Vanilla options prices are non-negative,

$$V_c^{Eu}(t) \geq 0, \quad V_p^{Eu}(t) \geq 0, \quad V_c^{Am}(t) \geq 0, \quad V_p^{Am}(t) \geq 0.$$

- American vanilla options values are greater (or equal) than to their *exercise values*,

$$V_c^{Am}(t) \geq (S - K)_+, \quad V_p^{Am}(t) \geq (K - S)_+.$$

- American vanilla options are worth, at least, their European counterparts:

$$V_c^{Am}(t) \geq V_c^{Eu}(t), \quad V_p^{Am}(t) \geq V_p^{Eu}(t).$$

In fact, the previous properties are satisfied by general options, not only by the vanilla ones. The following lower/upper bounds and *put-call parity* relation hold only for European vanilla options:

- Lower and upper bounds,

$$\left(S(t) - Ke^{\int_{t_0}^t r(\tau) d\tau} \right)_+ \leq V_c^{Eu}(t) \leq S(t), \quad (1.5)$$

$$\left(Ke^{\int_{t_0}^t r(\tau) d\tau} - S(t) \right)_+ \leq V_p^{Eu}(t) \leq Ke^{\int_{t_0}^t r(\tau) d\tau}. \quad (1.6)$$

- Put-call parity relation,

$$V_c^{Eu}(t) - V_p^{Eu}(t) = S(t) - Ke^{\int_{t_0}^t r(\tau) d\tau}. \quad (1.7)$$

1.3.2 Stochastic model for the underlying asset

The original idea of efficient market arises with Louis Bachelier in 1900 [3] to describe the behavior of stock markets. However, the concept was partly developed in the 1960s. by Eugene Fama. The Efficient Market Hypothesis (EMH) postulates that, at any given time, the price of any particular stock reflects all available information. In other words, in an active market where there are many well-informed and intelligent investors, stocks would be priced such that they would reflect all available information known about them.

One implication of this hypothesis is that the movement of stock prices is essentially random in nature and cannot be predicted in advance, i.e., the movement follows a random walk. This motivates asset prices to be modelled by stochastic processes (see Definition A.1). Furthermore, in order to be consistent with the EMH, these stochastic processes are demanded to satisfy the Markov property (see Definition A.12). If the asset prices follow a Markovian process, then only the present asset prices are relevant for predicting their future values. This is an important point in option pricing, because the value of an option on a Markovian underlying at time t only depends on the asset price at time t .

In general, stochastic processes for modelling underlying assets price are governed by stochastic differential equations on Brownian motions as (A.8) and thus they satisfy Markov property (Theorem A.6).

Bachelier [3] was the first proposing a stochastic model, a posteriori called Brownian motion (see Definition A.6), for the asset prices. Bachelier's idea was not noticed until, around the middle of the last century, a series of empirical studies refined his initial insights in various ways.

- The original Bachelier's model allowed for negative prices and thus option prices could be higher than asset prices. It was observed in practice that logarithms of assets prices fit better the Brownian motion model than prices themselves.
- The increments of the Brownian motion have zero expectation function, which suggests a null expectation of growth. This fact does not agree with the observed feature that stock prices tend to increase in value over the years.

A stochastic model that avoids the above drawbacks is the geometric Brownian motion,

$$S(t) = S_0 + \int_0^t \mu S(\tau) d\tau + \int_0^t \sigma S(\tau) dW(\tau), \quad (1.8)$$

where $(W(t), 0 \leq t \leq T)$ is a standard Brownian motion (see Definitions A.6 and A.7), and μ and σ are constant parameters. The *drift rate*, also called *expected return* or *growth rate*, μ , contains the deterministic part of the random walk, and it is defined as the annualized standard expectation of the relative change of S . The *volatility* parameter, σ , is the annualized standard deviation of the relative change and contains the non-predictable part of the random walk.

Using the unidimensional Ito's formula (A.4) it is easy to show that the solution of (1.8) is

$$S(t) = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)},$$

and using Theorem (A.5) the above solution is unique.

The geometric Brownian motion is one of the most widely extended models in financial theory when modelling share prices, since it was proposed by Black and Scholes [29] and Merton [77], and it is the model that we will adopt all over this document. Easy generalizations have been also considered, with time-dependent drift and volatility parameters, or time/share-dependent parameters (see, for instance, [19]).

However, in the last years, more complicated models have been also studied as stochastic volatility models (see Chapter 23 of [110]), jump diffusion models (see Chapter 26 of [110], or [39]), etc. Moreover, for underlyings different from shares, such as interest rates, other models have been proposed (see [110]).

1.3.3 Black-Scholes and European path dependent options. Eurasian options

In the present section we apply Black Scholes [29] and Merton [77] methodology to price European path-dependent contingent claims, i.e., options whose payoffs depend upon the past history of the stock price. Moreover, exercise is allowed only at expiry. We assume, then, that the underlying asset is driven by a geometric Brownian motion. More complicated models have been considered

in the literature, as the Levy processes in [109] or the jump diffusion models in [43], but they are out of the scope of this document.

The new idea here is the elimination of risk (*delta-hedging*) by using a *self-financed* strategy. In order to clarify these techniques we have included in Appendix B similar deductions for vanilla options, together with a more probabilistic approach. For further study see, for instance, [71] or [111].

Path-dependent options are hedgeable as vanilla call and puts (see Appendix B), but whereas the vanilla options prices depend only upon the underlying asset and time, path-dependent option price functions depend also on a new factor which contains or measures the path-dependent quantity. The idea of introducing a new variable was first developed in [60], where the model for Asian options was deduced.

Let us assume the following hypotheses on the market:

Hypothesis 1.1 *There exists a theoretical financial market with two traded assets:*

- A risk-free asset or cash, B , whose evolution is given by

$$dB(t) = rB(t)dt, \quad B(t_0) = B_0.$$

Parameter r is the short-term interest rate, which is supposed to be a given constant.

- A risky asset following a geometric Brownian motion with constant parameters μ and σ , i.e.,

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad t \geq t_0, \quad S(t_0) = S_0. \quad (1.9)$$

The financial market is liquid (there are always buyers and sellers), no arbitrage (there can be no opportunity of riskless placement earning more than the risk-free interest rate of the economy, r) and perfect (there are no restriction of any kind on the sales, neither transactions costs). Then this market is complete and options can be replicated by combining cash and risky stocks (see [72]).

Moreover, a continuous dividend yield of rate d_0S is allowed.

Let us now assume that the option contract starts at time T_i and expires at time T_f and the payoff is given by

$$\Lambda(S(T_f), I(T_f)). \quad (1.10)$$

where $I(T_f)$ is a new state variable at time T_f , to be called the *path dependent variable*. We are considering the general form

$$I(t) = g(t) \int_{T_i}^t f(S(\tau), \tau) d\tau, \quad (1.11)$$

for f and g given functions. Let us assume that V , the function giving the option price, satisfies:

- V is a deterministic function of S, I and t .
- $V \in C^{2,2,1}(\mathbb{R}_+ \times \mathbb{R}_+ \times [T_i, T_f])$.

By (formally) differentiating expression (1.11) with respect to time, the random walk for I is obtained, namely,

$$dI(t) = \left(g(t)f(S(t), t) + \frac{g'(t)I(t)}{g(t)} \right) dt. \quad (1.12)$$

Since the evolution of I , given by (1.12), is (locally) deterministic, a riskless hedge for this option will only require eliminating the stock-induced risk. For proving this, let us consider a self-financed portfolio (see Appendix B.1 for a definition), Π , consisting of one option and an amount of the underlying asset equal to $-\Delta$. In this case,

$$\Pi(t) = V(S(t), I(t), t) - \Delta S(t),$$

and the variation of the portfolio between t and $t + \Delta t$ (with Δt a small variation of time) is

$$d\Pi(t) = dV(S(t), I(t), t) - \Delta dS(t).$$

By applying two-dimensional Ito's formula (A.6) to dV and taking into account (1.12) and (1.9), we obtain

$$\begin{aligned} dV(S(t), I(t), t) &= \sigma S(t) \frac{\partial V}{\partial S} dW(t) \\ &+ \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S(t)^2 \frac{\partial^2 V}{\partial S^2} + \mu S(t) \frac{\partial V}{\partial S} + \left(g(t)f(S(t), t) + \frac{g'(t)}{g(t)} I(t) \right) \frac{\partial V}{\partial I} \right) dt, \end{aligned}$$

and thus

$$\begin{aligned} d\Pi(S(t), I(t), t) &= \left(\sigma S(t) \frac{\partial V}{\partial S} - \sigma S(t) \Delta \right) dW(t) \\ &+ \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S(t)^2 \frac{\partial^2 V}{\partial S^2} + \mu S(t) \frac{\partial V}{\partial S} + \left(g(t)f(S(t), t) + \frac{g'(t)}{g(t)} I(t) \right) \frac{\partial V}{\partial I} - \mu S(t) \Delta \right) dt. \end{aligned}$$

In order to cancel the random part, we choose $\Delta = \frac{\partial V}{\partial S}$, so

$$d\Pi(S(t), I(t), t) = \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S(t)^2 \frac{\partial^2 V}{\partial S^2} + \left(g(t)f(S(t), t) + \frac{g'(t)}{g(t)} I(t) \right) \frac{\partial V}{\partial I} \right) dt. \quad (1.13)$$

On the other hand, the return of the riskless portfolio in the time interval $(t, t + \Delta t)$ results to be

$$r\Pi(S(t), I(t), t)dt + d_0\Delta S(t)dt. \quad (1.14)$$

In view of the arbitrage argument, we identify expressions (1.13) and (1.14). We get the deterministic partial differential equation governing the European path-dependent option price:

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - d_0)S \frac{\partial V}{\partial S} + \left(g(t)f(S, t) + \frac{g'(t)}{g(t)} I \right) \frac{\partial V}{\partial I} - rV = 0.$$

Moreover, the option value at expiry is given by (1.10), namely,

$$V(S, I, T_f) = \Lambda(S, I). \quad (1.15)$$

Then, the fair price of path-dependent options whose payoffs can be written in the form (1.10), solves the following final value problem:

$$\begin{cases} \mathcal{L}_{pd}[V] &= 0, & \text{in } \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \\ V(S, I, T_f) &= \Lambda(S, I), & \text{in } \mathbb{R}_+ \times \mathbb{R}_+, \end{cases}$$

where

$$\mathcal{L}_{pd}[\phi] := \frac{\partial \phi}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 \phi}{\partial S^2} + (r - d_0) S \frac{\partial \phi}{\partial S} + \left(g(t) f(S, t) + \frac{g'(t)}{g(t)} I \right) \frac{\partial \phi}{\partial I} - r \phi. \quad (1.16)$$

As we have mentioned, popular examples of path-dependent options are barrier options, lookback options and Asian options.

Barrier options have payoffs like those of vanilla options, but they cease to exist (or begin to exist) if the underlying asset crosses a value which is predefined by the contract. A particular study of this case (see, for instance, [111]) shows that path-dependency affects the domain of definition of the equation, but no new variables are added in the analysis, and the Black-Scholes operator appears (see appendix B.1 for the definition of the Black-Scholes operator).

For lookback options the path-dependent quantity is the maximum (or minimum) value achieved by the underlying asset. In order to accommodate the lookback case into our general setting more work is needed. Let us first define, for n integer,

$$I_n = \int_{T_i}^{T_f} (S(\tau))^n d\tau,$$

and

$$J_n = (I_n)^{\frac{1}{n}}.$$

Then our additional variable is the limit,

$$J = \lim_{n \rightarrow \infty} J_n = \max_{T_i \leq \tau \leq T_f} S(\tau).$$

For more details on the deduction of the equation governing lookback option prices see [111].

Since averages are usually expressed as integrals, it is easy to place Asian options with European feature (Eurasian options) in the setting described in the present section, or, in other words, to describe the path-dependent quantity by an integral as in (1.11). An exhaustive classification of Asian options is out of the scope of this document (see [114] for a textbook on the subject), but we can recall here some of the more common Eurasian options, characterized by their path-dependent quantity and the form of their payoffs. For instance, attending to the sampling average technique, we have continuous or discrete averaging, and attending to the averaging procedure, we have arithmetic or geometric averaging:

- Continuous arithmetic averaging

$$I(t) = \frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau, \quad g(t) = \frac{1}{t - T_i}, \quad f(S, t) = S.$$

- Continuous geometric averaging

$$I(t) = \frac{1}{t - T_i} \int_{T_i}^t \log(S(\tau)) d\tau, \quad g(t) = \frac{1}{t - T_i}, \quad f(S, t) = \log(S).$$

- Discrete arithmetic averaging

$$I_{n(t)} = \frac{1}{n(t)} \sum_1^{n(t)} S(t_i), \quad g(t) = \frac{1}{n(t)}, \quad f(S, t) = \sum_1^{n(t)} \delta(t - t_i) S.$$

- Discrete geometric averaging

$$I_{n(t)} = \frac{1}{n(t)} \left(\prod_1^{n(t)} S(t_i) \right)^{\frac{1}{n(t)}}, \quad g(t) = \frac{1}{n(t)}, \quad f(S, t) = \sum_1^{n(t)} \delta(t - t_i) \log(S).$$

In discrete sampling, sequence $\{t_j\}_{j=1}^{n(t)}$ is a partition of the interval $[T_i, t_{n(t)}]$ and $n(t)$ is the largest integer such that $t_{n(t)} < t$ (for more details see [111]).

Remark 1.1 *In the above list the choice of functions g and f is not unique.*

Another classification is based on the form of the payoff.

- Fixed strike or average rate call (respectively, put)

$$(I(T_f) - K)_+ \quad \left(\text{respectively, } (K - I(T_f))_+ \right).$$

- Floating-strike or average strike call (respectively, put)

$$(S - I(T_f))_+ \quad \left(\text{respectively, } (I(T_f) - S)_+ \right).$$

1.3.4 Black-Scholes and American path dependent options. Amerasian options

The classification given in the previous section also holds if we add the early exercise feature, given rise to different types of *Amerasian* options, analogous to *Eurasian* ones.

The model for American path-dependent options is deduced from combining the ideas of American vanilla option pricing of Section B.2 and European path-dependent option pricing of Section 1.3.3. It can be seen in [42], for instance.

We will use the notation and definitions introduced in the previous section, in particular, (1.11), (1.10) and (1.16). In the American case we know in advance the exercise value function $\underline{\Lambda} = \underline{\Lambda}(S, I, t)$ for $t \in [T_i, T_f]$, because it is specified in the particular option contract.

Arbitrage considerations (see Section 1.3.1) give the first inequality constraint satisfied by the option price $V = V(S, I, t)$:

$$V(S, I, t) \geq \underline{\Lambda}(S, I, t). \quad (1.17)$$

Then we construct the same delta-hedging portfolio leading to the partial differential operator defined in (1.16), which measures the difference between the rate of return of a risk-free Δ -hedged portfolio and a bank deposit of the same value. However, in the American case perhaps it is not always possible to take both long and short positions on the option during $(t, t + \Delta t)$, because maybe it is optimal to exercise the option by the holder. Thus, non arbitrage arguments only

claim that the rate of return from a Δ -hedged portfolio cannot exceed the rate of return from a bank deposit, leading to the inequality:

$$\mathcal{L}_{pd}[V] \leq 0. \quad (1.18)$$

In fact, two possibilities can be distinguished:

- If $V > \underline{\Delta}$, then it is optimal to hold the option and the equality $\mathcal{L}_{pd}[V] = 0$ holds.
- Otherwise, if $V = \underline{\Delta}$, it is optimal to exercise the option.

Thus, the pricing of American path dependent options can be written as a parabolic linear complementarity problem in the form

$$\left. \begin{array}{rcl} \mathcal{L}_{pd}[V] (V - \underline{\Delta}) & = & 0 \\ \mathcal{L}_{pd}[V] & \leq & 0 \\ V - \underline{\Delta} & \geq & 0 \end{array} \right\} \quad \text{in } \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \quad (1.19)$$

with V and $\partial V / \partial S$ continuous, if $\underline{\Delta}$ is continuous (see [42]), and with final condition

$$V(S, I, T_f) = \underline{\Delta}(S, I) \quad \text{in } \mathbb{R}_+ \times \mathbb{R}_+. \quad (1.20)$$

At each time $t \in (T_i, T_f)$ the spatial domain can be divided into two sets; the open set

$$\mathcal{D}(t) := \{(S, M), V(S, M, t) > \underline{\Delta}(S, M, t)\} \quad (1.21)$$

and its complementary

$$\mathcal{C}(t) := \mathcal{D}(t)^C = \{(S, M), V(S, M, t) = \underline{\Delta}(S, M, t)\}, \quad (1.22)$$

which is known as coincidence set (or exercise region). Finally, the set of points $\{\Sigma(t) = (S_f(t), I_f(t), t \in (T_i, T_f))\}$, separating both regions

$$\Sigma(t) = \overline{\mathcal{C}(t)} \cap \mathcal{D}(t) \quad (1.23)$$

is known as the free boundary (or the optimal exercise boundary).

1.4 Fixed strike Asian options with continuous arithmetic averaging

As we have said, depending on the choice of the averaging and payoff, different Asian options can be defined. Discretely sampled models can be reduced to one-dimensional Black-Scholes equations with jump conditions. With respect to the form of averaging, closed-form solutions for geometric averaging Asian options are known. Moreover, since traders almost exclusively use arithmetic averages to construct Asian options, the geometric averaging options are used to approximate the values of their arithmetic counterparts (see [114]). On the other hand, two-factors models for pricing arithmetically averaged Asian options can be reduced to one-factor ones if the payoff has the form

$$S^\alpha F(I/S, t),$$

for some constant α and function F (see, for instance, [71]). This is the case of continuously sampled floating strike options. In this document we focus our study on fixed strike Asian options with continuous arithmetic averaging. Some changes of variable have been also proposed for this kind of options which reduce the model to one factor (see [97, 108]), but they only apply to the European case.

The fixed-strike Amerasian call option with continuous arithmetic averaging gives the holder the right to exercise at any time prior to maturity date T_f (American feature) and get the value

$$\left(\frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau - K \right)_+ . \quad (1.24)$$

In expression (1.24), K is the fixed strike price. S is the asset price, and (T_i, T_f) is the averaging time interval. The corresponding put option is characterized by the exercise value

$$\left(K - \frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau \right)_+ . \quad (1.25)$$

The European counterparts (Eurasian) only allow to exercise at maturity.

Remark 1.2 *In this memory the fixed-strike Amerasian options with continuous arithmetic averaging (respectively, fixed-strike Eurasian options with continuous arithmetic averaging) will be also named as fixed-strike Amerasian options or Amerasian options (respectively, fixed-strike Eurasian options or Eurasian options).*

1.4.1 Some option price properties from the no arbitrage assumption

Let us assume the Black-Scholes hypotheses on the market, i.e. Hypothesis 1.1, except for the stochastic value of the underlying asset. Some option price properties of fixed-strike Eurasian and Amerasian call and put option prices are deduced in this section by mainly using the no arbitrage argument.

Let us denote by $V_c^{Eu}(t)$ (respectively, by $V_p^{Eu}(t)$) the value of an Eurasian call option (respectively, an Eurasian put option) at time t . Firstly, we recall from [71] the following result:

Proposition 1.1 *If*

$$\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K \geq 0 \quad \forall t \in [T_i, T_f],$$

then the Eurasian call option value is

$$V_c^{Eu}(t) = \left(\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K \right) e^{-r(T_f - t)} + \frac{1 - e^{-r(T_f - t)}}{r(T_f - T_i)} S(t). \quad (1.26)$$

PROOF. Firstly the terminal payoff of an Eurasian call option is decomposed as follows

$$\frac{1}{T_f - T_i} \int_{T_i}^{T_f} S(\tau) d\tau - K = \left(\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K \right) + \frac{1}{T_f - T_i} \int_t^{T_f} S(\tau) d\tau.$$

Assume that at time t_0 , the term between parentheses is positive, then it will be positive for all $t \geq t_0$ because the asset value is always positive.

Secondly, a self-financed strategy replicating the above payoff is built:

- The quantity $\left(\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K\right) e^{-r(T_f - t_0)}$ is invested into riskless bonds, so the return will be $\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K$ at time T_f .
- The investor transfers $\frac{1}{T_f - T_i} e^{-r(T_f - \tau)} \Delta\tau$ units of asset to a riskless bond every time in a time interval $(\tau, \tau + \Delta\tau)$ for $t_0 \leq \tau \leq T_f$, and thus the total units of asset required at time t_0 is

$$\frac{1}{T_f - T_i} \int_{t_0}^{T_f} e^{-r(T_f - \tau)} d\tau = \frac{1 - e^{-r(T_f - t_0)}}{r(T_f - T_i)}.$$

By no arbitrage arguments the two portfolios must be equal and then (1.26) holds. $\square$

Let us now establish an analytic formula for the difference between Eurasian call and Eurasian put options, i.e., a put-call parity relation similar to the one given for vanilla options in (1.7) by using similar techniques to those of the previous lemma.

Proposition 1.2 *The put-call parity, at any time $t \in [T_i, T_f]$, is given by*

$$V_c^{Eu}(t) - V_p^{Eu}(t) = \left(\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K\right) e^{-r(T_f - t)} + \frac{1 - e^{-r(T_f - t)}}{r(T_f - T_i)} S(t). \quad (1.27)$$

PROOF. Let us construct two self-financed strategies starting at time $t_0 \in [T_i, T_f]$.

- The first portfolio is long one Eurasian call and invest a quantity $K e^{-r(T_f - t_0)}$ in riskless bonds. At time T_f the return is

$$\left(\frac{1}{T_f - T_i} \int_{T_i}^{T_f} S(\tau) d\tau - K\right)_+ - K.$$

- The second portfolio is long one Eurasian put, invests $\frac{e^{-r(T_f - t_0)}}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau$ in a riskless bond, and transfers $\frac{1}{T_f - T_i} e^{-r(T_f - \tau)} \Delta\tau$ units of asset to a riskless bond every time during a time interval $(\tau, \tau + \Delta\tau)$ for $t_0 \leq \tau \leq T_f$. At time T_f the return is

$$\left(K - \frac{1}{T_f - T_i} \int_{T_i}^{T_f} S(\tau) d\tau\right)_+ + \frac{1}{T_f - T_i} \int_{T_i}^{t_0} S(\tau) d\tau + \frac{1}{T_f - T_i} \int_{t_0}^{T_f} S(\tau) d\tau.$$

By no arbitrage arguments the two portfolios must be equal, and then (1.27) holds. $\square$

Next, by using the previous propositions, we deduce the solution for the Eurasian pricing problem when the strike is equal to zero. The same result has been obtained in [12] for the particular case where the underlying asset follows a lognormal random walk.

Proposition 1.3 *For an Eurasian option with $K = 0$ we have*

$$V_c^{Eu}(t) = \left(\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau\right) e^{-r(T_f - t)} + \frac{1 - e^{-r(T_f - t)}}{r(T_f - T_i)} S(t), \quad (1.28)$$

$$V_p^{Eu}(t) = 0. \quad (1.29)$$

PROOF. Firstly, for the Eurasian put we have a sure zero return at maturity, thus (1.29) directly follows. Secondly, by using the put-call parity and (1.29), equation (1.28) is deduced. $\square$

Let us now introduce the notation $V_c^{Am}(t)$ (respectively, by $V_p^{Am}(t)$) for the value of an Amerasian call option (respectively, an Amerasian put option) at time t .

The following proposition is valid for all options, not only Asian ones, and it follows from the fact that American options offer more rights than their European counterparts (see, for instance, [71]).

Proposition 1.4 *Amerasian option price is higher than its European counterpart, namely,*

$$V_c^{Eu}(t) \leq V_c^{Am}(t), \quad V_p^{Eu}(t) \leq V_p^{Am}(t).$$

Proposition 1.5 *Assume that the price of the underlying asset S satisfies the following property: if at any time $S(t_0) = 0$, then $S(t) = 0 \quad \forall t > t_0$. In that case, if $S(t) = 0$ then the Asian option price at time t is given by*

$$V_c^{Eu}(t) = \left(\frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau - K \right)_+ e^{-r(T_f - t)}, \quad (1.30)$$

$$V_p^{Eu}(t) = \left(K - \frac{1}{T_f - T_i} \int_{T_i}^t S(\tau) d\tau \right)_+ e^{-r(T_f - t)}, \quad (1.31)$$

$$V_c^{Am}(t) = \left(\frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau - K \right)_+, \quad (1.32)$$

$$V_p^{Am}(t) = \left(K - \frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau \right)_+. \quad (1.33)$$

PROOF. Let us assume $S(t_0) = 0$ for some $t_0 \in [T_i, T_f]$. Then $\int_{T_i}^t S(\tau) d\tau = \int_{T_i}^{t_0} S(\tau) d\tau$ for any $t \geq t_0$. In this case, the payoff at T_f for the European option becomes a deterministic quantity whose discounted value gives (1.30) and (1.31).

For the American case, the exercise value function becomes also a deterministic function which decreases with time. Thus, it is optimal to exercise at t_0 and (1.32) and (1.33) follow. $\square$

It is clear that, if the exercise value function of an American option is zero, then it is optimal to hold it. The following two propositions apply this result to Amerasian options.

Proposition 1.6 *If $\frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau \leq K$ at time $t \in [T_i, T_f]$, then it is optimal to hold the Amerasian call option.*

Proposition 1.7 *If $\frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau \geq K$ at time $t \in [T_i, T_f]$, then it is optimal to hold the Amerasian put option.*

With the only assumption that the stochastic process for the underlying asset has continuous sample paths we have stated a set where it is optimal to hold the option (a subset of the set $\overline{\mathcal{D}}(t)$ for $t \in [T_i, T_f]$, defined in (1.21)) for both Amerasian call and put options.

Proposition 1.8 *Let us assume that the samples of the underlying asset value, $(S(t), T_i \leq t \leq T_f)$, are continuous. If*

$$S(t) > (1 + r(t - T_i)) \frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau, \quad (1.34)$$

at any time $t \in [T_i, T_f]$, then it is optimal to hold the Amerasian call option.

PROOF. Let us consider a fixed time $t_0 \in [T_i, T_f]$ where condition (1.34) is satisfied. For $\varepsilon > 0$ it will be useful to introduce the function

$$\alpha(\varepsilon) := e^{r\varepsilon} + \frac{e^{r\varepsilon} - 1}{\varepsilon}(t_0 - T_i),$$

which is continuous increasing and satisfies

$$\lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) = 1 + r(t_0 - T_i).$$

Firstly, from the continuity of α at $\varepsilon = 0$, we can find a small enough positive number ε_1^* such that $t_0 + \varepsilon_1^* < T_f$ and

$$S(t_0) > \left(e^{r\varepsilon_1^*} + \frac{e^{r\varepsilon_1^*} - 1}{\varepsilon_1^*}(t_0 - T_i) \right) \frac{1}{t_0 - T_i} \int_{T_i}^{t_0} S(\tau) d\tau.$$

Secondly, by using the continuity of the sample paths of the underlying asset we can find $\varepsilon_2^* < \varepsilon_1^*$ such that $t_0 + \varepsilon_2^* < T_f$ and satisfying

$$\begin{aligned} S(t_0 + \varepsilon) &> \left(e^{r\varepsilon_1^*} + \frac{e^{r\varepsilon_1^*} - 1}{\varepsilon_1^*}(t_0 - T_i) \right) \frac{1}{t_0 - T_i} \int_{T_i}^{t_0} S(\tau) d\tau \\ &> \left(e^{r\varepsilon} + \frac{e^{r\varepsilon} - 1}{\varepsilon}(t_0 - T_i) \right) \frac{1}{t_0 - T_i} \int_{T_i}^{t_0} S(\tau) d\tau \quad \forall \varepsilon \in (0, \varepsilon^*), \end{aligned} \quad (1.35)$$

where the second inequality comes from the fact that α is an increasing function. Now we write the average at time $t_0 + \varepsilon_2^*$:

$$M(t_0 + \varepsilon_2^*) = \frac{t_0 - T_i}{t_0 - T_i + \varepsilon_2^*} \frac{1}{t_0 - T_i} \int_{T_i}^{t_0} S(\tau) d\tau + \frac{1}{t_0 - T_i + \varepsilon_2^*} \int_{t_0}^{t_0 + \varepsilon_2^*} S(\tau) d\tau.$$

Using (1.35) and simplifying we get

$$\frac{1}{t_0 - T_i} \int_{T_i}^{t_0 + \varepsilon_2^*} S(\tau) d\tau > e^{r\varepsilon_2^*} \frac{1}{t_0 - T_i} \int_{T_i}^{t_0} S(\tau) d\tau.$$

Comparing the gain of exercising at time t_0 with the gain of waiting until $t_0 + \varepsilon_2^*$ we have at time $t_0 + \varepsilon_2^*$,

$$\left(\frac{1}{t_0 - T_i} \int_{T_i}^{t_0 + \varepsilon_2^*} S(\tau) d\tau - K \right)_+ \geq e^{r\varepsilon_2^*} \left(\frac{1}{t_0 - T_i} \int_{T_i}^{t_0} S(\tau) d\tau - K \right)_+.$$

Summing up, we have found a later time, $t_0 + \varepsilon_2^*$, at which the exercise gain is, at least, equal to the gain of exercising at t_0 . Therefore, it is optimal to hold the option. $\square$

Proposition 1.9 *Let us assume that the sample paths of the underlying asset value, $(S(t), T_i \leq t \leq T_f)$, are continuous. If*

$$S(t) < \frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau, \quad (1.36)$$

at any time $t \in [T_i, T_f]$, then it is optimal to hold the Amerasian put option.

PROOF. Using the continuity of the trajectories of the asset values we can find a positive number $\varepsilon_1^* > 0$, with $t_0 + \varepsilon_1^* < T_f$ such that $S(t_0 + \varepsilon) < M(t_0)$, for $0 < \varepsilon < \varepsilon_1^*$. In that case, inequality $M(t_0 + \varepsilon_1^*) < M(t_0)$ also holds. Next, we can find a small enough number $0 < \varepsilon_2^* \leq \varepsilon_1^*$ such that

$$M(t_0 + \varepsilon_1^*) < M(t_0) - K \left(e^{r\varepsilon_2^*} - 1 \right) \leq M(t_0) - K \left(e^{r\varepsilon_1^*} - 1 \right).$$

Thus, we have

$$K - M(t_0 + \varepsilon_1^*) > K - M(t_0) + K \left(e^{r\varepsilon_1^*} - 1 \right) = e^{r\varepsilon_1^*} K - M(t_0) \geq e^{r\varepsilon_1^*} (K - M(t_0)).$$

We have found a later time, $t_0 + \varepsilon_1^*$ where the payoff exercise value function is greater than or equal to the actualized value of the payoff function at t_0 . Thus, it is optimal to hold the option. $\square$

1.4.2 Black-Scholes model for fixed-strike Eurasian options

We follow the analysis of Section 1.3.3 for a particular path dependent option, the fixed strike Eurasian option. Different choices for variable $I(t)$ are possible. Let us consider the following:

$$M(t) = \frac{1}{t - T_i} \int_{T_i}^t S(\tau) d\tau, \quad (\text{i.e. } g(t) = 1/(t - T_i) \quad \text{and} \quad f(S, t) = S),$$

where we are using the notation of Section 1.3.3. Let us introduce

$$\mathcal{L}_M[\phi] := \frac{\partial \phi}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 \phi}{\partial S^2} + (r - d_0) S \frac{\partial \phi}{\partial S} + \frac{S - M}{t - T_i} \frac{\partial \phi}{\partial M} - r\phi.$$

The final value problem for the Eurasian option value function, $V = V(S, M, t)$, reads

$$\begin{cases} \mathcal{L}_M[V] &= 0 & \text{for } (S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \\ V(S, M, T_f) &= (M - K)_+ & \text{for } (S, M) \in \mathbb{R}_+ \times \mathbb{R}_+. \end{cases} \quad (1.37)$$

Remark 1.3 *We could also consider the cumulative integral variable*

$$A(t) = \int_{T_i}^t S(\tau) d\tau, \quad g(t) = 1 \quad \text{and} \quad f(S, t) = S.$$

In this case, the operator reads

$$\mathcal{L}_A[\phi] = \frac{\partial \phi}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 \phi}{\partial S^2} + (r - d_0) S \frac{\partial \phi}{\partial S} + S \frac{\partial \phi}{\partial A} - r\phi.$$

The final value problem for the option value, $U = U(S, A, t)$, is

$$\begin{cases} \mathcal{L}_A[U] &= 0 & \text{for } (S, A, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \\ U(S, A, T_f) &= (A/(T_f - T_i) - K)_+ & \text{for } (S, A) \in \mathbb{R}_+ \times \mathbb{R}_+. \end{cases} \quad (1.38)$$

Notice that the assumptions required for the propositions given in Section 1.4.1 are satisfied by the lognormal model for the underlying asset. Let us rewrite them applied to function V , for instance, and considering the case of continuous dividend yield d_0 :

- Put-call parity for $(S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f)$,

$$V_c(S, M, t) - V_p(S, M, t) = \left(\frac{t - T_i}{T_f - T_i} M - K \right) e^{-r(T_f - t)} + \frac{e^{-d_0(T_f - t)} - e^{-r(T_f - t)}}{(r - d_0)(T_f - T_i)} S. \quad (1.39)$$

- For $(S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f)$ such that $K(T_f - T_i) \leq (t - T_i)M$ we have

$$V_c(S, M, t) = \left(\frac{t - T_i}{T_f - T_i} M - K \right) e^{-r(T_f - t)} + \frac{e^{-d_0(T_f - t)} - e^{-r(T_f - t)}}{(r - d_0)(T_f - T_i)} S. \quad (1.40)$$

- For $(S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f)$ such that $S = 0$ we have

$$V_c(0, M, t) = \left(\frac{t - T_i}{T_f - T_i} M - K \right)_+ e^{-r(T_f - t)}.$$

Remark 1.4 Notice that if we divide the first equation of (1.37) by S^2 , we deduce that

$$\lim_{S \rightarrow \infty} \frac{\partial^2 V}{\partial S^2}(S, M, t) = 0.$$

More information about the behavior of the call option price at infinity can be deduced from (1.40) when $M/(T_f - T_i) \geq (t - T_i)K$. Otherwise, i.e., when $M/(T_f - T_i) < K$, we have the following property proposed in [70],

$$\lim_{S \rightarrow \infty} \frac{\partial V_c}{\partial S}(S, M, t) = \frac{T_f - t}{T_f - T_i} e^{-r(T_f - t)}. \quad (1.41)$$

Analogous conditions for put options follow from the put-call parity (1.39).

Remark 1.5 Notice that operator $\mathcal{L}_M$ degenerates at $t = T_i$. In fact, regarding the definition of the new path-dependent variable, we have

$$\lim_{t \rightarrow T_i} M(t) = S(t). \quad (1.42)$$

Thus, the two-factor problem becomes a one-factor problem at $t = T_i$, and the only not meaningless values are $V(S, S, T_i)$. In fact, a one-factor model can be formulated for times before the averaging time, as it is explained in Section 1.4.4.

Remark 1.6 Similarly to the vanilla options pricing problem, the Eurasian option value has a probabilistic representation (see Appendix B.1). The pair $(S(t), M(t), t \geq 0)$ is a diffusion process. Suppose that there exists a function ϕ satisfying some regularity hypotheses and furthermore

$$\begin{aligned} \mathcal{L}_M[\phi] &= 0, \\ \phi(x_1, x_2, T_f) &= \Lambda(x_1, x_2), \quad \text{with } \Lambda(x_1, x_2) = (x_2 - K)_+. \end{aligned}$$

Then, the option price has the probabilistic representation (see [74])

$$\mathbf{E} \left(e^{-r(T_f - t)} \Lambda(S(T_f), M(T_f)) | \mathcal{F}_t \right)_+,$$

where $\mathcal{F}_t$ denotes the natural filtration, and the mean is computed with respect to the “risk-neutral” probability (see Appendix A).

1.4.3 Black-Scholes model for fixed-strike Amerasian options

The general analysis for path-dependent Amerasian options in Section 1.3.4 directly applies to the fixed strike case. Let $V : \mathbb{R}_+ \times \mathbb{R}_+ \times [T_i, T_f] \rightarrow \mathbb{R}_+$ be the function giving the Amerasian option value, $\underline{\Lambda} : \mathbb{R}_+ \times \mathbb{R}_+ \times [T_i, T_f] \rightarrow \mathbb{R}_+$ the exercise value function and $\Lambda : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ the payoff function. For a fixed-strike call option $\underline{\Lambda}(S, M, t) = (\frac{M}{t-T_i} - K)_+$, and for the corresponding put option $\underline{\Lambda}(S, M, t) = (K - \frac{M}{t-T_i})_+$. In both cases, $\Lambda(S, M) = \underline{\Lambda}(S, M, T_f)$.

Following the notation of Section 1.3.4, we have two possibilities at each time $t \in (T_i, T_f)$:

$$\begin{aligned} V(S, M, t) &= \underline{\Lambda}(S, M, t) & \text{and} & \quad \mathcal{L}_M \leq 0 & \text{for } (S, M) \in \mathcal{C}(t), \\ V(S, M, t) &> \underline{\Lambda}(S, M, t) & \text{and} & \quad \mathcal{L}_M = 0 & \text{for } (S, M) \in \mathcal{D}(t). \end{aligned} \quad (1.43)$$

We also have the final condition

$$V(S, M, T_f) = \Lambda(S, M), \text{ for } (S, M) \in \mathbb{R}_+ \times \mathbb{R}_+. \quad (1.44)$$

We notice that (1.43) can be written as a linear complementarity problem:

$$\left. \begin{aligned} \mathcal{L}_M[V] (V - \underline{\Lambda}) &= 0 \\ \mathcal{L}_M[V] &\leq 0 \\ V - \underline{\Lambda} &\geq 0 \end{aligned} \right\} \quad \text{in } \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f). \quad (1.45)$$

The properties deduced in the Section 1.4.1 also hold under Black-Scholes hypotheses (i.e. Hypothesis 1.1). They are also valid if we include continuous dividends. Let us then rewrite them applied to function V :

- For $(S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f)$ such that $S = 0$,

$$V_c(0, M, t) = (M - K)_+, \quad (1.46)$$

$$V_p(0, M, t) = (K - M)_+. \quad (1.47)$$

- For an Amerasian call option,

$$\{(S, M), M \leq K\} \cup \{(S, M), S > (1 + rt)M\} \subset \overline{\mathcal{D}(t)}, \quad (1.48)$$

at each time $t \in (T_i, T_f)$.

- For an Amerasian put option,

$$\{(S, M), M \geq K\} \cup \{(S, M), S < M\} \subset \overline{\mathcal{D}(t)}, \quad (1.49)$$

at each time $t \in (T_i, T_f)$.

1.4.4 One dimensional model before averaging

Let us study a more general Asian option contract whose life begins before the averaging interval, i.e. it starts at T_0 with $T_0 < T_i$. For this purpose, we denote by $\widehat{V}$ the option price for the period

(T_0, T_i) . We consider, for instance, an European contract in (T_0, T_i) . By using dynamic hedging techniques, as in Section B.1, $\widehat{V} = \widehat{V}(S, t)$ satisfies the one-factor Black-Scholes equation,

$$\frac{\partial \widehat{V}}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \widehat{V}}{\partial S^2} + (r - d_0)S \frac{\partial \widehat{V}}{\partial S} - r\widehat{V} = 0, \quad \text{for } (S, t) \in \mathbb{R}_+ \times (T_0, T_i), \quad (1.50)$$

subject to final condition

$$\widehat{V}(S, T_i) = \lim_{t \rightarrow T_i^+} V(S, S, t), \quad (1.51)$$

where V denotes the option value after T_i (which is solution of (1.37) in the option is European after T_i and the solution of (1.45)-(1.44) if it is American). Condition (1.51) is easily deduced from identity (1.42), which, in its turn, is obtained by localization arguments from the definition of M .

An explicit relationship between call and put prices can also be found in this case, namely,

$$\widehat{V}_c(S, t) - \widehat{V}_p(S, t) = -Ke^{-r(T_f-t)} + \frac{e^{-d_0(T_f-t)} - e^{-r(T_f-T_i)}e^{-d_0(T_i-t)}}{(r - d_0)(T_f - T_i)}S. \quad (1.52)$$

Chapter 2

Functional framework for option pricing

2.1 Introduction

In Chapter 1 we have discussed the mathematical modelling of some option pricing problems (see also Appendix B) and we have developed a deeper analysis on the fixed strike Eurasian and Amerasian option pricing problem. In the present chapter we formulate the above problems in a suitable functional framework, not only for studying theoretical issues but also for carrying out a numerical discretization in a later step.

We first consider general problems, initially formulated in unbounded domains. Classical results (see, for instance, [54, 75, 87, 15, 57]) together with some particular applications to finance (such as [66, 62, 8]) are recalled.

Having in view the numerical solution, a truncation of the unbounded spatial domain is required. Thus, a discussion on this procedure, called localization process, is included (see [50, 66, 112, 62, 8]). Moreover, a weak formulation of the “localized problems” is an unavoidable step, since we will use the finite element method for spatial discretization.

Next, the general framework is particularized to the case of path-dependent options, which are characterized by a strong degeneration of the diffusion term. The ultraparabolic equations theory (see [96, 92, 73]) and the viscosity solutions theory (see [40, 9, 8]) seem to be the correct framework where to study such degenerate problems.

Next, the fixed strike Eurasian pricing problem is analyzed. A suitable change of variable allows us to treat an almost equivalent problem to the one studied in [13], where existence, uniqueness and regularity of solution are proved. Moreover, within this framework, the properties of the solution deduced in Chapter 1 are formally stated. As an example, the put-call parity relation is reconsidered. Moreover, it justifies the study of the Eurasian option pricing problem only in the call case.

The Amerasian option pricing problem is formulated in the functional framework proposed in [76]. However, the existence result remains to be an open problem.

Finally, the localization process for both Eurasian and Amerasian options is developed. Two points are studied: firstly, where it is necessary to impose a boundary condition (see [83]), and secondly, what are the suitable boundary conditions.

2.2 General problems

2.2.1 Unbounded domains

As we have seen in Chapter 1 (see also Appendix B), some pricing problems in finance are formulated as final value problems for general parabolic operators of the form,

$$\mathcal{L}_{nd}[\phi(\mathbf{x}, t)] := \frac{\partial \phi}{\partial t}(\mathbf{x}, t) + \sum_{i,j=1}^m a_{ij}(\mathbf{x}, t) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(\mathbf{x}, t) + \sum_{j=1}^m b_j(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j} + a_0(\mathbf{x}, t) \phi(\mathbf{x}, t). \quad (2.1)$$

The function ϕ is defined for $(\mathbf{x}, t) = (x_1, \dots, x_m, t) \in \Omega \times (T_i, T_f)$, where $\Omega \subset \mathbb{R}^m$ is an open domain. Moreover a_{ij}, b_i, a_0 and f are given measurable functions defined in $\Omega \times (T_i, T_f)$, and Λ is a given measurable function in Ω .

Let us reverse the direction of time by introducing a new variable, $T_f - t$, still denoted by t . The time domain is now $(0, T)$ with $T := T_f - T_i$. Moreover, let us write operator (2.1) in divergence form, namely,

$$\mathcal{L}[\phi(\mathbf{x}, t)] = \frac{\partial \phi}{\partial t}(\mathbf{x}, t) - \text{Div}(\mathbf{A}(\mathbf{x}, t) \nabla \phi(\mathbf{x}, t)) + \mathbf{v}(\mathbf{x}, t) \cdot \nabla \phi(\mathbf{x}, t) + l(\mathbf{x}, t) \phi(\mathbf{x}, t), \quad (2.2)$$

where the new coefficients A_{ij}, v_i, l are given by

$$A_{ii} = a_{ii}, \quad A_{ij} = A_{ji} = \frac{1}{2}(a_{ij} + a_{ji}), \quad (2.3)$$

$$v_j = \sum_{i=1}^n \frac{\partial A_{ij}}{\partial x_i} - b_j = \frac{\partial a_{jj}}{\partial x_j} + \frac{1}{2} \sum_{i=1}^n \frac{\partial (a_{ij} + a_{ji})}{\partial x_i} - b_j, \quad (2.4)$$

$$l = -a_0. \quad (2.5)$$

Notice that in (2.3) we have imposed symmetry to matrix $\mathbf{A}$. Equation (2.2) is simply a two-dimensional linear convection-diffusion-reaction equation, with diffusion tensor $\mathbf{A}$, velocity vector $\mathbf{v}$ (convection) and reaction coefficient l .

In finance, the “space-like” variables $x_1, x_2, \dots, x_m$ represent quantities (factors) such as the value of an underlying asset, a stochastic interest rate, an artificial state variable measuring averaging, etc. Very often, the spatial domain Ω is unbounded and there isn’t any boundary condition prescribed on its boundary. However, let us denote by Γ the boundary of Ω and let us consider the boundary conditions

$$\frac{\partial \phi}{\partial n_A}(\mathbf{x}, t) + \alpha \phi(\mathbf{x}, t) = g(\mathbf{x}, t) \text{ on } \Gamma_R, \quad (2.6)$$

$$\phi(\mathbf{x}, t) = h(\mathbf{x}, t) \text{ on } \Gamma_D, \quad (2.7)$$

where

$$\frac{\partial \phi}{\partial n_A}(\mathbf{x}, t) = \sum_{i,j=1}^m A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) n_i(\mathbf{x}),$$

and $\mathbf{n} = (n_1, \dots, n_m)$ denotes a unit outward normal vector to Γ . We are denoting by Γ_D (respectively, by Γ_R) the subset of Γ where Dirichlet boundary conditions (respectively, Robin boundary conditions) are prescribed. We always assume $\Gamma_D \cap \Gamma_R = \emptyset$, and we could also have $\Gamma_R = \emptyset$ and/or $\Gamma_D = \emptyset$.

With the previous notation and definitions, we can formulate some (European) pricing problems as initial value problems of the form:

Find $\phi : \Omega \times [0, T] \longrightarrow \mathbb{R}$ such that

$$\begin{cases} \mathcal{L}[\phi(\mathbf{x}, t)] = f(\mathbf{x}, t) & \text{in } \Omega \times (0, T), \\ \phi(\mathbf{x}, 0) = \Lambda(\mathbf{x}) & \text{in } \Omega, \end{cases} \quad (2.8)$$

and subjected to the boundary conditions (2.6)-(2.7).

The existence and uniqueness of classical solutions of problem (2.8) is established, for instance, in [54], in the particular case of $\mathcal{L}$ being a uniformly parabolic operator. Weak solutions are studied in [75]. The analytic semi-groups theory provides a stronger framework where to study the same problem (see, for instance, [87]). Unfortunately, none of these results applies to our Asian options pricing problem.

When the holder of the financial product has the additional right to exercise prior to expiry the pricing problem becomes nonlinear. The following unilateral-obstacle formulation covers a wide family of financial pricing problems.

Find $\phi : \Omega \times [0, T] \longrightarrow \mathbb{R}$ such that

$$\begin{cases} (\mathcal{L}[\phi(\mathbf{x}, t)] - f(\mathbf{x}, t))(\phi(\mathbf{x}, t) - \underline{\Lambda}(\mathbf{x}, t)) = 0 & \text{in } \Omega \times (0, T) \\ \mathcal{L}[\phi(\mathbf{x}, t)] \leq f(\mathbf{x}, t) & \text{in } \Omega \times (0, T), \\ \underline{\Lambda}(\mathbf{x}, t) \leq \phi(\mathbf{x}, t) & \text{in } \Omega \times (0, T), \\ \phi(\mathbf{x}, 0) = \Lambda(\mathbf{x}) & \text{in } \Omega, \end{cases} \quad (2.9)$$

and subjected to the boundary conditions (2.6)-(2.7).

In (2.9), operator $\mathcal{L}$ and the given functions f, Λ are defined as in (2.8), and $\underline{\Lambda}$ is a given measurable function in $\Omega \times (T_i, T_f)$.

See [57] for existence results under ellipticity hypothesis. However, sometimes financial pricing problems present particular features so that neither the classical theory nor the weak theory of partial differential equations seem to be the correct framework to study them. They concern, for instance, the degeneration of some diffusion operators at the axis (weak degeneration), the strong degeneration of the diffusion operator of the path dependent options, the lack of the regularity of the initial or constraint conditions, etc. In section 2.4.1 we shall see that the Eurasian options pricing problem can be analyzed in the framework of ultraparabolic equations (see, for instance, [96, 92]). For more general problems, the theory of viscosity solutions could provide a framework where to state the existence of solution (see [40]).

2.2.2 Truncated domains

As we have said, very often the pricing problems are pure Cauchy problems associated to a parabolic operator, i.e., with $\Gamma_D \cup \Gamma_R = \emptyset$. In general, only an initial condition is needed to ensure existence and uniqueness of solution. In some cases, the growth of the solution at infinity has to be prescribed too. However, numerical discretization by using finite-difference, finite-element or finite-volume methods makes it necessary to cut the unbounded spatial domain Ω at finite distance and to introduce there artificial boundary conditions. These are generally obtained by financial arguments, but also by pure mathematical reasoning, and they have to be included in the weak formulation. The process of approximating the solution of a problem posed in an unbounded domain by a solution of a problem posed in a bounded domain is called “localization process”, and it will be explained in the present section.

Now, let us denote by Ω^k a sequence of bounded domains, and by Γ^k their respective boundaries. We assume that $\Omega^k \subset \Omega$ and $\Omega^k \subset \Omega^{k+1} \forall k$, and $\Omega = \bigcup \Omega^k$. We denote by Γ_D^k (respectively Γ_R^k) the subset of Γ^k where Dirichlet (respectively Robin) boundary conditions are imposed. More specifically, given measurable functions α^k and g^k defined on $\Gamma_R^k \times [0, T]$ and h^k , defined on Γ_D^k , the linear problem (2.8) is replaced by the following one:

Find $\phi^k : \Omega^k \times [0, T] \longrightarrow \mathbb{R}$ such that

$$\begin{cases} \mathcal{L}[\phi^k(\mathbf{x}, t)] = f(\mathbf{x}, t) & \text{in } \Omega^k \times [0, T], \\ \phi^k(\mathbf{x}, 0) = \Lambda(\mathbf{x}) & \text{in } \Omega^k, \end{cases} \quad (2.10)$$

and subjected to the boundary conditions

$$\frac{\partial \phi^k}{\partial n_A}(\mathbf{x}, t) + \alpha^k \phi^k(\mathbf{x}, t) = g^k(\mathbf{x}, t) \text{ on } \Gamma_R^k, \quad (2.11)$$

$$\phi^k(\mathbf{x}, t) = h^k(\mathbf{x}, t) \text{ on } \Gamma_D^k, \quad (2.12)$$

where

$$\frac{\partial \phi^k}{\partial n_A}(\mathbf{x}, t) = \sum_{i,j=1}^m A_{ij}(\mathbf{x}, t) \frac{\partial \phi^k}{\partial x_j}(\mathbf{x}, t) n_i(\mathbf{x}),$$

and $\mathbf{n} = (n_1, \dots, n_m)$ denotes a unit outward normal vector to Γ^k .

Notice that functions α^k , g^k and h^k are “new” additional data, not necessarily present in (2.8). In general this “new” or “artificial” boundary conditions are associated to the new boundaries of the truncated domain, i.e., associated to $\Gamma^k \setminus (\Gamma \cap \Gamma^k)$. Moreover, similarly to the non-truncated problem, in some cases $\Gamma^k \setminus (\Gamma_D^k \cup \Gamma_R^k)$ is a non-empty set where no boundary conditions are needed because the natural condition is automatically satisfied.

In order to get to a weak formulation we multiply by a test function vanishing on Γ_D^k . Next, by integrating in Ω^k , applying Green’s formula and replacing the boundary conditions, the following formulation is obtained:

$$\begin{aligned} & \int_{\Omega^k} \frac{\partial \phi^k}{\partial t}(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} + \sum_{i,j=1}^m \int_{\Omega^k} A_{ij}(\mathbf{x}, t) \frac{\partial \phi^k}{\partial x_j}(\mathbf{x}, t) \frac{\partial \psi}{\partial x_i}(\mathbf{x}) d\mathbf{x} \\ & + \sum_{j=1}^m \int_{\Omega^k} v_j(\mathbf{x}, t) \frac{\partial \phi^k}{\partial x_j}(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^k} l(\mathbf{x}, t) \phi^k(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Gamma_R^k} \alpha^k \phi^k(\mathbf{x}, t) \psi(\mathbf{x}) dA_{\mathbf{x}} \\ & = \int_{\Omega^k} f(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Gamma_R^k} g^k(\mathbf{x}, t) \psi(\mathbf{x}) dA_{\mathbf{x}}. \end{aligned}$$

Remark 2.1 Very often in financial models, the truncated spatial domain is rectangular, i.e.,

$$\Omega^k = (a_1, b_1) \times \dots \times (a_m, b_m),$$

so that its boundary can be written as:

$$\Gamma^k = \bigcup_{i=1}^m \left(\Gamma_{i,+}^k \cup \Gamma_{i,-}^k \right),$$

where $\Gamma_{i,+}^k$ (respectively, $\Gamma_{i,-}^k$) is characterized by the unit outward normal vector

$$(0, \dots, 1^{(i)}, \dots, 0)$$

(respectively, $(0, \dots, -1^{(i)}, \dots, 0)$) (see Figure 2.1). With this notation Green's formula reads

$$\begin{aligned} \int_{\Omega^k} \text{Div}(\mathbf{A}(\mathbf{x}, t) \nabla \phi(\mathbf{x}, t)) \psi(\mathbf{x}) d\mathbf{x} &= - \sum_{i,j=1}^m \int_{\Omega^k} A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \frac{\partial \psi}{\partial x_i}(\mathbf{x}) d\mathbf{x} \\ &+ \sum_{i,j=1}^m \int_{\Gamma_{i,+}^k} A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \psi(\mathbf{x}) dA_{\mathbf{x}} \\ &- \sum_{i,j=1}^m \int_{\Gamma_{i,-}^k} A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \psi(\mathbf{x}) dA_{\mathbf{x}}. \end{aligned}$$

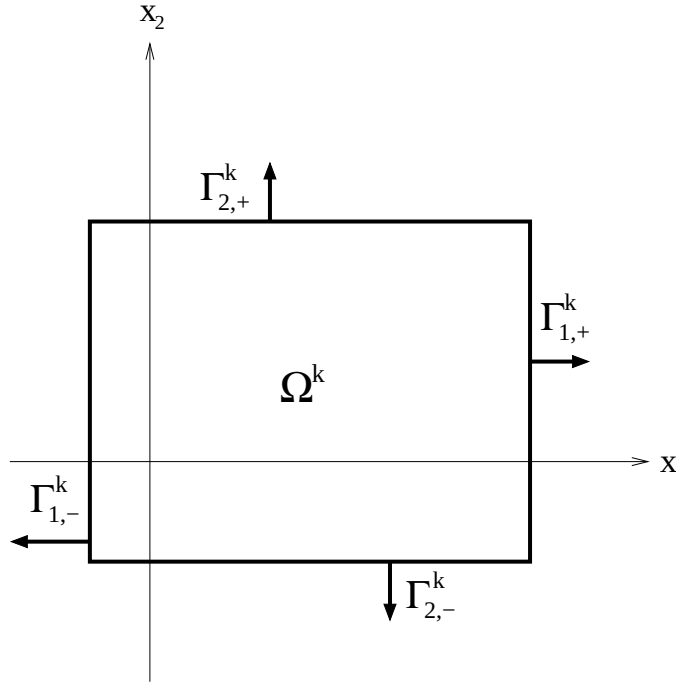


Figure 2.1: Two dimensional ($m = 2$) truncated spatial domain.

The same truncation technique is applied to the numerical solution of early exercise options pricing problems, leading to the formulation of obstacle problems in bounded domains.

The question of existence of the truncated solution, ϕ^k , defined in Ω^k and its convergence to the exact one, ϕ , defined in Ω , arises. In general, existence is gained for Dirichlet boundary conditions derived from the payoff function under similar conditions to existence in the unbounded case. Concerning the error introduced by the localization and regarding the linear case, it is commonly accepted for Black-Scholes type equations (see Appendix B) that, due to the rapid decay of the Green's function, the mathematical effect of imposing any non-exact condition on the artificial boundary can be made arbitrarily small by using a sufficiently large computational domain (see [50, 66, 112]). More specific results have been obtained for specific problems (see [66, 62, 72, 8]).

In [66] a particular study of the European multiasset vanilla options pricing problem (see Appendix B.3 for this mathematical model) has been developed. Existence and uniqueness of the solution is gained under some regularity hypotheses on the coefficients, by including Dirichlet

boundary conditions derived from the payoff and using a logarithmic change of the spatial variable that allows to use classical theory [54]. Moreover, from a comparison principle for parabolic PDEs, pointwise error estimates in the interior of the domain are stated in terms of the error on the artificial boundary condition. The logarithmic change of variable is also applied in [62] to study multiasset American vanilla options. Existence and uniqueness results for the unilateral constrained variational inequality is gained in the framework of weighted Sobolev spaces (for unbounded domains). The convergence of the truncated solution to the non-truncated one is also studied in [62] by a probabilistic approach (the same method is applied in [72] to the one factor linear Black-Scholes equation).

More general equations have been studied in the framework of viscosity solutions. In [8], Dirichlet or Neumann boundary conditions deduced from the payoff function are proposed and uniform convergence (at exponential rate) on compact subsets is shown. For uniform convergence on compact subsets a maximum principle is used and the rate of convergence is established by a probabilistic approach (large deviations theory).

2.3 Path-dependent options

2.3.1 Unbounded domain

A generalization of the model studied in Section 1.3.3 for an option on one underlying asset and one path-dependent variable is considered here. Assume we have m_1 underlying assets, $\{x_i\}_{i=1}^{m_1}$, and m_2 path-dependent variables, $\{x_i\}_{i=m_1+1}^{m_1+m_2}$. This case is clearly included in the more general one treated at the beginning of the present chapter. Indeed, we have the general operator (2.1) for $m = m_1 + m_2$, or, more precisely,

$$\mathcal{L}_{pd}[\phi(\mathbf{x}, t)] := \frac{\partial \phi}{\partial t}(\mathbf{x}, t) + \sum_{i,j=1}^{m_1} a_{ij}(\mathbf{x}, t) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(\mathbf{x}, t) + \sum_{j=1}^{m_1+m_2} b_j(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j} + a_0(\mathbf{x}, t) \phi(\mathbf{x}, t), \quad (2.13)$$

for $\mathbf{x}$ belonging to the unbounded domain $\Omega \in \mathbb{R}^m$ and $t \in (0, T)$. Thus, in the divergence form (2.2) the diffusion matrix takes the form

$$\begin{pmatrix} \mathbf{A}_{m_1}(\mathbf{x}, t) & \Theta_{m_1, m_2} \\ \Theta_{m_2, m_1} & \Theta_{m_2, m_2} \end{pmatrix},$$

where $\Theta_{i,j}$ denotes the $i \times j$ zero matrix and $\mathbf{A}_{m_1}$ is the reduced $m_1 \times m_1$ diffusion matrix. The lack of uniform ellipticity (or coercivity when regarding the weak formulation) leads to the analysis in the framework of equations with nonnegative characteristic form (see [83]), in particular, in the framework of ultraparabolic equations, for which the theory continues to be developed (see, for instance, [96, 92]). Some financial problems have been studied in this framework (see [73] for a survey paper). For instance, a nonlinear ultraparabolic equation is studied in [36]. It arises when agent decisions under risk are modelled. Moreover, an Asian option pricing problem is considered in [13].

2.3.2 Truncated domain

Having in view the numerical solution of path dependent options, the previously quoted localization techniques seem to be a first reasonable step. Thus, if we write the weak formulation of

(2.13) for a general test function ψ in a rectangular truncated domain Ω^k (see Remark 2.1 and Figure 2.1), we obtain

$$\begin{aligned}
& \int_{\Omega^k} \frac{\partial \phi}{\partial t}(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} + \sum_{i,j=1}^{m_1} \int_{\Omega^k} A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \frac{\partial \psi}{\partial x_i}(\mathbf{x}) d\mathbf{x} \\
& + \sum_{j=1}^{m_1+m_2} \int_{\Omega^k} v_j(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^k} l(\mathbf{x}, t) \phi(\mathbf{x}, t) \psi(\mathbf{x}) d\mathbf{x} \\
& = \sum_{i=1}^{m_1} \int_{\Gamma_{i,+}^k} \sum_{j=1}^{m_1} A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \psi(\mathbf{x}) dA_{\mathbf{x}} - \sum_{i=1}^{m_1} \int_{\Gamma_{i,-}^k} \sum_{j=1}^{m_1} A_{ij}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, t) \psi(\mathbf{x}) dA_{\mathbf{x}}.
\end{aligned} \tag{2.14}$$

Notice that the above equation suggests the need of boundary conditions on the first $2m_1$ boundaries, i.e., on $\Gamma_{i,+}^k$ and $\Gamma_{i,-}^k$, $i = 1, \dots, m_1$.

2.4 Fixed-strike Eurasian options

2.4.1 Existence of solution in the unbounded domain

Now, let us consider the particular parabolic linear operator for Asian options

$$\mathcal{L}_M[V] = \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - d_0) S \frac{\partial V}{\partial S} + \frac{S - M}{t - T_i} \frac{\partial V}{\partial M} - rV, \tag{2.15}$$

where V is a function defined on $\mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f)$. Operator $\mathcal{L}_M$ is a clear example of ultraparabolic operator, with no diffusion in one of the spatial directions. In the present section we study the Cauchy problem

$$\begin{cases} \mathcal{L}_M[V] = 0 & \text{for } (S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \\ \phi(S, M, T_f) = (M - K)_+ & \text{for } (S, M) \in \mathbb{R}_+ \times \mathbb{R}_+. \end{cases} \tag{2.16}$$

We can lead to a problem addressed in [13] by introducing the following change of variable and function

$$\begin{aligned}
y_1 &= S, y_2 = (t - T_i) \frac{\sigma^2}{2} M, \tau = \frac{\sigma^2}{2} (T_f - t), \\
\phi(y_1, y_2, \tau) &= y_1^m e^{q\tau} U \left(y_1, \frac{2y_2}{\sigma^2}, T_f - \frac{2\tau}{\sigma^2} \right), \\
m &= \frac{r-d_0}{\sigma^2}, q = m^2 + m - d_0, T = \frac{\sigma^2}{2} (T_f - T_i).
\end{aligned} \tag{2.17}$$

Under the above change, operator $\mathcal{L}_M$ becomes

$$\mathcal{L}_1[\phi] := \frac{\partial \phi}{\partial \tau} - y_1^2 \frac{\partial^2 \phi}{\partial y_1^2} - y_1 \frac{\partial \phi}{\partial y_2} = \frac{\partial \phi}{\partial \tau} - \frac{\partial}{\partial y_1} \left(y_1^2 \frac{\partial \phi}{\partial y_1} \right) + 2y_1 \frac{\partial \phi}{\partial y_1} - y_1 \frac{\partial \phi}{\partial y_2}, \tag{2.18}$$

for a function ϕ defined on $\mathbb{R}_+ \times \mathbb{R}_+ \times (0, T)$; and function Λ becomes

$$\Lambda(y_1, y_2) = y_1^m \left(\frac{2y_2}{\sigma^2 T} - K \right)_+ \quad \text{for } (y_1, y_2) \in \Omega = \mathbb{R}_+ \times \mathbb{R}_+. \tag{2.19}$$

If we additionally develop the change of variable $(z_1, z_2, \tau) = (\ln y_1, y_2, \tau)$, the following operator is obtained

$$\mathcal{L}_2[\phi] := \frac{\partial \phi}{\partial \tau} - \frac{\partial^2 \phi}{\partial z_1^2} + \frac{\partial \phi}{\partial z_1} - e^{z_1} \frac{\partial \phi}{\partial z_2}, \tag{2.20}$$

for a function ϕ defined on $\mathbb{R} \times \mathbb{R}_+ \times (0, T)$.

Using the notation of Section 2.3 we have, for both operators (2.18) and (2.20), an ultra-parabolic equation with $m_1 = 1, m_2 = 1$. Neither of them fits in the “general” framework of the bibliography of ultraparabolic problems, because the diffusive term in (2.18) degenerates completely at $y_1 = 0$, and the coefficients of the convective term in (2.20) are nonlinear in y_1 and y_2 . However, the problem has been separately studied in [13] by using operator $\mathcal{L}_1$ defined in (2.18). Existence, uniqueness, and classical regularity properties of the solution of

$$\begin{cases} \mathcal{L}_1[\phi] = 0 & \text{for } (y_1, y_2, \tau) \in \Omega \times (0, T), \\ \phi(y_1, y_2, 0) = \Lambda(y_1, y_2) & \text{for } (y_1, y_2) \in \Omega, \end{cases} \quad (2.21)$$

with Λ defined in (2.19), have been obtained. The problem is studied in the extended domain $\mathbb{R}_+ \times \mathbb{R}$ that we still note Ω . Notice that, if existence and uniqueness are established in the extended domain, they also hold true in the initial one.

Remark 2.2 *There exist two main points where we have to take special care in order to use the theory developed in [13] for our problem (2.16).*

- *We use the averaging variable, M , instead of the cumulative sum variable. For this, our change of variable (2.17) becomes singular for $t = T_f$, so that the results only apply in the open time interval.*
- *We allow for continuous dividend yield, thus the constant m defined in (2.17) can take negative values, whereas in [13] it is always positive. Although, from the financial point of view, it is not much realistic to have $d_0 > r$, we have considered also this case. This implies that, after developing the change of variable, function Λ defined in (2.19) is singular at $y_1 = 0$ if $m < 0$.*

In the following, we recall the main results in [13], pointing out the particular application to our problem, as it is explained in Remark 2.2.

Firstly, let us define a supersolution $\bar{\phi}$ of (2.21) as a function satisfying

$$\begin{cases} \mathcal{L}_1[\bar{\phi}] \geq 0 & \text{for } (y_1, y_2, \tau) \in \Omega \times (0, T), \\ \bar{\phi}(y_1, y_2, 0) \geq \Lambda(y_1, y_2) & \text{for } (y_1, y_2) \in \Omega, \end{cases} \quad (2.22)$$

and similarly for a subsolution $\underline{\phi}$ but with the reverse inequalities. Next, we have the following equivalence result, which claims that existence for the Cauchy problem (2.21) is equivalent to find a subsolution and a supersolution of the same problem.

Theorem 2.1 *Let $\Lambda \in C(\Omega)$ and $\bar{\phi}$ and $\underline{\phi}$ be super and subsolution, respectively, of problem (2.21) satisfying $\underline{\phi} \leq \bar{\phi}$ in $\Omega \times [0, T]$. Then, there exists a (classical) solution ϕ to problem (2.21), such that*

$$\underline{\phi} \leq \phi \leq \bar{\phi}, \text{ in } \Omega \times (0, T). \quad (2.23)$$

The proof given in [13] is constructive in the sense that two sequences of functions, defined in bounded domains and converging to “two” solutions of (2.21), are built. The axis $y_1 = 0$ is not included in any of the truncated domains (see [13] or Section 2.4.3), so these sequences can be defined also when $m < 0$.

In [13] a pair of sub/supersolutions is proposed. In the following we propose a parameter-dependent pair of functions, and verify that they are sub/supersolutions. We will see that the cases $m \geq 0$ and $m < 0$ lead to different functions. We propose:

- subsolution: $\underline{\phi} = 0$,
- supersolution: $\bar{\phi} = \alpha_0 e^{\tau(m^2 + (\alpha_1 - 1)m + \alpha_2)} y_1^m \sqrt{y_1^2 + y_2^2}$, for constant parameters α_i , $i = 0, 1, 2$.

The verifications for $\underline{\phi}$ are straightforward. The second condition of (2.22) for the supersolution is satisfied by choosing $\alpha_0 = 2\sigma^2/T$. Indeed, we have

$$\bar{\phi}(y_1, y_2, 0) = \alpha_0 y_1^m \sqrt{y_1^2 + y_2^2} \geq x^m (\alpha_0 y_2)_+ \geq x^m (\alpha_0 y_2 - K)_+.$$

Next, the application of operator $\mathcal{L}_1$ to $\bar{\phi}$ gives

$$\begin{aligned} \mathcal{L}_1[\bar{\phi}] = \alpha_0 e^{\tau(m^2 + (\alpha_1 - 1)m + \alpha_2)} & \left(\sqrt{y_1^2 + y_2^2} (\alpha_1 m + \alpha_2 - m^2 + m) \right. \\ & \left. - \frac{2m y_1^2 + y_1^2 + y_1 y_2}{\sqrt{y_1^2 + y_2^2}} + \frac{y_1^4}{(y_1^2 + y_2^2)^{\frac{3}{2}}} \right). \end{aligned}$$

Now, by developing some computations and reordering the terms, the above expression can be written as

$$\begin{aligned} \mathcal{L}_1[\bar{\phi}] = & \alpha_0 e^{\tau(m^2 + (\alpha_1 - 1)m + \alpha_2)} (m(y_1^2 + y_2^2) ((\alpha_1 - 2)y_1^2 + \alpha_1 y_2^2) \\ & + \frac{(y_1 - y_2)^4}{4} + \left(\alpha_2 - \frac{1}{4}\right) y_1^4 + \left(\alpha_2 - \frac{1}{4}\right) y_2^4 + \left(2\alpha_2 - \frac{5}{2}\right) y_1^2 y_2^2). \end{aligned}$$

We can now discuss the sign of $\mathcal{L}_1[\bar{\phi}]$ in function of m and α_i , $i = 0, 1, 2$. More precisely, for fixed $\alpha_0 = 2\sigma^2/T$, we choose parameters α_i , $i = 1, 2$ in order to satisfy the first condition of (2.22):

- If $m \geq 0$, $\alpha_1 > 2$ and $\alpha_2 > 5/4$ then $\mathcal{L}_1[\bar{\phi}] \geq 0$.
- If $m < 0$, $\alpha_1 < 0$ and $\alpha_2 > 5/4$ then $\mathcal{L}_1[\bar{\phi}] \geq 0$.

Remark 2.3 *The supersolution proposed in [13] corresponds to the choice $\alpha_1 = 3$ and $\alpha_2 = 2$. Thus, it satisfies the condition when $m \geq 0$, but it is not a supersolution when $m < 0$.*

Uniqueness of solution is deduced under some additional conditions from the following maximum principle proposition (see [13]).

Proposition 2.1 *Let $\psi \in C(\overline{\Omega \times [0, T]})$ be such that*

$$\begin{array}{lll} \mathcal{L}_1[\psi] & \geq 0 & \text{in } \Omega \times (0, T), \\ \psi(y_1, y_2, 0) & \geq 0 & \text{in } \Omega, \\ \psi(y_1, y_2, t) & \geq -c_1 e^{c_2(\ln(y_1^2 + y_2^2 + 1) - \ln(y_1))^2} & \text{in } \Omega \times (0, T), \end{array}$$

for two positive constants c_1 and c_2 . Then, $\psi \geq 0$ in $\Omega \times (0, T)$.

In [13] two solutions of problem (2.21) satisfying condition (2.23) are built. By using Proposition 2.1 and the pair of proposed sub/supersolutions, it follows that the two solutions are the same.

With respect to the regularity of the solution, as it is pointed out in [13], when the diffusion matrix is positive semidefinite and with an eigenvalue equal to zero, then it may be possible

to reduce in one the dimension of the problem through a change of variable, eliminating the degeneration of the initial problem. In such a case, the solution would be not necessarily smooth. Otherwise, by using results from [59], the solution is a C^∞ function. In [12] authors give necessary and sufficient conditions to characterize whether or not the reduction of the dimension is available for degenerate diffusions by using Lie algebra tools. In particular, they prove that the fixed strike Eurasian options PDE “cannot be reduced to a one-dimensional non-degenerate PDE”. Moreover, the result is independent of the sign of $r - d_0$. Thus, the solution of (2.16) is C^∞ .

Remark 2.4 *In Chapter 5 we will treat the numerical solution of the Eurasian options pricing problems. Notice that the previously described analysis provides enough regularity so that numerical methods based on the smoothness of the solution can be successfully applied, particularly, the one proposed in Chapter 5.*

2.4.2 Put-call parity relation and other properties

Properties deduced by no arbitrage arguments in Section 1.4.1 can be now established with a pure mathematical reasoning by using the results of the last section. For instance, if we denote by ϕ_c the call option value, we have

$$\phi_c(y_1, y_2, \tau) = y_1^m e^{q\tau} \left(\left(\frac{2}{\sigma^2 T} y_2 - K \right) e^{-\frac{2r}{\sigma^2} \tau} + \frac{e^{-\frac{2d_0}{\sigma^2} \tau} - e^{-\frac{2r}{\sigma^2} \tau}}{(r - d_0)T} y_1 \right),$$

for $y_2 \geq \frac{\sigma^2 KT}{2}$, and it is the unique solution in that region.

Now, let ϕ_p be the corresponding Eurasian put value, and let $\phi_{cp} := \phi_c - \phi_p$. By linearity, ϕ_{cp} solves (2.21) with the initial condition

$$\Lambda(y_1, y_2) = y_1^m \left(\frac{2y_2}{\sigma^2 T} - K \right).$$

Existence of solution ϕ_{cp} can be obtained by using Theorem 2.1 for the pair:

- subsolution: $\underline{\phi} = -K \frac{2}{\sigma^2 T} e^{\tau(m^2 + (\alpha_1 - 1)m + \alpha_2)} y_1^m \sqrt{y_1^2 + y_2^2}$,
- supersolution: $\overline{\phi} = \frac{2}{\sigma^2 T} e^{\tau(m^2 + (\alpha_1 - 1)m + \alpha_2)} y_1^m \sqrt{y_1^2 + y_2^2}$,

with α_1 and α_2 suitably chosen depending on the sign of m . The uniqueness of solution can be stated by using Proposition 2.1 together with the pair of sub/supersolutions proposed.

Furthermore, we can also obtain an analytical expression of the solution, namely,

$$\phi_{cp}(y_1, y_2, \tau) = y_1^m e^{q\tau} \left(\left(\frac{2}{\sigma^2 T} y_2 - K \right) e^{-\frac{2r}{\sigma^2} \tau} + \frac{e^{-\frac{2d_0}{\sigma^2} \tau} - e^{-\frac{2r}{\sigma^2} \tau}}{(r - d_0)T} y_1 \right).$$

Remark 2.5 *Notice that the put-call parity for Eurasian options allows us to study only the Eurasian call option pricing problem. This does not occur for Amerasian options.*

2.4.3 Truncated domain: boundary conditions and weak formulation

In Section 2.2.2 we have introduced a localization process applied to general models arising in finance, as an intermediate step previous to the numerical solution by PDE methods. In the present section we discuss some questions related to the localization process when applied to the Eurasian options pricing problem.

Firstly, let us point out that the proof of existence of Theorem 2.1 in [13] provides a tool for defining a well-posed problem in a bounded spatial domain, whose solution converges (uniformly in compact sets) to the solution of the Cauchy problem (2.21).

More precisely, for any natural number k , let Ω^k be defined by

$$\Omega^k = \left(\frac{1}{k+1}, k+1 \right) \times (-(k+1), k+1), \quad (2.24)$$

Let $\chi^k : \mathbb{R}_+ \times \mathbb{R} \longrightarrow [0, 1]$ be a continuous function such that

$$\begin{aligned} \chi^k(\mathbf{y}) &= 0 \quad \forall \mathbf{y} \notin \Omega^k \\ \chi^k(\mathbf{y}) &= 1 \quad \forall \mathbf{y} \in \Omega^{k-1} \end{aligned}$$

and let ψ^k be the solution of

$$\begin{cases} \frac{\partial \psi^k}{\partial \tau} - y_1^2 \frac{\partial^2 \psi^k}{\partial y_1^2} - y_1 \frac{\partial \psi^k}{\partial y_2} = 0 & \text{for } (y_1, y_2, \tau) \in \Omega^k \times (0, T), \\ \psi^k(y_1, y_2, 0) = \Lambda(y_1, y_2) & \text{for } (y_1, y_2) \in \Omega^k, \\ \psi^k(y_1, y_2, \tau) = \widetilde{\Lambda}^k(y_1, y_2, \tau) & \text{for } (y_1, y_2, \tau) \in \partial\Omega^k \times (0, T). \end{cases} \quad (2.25)$$

By choosing

$$\widetilde{\Lambda}^k(\mathbf{y}, t) = \chi^k(\mathbf{y})\Lambda(\mathbf{y}) + (1 - \chi^k(\mathbf{y}))\underline{\phi}(\mathbf{y}, \tau),$$

an increasing sequence of “generalized” solutions of problem (2.25), $\{\psi^k\}_k$, satisfying

$$\underline{\phi} \leq \psi^k \leq \overline{\phi} \quad (2.26)$$

is found. A generalized solution is a C^∞ solution in the interior of the domain, but it is not a priori clear in what part of the boundary takes the prescribed value (see [13] for a more precise definition). Analogously, a decreasing sequence is found for

$$\widetilde{\Lambda}^k(\mathbf{y}, \tau) = \chi^k(\mathbf{y})\Lambda(\mathbf{y}) + (1 - \chi^k(\mathbf{y}))\overline{\phi}(\mathbf{y}, \tau).$$

Each of them converges uniformly on compact sets to a function, $\varphi_i, i = 1, 2$, satisfying (2.21) and $\underline{\phi} \leq \varphi_i \leq \overline{\phi}, i = 1, 2$.

However, from the numerical efficiency point of view, we prefer to work with operator $\mathcal{L}_M$ (instead of $\mathcal{L}_1$) and with the spatial domain $\Omega^* := (0, x_1^*) \times (0, x_2^*)$ with boundary Γ^* (instead of Ω^k and Γ^k). In this case

$$\Gamma^* = \bigcup_{i=1}^2 (\Gamma_{i,+}^* \cup \Gamma_{i,-}^*),$$

where

$$\begin{aligned}\Gamma_{1,+}^* &= \{(x_1, x_2), x_1 = x_1^*, 0 \leq x_2 \leq x_2^*\}, \\ \Gamma_{1,-}^* &= \{(x_1, x_2), x_1 = 0, 0 \leq x_2 \leq x_2^*\}, \\ \Gamma_{2,+}^* &= \{(x_1, x_2), x_2 = x_2^*, 0 \leq x_1 \leq x_1^*\}, \\ \Gamma_{1,+}^* &= \{(x_1, x_2), x_2 = 0, 0 \leq x_1 \leq x_1^*\}.\end{aligned}$$

Moreover, by using the classical theory of second order equations with nonnegative characteristic form, we will have a well posed problem even with a portion of the boundary free from boundary conditions (see [83]). For this purpose, let us consider the elliptic operator $\mathcal{L}_*$ associated to the parabolic operator $\mathcal{L}_M$ written in non-divergence form

$$\mathcal{L}_*[\phi] := \sum_{ij} a_{ij}^* \frac{\partial^2 \phi}{\partial x_i \partial x_j} + \sum_j b_j^* \frac{\partial \phi}{\partial x_j} + l^* \phi,$$

with

$$\mathbf{A}^* := (a_{ij}^*) = \begin{pmatrix} \frac{1}{2}\sigma^2 x_1^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{v}^* := (b_j^*) = \begin{pmatrix} (r - d_0)x_1 \\ \frac{x_1 - x_2}{T - \tau} \end{pmatrix}, \quad l^* = -r. \quad (2.27)$$

The Dirichlet problem for operator $\mathcal{L}^*$ is well posed when imposing the boundary condition only on $\Sigma_2 \cup \Sigma_3$, where

$$\Sigma_3 := \left\{ (x_1, x_2) \in \Gamma^*, \sum_{ij} a_{ij}^* m_i m_j > 0 \right\},$$

and

$$\Sigma_2 := \left\{ (x_1, x_2) \in \Gamma^* \setminus \Sigma_3, \sum_j \left(b_j^* - \sum_r \frac{\partial a_{jr}^*}{\partial x_r} \right) m_j < 0 \right\},$$

with $\mathbf{m} = (m_1, m_2)$ the inward normal vector. By replacing coefficients (2.27) in each of the boundaries, we obtain

- $\Gamma_{1,+}^*$

$$a_{11}^* m_1 m_1 = \frac{1}{2} \sigma^2 x_1^2 (-1)^2|_{x_1=x_1^*} > 0.$$

- $\Gamma_{1,-}^*$

$$a_{11}^* m_1 m_1 = \frac{1}{2} \sigma^2 x_1^2|_{x_1=0} = 0, \quad \sum_j \left(b_j^* - \sum_r \frac{\partial a_{jr}^*}{\partial x_r} \right) m_j = (r - d_0 - \sigma^2) x_1|_{x_1=0} = 0.$$

- $\Gamma_{2,+}^*$

$$a_{11}^* m_1 m_1 = 0, \quad \sum_j \left(b_j^* - \sum_r \frac{\partial a_{jr}^*}{\partial x_r} \right) m_j = \frac{x_2^* - x_1}{T - \tau} > 0 \text{ if } x_2^* \geq x_1^*.$$

- $\Gamma_{2,-}^*$

$$a_{11}^* m_1 m_1 = 0, \quad \sum_j \left(b_j^* - \sum_r \frac{\partial a_{jr}^*}{\partial x_r} \right) m_j = \frac{x_1}{T - \tau} > 0.$$

Therefore $\Sigma_3 = \Gamma_{1,+}^*$ and, by choosing $x_2^* \geq x_1^*$, the set Σ_2 is empty.

Let us now consider the parabolic problem for operator $\mathcal{L}_M$ and write a weak formulation of the problem. Firstly, by using the following variables

$$x_1 = S, x_2 = M, \tau = T_f - t,$$

then operator $\mathcal{L}_M$ written in divergence form reads

$$\mathcal{L}_M[\phi] = \frac{\partial \phi}{\partial t} - \text{Div}(\mathbf{A}(\mathbf{x}, t) \nabla \phi(\mathbf{x}, t)) + \mathbf{v}(\mathbf{x}, t) \cdot \nabla \phi(\mathbf{x}, t) + l(\mathbf{x}, t),$$

with

$$\mathbf{A}(\mathbf{x}, t) = \begin{pmatrix} \frac{1}{2}\sigma^2 x_1^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{v}(\mathbf{x}, t) = \begin{pmatrix} (\sigma^2 - r + d_0)x_1 \\ \frac{x_2 - x_1}{T - \tau} \end{pmatrix}, \quad l(\mathbf{x}, t) = r, \quad (2.28)$$

and ϕ defined in $\mathbb{R}_+ \times \mathbb{R}_+ \times [0, T]$ with $T := T_f - T_i$. Next, let us introduce the bounded domain $\Omega^* = (0, x_1^*) \times (0, x_2^*)$, for fixed numbers x_1^* and x_2^* . For the sake of simplicity we still denote ϕ the solution of the truncated problem.

The weak formulation (2.14) applied to this equation reads

$$\begin{aligned} & \int_{\Omega^*} \frac{\partial \phi}{\partial \tau}(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^*} A_{11}(\mathbf{x}, t) \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \frac{\partial \psi}{\partial x_1}(\mathbf{x}) d\mathbf{x} \\ & + \sum_{j=1}^2 \int_{\Omega^*} v_j(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^*} l(\mathbf{x}, \tau) \phi(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} \\ & = \int_{\Gamma_{1,+}^*} A_{11}(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \psi(\mathbf{x}) dA_{\mathbf{x}}, \end{aligned}$$

for a general test function ψ , where we have taken into account that $A_{11}(\mathbf{x}, \tau) = 0$ if $\mathbf{x} \in \Gamma_{1,-}$.

Remark 2.6 We notice that the weak formulation also suggests that the only portion of the boundary not free from the prescription of boundary conditions is $\Gamma_{1,+}^*$. However, the boundary integral term in the weak formulation derives only from the form of the diffusion matrix, whereas results appearing in [83] also involve the velocity field. In fact, we will see in Chapter 5 that, under condition $x_2^* \geq x_1^*$, the “characteristic lines” cross the boundary $\Gamma^* \setminus \Gamma_{1,+}^*$ tangentially or outwards (this behavior is desirable from the numerical point of view).

We could prescribe on $\Gamma_{1,+}^*$ the Dirichlet boundary condition deduced from the payoff function (or initial condition). However, in practice, we impose a Neumann boundary condition given by the following function:

$$\frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) = g(\mathbf{x}, \tau) = \begin{cases} \frac{\tau}{T} e^{-r\tau} & \text{if } 0 < x_2 < TK, \\ \frac{e^{-d_0\tau} - e^{-r\tau}}{T(r - d_0)} & \text{if } x_2 > TK, \end{cases} \quad (2.29)$$

where the first expression is taken from [70] and the second one is motivated by equation (1.41).

Then we are led to the approximated problem

$$\begin{aligned} & \int_{\Omega^*} \frac{\partial \phi}{\partial \tau}(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^*} A_{11}(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \frac{\partial \psi}{\partial x_1}(\mathbf{x}) d\mathbf{x} \\ & + \sum_{j=1}^2 \int_{\Omega^*} v_j(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^*} l(\mathbf{x}, \tau) \phi(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} \\ & = \int_{\Gamma_{1,+}^*} A_{11}(\mathbf{x}, \tau) g(\mathbf{x}, \tau) \psi(\mathbf{x}) dA_{\mathbf{x}}. \end{aligned} \quad (2.30)$$

Although the Cauchy problem for the Neumann condition on $\Gamma_{1,+}$ does not fit the framework of the theory, numerical evidence suggests that the problem remains well posed (see also [78]). We will address its numerical solution in Chapter 5.

2.5 Fixed strike Amerasian options

2.5.1 Unbounded domain

Some recent works in the literature, as [36] and [73], study a nonlinear problem for an ultraparabolic operator arising in finance. In this studies the nonlinearity comes from a convection term $(b_1, b_2) = (0, \phi)$, where ϕ denotes the solution. Variational inequalities of ultraparabolic type, i.e., where the nonlinearity is of obstacle type, are less studied. It seems that the viscosity solutions theory is the correct framework where these problems should be studied. The question of existence and regularity of Amerasian call option prices can be addressed in this context.

In [76] a functional framework where to formulate the Amerasian options pricing problem has been proposed. The formulation involves operator (2.20), and the weighted Sobolev spaces $W^{m,p}(\mathbb{R}, e^{-c|x_1|})$, for a positive constant c . Moreover, if $Q := (0, \infty) \times (0, T) \subset \mathbb{R}^2$ and X is a Banach space, the space $L^p(\overline{Q}, X)$ is the set of measurable functions $g : \overline{Q} \rightarrow X$ such that $\int_{\overline{Q}} \|g(x_2, t)\|_X^p dx_2 dt < \infty$.

If $\underline{\Lambda}$ satisfies the growth constraint $\underline{\Lambda}(\mathbf{x}) = O(e^{-c_2|x|})$ for some $c_2 > 0$, then for any $p > 3/2$ and c large enough we have $\underline{\Lambda} \in L^2(\overline{Q}, W^{2,p}(\mathbb{R}, e^{-c|x_1|}))$ and the solution is found in $\phi \in L^2(\overline{Q}, H_{loc}^1(\mathbb{R})) \cap L^p(\overline{Q}, W^{2,p}(\mathbb{R}, e^{-c|x_1|}))$, with $\frac{\partial \phi}{\partial t}$ and $\frac{\partial \phi}{\partial x_2} \in L^2(\overline{Q}, L_{loc}^2(\mathbb{R})) \cap L^p(\overline{Q}, L^p(\mathbb{R}, e^{-c|x_1|}))$ satisfying:

$$\begin{cases} \mathcal{L}_2[\phi] (\phi - \underline{\Lambda}) &= 0 & \text{in } \mathbb{R} \times Q, \\ \mathcal{L}_2[\phi] &\geq 0 & \text{in } \mathbb{R} \times Q, \\ \phi &\geq \underline{\Lambda} & \text{in } \mathbb{R} \times Q, \\ \phi(\mathbf{x}, 0) &= \Lambda(\mathbf{x}) & \text{in } \mathbb{R} \times \mathbb{R}_+. \end{cases} \quad (2.31)$$

Operator $\mathcal{L}_2$ is defined by (2.20),

$$\underline{\Lambda}(x_1, x_2, t) = e^{mx_1} \left(\frac{2x_2}{\sigma^2(T-t)} - K \right)_+$$

for a call option, and

$$\underline{\Lambda}(x_1, x_2, t) = e^{mx_1} \left(K - \frac{2x_2}{\sigma^2(T-t)} \right)_+$$

for a put option. Furthermore $\Lambda(\mathbf{x}) = \underline{\Lambda}(\mathbf{x}, 0)$ in both cases. In [76] the problem is also formulated in a bounded domain, and convergence considerations are given. Nevertheless, it seems that the existence of solution is an open problem for both unbounded and bounded domains.

2.5.2 Truncated domain: boundary conditions and weak formulation

In practice we solve the extension of the problem (2.30) given in Section 2.4.3 to the unilateral constrained case. Besides the notations used in that section, we introduce the constraint function, $\underline{\Delta}$, defined in $\Omega^* \times (0, T)$. In this case $\mathcal{K}_\tau$ is

$$\mathcal{K}_\tau := \{\psi : \Omega^* \longrightarrow \mathbb{R}, \psi(\mathbf{x}) \geq \underline{\Delta}(\mathbf{x}, \tau)\}.$$

Assuming that we know a function g such that

$$g(\mathbf{x}, \tau) = \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \quad \mathbf{x} \in \Gamma_{1,+}^*,$$

the Amerasian problem, formulated as a variational inequality, reads:

For all $\tau \in (0, T)$, find $\phi(\tau) \in \mathcal{K}_\tau$ such that

$$\begin{aligned} & \int_{\Omega^*} \frac{\partial \phi}{\partial \tau}(\mathbf{x}, \tau) (\psi(\mathbf{x}) - \phi(\mathbf{x}, \tau)) d\mathbf{x} + \int_{\Omega^*} A_{11}(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \left(\frac{\partial \psi}{\partial x_1}(\mathbf{x}) - \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \right) d\mathbf{x} \\ & + \sum_{j=1}^2 \int_{\Omega^*} v_j(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, \tau) (\psi(\mathbf{x}) - \phi(\mathbf{x}, \tau)) d\mathbf{x} + \int_{\Omega^*} l(\mathbf{x}, \tau) \phi(\mathbf{x}, \tau) (\psi(\mathbf{x}) - \phi(\mathbf{x}, \tau)) d\mathbf{x} \\ & \geq \int_{\Gamma_{1,+}^*} A_{11}(\mathbf{x}, \tau) g(\mathbf{x}, \tau) (\psi(\mathbf{x}) - \phi(\mathbf{x}, \tau)) dA_{\mathbf{x}}, \end{aligned}$$

for all $\psi \in \mathcal{K}_\tau$.

The exercise function for a call option is

$$\underline{\Delta}(\mathbf{x}, \tau) = \left(\frac{x_2}{T - \tau} - K \right)_+,$$

and for a put option is

$$\underline{\Delta}(\mathbf{x}, \tau) = \left(K - \frac{x_2}{T - \tau} \right)_+.$$

With respect to the choice of function g we can derive it from the exercise value function, so

$$g(\mathbf{x}, t) = 0,$$

for both Amerasian call and put options. However, in the call option case, we can combine our knowledge about the regions of the spatial domain where it is a priori optimal to hold the option (i.e., where a European-like behavior is expected) with the boundary condition given for Eurasian call options. In fact, regarding Figure 2.3, we notice that $\Gamma_{1,+}^*$ is “almost” inside the optimal-to-hold region. This is the reason why we propose, in practice, to use the same function g for Eurasian and Amerasian call options.

Following Proposition 1.7, in Figure 2.2 we have, colored the regions where we know a priori that it is optimal to hold the Amerasian put option. Notice that this region is time independent. In Figure 2.3 we have colored the regions where we know a priori that it is optimal to hold the Amerasian call option, for three different times. Proposition 1.6 has been used in this second case.

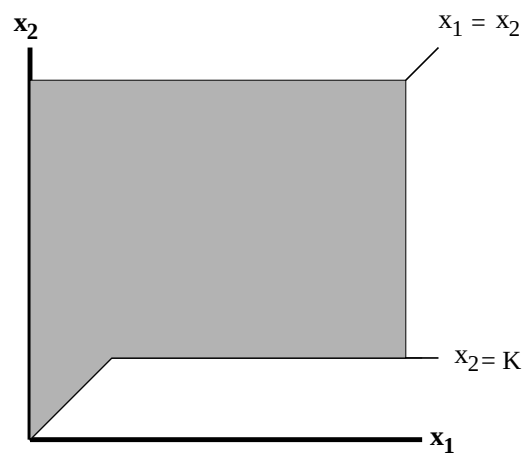


Figure 2.2: Spatial domain $\Omega = (0, \infty) \times (0, \infty)$ for the Amerasian put option pricing problem derived from Proposition 1.9. The region where it is optimal to hold the option is colored.

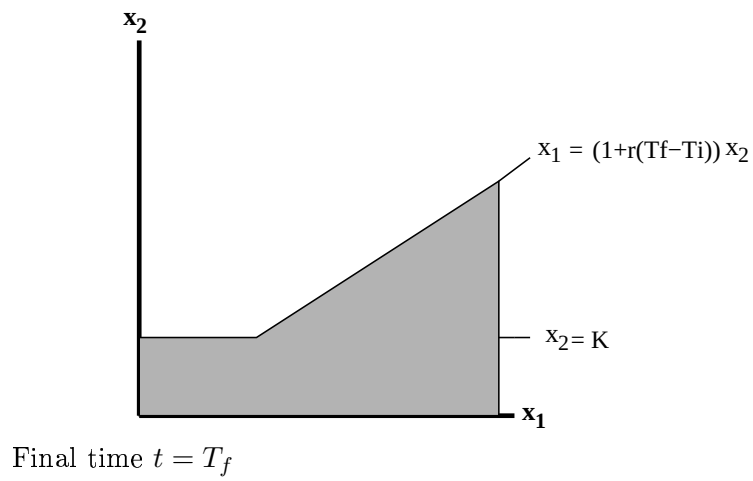
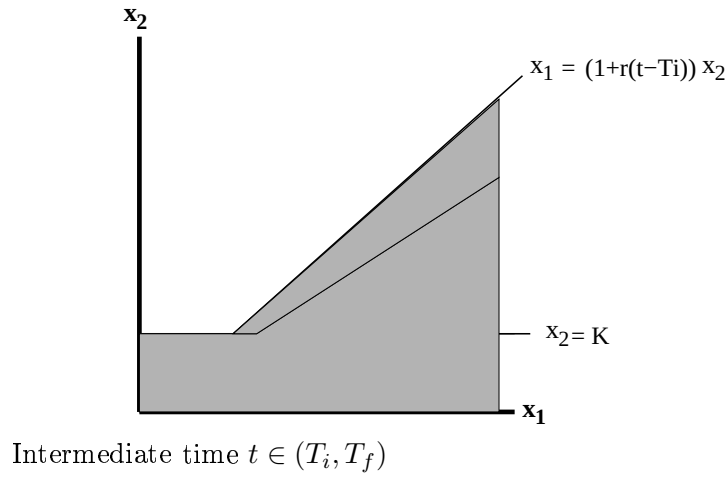
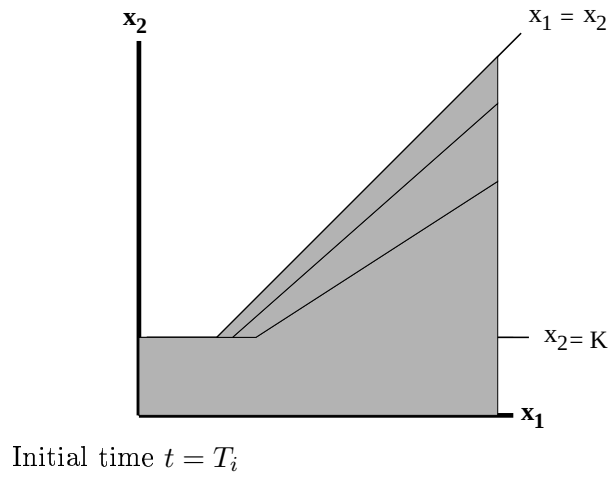


Figure 2.3: Spatial domain $\Omega = (0, \infty) \times (0, \infty)$ for the Amerasian call option pricing problem depending on the running time t derived from Proposition 1.8. The region where it is optimal to hold the options is colored.

Chapter 3

Semi-Lagrangian time discretization of convection-diffusion-reaction equations

3.1 Introduction

The linear convection-diffusion-reaction equation arises in mathematical modelling of many important problems from different fields of engineering and applied sciences, such as heat transfer, fluid mechanics, finance, etc (see [81], for example). Nonlinear versions of the equations are commonly used in more complex problems: multi-phase flows, aluminium industry, lubrication, glaciology, etc.

In many cases the diffusive term is smaller than the convective one, giving rise to the so called convection dominated problems (see the review paper [47]). Furthermore, in some convection dominated cases, the diffusive term becomes degenerated. This is the case of some financial models, where we can find weak degeneration (as in the Black-Scholes equation formulated in Appendix B, with the diffusion vanishing at the edges), but also strong degeneration of ultraparabolic type (as in the path dependent options studied in Section 2.3).

The almost hyperbolic nature of convection-dominated problems makes its numerical solution a more complex task than the solution of fully elliptic or parabolic equations. This is partly due to the fact that the solution of linear hyperbolic problems will only be as smooth as the initial solution. Moreover, the natural dissipation contained in the parabolic partial differential equations helps to make the numerical schemes stable. In this framework, a possible strategy is provided by a classical method for hyperbolic problems: the method of characteristics for time discretization (see, for instance, [89]). This approach is included in the more general family of *upwinding methods* which take into account the local flow direction. In particular, it is based on the discretization of the total (or material) derivative, i.e., the time derivative along the characteristics lines of the convective part of the equation, by a finite difference scheme. We use schemes based on integrating transport and transport-diffusion equations by tracing backwards the position at time t_n of the characteristics and evaluate the quantities at these point by an interpolation procedure. This method, first described in [45, 88], is known as *modified method of characteristics* or *semi-Lagrangian method*. We refer to it also by *characteristics method* and, when combined with finite elements, by *characteristics finite elements* or *Lagrange-Galerkin method*. It has been mathematically analyzed and applied to different problems by many authors. In general, they remark:

- Unconditional stability of the method, even when applied to the transport equation.
- In problems with large convection, the solution changes less rapidly along the characteristic lines, thus, the characteristics method allows for larger time steps without losing accuracy.
- Since the convection operator is treated explicitly, it yields symmetrical systems.
- One weak point is on the preservation of local and global mass balance. A variant of the method that preserves the conservation laws at a minor additional computational cost is proposed in [44], but we do not consider it in this work.

The (classical) semi-Lagrangian or characteristics method, introduced at the beginning of the eighties, is first order accurate in time. It has been applied to convection-diffusion equations combined with finite elements [45, 88], finite differences [45], spectral Galerkin [4], etc. Its adaptation to steady state convection-diffusion equations has been proposed in [20] and, more recently, the combination of the classical first order scheme with discontinuous Galerkin methods has been used to solve first-order stationary hyperbolic equations in [6, 5, 7]. Furthermore, it has been also applied to the Navier-Stokes equations in [88, 102].

We want to achieve better time error accuracy, i.e., we are interested in higher order characteristics methods. The increase in the order of time approximation can be achieved by using higher order schemes for the discretization of the material derivative. In [46] multi-step Lagrange-Galerkin methods for (time independent coefficients) convection-diffusion problems are studied and the need of analyzing the variable coefficient case is pointed out. Also, in [30, 31] multi-step methods to approximate the material time derivative, combined with either mixed finite elements or spectral methods for spatial discretization, are proposed to solve incompressible Navier-Stokes equations. Stability is proved and optimal error estimates for the fully discretized problem are obtained. The performance of this method is also shown in [53] when applied to wind engineering. In [99], a second order Crank-Nicholson characteristics method is proposed for solving a constant coefficient convection-diffusion equation with Dirichlet boundary conditions. The idea of this method was given in [16], but it failed in the adequate upwinding of the diffusive term.

Moreover, we are interested in methods which work well when the diffusion is degenerate. In [63], first order characteristics methods are applied to a class of nonlinear degenerate convection-diffusion problems, where the degeneration comes from the coefficient of the accumulation term. In their general setting they only prove convergence, without any rate. The same scheme is extended to the numerical solution of variational inequalities in [64, 65]. Another application to degenerate equations can be seen in [34] an error analysis for characteristics based methods applied to advection (nonlinear degenerate) diffusion equations. The analysis is developed under minimal regularity assumptions on the solution. In this context (assuming minimal regularity of the solution), the analysis of higher order spectral Lagrange-Galerkin schema in the framework of viscosity solutions and applied to the advection equation has been developed in [48]. In [113] numerical results, extended to convection-diffusion and Navier-Stokes equations, validate their error estimates.

In this chapter, after having formulated the Cauchy problem in a bounded domain, we first study the characteristic lines associated to a velocity field and some approximate schemes to compute them. The mathematical formalism of continuum mechanics (see for instance [58]) is used to express the results and notations related not only to the approximated characteristics proposed in [99] but also to the exact ones. Indeed, by using Taylor expansions we rigorously justify some approximations of the characteristics, velocity gradients and their determinants and inverses.

Secondly, we give an appropriate variational formulation of the problem in terms of the characteristic lines and propose, for solving it, two second order semidiscretized schemes: a Crank-Nicholson and a two-step characteristics method, jointly with the (first order) classical one.

Thirdly, we analyze the second order Crank-Nicholson semidiscretized scheme. Results of [99] are extended in some aspects: firstly, we deal with a (possibly degenerated) variable coefficient diffusive term instead of the more classical Laplacian one. Secondly, non divergence free velocity field and nonzero reaction function are allowed. Thirdly, a general mixed Dirichlet-Robin boundary condition is considered. Our analysis is developed assuming enough regularity on the solution of the problem.

Fourthly, we recall similar results for other semi-Lagrangian methods.

3.2 Cauchy problem and notations

Let Ω be a bounded domain in $\mathbb{R}^m$ ($m = 2, 3$) with Lipschitz boundary, Γ , divided into two parts: $\Gamma = \Gamma_D \cup \Gamma_R$, with $\Gamma_D \cap \Gamma_R = \emptyset$. Let T be a positive constant. We consider the following initial boundary value problem.

Strong problem: Find a function $\phi : \Omega \times (0, T) \longrightarrow \mathbb{R}$ such that

$$\frac{\partial \phi}{\partial t}(\mathbf{x}, t) - \text{Div}(\mathbf{A}(\mathbf{x})\nabla \phi(\mathbf{x}, t)) + \mathbf{v}(\mathbf{x}, t) \cdot \nabla \phi(\mathbf{x}, t) + l(\mathbf{x})\phi(\mathbf{x}, t) = f(\mathbf{x}, t), \quad (3.1)$$

for $(\mathbf{x}, t) \in \Omega \times (0, T)$, subject to the boundary conditions

$$\phi(\mathbf{x}, t) = 0 \text{ on } \Gamma_D \times (0, T), \quad (3.2)$$

$$\alpha \phi(\mathbf{x}, t) + \mathbf{A}(\mathbf{x})\nabla \phi(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}) = g(\mathbf{x}, t) \text{ on } \Gamma_R \times (0, T), \quad (3.3)$$

and to the initial condition

$$\phi(\mathbf{x}, 0) = \phi^0(\mathbf{x}) \text{ in } \Omega. \quad (3.4)$$

In the above equations, $\mathbf{A} : \overline{\Omega} \longrightarrow \mathcal{S}_d$ denotes the diffusion matrix function where $\mathcal{S}_d$ is the space of symmetric $m \times m$ matrices, $\mathbf{v} : \overline{\Omega} \times [0, T] \longrightarrow \mathbb{R}^m$ is the convection vector field, $l : \overline{\Omega} \longrightarrow \mathbb{R}$ is the reaction function, $f : \overline{\Omega} \times [0, T] \longrightarrow \mathbb{R}$ and $g : \Gamma_R \times [0, T] \longrightarrow \mathbb{R}$ are given scalar functions, and $\mathbf{n}$ is the outward unit normal vector to Γ .

Let us introduce the Lebesgue spaces $L^p(\Omega)$ and the Sobolev spaces $W^{n,p}(\Omega)$, for $p = 1, 2, \dots, \infty$, n being an integer. For the particular case $p = 2$, we consider the Hilbert space $L^2(\Omega)$ with the usual inner product $\langle \cdot, \cdot \rangle$ which induces the norm $\|\cdot\|_0$ and the spaces $H^n(\Omega) = W^{n,2}(\Omega)$ equipped with the usual norms $\|\cdot\|_n$ (see [1] for details). Moreover, we denote by $H_{\Gamma_D}^1(\Omega)$ the closed subspace of $H^1(\Omega)$ defined by

$$H_{\Gamma_D}^1(\Omega) := \{\varphi \in H^1(\Omega), \varphi|_{\Gamma_D} = 0\}.$$

If X denotes a Banach space and n is an integer, spaces $C^n([0, T], X)$ and $H^n((0, T), X)$ will be shortly denoted by $C^n(X)$ and $H^n(X)$, respectively. They are endowed with the norms

$$\|\varphi\|_{C^n(X)} := \max_{t \in [0, T]} \left\{ \max_{j=0, \dots, n} \|\varphi^{(j)}(t)\|_X \right\}, \quad \|\varphi\|_{H^n(X)} := \left(\int_0^T \sum_{j=0}^n \|\varphi^{(j)}(t)\|_X^2 dt \right)^{\frac{1}{2}}.$$

In the above definitions, $\varphi^{(j)}$ denotes the j -th derivative of φ with respect to time. Moreover, for a non-negative integer n we introduce the Banach space

$$Z^n = \{\varphi \in C^j(H^{n-j}(\Omega)); 0 \leq j \leq n\},$$

equipped with the norm

$$\|\varphi\|_{Z^n} := \max \left\{ \|\varphi\|_{C^j(H^{n-j})}; 0 \leq j \leq n \right\}.$$

Similar spaces are considered on the boundary Γ_R . In the sequel, when the set is different from Ω , it will be specified in inner products and norms. For instance, we will use the notations $\|\cdot\|_{n,\Gamma_R}$, $\|\cdot\|_{Z^n,\Gamma_R}$, etc.

Finally, vector-valued space functions will be distinguished by “sans serif” fonts, i.e., $C^n(\Omega)$, $H^n(\Omega)$ and

$$Z^n = \{\varphi \in C^j(H^{n-j}(\Omega)); 0 \leq j \leq n\};$$

and matrix-valued function spaces will be distinguished by bold fonts, i.e., $\mathbf{C}^n(\Omega)$, $\mathbf{H}^n(\Omega)$ and

$$\mathbf{Z}^n = \{\varphi \in C^j(\mathbf{H}^{n-j}(\Omega)); 0 \leq j \leq n\}.$$

3.3 Characteristic curves

In this section, we first define the characteristic lines associated with vector field $\mathbf{v}$ and, using the classical theory of ordinary differential equations, we recall some properties satisfied by them that will be used in the sequel. Secondly, we consider some approximation schemes of these curves, and obtain similar results.

Thus, for given $(\mathbf{x}, t) \in \overline{\Omega} \times [0, T]$, the characteristic line through $(\mathbf{x}, t)$ is the vector function

$$\begin{aligned} X_e(\mathbf{x}, t; \cdot) : (0, T) &\longrightarrow \mathbb{R}^m \\ \tau &\longrightarrow X_e(\mathbf{x}, t; \tau), \end{aligned}$$

solving the initial value problem

$$\begin{cases} \frac{\partial X_e}{\partial \tau}(\mathbf{x}, t; \tau) = \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau), \\ X_e(\mathbf{x}, t; t) = \mathbf{x}. \end{cases} \quad (3.5)$$

It represents the trajectory described by a material point that is placed at position $\mathbf{x}$ at time t and is driven by the velocity field $\mathbf{v}$. Provided that $\mathbf{v}$ is continuous, we have existence of solution of (3.5) and the following integral representation

$$X_e(\mathbf{x}, t; \tau) = \mathbf{x} + \int_t^\tau \mathbf{v}(X_e(\mathbf{x}, t; s), s) ds,$$

for τ belonging to a neighborhood of t . If $\mathbf{v} \in C^0(C(\overline{\Omega}))$ and it is Lipschitz continuous with respect to the first variable, then there exists a maximal neighborhood of t , $U_{\mathbf{x},t}$, open in $[0, T]$, such that the characteristic line,

$$X_e(\mathbf{x}, t; \cdot) : U_{\mathbf{x},t} \subset [0, T] \longrightarrow \overline{\Omega},$$

solving (3.5) exists, and it is unique and C^1 with respect to τ in $U_{\mathbf{x},t}$ (see, for instance, [2]). Next, in order to extend the existence result to the entire interval $[0, T]$, more hypotheses on the velocity field are required.

Proposition 3.1 *Let the velocity field, $\mathbf{v} \in C^0(C^0(\overline{\Omega}))$, be Lipschitz continuous with respect to the spatial variable and vanish on Γ . Then,*

$$U_{\mathbf{x},t} = [0, T] \quad \forall (\mathbf{x}, t) \in \overline{\Omega} \times [0, T].$$

PROOF. Let us fix $\mathbf{x} \in \Omega$ and $t \in (0, T]$. We want to prove that the number

$$\tilde{\tau} = \inf_{\tau \in [0, T]} \{ \tau : \exists X_e(\mathbf{x}, t; s) \text{ for } s \in [\tau, t] \}, \quad (3.6)$$

is equal to zero.

Firstly, since $U_{\mathbf{x},t}$ is nonempty, the infimum exists. Now, let us consider a sequence $\{\tau_n\}_n$, with $t \geq \tau_n > \tilde{\tau}$, converging to $\tilde{\tau}$. Let $\mathbf{y} \in \overline{\Omega}$ be an accumulation point of the corresponding sequence $\{X_e(\mathbf{x}, t; \tau_n)\}_n$. We notice that $\mathbf{y}$ exists because $\overline{\Omega}$ is compact.

Now, we claim that, if $\tilde{\tau} > 0$, then $\mathbf{y} \in \Gamma$. Indeed, otherwise $\mathbf{y} \in \Omega$ and we could define a new initial value problem with initial condition

$$X_e(\mathbf{y}, \tilde{\tau}; \tilde{\tau}) = \mathbf{y}.$$

By applying to this new problem the local existence theorem, we could extend the interval of definition of $X_e(\mathbf{x}, t; \tau)$ for $\tau < \tilde{\tau}$, which is a contradiction with (3.6). At this point, taking into account that the velocity field vanishes on the boundary, we extend the solution for times $\tau < \tilde{\tau}$ as follows:

$$X_e(\mathbf{x}, t; \tau) = \mathbf{y}, \text{ for } \tau \in [0, \tilde{\tau}].$$

Thus, we have a continuous solution that is well defined on $[0, t]$.

An analogous argument leads to have a well defined solution on $[t, T]$. The combination of both cases implies that $U_{\mathbf{x},t} = [0, T]$ for all $\mathbf{x} \in \Omega$ and $t \in [0, T]$.

Finally, if $\mathbf{x} \in \Gamma$, then $X_e(\mathbf{x}, t; \tau) = \mathbf{x}$ for all $\tau \in [0, T]$ and the result follows. $\square$

Throughout this section we assume that the hypotheses on $\mathbf{v}$ in Proposition 3.1 hold and we only specify the regularity of the velocity if more smoothness is required. Under these assumptions, the characteristic line solving (3.5) is well defined in the whole domain $[0, T]$,

$$X_e(\mathbf{x}, t; \cdot) : \tau \in [0, T] \longrightarrow \overline{\Omega},$$

and it is unique for each initial condition $(\mathbf{x}, t)$. Moreover, considered as a function of $(\mathbf{x}, t; \tau)$, it is Lipschitz continuous in $\Omega \times [0, T] \times [0, T]$ (see [2]). The following properties can be easily deduced (see [95]):

- If $\tau_1, \tau_2, \tau_3 \in [0, T]$ and $\mathbf{x} \in \Omega$ then

$$X_e(\mathbf{x}, \tau_1; \tau_3) = X_e(X_e(\mathbf{x}, \tau_1; \tau_2), \tau_2; \tau_3). \quad (3.7)$$

- For fixed $\tau_1, \tau_2 \in [0, T]$ and $\mathbf{x} \in \Omega$, we have

$$X_e(X_e(\mathbf{x}, \tau_1; \tau_2), \tau_2; \tau_1) = \mathbf{x}. \quad (3.8)$$

Regarding Figure 3.1, the previous properties mean:

$$\begin{aligned} \mathbf{y} &= X_e(\mathbf{x}, \tau_1; \tau_2), \\ \mathbf{x} &= X_e(\mathbf{y}, \tau_2; \tau_1), \\ \mathbf{z} &= X_e(\mathbf{x}, \tau_1; \tau_3) = X_e(\mathbf{y}, \tau_2; \tau_3). \end{aligned}$$

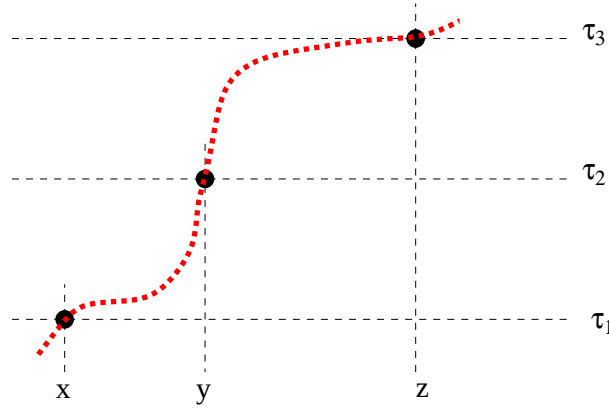


Figure 3.1: Characteristic line for three different times.

Corollary 3.1 *For $\tau_1, \tau_2 \in [0, T]$, the mapping*

$$X_e(\cdot, \tau_1; \tau_2) : \overline{\Omega} \longrightarrow \overline{\Omega}$$

is one-to-one, with inverse $X_e(\cdot, \tau_2; \tau_1)$.

We denote by $\mathbf{F}_e$ (respectively by $\mathbf{L}$) the gradient of X_e (respectively, of $\mathbf{v}$) with respect to the space variable $\mathbf{x}$, i.e.,

$$[F_e]_{rs}(\mathbf{x}, t; \tau) := \frac{\partial [X_e]_r}{\partial x_s}(\mathbf{x}, t; \tau), \quad [L]_{rs}(\mathbf{x}, t) := \frac{\partial v_r}{\partial x_s}(\mathbf{x}, t), \quad 1 \leq r, s \leq m,$$

assuming they exist.

If we differentiate equations (3.5) with respect to $\mathbf{x}$ and interchange these derivatives with $\partial/\partial\tau$, we can (formally) obtain the following initial value problem for $\mathbf{F}_e$ (see also [58] page 63):

$$\begin{cases} \frac{\partial \mathbf{F}_e}{\partial \tau}(\mathbf{x}, t; \tau) &= \mathbf{L}(X_e(\mathbf{x}, t; \tau), \tau) \mathbf{F}_e(\mathbf{x}, t; \tau), \\ \mathbf{F}_e(\mathbf{x}, t; t) &= \mathbf{I}. \end{cases} \quad (3.9)$$

Remark 3.1 *For the sake of clearness of the notation, in the expressions involving gradients and time derivatives we introduce the notation (see [58]):*

- If φ is a “material field” in the sense that $\varphi = \varphi(\mathbf{x}, t)$, we denote by $\nabla \varphi$ (respectively, by $\text{Div } \varphi$) the gradient (respectively, the divergence) with respect to the first argument.
- If ψ is a “spatial field” in the sense that $\psi = \psi(X_e(\mathbf{x}, t; \tau), \tau)$, we denote by $\text{grad } \psi$ (respectively, by $\text{div } \psi$) the gradient (respectively, the divergence) with respect to the first argument, and by ψ' the partial derivative with respect to the second argument (time). Moreover, the expression $\dot{\psi}$ denotes the total or material derivative of ψ , i.e.,

$$\begin{aligned} \dot{\psi}(X_e(\mathbf{x}, t; \tau), \tau) &:= \frac{\partial}{\partial \tau} (\psi(X_e(\mathbf{x}, t; \tau), \tau)) \\ &= \psi'(X_e(\mathbf{x}, t; \tau), \tau) + \text{grad } \psi(X_e(\mathbf{x}, t; \tau), \tau) \cdot \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau), \end{aligned} \quad (3.10)$$

where the chain rule and the definition of the characteristic lines have been used. Analogously, for a vector field $\mathbf{w} = \mathbf{w}(X_e(\mathbf{x}, t; \tau), \tau)$ we define

$$\begin{aligned}\dot{\mathbf{w}}(X_e(\mathbf{x}, t; \tau), \tau) &:= \frac{\partial}{\partial \tau} (\mathbf{w}(X_e(\mathbf{x}, t; \tau), \tau)) \\ &= \mathbf{w}'(X_e(\mathbf{x}, t; \tau), \tau) + \text{grad } \mathbf{w}(X_e(\mathbf{x}, t; \tau), \tau) \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau),\end{aligned}\tag{3.11}$$

Following [58] we will consider that $\mathbf{v}$ and $\mathbf{L}$ are spatial fields. Notice that, with the above notation,

$$\dot{\mathbf{v}}(X_e(\mathbf{x}, t; \tau), \tau) = (\mathbf{v}' + \mathbf{L}\mathbf{v})(X_e(\mathbf{x}, t; \tau), \tau).$$

Finally, for fields X_e and $\mathbf{F}_e$ depending on $(\mathbf{x}, t; \tau)$ we use, respectively, notations ∇ and Div for the gradient and the divergence operators with respect to $\mathbf{x}$.

Remark 3.2 If $g \in C^0(C^1(\Omega))$ and $Y : [0, T] \rightarrow \Omega$ is any solution of the differential equation $dY/d\tau = g(Y, \tau)$, then the linear differential equation

$$\frac{dZ}{d\tau}(\tau) = \text{grad } g(Y(\tau), \tau)Z(\tau)$$

is called the variational equation along the solution $Y(\cdot)$. Thus, (3.9) is the variational equation along $X_e(\mathbf{x}, t; \cdot)$.

Proposition 3.2 If $\mathbf{v} \in C^0(C^1(\overline{\Omega}))$, then $X_e \in C^1(\overline{\Omega} \times [0, T] \times [0, T])$ and $\mathbf{F}_e$ is the solution of problem (3.9). Moreover, $\partial \mathbf{F}_e / \partial \tau$ exists and it is continuous.

PROOF. See, for instance, [2]. □

Proposition 3.3 If $\mathbf{v} \in C^0(C^1(\overline{\Omega}))$, then

$$\|\mathbf{F}_e(\mathbf{x}, t; \tau)\| \leq e^{\|\mathbf{v}\|_{C^0(C^1(\overline{\Omega}))}|\tau-t|},\tag{3.12}$$

for all $\mathbf{x} \in \Omega$, $t, \tau \in [0, T]$.

PROOF. Firstly, by using Barrow's rule and (3.9), we have

$$\mathbf{F}_e(\mathbf{x}, t; \tau) = \mathbf{I} + \int_t^\tau \mathbf{L}(X_e(\mathbf{x}, t; s), s) \mathbf{F}_e(\mathbf{x}, t; s) ds.$$

Next, by applying norms, the following inequality is obtained:

$$\|\mathbf{F}_e(\mathbf{x}, t; \tau)\| \leq 1 + \left| \int_t^\tau \|\mathbf{L}(X_e(\mathbf{x}, t; s), s)\| \|\mathbf{F}_e(\mathbf{x}, t; s)\| ds \right|.$$

Finally, (3.12) is deduced by the continuous Gronwall's Lemma (see, for instance, [2]). □

Proposition 3.4 Assume $\mathbf{v} \in C^0(C^1(\overline{\Omega}))$, then

$$\det \mathbf{F}_e(\mathbf{x}, t; \tau) > 0.\tag{3.13}$$

PROOF. We first note that, by applying the *chain rule* to (3.8), we have

$$\det \mathbf{F}_e(\mathbf{x}, t; \tau) \neq 0 \quad \forall (\mathbf{x}, t; \tau) \in \Omega \times [0, T] \times [0, T].$$

Then, (3.13) follows from the continuity of $\det$ and $\mathbf{F}_e$ with respect to τ , and the fact that

$$\det \mathbf{F}_e(\mathbf{x}, t; t) = 1.$$

□

There exist analogous results to Proposition 3.2 for higher order derivatives with respect to time and space. Let us recall a result for higher order space derivatives.

Proposition 3.5 *If $\mathbf{v} \in C^0(C^n(\overline{\Omega}))$ for $n \geq 1$ an integer, then $X_e \in C^0(\overline{\Omega} \times [0, T] \times [0, T])$ and it is C^n with respect to the $\mathbf{x}$ variable.*

PROOF. See, for instance, [2].

□

Remark 3.3 *Notice that, if $\mathbf{v}$ is regular enough (e.g. $\mathbf{v} \in C^2(C^3(\overline{\Omega}))$), then, for each $t \in [0, T]$, the mapping*

$$X_e(\cdot, t; \cdot) : (\mathbf{x}; \tau) \in \Omega \times [0, T] \longrightarrow \Omega$$

defines a motion in the sense of [58].

It will be useful to compute a second order approximation of matrices $\mathbf{F}_e$ and $\mathbf{F}_e^{-1}$. For this, let us first (formally) differentiate the first equation of (3.9) with respect to τ . We obtain a second order initial value problem for $\mathbf{F}_e$ (see also [58] page 63),

$$\begin{cases} \frac{\partial^2 \mathbf{F}_e}{\partial \tau^2}(\mathbf{x}, t; \tau) &= \text{grad } \dot{\mathbf{v}}(X_e(\mathbf{x}, t; \tau), \tau) \mathbf{F}_e(\mathbf{x}, t; \tau), \\ \frac{\partial \mathbf{F}_e}{\partial \tau}(\mathbf{x}, t; t) &= \mathbf{L}(\mathbf{x}, t), \\ \mathbf{F}_e(\mathbf{x}, t; t) &= \mathbf{I}. \end{cases} \quad (3.14)$$

Proposition 3.6 *If $\mathbf{v} \in C^0(C^2(\overline{\Omega})) \cap C^1(C^1(\overline{\Omega}))$, then $\mathbf{F}_e$ is the solution of (3.14) and $\partial^2 \mathbf{F}_e / \partial \tau^2 \in C^0(\overline{\Omega} \times [0, T] \times [0, T])$.*

PROOF. See, for instance, [2].

□

Proposition 3.7 *If $\mathbf{v} \in C^0(C^2(\overline{\Omega})) \cap C^1(C^1(\overline{\Omega}))$, then $\mathbf{F}_e$ satisfies*

$$\mathbf{F}_e(\mathbf{x}, t; s) = \mathbf{I} + (s - t) \mathbf{L}(\mathbf{x}, t) + \int_s^t (\tau - s) \text{grad } \dot{\mathbf{v}}(X_e(\mathbf{x}, t; \tau), \tau) \mathbf{F}_e(\mathbf{x}, t; \tau) d\tau, \quad (3.15)$$

and its inverse, $\mathbf{F}_e^{-1}$, satisfies:

$$\begin{aligned} \mathbf{F}_e^{-1}(\mathbf{x}, t; s) &= \mathbf{I} + (t - s) \mathbf{L}(X_e(\mathbf{x}, t; s), s) \\ &\quad - \int_s^t (\tau - t) \text{grad } \dot{\mathbf{v}}(X_e(\mathbf{x}, t; \tau), \tau) \mathbf{F}_e(X_e(\mathbf{x}, t; s), s; \tau) d\tau. \end{aligned} \quad (3.16)$$

PROOF. Expression (3.15) comes from the Taylor expansion of $\mathbf{F}_e(\mathbf{x}, t; s)$ in variable s around $(\mathbf{x}, t; t)$, using (3.14).

In order to prove (3.16) we first differentiate (3.8) for $\tau_1 = s$ and $\tau_2 = t$ to obtain

$$\mathbf{F}_e^{-1}(\mathbf{x}, t; s) = \mathbf{F}_e(X_e(\mathbf{x}, t; s), s; t).$$

Then, we apply (3.15) and take the equality $X_e(X_e(\mathbf{x}, t; s), s; \tau) = X_e(\mathbf{x}, t; \tau)$ into account. $\square$

Remark 3.4 If $\mathbf{v} \in C^0(C^2(\Omega))$ then $\partial \mathbf{F}_e / \partial x_i$ exists and it is continuous; moreover, we can take the derivative of the first equation of (3.9) with respect to x_i obtaining

$$\frac{\partial^2 \mathbf{F}_e}{\partial x_i \partial \tau} = \left(\sum_{j=1}^m \frac{\partial \mathbf{L}}{\partial x_j}(X_e(\mathbf{x}, t; \tau), \tau) [\mathbf{F}_e]_{ji}(\mathbf{x}, t; \tau) \right) \mathbf{F}_e(\mathbf{x}, t; \tau) + \mathbf{L}(X_e(\mathbf{x}, t; \tau), \tau) \frac{\partial \mathbf{F}_e}{\partial x_i}(\mathbf{x}, t; \tau).$$

If we do the same with the first equation of (3.14) we get

$$\begin{aligned} \frac{\partial^3 \mathbf{F}_e}{\partial x_i \partial \tau^2}(\mathbf{x}, t; \tau) &= \left(\sum_{j=1}^m \frac{\partial}{\partial x_j} \text{grad } \dot{\mathbf{v}}(X_e(\mathbf{x}, t; \tau), \tau) [\mathbf{F}_e]_{ji}(\mathbf{x}, t; \tau) \right) \mathbf{F}_e(\mathbf{x}, t; \tau) \\ &\quad + \text{grad } \dot{\mathbf{v}}(X_e(\mathbf{x}, t; \tau), \tau) \frac{\partial \mathbf{F}_e}{\partial x_i}(\mathbf{x}, t; \tau), \end{aligned}$$

which is continuous for $\mathbf{v} \in C^0(C^3(\Omega)) \cap C^1(C^2(\Omega))$.

Now, we develop analogous computations for $\det \mathbf{F}_e^{-1}$. To do that, we first use Liouville's Theorem (see [2]), the *chain rule*, and the first equation of (3.9), obtaining the initial value problem (see [58] page 77)

$$\begin{cases} \frac{\partial}{\partial \tau} \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) &= -\det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) \operatorname{div} \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau), \\ \det \mathbf{F}_e^{-1}(\mathbf{x}, t; t) &= 1. \end{cases} \quad (3.17)$$

Proposition 3.8 If $\mathbf{v} \in C^0(C^1(\overline{\Omega}))$, then

$$\det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) \leq e^{\|\mathbf{v}\|_{C^0(C^1(\overline{\Omega}))} |\tau - t|}, \quad (3.18)$$

for all $\mathbf{x} \in \Omega$, $t, \tau \in [0, T]$.

PROOF. The solution of (3.17) is

$$\det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) = e^{-\int_t^\tau \operatorname{div} \mathbf{v}(X_e(\mathbf{x}, t; s), s) ds},$$

from which the result follows. $\square$

Moreover, we can obtain a second order approximation of $\det \mathbf{F}_e^{-1}$ for smoother $\mathbf{v}$.

Proposition 3.9 If $\mathbf{v} \in C^0(C^2(\Omega)) \cap C^1(C^1(\overline{\Omega}))$, then $\det \mathbf{F}_e^{-1}$ satisfies

$$\begin{aligned} \det \mathbf{F}_e^{-1}(\mathbf{x}, t; s) &= 1 - (s - t) \operatorname{div} \mathbf{v}(\mathbf{x}, t) \\ &\quad + \int_s^t (\tau - s) \frac{\partial^2}{\partial \tau^2} \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) d\tau. \end{aligned} \quad (3.19)$$

PROOF. It is an immediate consequence of Taylor expansion of $\det \mathbf{F}_e^{-1}(\mathbf{x}, t; s)$ around $s = t$.
 $\square$

By taking now the derivative of (3.17) with respect to τ and replacing (3.17) we obtain

$$\begin{aligned} \frac{\partial^2}{\partial \tau^2} \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) &= \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) (\operatorname{div} \mathbf{v})^2(X_e(\mathbf{x}, t; \tau), \tau) \\ &\quad - \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) \frac{\partial}{\partial \tau} (\operatorname{div} \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau)). \end{aligned}$$

Moreover, by using the definition of the trace operator and developing the derivative of a product we get

$$\begin{aligned} \frac{\partial}{\partial \tau} (\operatorname{div} \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau)) &= \frac{\partial}{\partial \tau} \operatorname{tr} \left(\frac{\partial \mathbf{F}_e}{\partial \tau} \mathbf{F}_e^{-1} \right) (\mathbf{x}, t; \tau) \\ &= \operatorname{tr} \left(\frac{\partial^2 \mathbf{F}_e}{\partial \tau^2} \mathbf{F}_e^{-1} - \frac{\partial \mathbf{F}_e}{\partial \tau} \mathbf{F}_e^{-1} \frac{\partial \mathbf{F}_e}{\partial \tau} \mathbf{F}_e^{-1} \right) (\mathbf{x}, t; \tau) \\ &= \operatorname{tr} (\operatorname{grad} \dot{\mathbf{v}} + \mathbf{L} \mathbf{L}) (\mathbf{x}, t; \tau), \end{aligned}$$

where the first equations of (3.9) and (3.14) have been used in the last line. Thus, by replacing $\mathbf{F}_e(\mathbf{x}, t; t) = \mathbf{I}$, the following second order initial value problem is obtained

$$\begin{cases} \frac{\partial^2}{\partial \tau^2} \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) = -\det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) (-(\operatorname{div} \mathbf{v})^2 + \operatorname{div} \dot{\mathbf{v}} - \mathbf{L} \cdot \mathbf{L}^T) (X_e(\mathbf{x}, t; \tau), \tau), \\ \frac{\partial}{\partial \tau} \det \mathbf{F}_e^{-1}(\mathbf{x}, t; t) = -\operatorname{div} \mathbf{v}(\mathbf{x}, t), \\ \det \mathbf{F}_e^{-1}(\mathbf{x}, t; t) = 1. \end{cases} \quad (3.20)$$

3.3.1 Approximate characteristic lines

Now we introduce the number of time steps, N , the time step, $\Delta t = T/N$, and the mesh-points, $t_n = n\Delta t$ for $n = 0, 1/2, 1, 3/2, \dots, N$. In what follows, we use the notation $\psi^n(\mathbf{x}) := \psi(\mathbf{x}, t_n)$ for a function $\psi(\mathbf{x}, t)$. Moreover, for $n = 0, 1, 2, \dots$, we define

$$X_e^n(\mathbf{x}) := X_e(\mathbf{x}, t_{n+1}; t_n), \quad \mathbf{F}_e^n(\mathbf{x}) := \mathbf{F}_e(\mathbf{x}, t_{n+1}; t_n),$$

We recall that $X_e(\mathbf{x}, t_{n+1}; \tau)$ is the unique solution of the Cauchy's problem

$$\begin{cases} \frac{dX_e}{d\tau}(\mathbf{x}, t_{n+1}; \tau) = \mathbf{v}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau), \\ X_e(\mathbf{x}, t_{n+1}; t_{n+1}) = \mathbf{x}. \end{cases} \quad (3.21)$$

In most cases the solution of (3.21) is not easy to compute analytically. Instead, we consider the following numerical schemes to approximate $X_e^n(\mathbf{x})$:

- *First order explicit Euler scheme:*

$$X_E^n(\mathbf{x}) := \mathbf{x} - \Delta t \mathbf{v}^{n+1}(\mathbf{x}). \quad (3.22)$$

- *Second order explicit Runge-Kutta scheme:*

$$X_{RK}^n(\mathbf{x}) := \mathbf{x} - \Delta t \mathbf{v}^{n+\frac{1}{2}}(Y^n(\mathbf{x})), \quad (3.23)$$

where

$$Y^n(\mathbf{x}) := \mathbf{x} - \frac{\Delta t}{2} \mathbf{v}^{n+1}(\mathbf{x}). \quad (3.24)$$

- *Second order explicit two-step scheme:*

$$X_{TS}^n(\mathbf{x}) := \mathbf{x} - \Delta t (2\mathbf{v}^n(\mathbf{x}) - \mathbf{v}^{n-1}(\mathbf{x})). \quad (3.25)$$

Remark 3.5 *The above definitions are valid for $n = 0, 1, \dots, N-1$ in the first two cases, and for $n = 2, 3, \dots, N$ in the third one.*

Similar notations are used for the Jacobian matrices of mappings X_E^n , X_{RK}^n and X_{TS}^n , namely,

$$\mathbf{F}_E^n(\mathbf{x}) := \nabla X_E^n(\mathbf{x}) = \mathbf{I}(\mathbf{x}) - \Delta t \mathbf{L}^{n+1}(\mathbf{x}), \quad (3.26)$$

$$\mathbf{F}_{RK}^n(\mathbf{x}) := \nabla X_{RK}^n(\mathbf{x}) = \mathbf{I}(\mathbf{x}) - \Delta t \mathbf{L}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) \left(\mathbf{I}(\mathbf{x}) - \frac{\Delta t}{2} \mathbf{L}^{n+1}(\mathbf{x}) \right), \quad (3.27)$$

$$\mathbf{F}_{TS}^n(\mathbf{x}) := \nabla X_{TS}^n(\mathbf{x}) = \mathbf{I}(\mathbf{x}) - \Delta t (2\mathbf{L}^n(\mathbf{x}) - \mathbf{L}^{n-1}(\mathbf{x})). \quad (3.28)$$

In order to state some properties of the approximate characteristic curves similar to those satisfied by the exact characteristics, we require the velocity to satisfy the following assumption:

Hypothesis 3.1 *The velocity field $\mathbf{v} \in C^0(\mathcal{W}^{1,\infty}(\Omega))$ and satisfies $\mathbf{v} \equiv 0$ on Γ .*

Lemma 3.1 *Under Hypothesis 3.1, we have:*

$$\begin{aligned} \text{If } \|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))} \Delta t < 1 & \text{ then } X_E^n(\overline{\Omega}) = \overline{\Omega}. \\ \text{If } \|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))} \Delta t < \frac{1}{2} & \text{ then } X_{RK}^n(\overline{\Omega}) = \overline{\Omega}. \\ \text{If } \|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))} \Delta t < \frac{1}{3} & \text{ then } X_{TS}^n(\overline{\Omega}) = \overline{\Omega}. \end{aligned} \quad (3.29)$$

PROOF. Since $\mathbf{v}$ vanishes on Γ , then $X_i^n(\Gamma) = \Gamma, i = E, RK, TS$. Thus, if we prove that $X_i^n(\Omega) = \Omega$, for $i = E, RK, TS$, then result (3.29) is concluded.

- *Euler approximation.* Let $\mathbf{x} \in \Omega$, then $\text{dist}(\mathbf{x}, \Gamma) = |\mathbf{x} - \overline{\mathbf{x}}| > 0$ for some $\overline{\mathbf{x}} \in \Gamma$. Since $\mathbf{v}$ is Lipschitz continuous and vanishes on the boundary by Hypothesis 3.1, we have

$$\begin{aligned} |\mathbf{v}^{n+1}(\mathbf{x})| &= |\mathbf{v}^{n+1}(\mathbf{x}) - \mathbf{v}^{n+1}(\overline{\mathbf{x}})| \leq \|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))} |\mathbf{x} - \overline{\mathbf{x}}| \\ &= \text{dist}(\mathbf{x}, \Gamma) \|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))}. \end{aligned}$$

Thus, if $\Delta t < 1/(2\|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))}) < 1/\|\mathbf{v}\|_{C^0(\mathcal{W}^{1,\infty}(\Omega))}$, then

$$|\mathbf{x} - X_E^n(\mathbf{x})| = \Delta t |\mathbf{v}^{n+1}(\mathbf{x})| < \text{dist}(\mathbf{x}, \Gamma),$$

so we have proved that $X_E^n(\Omega) \subset \Omega$. In order to prove the reverse inclusion, we show that for $\mathbf{y} \in \Omega$ there exists $\mathbf{x} \in \Omega$ satisfying

$$\mathbf{y} - (\mathbf{x} - \Delta t \mathbf{v}^{n+1}(\mathbf{x})) = 0.$$

This follows from applying the implicit function theorem to the above equation by taking into account that, for $\Delta t < 1/\|\mathbf{v}\|_{C^0(\mathbf{W}^{1,\infty}(\Omega))}$, the Jacobian with respect to $\mathbf{x}$,

$$-(\mathbf{I} - \Delta t \mathbf{L}^{n+1}(\mathbf{x})),$$

is non singular.

- *Runge-Kutta approximation.* From the previous result for the Euler approximation, we deduce that $Y^n(\Omega) \subset \Omega$, with Y^n defined in (3.24). Indeed,

$$|\mathbf{x} - Y^n(\mathbf{x})| = \frac{1}{2}|\mathbf{x} - X_E^n(\mathbf{x})| < \frac{1}{2}|\mathbf{x} - \bar{\mathbf{x}}|.$$

Thus, we also have $X_{RK}^n(\Omega) \subset \Omega$ because

$$\begin{aligned} |\mathbf{x} - X_{RK}^n(\mathbf{x})| &= \Delta t |\mathbf{v}^{n+\frac{1}{2}}(Y^n(\mathbf{x}))| = \Delta t |\mathbf{v}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) - \mathbf{v}^{n+\frac{1}{2}}(\bar{\mathbf{x}})| \\ &\leq \Delta t \|\mathbf{v}\|_{C^0(\mathbf{W}^{1,\infty}(\Omega))} |Y^n(\mathbf{x}) - \bar{\mathbf{x}}| \\ &\leq \Delta t \|\mathbf{v}\|_{C^0(\mathbf{W}^{1,\infty}(\Omega))} \frac{3}{2} |\mathbf{x} - \bar{\mathbf{x}}| < |\mathbf{x} - \bar{\mathbf{x}}|. \end{aligned}$$

Finally, similar to the Euler approximation, surjectivity of X_{RK}^n follows from the implicit function theorem.

- *Second order two-step approximation.* It is enough to take into account that approximation (3.25) is analogous to (3.22) for the new velocity field

$$\mathbf{w}^{n+1}(\mathbf{x}) := 2\mathbf{v}^n(\mathbf{x}) - \mathbf{v}^{n-1}(\mathbf{x}). \quad \square$$

Remark 3.6 In the following lemmas we will assume that the velocity field is smooth and the time step satisfies an estimate as follows

$$\Delta t \leq \frac{c}{\|\mathbf{v}\|_{C^0(\mathbf{W}^{1,\infty}(\Omega))}},$$

for $c = 1/3$. In fact, concerning Euler approximation only $c = 1$ is needed, while for Runge-Kutta approximation $c = 1/2$.

Lemma 3.2 Under Hypothesis 3.1, if $\|\mathbf{v}\|_{C^0(\mathbf{W}^{1,\infty}(\Omega))}\Delta t < 1/3$, then $\mathbf{F}_i^n$ are invertible for $i = E, RK, TS$. Their respective inverses satisfy

$$(\mathbf{F}_E^n)^{-1}(\mathbf{x}) = \mathbf{I}(\mathbf{x}) + \Delta t \mathbf{L}^{n+1}(\mathbf{x}) + \Delta t^2 (\mathbf{L}^{n+1}(\mathbf{x}))^2 + O(\Delta t^3), \quad (3.30)$$

$$\begin{aligned} (\mathbf{F}_{RK}^n)^{-1}(\mathbf{x}) &= \mathbf{I}(\mathbf{x}) + \Delta t \mathbf{L}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) - \frac{\Delta t^2}{2} \mathbf{L}^{n+\frac{1}{2}}(Y^n(\mathbf{x}))\mathbf{L}^{n+1}(\mathbf{x}) \\ &\quad + \Delta t^2 \left(\mathbf{L}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) \right)^2 + O(\Delta t^3), \end{aligned} \quad (3.31)$$

$$\begin{aligned} (\mathbf{F}_{TS}^n)^{-1}(\mathbf{x}) &= \mathbf{I}(\mathbf{x}) + \Delta t (2\mathbf{L}^n(\mathbf{x}) - \mathbf{L}^{n-1}(\mathbf{x})) \\ &\quad + \Delta t^2 (2\mathbf{L}^n(\mathbf{x}) - \mathbf{L}^{n-1}(\mathbf{x}))^2 + O(\Delta t^3), \quad a.e. \mathbf{x} \in \Omega. \end{aligned} \quad (3.32)$$

PROOF. By applying norms to expressions (3.26)-(3.28) it is easy to prove that

$$\|\mathbf{I} - \mathbf{F}_i^n\|_{\mathbf{L}^\infty(\Omega)} < 1 \quad \text{for } i = E, RK, TS.$$

Notice that the above bound is uniform in $\mathbf{x} \in \Omega$. Thus, they are invertible with inverses given by the series

$$(\mathbf{F}_i^n)^{-1}(\mathbf{x}) = \sum_{j=0}^{\infty} (\mathbf{I}(\mathbf{x}) - \mathbf{F}_i^n(\mathbf{x}))^j, \text{ for } i = E, RK, TS \text{ and a.e. } \mathbf{x} \in \Omega,$$

from which (3.30), (3.31) and (3.32) follow. $\square$

Corollary 3.2 *Under the assumptions of Lemma 3.2 the determinants of the inverses of Jacobian matrices satisfy*

$$\det(\mathbf{F}_E^n)^{-1}(\mathbf{x}) = 1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}) + O(\Delta t^2), \quad (3.33)$$

$$\det(\mathbf{F}_{RK}^n)^{-1}(\mathbf{x}) = 1 + \Delta t \operatorname{div} \mathbf{v}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) + O(\Delta t^2), \quad (3.34)$$

$$\det(\mathbf{F}_{TS}^n)^{-1}(\mathbf{x}) = 1 + \Delta t \operatorname{div} (2\mathbf{v}^n(\mathbf{x}) - \mathbf{v}^{n-1}(\mathbf{x})) + O(\Delta t^2), \quad (3.35)$$

a.e. $\mathbf{x} \in \Omega$. Moreover, there exists a constant c depending on $\|\mathbf{v}\|_{C^0(W^{1,\infty}(\Omega))}$ such that

$$|\det(\mathbf{F}_i^n)^{-1}(\mathbf{x})| \leq 1 + \Delta t c, \quad (3.36)$$

for $i = E, RK, TS$, a.e. $\mathbf{x} \in \Omega$.

PROOF. Firstly we have

$$\det(\mathbf{I} + \mathbf{D}) = 1 + \operatorname{tr} \mathbf{D} + \frac{1}{2} ((\operatorname{tr} \mathbf{D})^2 - \operatorname{tr} \mathbf{D}^2) + \det \mathbf{D},$$

for every matrix $\mathbf{D}$ (see, for instance, [58]). Thus, the result straightly follows by replacing, in the previous formula,

$$\begin{aligned} \mathbf{D} &= \Delta t \left(\mathbf{L}^{n+1}(\mathbf{x}) + \Delta t (\mathbf{L}^{n+1}(\mathbf{x}))^2 + \dots \right) & \text{for } i = E, \\ \mathbf{D} &= \Delta t \left(\mathbf{L}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) - \frac{\Delta t}{2} \mathbf{L}^{n+\frac{1}{2}}(Y^n(\mathbf{x})) \mathbf{L}^{n+1}(\mathbf{x}) + \dots \right) & \text{for } i = RK, \\ \mathbf{D} &= \Delta t \left(2\mathbf{L}^n(\mathbf{x}) - \mathbf{L}^{n-1}(\mathbf{x}) + \Delta t (2\mathbf{L}^n(\mathbf{x}) - \mathbf{L}^{n-1}(\mathbf{x}))^2 + \dots \right) & \text{for } i = TS. \end{aligned}$$

The bounds for the determinants directly follows from expressions (3.33), (3.34) and (3.35). $\square$

Lemma 3.3 *Under Hypothesis 3.1, if $\psi \in L^2(\Omega)$ and $\|\mathbf{v}\|_{C^0(W^{1,\infty}(\Omega))} \Delta t < 1/3$, then there exists a positive constant c such that*

$$\|\psi \circ X_i^n\|_0^2 \leq (1 + c \Delta t) \|\psi\|_0^2, \quad (3.37)$$

for $n = 0, \dots, N$ and $i = E, RK, TS$.

PROOF. By using Lemma 3.1 and the change of variable $\mathbf{y} = X_i^n(\mathbf{x})$, we have

$$\|\psi \circ X_i^n\|_0^2 = \int_{\Omega} (\psi(X_i^n(\mathbf{x})))^2 d\mathbf{x} = \int_{\Omega} \psi^2(\mathbf{y}) |\det(\mathbf{F}_i^n)^{-1}(\mathbf{y})| d\mathbf{y},$$

for $i = E, RK, TS$. Finally, by virtue of (3.36), estimate (3.37) is proved. $\square$

3.4 Weak formulation

We are going to develop some formal computations in order to give a weak formulation of problem (3.1)-(3.4). Recall first the definition of the material derivative given in (3.10) which, applied to function ϕ , gives:

$$\dot{\phi}(X_e(\mathbf{x}, t; \tau), \tau) = \phi'(X_e(\mathbf{x}, t; \tau), \tau) + \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau) \cdot \text{grad } \phi(X_e(\mathbf{x}, t; \tau), \tau). \quad (3.38)$$

Remark 3.7 *We will consider that the solution and the test functions are spatial fields and use their corresponding notations for the differential operators.*

By writing equation (3.1) at point $X_e(\mathbf{x}, t; \tau)$ and time τ , and using (3.38), we have

$$\begin{aligned} \dot{\phi}(X_e(\mathbf{x}, t; \tau), \tau) - \text{div}(\mathbf{A}(X_e(\mathbf{x}, t; \tau)) \text{grad } \phi(X_e(\mathbf{x}, t; \tau), \tau)) \\ + l(X_e(\mathbf{x}, t; \tau))\phi(X_e(\mathbf{x}, t; \tau), \tau) = f(X_e(\mathbf{x}, t; \tau), \tau). \end{aligned} \quad (3.39)$$

Before giving a weak formulation of equation (3.39), we state the following two lemmas. The first one can be considered as a Green's formula.

Lemma 3.4 *Let $X : \overline{\Omega} \longrightarrow \overline{X(\Omega)}$, $X \in C^2(\overline{\Omega})$, be an invertible vector valued function. Let $\mathbf{F} = \nabla X$ and assume that $\mathbf{F}^{-1} \in \mathbf{C}^1(\overline{\Omega})$. Then,*

$$\begin{aligned} \int_{\Omega} \text{div } \mathbf{w}(X(\mathbf{x})) \psi(\mathbf{x}) \, d\mathbf{x} = \int_{\Gamma} \mathbf{F}^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot \mathbf{w}(X(\mathbf{x})) \psi(\mathbf{x}) \, dA_{\mathbf{x}} \\ - \int_{\Omega} \mathbf{F}^{-1}(\mathbf{x}) \mathbf{w}(X(\mathbf{x})) \cdot \nabla \psi(\mathbf{x}) \, d\mathbf{x} - \int_{\Omega} \text{Div } \mathbf{F}^{-T}(\mathbf{x}) \cdot \mathbf{w}(X(\mathbf{x})) \psi(\mathbf{x}) \, d\mathbf{x}, \end{aligned} \quad (3.40)$$

being $\mathbf{w} \in H^1(X(\Omega))$ a vector valued function and $\psi \in H^1(\Omega)$ a scalar function.

PROOF. Firstly, by Gauss theorem, we have

$$\int_{\Gamma} \mathbf{F}^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot \mathbf{w}(X(\mathbf{x})) \psi(\mathbf{x}) \, dA_{\mathbf{x}} = \int_{\Omega} \text{Div}(\mathbf{F}^{-1}(\mathbf{w} \circ X)\psi)(\mathbf{x}) \, d\mathbf{x}. \quad (3.41)$$

Next we give three formulas developing the divergence term appearing in the previous integral:

$$\begin{aligned} \text{Div}(\mathbf{F}^{-1}(\mathbf{w} \circ X)\psi)(\mathbf{x}) &= \psi(\mathbf{x}) \text{Div}(\mathbf{F}^{-1}(\mathbf{w} \circ X))(\mathbf{x}) \\ &\quad + \nabla \psi(\mathbf{x}) \cdot \mathbf{F}^{-1}(\mathbf{x}) \mathbf{w}(X(\mathbf{x})), \end{aligned} \quad (3.42)$$

$$\begin{aligned} \text{Div}(\mathbf{F}^{-1}(\mathbf{w} \circ X))(\mathbf{x}) &= \mathbf{F}^{-T}(\mathbf{x}) \cdot \nabla(\mathbf{w} \circ X)(\mathbf{x}) \\ &\quad + \mathbf{w}(X(\mathbf{x})) \cdot \text{Div } \mathbf{F}^{-T}(\mathbf{x}), \end{aligned} \quad (3.43)$$

$$\begin{aligned} \mathbf{F}^{-T}(\mathbf{x}) \cdot \nabla(\mathbf{w} \circ X)(\mathbf{x}) &= \text{tr}(\nabla(\mathbf{w} \circ X)(\mathbf{x}) \mathbf{F}^{-1}(\mathbf{x})) \\ &= \text{tr}(\text{grad } \mathbf{w}(X(\mathbf{x}))) = \text{div } \mathbf{w}(X(\mathbf{x})). \end{aligned} \quad (3.44)$$

Finally, identity (3.40) is obtained by using (3.42)-(3.44) in (3.41). $\square$

Lemma 3.5 *Let $X : \overline{\Omega} \rightarrow \overline{\Omega}$, $X \in C^2(\overline{\Omega})$ be a vector valued invertible function satisfying*

$$X(\mathbf{x}) = \mathbf{x} \quad \forall \mathbf{x} \in \Gamma.$$

Let $\mathbf{F} = \nabla X$ and assume that $\mathbf{F}^{-1} \in \mathbf{C}^1(\overline{\Omega})$. Then, for $\mathbf{w} \in H^1(\Omega)$ and $\psi \in H^1(\Omega)$, we have

$$\int_{\Gamma} \mathbf{F}^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot \mathbf{w}(\mathbf{x}) \psi(\mathbf{x}) dA_{\mathbf{x}} = \int_{\Gamma} \mathbf{n}(\mathbf{x}) \cdot \mathbf{w}(\mathbf{x}) \psi(\mathbf{x}) \det \mathbf{F}^{-1}(\mathbf{x}) dA_{\mathbf{x}}, \quad (3.45)$$

where $\mathbf{n}$ denotes the outward unit normal vector to Γ .

PROOF.

Firstly we apply the change of variable $\mathbf{x} = X^{-1}(\mathbf{y})$ (see, for instance, [58] page 53). We get

$$\begin{aligned} & \int_{\Gamma} \mathbf{F}^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot \mathbf{w}(\mathbf{x}) \psi(\mathbf{x}) dA_{\mathbf{x}} \\ &= \int_{\partial X(\Omega)} \mathbf{m}(\mathbf{y}) \cdot \mathbf{w}(X^{-1}(\mathbf{y})) \psi(X^{-1}(\mathbf{y})) \det \mathbf{F}^{-1}(X^{-1}(\mathbf{y})) dA_{\mathbf{y}}, \end{aligned}$$

where $\partial X(\Omega)$ denotes the boundary of $X(\Omega)$ and $\mathbf{n}$ (respectively, $\mathbf{m}$) is the unit normal vector to Γ (respectively, to $\partial X(\Omega)$). Thus, (3.45) follows by using that X is the identity function on the boundary. $\square$

Now, multiplying equation (3.39) by a test function $\psi \in H_{\Gamma_D}^1(\Omega)$, integrating in Ω , applying the usual Green's formula, and formulas (3.40) and (3.45) with $X(\mathbf{x}) = X_e(\mathbf{x}, t; \tau)$ and $\mathbf{w} = \text{grad } \phi$, and replacing Robin condition (3.3), we have

$$\begin{aligned} & \int_{\Omega} \dot{\phi}(X_e(\mathbf{x}, t; \tau), \tau) \psi(\mathbf{x}) d\mathbf{x} \\ &+ \int_{\Omega} \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) \mathbf{A}(X_e(\mathbf{x}, t; \tau)) \text{grad } \phi(X_e(\mathbf{x}, t; \tau)) \cdot \text{grad } \psi(\mathbf{x}) d\mathbf{x} \\ &+ \int_{\Omega} \text{Div } \mathbf{F}_e^{-T}(\mathbf{x}, t; \tau) \cdot \mathbf{A}(X_e(\mathbf{x}, t; \tau)) \text{grad } \phi(X_e(\mathbf{x}, t; \tau), \tau) \psi(\mathbf{x}) d\mathbf{x} \\ &+ \int_{\Omega} l(X_e(\mathbf{x}, t; \tau)) \phi(X_e(\mathbf{x}, t; \tau), \tau) \psi(\mathbf{x}) d\mathbf{x} \\ &+ \int_{\Gamma_R} \alpha \phi(\mathbf{x}, \tau) \psi(\mathbf{x}) \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) dA_{\mathbf{x}} \\ &= \int_{\Omega} f(X_e(\mathbf{x}, t; \tau), \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Gamma_R} g(\mathbf{x}, \tau) \psi(\mathbf{x}) \det \mathbf{F}_e^{-1}(\mathbf{x}, t; \tau) dA_{\mathbf{x}}. \end{aligned} \quad (3.46)$$

For these computations, we have assumed appropriate regularity on the involved data and that velocity field $\mathbf{v}$ vanishes on the boundary, so the differentiable mapping,

$$\mathbf{x} \longrightarrow X_e(\mathbf{x}, t; \tau),$$

satisfies the assumptions required in Lemmas 3.4 and 3.5.

3.5 Time discretization: characteristics method

In this section we present the characteristics method for time semidiscretization of (3.46). Formally, it can be interpreted as a splitting method with one step of transport and one step of reaction-diffusion.

We will consider the classical characteristics method [45, 88], known also as *modified method of characteristics* or *semi-Lagrangian method*, which is first order in time, as well as two second order methods:

- The first one consists of approximating the material derivative with a two-point scheme and applying the θ -method. When $\theta = 1$ the classical first order characteristics method is recovered and, when $\theta = 1/2$, the second order scheme proposed in [99] for a simpler equation is obtained.
- The second one consists of applying a backward two-step method for discretizing the material derivative and an implicit evaluation of the equation. In this way, we are led to the scheme studied in [47, 31].

3.5.1 Approximation of the material derivative

We use the notations introduced at the beginning of Section 3.3.1.

For fixed $(\mathbf{x}, t) = (\mathbf{x}, t_{n+1})$ different formulas are considered to approximate the material derivative

$$\frac{d\phi}{d\tau}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau).$$

- *First order backward formula:*

$$\frac{\phi(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) - \phi(X_e(\mathbf{x}, t_{n+1}; \tau - \Delta t), \tau - \Delta t)}{\Delta t}. \quad (3.47)$$

- *Second order centered formula:*

$$\frac{\phi(X_e(\mathbf{x}, t_{n+1}; \tau + \frac{\Delta t}{2}), \tau + \frac{\Delta t}{2}) - \phi(X_e(\mathbf{x}, t_{n+1}; \tau - \frac{\Delta t}{2}), \tau - \frac{\Delta t}{2})}{\Delta t}. \quad (3.48)$$

- *Second order backward formula:*

$$\frac{1}{2\Delta t} (3\phi(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) - 4\phi(X_e(\mathbf{x}, t_{n+1}; \tau - \Delta t), \tau - \Delta t) + \phi(X_e(\mathbf{x}, t_{n+1}; \tau - 2\Delta t), \tau - 2\Delta t)). \quad (3.49)$$

Moreover, the $O(\Delta t)$ error of (3.47) does not change if we replace the exact characteristics by a first order approximation as, for instance, (3.22). Similarly, in (3.48) and (3.49) the exact characteristic lines can be replaced by second order approximations as (3.23) or (3.25), obtaining again an $O(\Delta t^2)$ error.

In what follows we first develop some computations to motivate the schemes, assuming that the characteristic lines are exactly computed. Then the schemes are proposed for approximate characteristic lines.

Remark 3.8 *In each of the terms we will always demand the less accurate characteristics approximation preserving the global error estimate. The stability and consistency properties of each of the schemes do not change if we replace the proposed approximation for a more accurate one (for instance the mapping X_E^n can be replaced by X_{RK}^n , X_{TS}^n or X_e^n ; approximations X_{RK}^n and X_{TS}^n can be interchanged, and always replaced by X_e^n , etc.)*

3.5.2 Theta semi-Lagrangian scheme

The semidiscretization scheme we are going to study arises from fixing $t = t_{n+1}, n = 0, 1, \dots, N-1$, in (3.46), approximating the material time derivative at $\tau = t_{n+\frac{1}{2}}$ by the centered formula (3.48), and using an interpolation formula involving values at $\tau = t_n$ and $\tau = t_{n+1}$ to approximate the rest of the terms. More precisely,

$$\begin{aligned}
& \int_{\Omega} \frac{\phi^{n+1}(X_e(\mathbf{x}, t_{n+1}; t_{n+1})) - \phi^n(X_e(\mathbf{x}, t_{n+1}; t_n))}{\Delta t} \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \theta \int_{\Omega} \mathbf{A}(X_e(\mathbf{x}, t_{n+1}; t_{n+1})) \operatorname{grad} \phi^{n+1}(X_e(\mathbf{x}, t_{n+1}; t_{n+1})) \cdot \operatorname{grad} \psi(\mathbf{x}) \, d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; t_n) \mathbf{A}(X_e(\mathbf{x}, t_{n+1}; t_n)) \operatorname{grad} \phi^n(X_e(\mathbf{x}, t_{n+1}; t_n)) \cdot \operatorname{grad} \psi(\mathbf{x}) \, d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} \operatorname{Div} \mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; t_n) \cdot \mathbf{A}(X_e(\mathbf{x}, t_{n+1}; t_n)) \operatorname{grad} \phi^n(X_e(\mathbf{x}, t_{n+1}; t_n)) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \theta \int_{\Omega} l(X_e(\mathbf{x}, t_{n+1}; t_{n+1})) \phi^{n+1}(X_e(\mathbf{x}, t_{n+1}; t_{n+1})) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} l(X_e(\mathbf{x}, t_{n+1}; t_n)) \phi^n(X_e(\mathbf{x}, t_{n+1}; t_n)) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \int_{\Gamma_R} \alpha (\theta \phi(\mathbf{x}, t_{n+1}) + (1 - \theta) \phi(\mathbf{x}, t_n) \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; t_n)) \psi(\mathbf{x}) \, dA_{\mathbf{x}} \\
& = \int_{\Omega} (\theta f^{n+1}(X_e(\mathbf{x}, t_{n+1}; t_{n+1})) + (1 - \theta) f^n(X_e(\mathbf{x}, t_{n+1}; t_n))) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \int_{\Gamma_R} (\theta g^{n+1}(\mathbf{x}) + (1 - \theta) g^n(\mathbf{x}) \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; t_n)) \psi(\mathbf{x}) \, dA_{\mathbf{x}},
\end{aligned}$$

where we have used that $\mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; t_{n+1}) = \mathbf{I}$. Next by replacing $X_e(\mathbf{x}, t_{n+1}; t_{n+1})$ and $X_e(\mathbf{x}, t_{n+1}; t_n)$ by $\mathbf{x}$ and $X_e^n(\mathbf{x})$, respectively, and $\mathbf{F}_e(\mathbf{x}, t_{n+1}; t_n)$ by $\mathbf{F}_e^n(\mathbf{x})$, equation (3.50) becomes

$$\begin{aligned}
& \int_{\Omega} \frac{\phi^{n+1}(\mathbf{x}) - \phi^n(X_e^n(\mathbf{x}))}{\Delta t} \psi(\mathbf{x}) \, d\mathbf{x} + \theta \int_{\Omega} \mathbf{A}(\mathbf{x}) \operatorname{grad} \phi^{n+1}(\mathbf{x}) \cdot \operatorname{grad} \psi(\mathbf{x}) \, d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} (\mathbf{F}_e^n)^{-1}(\mathbf{x}) \mathbf{A}(X_e^n(\mathbf{x})) \operatorname{grad} \phi^n(X_e^n(\mathbf{x})) \cdot \operatorname{grad} \psi(\mathbf{x}) \, d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} \operatorname{Div} (\mathbf{F}_e^n)^{-T}(\mathbf{x}) \cdot \mathbf{A}(X_e^n(\mathbf{x})) \operatorname{grad} \phi^n(X_e^n(\mathbf{x})) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \theta \int_{\Omega} l(\mathbf{x}) \phi^{n+1}(\mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x} + (1 - \theta) \int_{\Omega} l(X_e^n(\mathbf{x})) \phi^n(X_e^n(\mathbf{x})) \psi(\mathbf{x}) \, d\mathbf{x} \tag{3.50} \\
& + \int_{\Gamma_R} \alpha (\theta \phi^{n+1}(\mathbf{x}) + (1 - \theta) \phi^n(\mathbf{x}) \det (\mathbf{F}_e^n)^{-1}(\mathbf{x})) \psi(\mathbf{x}) \, dA_{\mathbf{x}} \\
& = \int_{\Omega} (\theta f^{n+1}(\mathbf{x}) + (1 - \theta) f^n(X_e^n(\mathbf{x}))) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \int_{\Gamma_R} (\theta g^{n+1}(\mathbf{x}) + (1 - \theta) g^n(\mathbf{x}) \det (\mathbf{F}_e^n)^{-1}(\mathbf{x})) \psi(\mathbf{x}) \, dA_{\mathbf{x}}.
\end{aligned}$$

Notice that, for $\theta = 1$ the gradient of the characteristics, $\mathbf{F}_e$, does not appear. On the other hand, in Section 3.6.3 we will prove that the approximations involved in scheme (3.50) are $O(\Delta t^2)$

at point $(X_e(\mathbf{x}, t_{n+1}; t_{n+\frac{1}{2}}), t_{n+\frac{1}{2}})$ when $\theta = 1/2$. In fact, the error of that semidiscretized scheme does not change if we replace both $\mathbf{F}_e^{-1}$ and $\det \mathbf{F}_e^{-1}$ by their $O(\Delta t^2)$ approximations given in (3.16) and (3.19). More precisely, by taking $(\mathbf{x}, t; s) = (\mathbf{x}, t_{n+1}; t_n)$ we have

$$(\mathbf{F}_e^n)^{-1}(\mathbf{x}) = \mathbf{I}(\mathbf{x}) + \Delta t \mathbf{L}^n(X_e^n(\mathbf{x})) + O(\Delta t^2), \quad (3.51)$$

$$\det (\mathbf{F}_e^n)^{-1}(\mathbf{x}) = 1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}) + O(\Delta t^2). \quad (3.52)$$

Moreover, since $\operatorname{div} (\operatorname{grad} \mathbf{w})^T = \operatorname{grad} (\operatorname{div} \mathbf{w})$ for any smooth vector field $\mathbf{w}$, we have

$$\begin{aligned} \operatorname{Div}(\mathbf{F}_e^n)^{-T}(\mathbf{x}) &= \operatorname{div} (\mathbf{I} + \Delta t (\mathbf{L}^n)^T) (X_e^n(\mathbf{x})) + O(\Delta t^2) \\ &= \Delta t \operatorname{grad} \operatorname{div} \mathbf{v}^n(X_e^n(\mathbf{x})) + O(\Delta t^2). \end{aligned} \quad (3.53)$$

Thus, scheme (3.50) with $\theta = 1/2$ becomes

$$\begin{aligned} & \int_{\Omega} \frac{\phi^{n+1}(\mathbf{x}) - \phi^n(X_e^n(\mathbf{x}))}{\Delta t} \psi(\mathbf{x}) d\mathbf{x} \\ & + \int_{\Omega} \frac{\mathbf{A}(\mathbf{x}) \operatorname{grad} \phi^{n+1}(\mathbf{x}) + \mathbf{A}(X_e^n(\mathbf{x})) \operatorname{grad} \phi^n(X_e^n(\mathbf{x}))}{2} \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x} \\ & + \frac{\Delta t}{2} \int_{\Omega} \mathbf{L}^n(X_e^n(\mathbf{x})) \mathbf{A}(X_e^n(\mathbf{x})) \operatorname{grad} \phi^n(X_e^n(\mathbf{x})) \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x} \\ & + \frac{\Delta t}{2} \int_{\Omega} \operatorname{grad} \operatorname{div} \mathbf{v}^n(X_e^n(\mathbf{x})) \cdot \mathbf{A}(X_e^n(\mathbf{x})) \operatorname{grad} \phi^n(X_e^n(\mathbf{x})) \psi(\mathbf{x}) d\mathbf{x} \\ & + \int_{\Omega} \frac{l(\mathbf{x})\phi^{n+1}(\mathbf{x}) + l(X_e^n(\mathbf{x}))\phi^n(X_e^n(\mathbf{x}))}{2} \psi(\mathbf{x}) d\mathbf{x} \\ & + \int_{\Gamma_R} \alpha \frac{\phi^{n+1}(\mathbf{x}) + \phi^n(\mathbf{x})(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}))}{2} \psi(\mathbf{x}) dA_{\mathbf{x}} \\ & = \int_{\Omega} \frac{f^{n+1}(\mathbf{x}) + f^n(X_e^n(\mathbf{x}))}{2} \psi(\mathbf{x}) d\mathbf{x} \\ & + \int_{\Gamma_R} \frac{g^{n+1}(\mathbf{x}) + g^n(\mathbf{x})(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}))}{2} \psi(\mathbf{x}) dA_{\mathbf{x}} + O(\Delta t^2). \end{aligned} \quad (3.54)$$

Remark 3.9 *The gain of using scheme (3.54) instead of scheme (3.50) with $\theta = 1/2$ is that, in the first case, we avoid the inversion of $\mathbf{F}_e^n$ matrix.*

Since very often the characteristic lines are not exactly computed, we propose the following scheme which results from (3.50) by replacing the exact characteristic lines and gradients by

accurate enough approximations,

$$\begin{aligned}
& \int_{\Omega} \frac{\phi^{n+1}(\mathbf{x}) - \phi^n(X_{RK}^n(\mathbf{x}))}{\Delta t} \psi(\mathbf{x}) d\mathbf{x} + \theta \int_{\Omega} \mathbf{A}(\mathbf{x}) \operatorname{grad} \phi^{n+1}(\mathbf{x}) \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} (\mathbf{F}_E^n)^{-1}(\mathbf{x}) \mathbf{A}(X_E^n(\mathbf{x})) \operatorname{grad} \phi^n(X_E^n(\mathbf{x})) \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x} \\
& + (1 - \theta) \int_{\Omega} \operatorname{Div} (\mathbf{F}_E^n)^{-T}(\mathbf{x}) \cdot \mathbf{A}(X_E^n(\mathbf{x})) \operatorname{grad} \phi^n(X_E^n(\mathbf{x})) \psi(\mathbf{x}) d\mathbf{x} \\
& + \theta \int_{\Omega} l(\mathbf{x}) \phi^{n+1}(\mathbf{x}) \psi(\mathbf{x}) d\mathbf{x} + (1 - \theta) \int_{\Omega} l(X_E^n(\mathbf{x})) \phi^n(X_E^n(\mathbf{x})) \psi(\mathbf{x}) d\mathbf{x} \\
& + \int_{\Gamma_R} \alpha (\theta \phi^{n+1}(\mathbf{x}) + (1 - \theta) \phi^n(\mathbf{x}) \det (\mathbf{F}_E^n)^{-1}(\mathbf{x})) \psi(\mathbf{x}) dA_{\mathbf{x}} \\
& = \int_{\Omega} (\theta f^{n+1}(\mathbf{x}) + (1 - \theta) f^n(X_E^n(\mathbf{x}))) \psi(\mathbf{x}) d\mathbf{x} \\
& + \int_{\Gamma_R} (\theta g^{n+1}(\mathbf{x}) + (1 - \theta) g^n(\mathbf{x}) \det (\mathbf{F}_E^n)^{-1}(\mathbf{x})) \psi(\mathbf{x}) dA_{\mathbf{x}}.
\end{aligned} \tag{3.55}$$

For $\theta = 1/2$ the above scheme is still second order in time accuracy. Moreover, similarly to the computations developed for the scheme with exact characteristics, we can avoid the inversion of matrix $\mathbf{F}_E^n$ by using the second order approximations given in (3.30) and (3.33) leading to

$$\begin{aligned}
& \int_{\Omega} \frac{\phi^{n+1}(\mathbf{x}) - \phi^n(X_E^n(\mathbf{x}))}{\Delta t} \psi(\mathbf{x}) d\mathbf{x} \\
& + \int_{\Omega} \frac{\mathbf{A}(\mathbf{x}) \operatorname{grad} \phi^{n+1}(\mathbf{x}) + \mathbf{A}(X_E^n(\mathbf{x})) \operatorname{grad} \phi^n(X_E^n(\mathbf{x}))}{2} \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x} \\
& + \frac{\Delta t}{2} \int_{\Omega} \mathbf{L}^n(\mathbf{x}) \mathbf{A}(X_E^n(\mathbf{x})) \operatorname{grad} \phi^n(X_E^n(\mathbf{x})) \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x} \\
& + \frac{\Delta t}{2} \int_{\Omega} \operatorname{grad} \operatorname{div} \mathbf{v}^n(\mathbf{x}) \cdot \mathbf{A}(X_E^n(\mathbf{x})) \operatorname{grad} \phi^n(X_E^n(\mathbf{x})) \psi(\mathbf{x}) d\mathbf{x} \\
& + \int_{\Omega} \frac{l(\mathbf{x}) \phi^{n+1}(\mathbf{x}) + l(X_E^n(\mathbf{x})) \phi^n(X_E^n(\mathbf{x}))}{2} \psi(\mathbf{x}) d\mathbf{x} \\
& + \int_{\Gamma_R} \alpha \frac{\phi^{n+1}(\mathbf{x}) + \phi^n(\mathbf{x}) (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}))}{2} \psi(\mathbf{x}) dA_{\mathbf{x}} \\
& = \int_{\Omega} \frac{f^{n+1}(\mathbf{x}) + f^n(X_E^n(\mathbf{x}))}{2} \psi(\mathbf{x}) d\mathbf{x} \\
& + \int_{\Gamma_R} \frac{g^{n+1}(\mathbf{x}) + g^n(\mathbf{x}) (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}))}{2} \psi(\mathbf{x}) dA_{\mathbf{x}} + O(\Delta t^2).
\end{aligned} \tag{3.56}$$

Instead of (3.56) we will analyze a similar scheme which results from replacing the exact characteristic lines by their approximate formulas in (3.54).

For this purpose, let us define the characteristic schemes (first and second order in time) with approximate characteristic lines. For $\phi \in C^0(H^1(\Omega))$, let us introduce

$$\mathcal{D}_E^{n+1}[\phi] := \phi^{n+1} - \phi^n \circ X_E^n, \tag{3.57}$$

and

$$\mathcal{D}_{RK}^{n+1}[\phi] := \phi^{n+1} - \phi^n \circ X_{RK}^n. \tag{3.58}$$

Let us define $\mathcal{M}_{\Delta t}^{n,\theta}[\phi] \in (H^1(\Omega))'$, for $\phi \in C^0(H^1(\Omega))$, and $\mathcal{N}_{\Delta t}^{n,\theta} \in (H^1(\Omega))'$ by

$$\begin{aligned} \langle \mathcal{M}_{\Delta t}^{n,\theta}[\phi], \psi \rangle := & \int_{\Omega} (\theta \mathbf{A} \operatorname{grad} \phi^{n+1} + (1-\theta)(\mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n) \cdot \operatorname{grad} \psi \, d\mathbf{x} \\ & + (1-\theta)\Delta t \int_{\Omega} (\mathbf{L}^n \mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n \cdot \operatorname{grad} \psi \, d\mathbf{x} \\ & + (1-\theta)\Delta t \int_{\Omega} (\operatorname{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n \psi \, d\mathbf{x} \\ & + \int_{\Omega} (\theta l \phi^{n+1} + (1-\theta)(l \phi^n) \circ X_E^n) \psi \, d\mathbf{x} \\ & + \alpha \int_{\Gamma_R} (\theta \phi^{n+1} + (1-\theta) \phi^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})) \psi \, dA_{\mathbf{x}}, \end{aligned} \quad (3.59)$$

$$\begin{aligned} \langle \mathcal{N}_{\Delta t}^{n,\theta}, \psi \rangle := & \int_{\Omega} (\theta f^{n+1} + (1-\theta) f^n \circ X_E^n) \psi \, d\mathbf{x} \\ & + \int_{\Gamma_R} (\theta g^{n+1} + (1-\theta) g^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})) \psi \, dA_{\mathbf{x}}, \quad \forall \psi \in H^1(\Omega). \end{aligned} \quad (3.60)$$

Remark 3.10 Regarding the definitions of $\mathcal{M}_{\Delta t}^{n,\theta}$ and $\mathcal{N}_{\Delta t}^{n,\theta}$, only the values of function ϕ at discrete time steps $\{t_n\}_{n=0}^N$ are required. Thus, these definitions can also be stated for a sequence of functions $\widehat{\phi} = \{\phi^n\}_{n=1}^N \in [H^1(\Omega)]^N$.

Now the characteristics or semi-Lagrangian semidiscretized schemes can be written in a more compact way.

- *Fully implicit (classical) scheme:*

$$\left\{ \begin{array}{l} \text{Given } \phi_{\Delta t}^0, \text{ find } \widehat{\phi}_{\Delta t} = \{\phi_{\Delta t}^n\}_{n=1}^N \in [H_{\Gamma_D}^1(\Omega)]^N \text{ such that} \\ \frac{1}{\Delta t} \langle \mathcal{D}_E^{n+1}[\phi], \psi \rangle + \langle \mathcal{M}_{\Delta t}^{n,1}[\phi], \psi \rangle = \langle \mathcal{N}_{\Delta t}^{n,1}, \psi \rangle \\ \text{for all } \psi \in H_{\Gamma_D}^1(\Omega) \text{ and } n = 0, \dots, N-1. \end{array} \right. \quad (3.61)$$

- *Crank-Nicholson scheme:*

$$\left\{ \begin{array}{l} \text{Given } \phi_{\Delta t}^0, \text{ find } \widehat{\phi}_{\Delta t} = \{\phi_{\Delta t}^n\}_{n=1}^N \in [H_{\Gamma_D}^1(\Omega)]^N \text{ such that} \\ \frac{1}{\Delta t} \langle \mathcal{D}_{RK}^{n+1}[\phi], \psi \rangle + \langle \mathcal{M}_{\Delta t}^{n,\frac{1}{2}}[\phi], \psi \rangle = \langle \mathcal{N}_{\Delta t}^{n,\frac{1}{2}}, \psi \rangle \\ \text{for all } \psi \in H_{\Gamma_D}^1(\Omega) \text{ and } n = 0, \dots, N-1. \end{array} \right. \quad (3.62)$$

3.5.3 Two-step semi-Lagrangian scheme

Other family of schemes that can be considered for semidiscretizing (3.46) consists of fixing $t = t_{n+1}$, $n = 1, \dots, N-1$ and approximate the time derivative by a N_{steps} -step formula. When $N_{\text{steps}} = 1$ we have the first order backward Euler approximation and we are led to the classical

fully implicit scheme studied in the previous section. When $N_{steps} = 2$ we use the second order backward approximation (3.49) which yields

$$\begin{aligned} & \int_{\Omega} \frac{3\phi^{n+1}(\mathbf{x}) - 4\phi^n(X_e^n(\mathbf{x})) + \phi^{n-1}(X_e^{n-1}(X_e^n(\mathbf{x})))}{2\Delta t} \psi(\mathbf{x}) d\mathbf{x} \\ & + \int_{\Omega} \mathbf{A}(\mathbf{x}) \cdot \text{grad } \phi^{n+1}(\mathbf{x}) \cdot \text{grad } \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega} l(\mathbf{x}) \phi^{n+1}(\mathbf{x}) d\mathbf{x} + \int_{\Gamma_R} \alpha \phi^{n+1}(\mathbf{x}) \psi(\mathbf{x}) dA_{\mathbf{x}} \\ & = \int_{\Omega} f^{n+1}(\mathbf{x}) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Gamma_R} g^{n+1}(\mathbf{x}) \psi(\mathbf{x}) dA_{\mathbf{x}}. \end{aligned} \quad (3.63)$$

The final scheme results from replacing in (3.63) the exact characteristic lines by the approximate ones. More precisely, point $X_e^n(\mathbf{x})$ is replaced by formula (3.25) and $X_e^{n-1}(X_e^n(\mathbf{x}))$ by

$$X_{TS}^{n-1,n}(\mathbf{x}) := \mathbf{x} - 2\Delta t (2\mathbf{v}^n(\mathbf{x}) - \mathbf{v}^{n-1}(\mathbf{x})). \quad (3.64)$$

The difference operator

$$\mathcal{D}_{TS}^{n+1}[\phi] \in (H^1(\Omega))'$$

is defined, for $\phi \in C^0(H^1(\Omega))$, by

$$\mathcal{D}_{TS}^{n+1}[\phi] := 3\phi^{n+1} - 4\phi^n \circ X_{TS}^n + \phi^{n-1} \circ X_{TS}^{n-1,n}. \quad (3.65)$$

• *Two-step scheme:*

$$\left\{ \begin{array}{l} \text{Given } \phi_{\Delta t}^0, \phi_{\Delta t}^1 \text{ find } \widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=2}^N \in [H_{\Gamma_D}^1(\Omega)]^{N-1} \text{ such that} \\ \frac{1}{2\Delta t} \langle \mathcal{D}_{TS}^{n+1}[\phi], \psi \rangle + \langle \mathcal{M}_{\Delta t}^{n,1}[\phi], \psi \rangle = \langle \mathcal{N}_{\Delta t}^{n,1}, \psi \rangle \\ \text{for all } \psi \in H_{\Gamma_D}^1(\Omega) \text{ and } n = 0, \dots, N-1. \end{array} \right. \quad (3.66)$$

Remark 3.11 *When dealing with m-steps methods, the problem of how to compute the first m steps, preserving the convergence rate, arises. See for instance [46].*

3.6 Analysis of the Crank-Nicholson semi-Lagrangian scheme

Let us introduce

$$\mathcal{L}_{\Delta t}^{n+\frac{1}{2}}\phi \in (H^1(\Omega))' \text{ and } \mathcal{F}_{\Delta t}^{n+\frac{1}{2}} \in (H^1(\Omega))'$$

defined, for $\phi \in C^0(H^1(\Omega))$, by

$$\left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}}\phi, \psi \right\rangle := \frac{1}{\Delta t} \langle \mathcal{D}_{RK}^{n+1}[\phi], \psi \rangle + \left\langle \mathcal{M}_{\Delta t}^{n,\frac{1}{2}}[\phi], \psi \right\rangle \quad \forall \psi \in H^1(\Omega) \quad (3.67)$$

and

$$\left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \psi \right\rangle := \left\langle \mathcal{N}_{\Delta t}^{n,\frac{1}{2}}, \psi \right\rangle \quad \forall \psi \in H^1(\Omega). \quad (3.68)$$

Then, the time semidiscretized scheme can be written as follows:

$$\left\{ \begin{array}{l} \text{Given } \phi_{\Delta t}^0, \text{ find } \widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N \in \left[H_{\Gamma_D}^1(\Omega) \right]^N \text{ such that} \\ \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \psi \right\rangle = \left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \psi \right\rangle \quad \forall \psi \in H_{\Gamma_D}^1(\Omega) \quad \text{and} \quad n = 0, \dots, N-1. \end{array} \right. \quad (3.69)$$

3.6.1 Stability of the semidiscretized scheme

In order to develop a stability analysis some assumptions on the different terms of equation (3.1) are required.

Hypothesis 3.2 *The velocity field $\mathbf{v} \in C^0(W^{2,\infty}(\Omega))$ and satisfies $\mathbf{v} \equiv 0$ on Γ .*

Remark 3.12 *The assumption of the velocity field vanishing on the boundary is required in our analysis and ensures, for instance, a fixed spatial domain all over the time (see [91] for an analysis of the classical Lagrange-Galerkin method for time-dependent domains). However, in practice we have also considered an extension of the second order Crank-Nicholson scheme for general velocity fields (see Chapter 5).*

Remark 3.13 *Throughout the present section c_1 denotes the maximum between the positive constant appearing in Lemma 3.3 and the norm of velocity $\mathbf{v}$ in $C^0(W^{2,\infty}(\Omega))$.*

Moreover, in order to apply Lemma 3.1 with both X_E^n and X_{RK}^n , we will impose $c_1 \Delta t < 1/2$. However, in most of the proofs below we will use $c_1 \Delta t < 1$, for the sake of simplicity.

Hypothesis 3.3 *The diffusion matrix coefficients, A_{ij} , belong to $W^{1,\infty}(\Omega)$. Moreover, $\mathbf{A}$ is a $m \times m$ symmetric matrix satisfying*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{m_1} & \mathbf{\Theta} \\ \mathbf{\Theta} & \mathbf{\Theta} \end{pmatrix},$$

with $\mathbf{A}_{m_1}$ being a positive definite symmetric $m_1 \times m_1$ matrix ($m_1 \geq 1$), and where $\mathbf{\Theta}$ denotes an appropriate zero matrix. Moreover, there exists a strictly positive constant δ which is a uniform lower bound for the eigenvalues of $\mathbf{A}_{m_1}$.

As a consequence of Hypothesis 3.3, there exists a unique positive definite symmetric $m_1 \times m_1$ matrix function, $\mathbf{C}_{m_1}$, such that $\mathbf{A}_{m_1} = (\mathbf{C}_{m_1})^2$. Let us denote by $\mathbf{C}$ the symmetric and positive semidefinite $m \times m$ matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{m_1} & \mathbf{\Theta} \\ \mathbf{\Theta} & \mathbf{\Theta} \end{pmatrix}.$$

Notice that $\mathbf{A} = \mathbf{C}^2$ and $C_{ij} \in W^{1,\infty}(\Omega)$. At this moment, let us introduce the following constant:

$$c_2 := \max_{i,j} \left\{ \|C_{ij}\|_{W^{1,\infty}(\Omega)}^2 \right\}.$$

Finally, let us denote by $\mathbf{B}$ the $m \times m$ matrix

$$\mathbf{B} = \begin{pmatrix} \mathbf{I}_{m_1} & \mathbf{\Theta} \\ \mathbf{\Theta} & \mathbf{\Theta} \end{pmatrix},$$

where $\mathbf{I}_{m_1}$ is the $m_1 \times m_1$ identity matrix. Under Hypothesis 3.3 and using the previous notation, we have

$$\delta \|\mathbf{B}\mathbf{w}\|_0^2 \leq \langle \mathbf{A}\mathbf{w}, \mathbf{w} \rangle = \|\mathbf{C}\mathbf{w}\|_0^2 \leq c_2 \|\mathbf{B}\mathbf{w}\|_0^2, \quad (3.70)$$

for all $\mathbf{w} \in \mathbb{R}^m$. Clearly,

$$\|\mathbf{A}\mathbf{w}\|_0 \leq c_2 \|\mathbf{w}\|_0. \quad (3.71)$$

Hypothesis 3.4 *The velocity field satisfies*

$$(\mathbf{I} - \mathbf{B})\mathbf{L}(\mathbf{x}, t)\mathbf{B} = \mathbf{0} \quad \forall (\mathbf{x}, t) \in \Omega \times [0, T].$$

Remark 3.14 *Hypothesis 3.4 is equivalent to have a velocity field $\mathbf{v}$ whose $m - m_1$ last components only depend on the $m - m_1$ last variables.*

Remark 3.15 *Under Hypotheses 3.3 and 3.4, for every $m \times m$ matrix $\mathbf{E}$ and vectors $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^m$ it is easy to see that*

$$\langle \mathbf{E}\mathbf{A}\mathbf{w}_1, \mathbf{w}_2 \rangle = \langle \mathbf{E}\mathbf{A}\mathbf{w}_1, \mathbf{B}\mathbf{w}_2 \rangle. \quad (3.72)$$

Hypothesis 3.5 *The reaction function, $l \in W^{1,\infty}(\Omega)$, satisfies*

$$0 < \gamma \leq l(\mathbf{x}) \text{ in } \Omega, \quad (3.73)$$

where γ is a constant.

Under the previous hypothesis, let us introduce $c_3 := \|\sqrt{l}\|_{W^{1,\infty}(\Omega)}^2$.

Hypothesis 3.6 *The source function $f \in C^0(L^2(\Omega))$.*

Hypothesis 3.7 *In the Robin boundary condition, $g \in C^0(L^2(\Gamma_R))$ and $\alpha > 0$.*

Corresponding to the semidiscretized scheme we have to deal with sequences of functions $\widehat{\psi} = \{\psi^n\}_{n=0}^N$. Thus, we consider spaces $l^\infty((0, T), L^2(\Omega))$, $l^2((0, T), L^2(\Omega))$ (shortly denoted by $l^\infty(L^2(\Omega))$ and $l^2(L^2(\Omega))$, respectively) equipped with their usual norms

$$\|\widehat{\psi}\|_{l^\infty(L^2(\Omega))} := \max_{n=0}^N \|\psi^n\|_0, \quad \|\widehat{\psi}\|_{l^2(L^2(\Omega))} := \sqrt{\Delta t \sum_{n=0}^N \|\psi^n\|_0^2}.$$

Similar definitions are used for functional spaces associated to the Robin boundary condition, namely, $l^\infty(L^2(\Gamma_R))$ and $l^2(L^2(\Gamma_R))$.

Moreover, let us introduce the notation

$$D_{\Delta t}^n \widehat{\psi} := \frac{\psi^{n+1} - \psi^n}{\Delta t}. \quad (3.74)$$

Thus, for the sequence $\|\widehat{\psi}\|_0 := \{\|\psi^n\|_0\}$, let us define

$$D_{\Delta t}^n \left(\|\widehat{\psi}\|_0 \right) := \frac{\|\psi^{n+1}\|_0 - \|\psi^n\|_0}{\Delta t}, \quad (3.75)$$

and

$$\left\| \widehat{D_{\Delta t} \psi} \right\|_{L^2(L^2(\Gamma_R))} = \sqrt{\Delta t \sum_{n=0}^{N-1} \left\| \frac{\psi^{n+1} - \psi^n}{\Delta t} \right\|_{0, \Gamma_R}^2}, \quad (3.76)$$

where $\widehat{D_{\Delta t} \psi} := \left\{ D_{\Delta t}^n \widehat{\psi} \right\}_{n=0}^{N-1}$.

Before establishing some technical lemmas, let us recall the inequality

$$ab \leq \frac{1}{2} \left(ca^2 + \frac{1}{c} b^2 \right), \quad (3.77)$$

for two real numbers a and b , and a positive number, c , which will be extensively used in what follows.

Lemma 3.6 *Let us assume Hypotheses 3.2, 3.3, 3.4 and 3.5. Let us suppose $c_1 \Delta t < 1/2$. If $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ denotes the solution of (3.69) and $\alpha > 0$, $\delta > 0$ are the constants appearing, respectively, in Hypothesis 3.3 and equation (3.3), then*

$$\begin{aligned} & \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \phi_{\Delta t}^{n+1} \right\rangle \\ & \geq D_{\Delta t}^n \left(\frac{1}{2} \left\| \widehat{\phi_{\Delta t}} \right\|_0^2 + \frac{\Delta t}{4} \left\| \mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}} \right\|_0^2 + \frac{\Delta t}{4} \left\| \sqrt{l} \widehat{\phi_{\Delta t}} \right\|_0^2 + \frac{\alpha \Delta t}{4} \left\| \widehat{\phi_{\Delta t}} \right\|_{0, \Gamma_R}^2 \right) \\ & \quad + \frac{1}{2\Delta t} \left\| \phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n \right\|_0^2 + \frac{1}{4} \left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 \\ & \quad + \frac{1}{4} \left\| \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 + \frac{\alpha}{4} \left\| \phi_{\Delta t}^{n+1} + \phi_{\Delta t}^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \right\|_{0, \Gamma_R}^2 \\ & \quad - \frac{c}{2} \left(\left\| \phi_{\Delta t}^n \right\|_0^2 + \left\| \phi_{\Delta t}^{n+1} \right\|_0^2 \right) - c\Delta t \delta \left(\left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^n \right\|_0^2 + \left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\|_0^2 \right) \\ & \quad - c\Delta t \left(\left\| \sqrt{l} \phi_{\Delta t}^n \right\|_0^2 + \left\| \sqrt{l} \phi_{\Delta t}^{n+1} \right\|_0^2 + \alpha \left\| \phi_{\Delta t}^n \right\|_{0, \Gamma_R}^2 \right), \end{aligned} \quad (3.78)$$

where $\mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}} := \{\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n\}$, $\mathbf{B} \operatorname{grad} \widehat{\phi_{\Delta t}} := \{\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\}$, and with

$$c = \max \left\{ 1, c_1, c_2, (2c_1 c_2 + c_1 c_2^2)/\delta, c_1 c_3/\gamma \right\}.$$

PROOF. Firstly, we decompose

$$\left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \phi_{\Delta t}^{n+1} \right\rangle = I_1 + I_2 + I_3 + I_4 + I_5 + I_6,$$

with

$$\begin{aligned}
I_1 &= \left\langle \frac{\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n}{\Delta t}, \phi_{\Delta t}^{n+1} \right\rangle, \\
I_2 &= \left\langle \frac{\mathbf{A} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{A} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n}{2}, \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle, \\
I_3 &= \Delta t \left\langle \frac{(\mathbf{L}^n \mathbf{A} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n}{2}, \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle, \\
I_4 &= \Delta t \left\langle \frac{(\operatorname{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n}{2}, \phi_{\Delta t}^{n+1} \right\rangle, \\
I_5 &= \left\langle \frac{l \phi_{\Delta t}^{n+1} + (l \phi_{\Delta t}^n) \circ X_E^n}{2}, \phi_{\Delta t}^{n+1} \right\rangle, \\
I_6 &= \alpha \left\langle \frac{\phi_{\Delta t}^{n+1} + \phi_{\Delta t}^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2}, \phi_{\Delta t}^{n+1} \right\rangle_{\Gamma_R}.
\end{aligned}$$

For I_1 we have

$$I_1 \geq D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{\phi_{\Delta t}}\|_0^2 \right) + \frac{1}{2\Delta t} \|\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n\|_0^2 - \frac{c_1}{2} \|\phi_{\Delta t}^n\|_0^2. \quad (3.79)$$

Indeed, from the definition of $D_{\Delta t}^n$, Lemma 3.3 and (3.77), we deduce

$$\begin{aligned}
&D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{\phi_{\Delta t}}\|_0^2 \right) + \frac{1}{2\Delta t} \|\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n\|_0^2 - \frac{c_1}{2} \|\phi_{\Delta t}^n\|_0^2 \\
&= \frac{1}{2\Delta t} \left(\|\phi_{\Delta t}^{n+1}\|_0^2 - \|\phi_{\Delta t}^n\|_0^2 \right) + \frac{1}{2\Delta t} \|\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n\|_0^2 - \frac{c_1}{2} \|\phi_{\Delta t}^n\|_0^2 \\
&= \frac{1}{2\Delta t} \left(\|\phi_{\Delta t}^{n+1}\|_0^2 - (1 + c_1 \Delta t) \|\phi_{\Delta t}^n\|_0^2 + \|\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n\|_0^2 \right) \\
&\leq \frac{1}{2\Delta t} \left(\|\phi_{\Delta t}^{n+1}\|_0^2 - \|\phi_{\Delta t}^n \circ X_{RK}^n\|_0^2 + \|\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_{RK}^n\|_0^2 \right) = I_1.
\end{aligned}$$

For I_2 the following lower bound can be stated:

$$\begin{aligned}
I_2 &\geq D_{\Delta t}^n \left(\frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}}\|_0^2 \right) + \frac{1}{4} \|\mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n\|_0^2 \\
&\quad - \frac{3c_1 c_2 \Delta t}{4} \left(\|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1}\|_0^2 \right). \quad (3.80)
\end{aligned}$$

In order to prove (3.80) let us first use the same arguments as for I_1 , namely,

$$\begin{aligned}
& D_{\Delta t}^n \left(\frac{\Delta t}{4} \left\| \mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}} \right\|_0^2 \right) + \frac{1}{4} \left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 \\
& - \frac{c_1 c_2 \Delta t}{4} \left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^n \right\|_0^2 \\
& \leq \frac{1}{4} \left(\left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\|_0^2 - \left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^n \right\|_0^2 \right) \\
& + \frac{1}{4} \left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 - \frac{c_1 \Delta t}{4} \left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^n \right\|_0^2 \\
& \leq \frac{1}{4} \left(\left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\|_0^2 - \left\| (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 + \right. \\
& \quad \left. \left\| \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 \right) \\
& = \frac{1}{2} \left\langle \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n, \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle.
\end{aligned} \tag{3.81}$$

Next, we introduce the function

$$\begin{aligned}
Y_E^n(\mathbf{x}, \cdot) : [t_n, t_{n+1}] & \longrightarrow \overline{\Omega} \\
s & \longrightarrow Y_E^n(\mathbf{x}, s) := \mathbf{x} - (t_{n+1} - s) \mathbf{v}^{n+1}(\mathbf{x}),
\end{aligned} \tag{3.82}$$

which satisfies $Y_E^n(\mathbf{x}, t_n) = X_E^n(\mathbf{x})$ and $Y_E^n(\mathbf{x}, t_{n+1}) = \mathbf{x}$. Moreover, since $\mathbf{A}_{m_1}$ is symmetric and positive definite, then $\mathbf{C}_{m_1} = \sqrt{\mathbf{A}_{m_1}}$ is a differentiable function. Notice that $\operatorname{grad} \mathbf{C}$ is the appropriate completion by zeros of $\operatorname{grad} \mathbf{C}_{m_1}$. Then, by Barrow's rule and the chain rule, the following identity holds:

$$\mathbf{C}(\mathbf{x}) = \mathbf{C}(X_E^n(\mathbf{x})) + \mathbf{D}^n(\mathbf{x}) \quad a.e. \quad \mathbf{x} \in \Omega \tag{3.83}$$

where we have denoted by $\mathbf{D}^n$ the $m \times m$ symmetric matrix defined by

$$D_{ij}^n(\mathbf{x}) := \int_{t_n}^{t_{n+1}} \operatorname{grad} C_{ij}(Y_E^n(\mathbf{x}, s)) \cdot \mathbf{v}^{n+1}(\mathbf{x}) \, ds \quad a.e. \quad \mathbf{x} \in \Omega \tag{3.84}$$

and bounded by

$$|D_{ij}^n(\mathbf{x})| \leq c_1 \sqrt{c_2} \Delta t. \tag{3.85}$$

With the previous notation, replacing (3.83) in (3.81), we obtain

$$\begin{aligned}
& \frac{1}{2} \left\langle \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} + (\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n, \mathbf{C} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle \\
& = \frac{1}{2} \left\langle \mathbf{C}^2 \operatorname{grad} \phi_{\Delta t}^{n+1} + \mathbf{C}(\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n, \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle \\
& = I_2 + \frac{1}{2} \left\langle \mathbf{D}^n(\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n, \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle.
\end{aligned} \tag{3.86}$$

Now, by using Cauchy-Schwartz inequality, Hypothesis 3.2 and 3.3, Lemma 3.3, inequality (3.77) and that $c_1 \Delta t < 1$ we get

$$\begin{aligned}
& \frac{1}{2} \left\langle \mathbf{D}^n(\mathbf{C} \operatorname{grad} \phi_{\Delta t}^n) \circ X_E^n, \operatorname{grad} \phi_{\Delta t}^{n+1} \right\rangle \\
& \leq \frac{1}{2} \sqrt{1 + c_1 \Delta t \sqrt{c_2}} \left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^n \right\|_0 \Delta t c_1 \sqrt{c_2} \left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\|_0 \\
& \leq \frac{\sqrt{2}}{4} c_1 c_2 \Delta t \left(\left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^n \right\|_0^2 + \left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\|_0^2 \right) \\
& \leq \frac{c_1 c_2 \Delta t}{2} \left(\left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^n \right\|_0^2 + \left\| \mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1} \right\|_0^2 \right).
\end{aligned} \tag{3.87}$$

Finally, result (3.80) is obtained by jointly considering (3.81) (3.86) and (3.87).

Next, by first using Remark 3.15 and then Cauchy-Schwartz again, Hypothesis 3.2 and 3.3, Lemma 3.3, inequality (3.77) and that $c_1\Delta t < 1$, we obtain

$$\begin{aligned} |I_3| &= \frac{\Delta t}{2} |\langle (\mathbf{L}^n \mathbf{A} \text{ grad } \phi_{\Delta t}^n) \circ X_E^n, \mathbf{B} \text{ grad } \phi_{\Delta t}^{n+1} \rangle| \\ &\leq \frac{\Delta t}{2} \|(\mathbf{L}^n \mathbf{A} \text{ grad } \phi_{\Delta t}^n) \circ X_E^n\|_0 \|\mathbf{B} \text{ grad } \phi_{\Delta t}^{n+1}\|_0 \\ &\leq \frac{\Delta t}{2} c_1 c_2 \sqrt{1 + c_1 \Delta t} \|\mathbf{B} \text{ grad } \phi_{\Delta t}^n\|_0 \|\mathbf{B} \text{ grad } \phi_{\Delta t}^{n+1}\|_0 \\ &\leq \frac{\Delta t}{2} \frac{c_1 c_2}{2} \left(\|\mathbf{B} \text{ grad } \phi_{\Delta t}^n\|_0^2 + \|\mathbf{B} \text{ grad } \phi_{\Delta t}^{n+1}\|_0^2 \right). \end{aligned}$$

Then, when both $I_3 \geq 0$ and $I_3 < 0$ we have

$$I_3 \geq -\frac{c_1 c_2 \Delta t}{2} \left(\|\mathbf{B} \text{ grad } \phi_{\Delta t}^n\|_0^2 + \|\mathbf{B} \text{ grad } \phi_{\Delta t}^{n+1}\|_0^2 \right). \quad (3.88)$$

For I_4 we obtain the estimate

$$I_4 \geq -\frac{c_1 c_2^2 \Delta t}{2} \|\mathbf{B} \text{ grad } \phi_{\Delta t}^n\|_0^2 - \frac{1}{4} \|\phi_{\Delta t}^{n+1}\|_0^2. \quad (3.89)$$

Indeed, we have

$$\begin{aligned} |I_4| &\leq \frac{1}{4} \left(\Delta t^2 \|(\text{grad div } \mathbf{v}^n \cdot \mathbf{A} \text{ grad } \phi_{\Delta t}^n) \circ X_E^n\|_0^2 + \|\phi_{\Delta t}^{n+1}\|_0^2 \right) \\ &\leq \frac{c_1^2 c_2^2 \Delta t^2 (1 + c_1 \Delta t)}{4} \|\mathbf{B} \text{ grad } \phi_{\Delta t}^n\|_0^2 + \frac{1}{4} \|\phi_{\Delta t}^{n+1}\|_0^2. \end{aligned}$$

For the reaction term, we can obtain

$$\begin{aligned} I_5 &\geq D_{\Delta t}^n \left(\frac{\Delta t}{4} \|\widehat{\sqrt{l} \phi_{\Delta t}}\|_0^2 \right) + \frac{1}{4} \left\| \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 \\ &\quad - \max\{c_1, c_1 c_3 / \gamma\} \frac{\Delta t}{2} \left(\|\sqrt{l} \phi_{\Delta t}^n\|_0^2 + \|\sqrt{l} \phi_{\Delta t}^{n+1}\|_0^2 \right). \end{aligned} \quad (3.90)$$

The proof of (3.90) is analogous to the one of (3.80). Indeed, we have

$$\begin{aligned} &D_{\Delta t}^n \left(\frac{\Delta t}{4} \|\widehat{\sqrt{l} \phi_{\Delta t}}\|_0^2 \right) + \frac{1}{4} \left\| \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 - \frac{c_1 \Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 \\ &= \frac{1}{4} \left(\|\sqrt{l} \phi_{\Delta t}^{n+1}\|_0^2 - \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 \right) + \frac{1}{4} \left\| \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 - \frac{\Delta t}{4} c_1 \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 \\ &\leq \frac{1}{4} \left(\|\sqrt{l} \phi_{\Delta t}^{n+1}\|_0^2 - \|(\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n\|_0^2 \right) + \frac{1}{4} \left(\left\| \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n \right\|_0^2 \right) \\ &= \frac{1}{2} \left\langle \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n, \sqrt{l} \phi_{\Delta t}^{n+1} \right\rangle. \end{aligned}$$

Now, by using the function introduced in (3.82), we have that

$$\sqrt{l}(\mathbf{x}) = \sqrt{l}(X_E^n(\mathbf{x})) + \int_{t_n}^{t_{n+1}} \text{grad } \sqrt{l}(Y_E^n(\mathbf{x}, s)) \cdot \mathbf{v}^{n+1}(\mathbf{x}) \, ds \quad \text{a.e. } \mathbf{x} \in \Omega. \quad (3.91)$$

From this equality we get

$$\begin{aligned} & \frac{1}{2} \left\langle \sqrt{l} \phi_{\Delta t}^{n+1} + (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n, \sqrt{l} \phi_{\Delta t}^{n+1} \right\rangle \\ &= I_5 + \left\langle (\sqrt{l} \phi_{\Delta t}^n) \circ X_E^n, \left(\int_{t_n}^{t_{n+1}} \text{grad } \sqrt{l}(Y_E^n(\mathbf{x}, s)) \cdot \mathbf{v}^{n+1}(\mathbf{x}) \, ds \right) \phi_{\Delta t}^{n+1} \right\rangle \\ &\leq I_5 + \frac{c_1 c_3 \Delta t}{2} \left(\|\phi_{\Delta t}^n\|_0^2 + \|\phi_{\Delta t}^{n+1}\|_0^2 \right). \end{aligned}$$

Moreover, since the reaction function is bounded from below by Hypothesis 3.5, we have

$$\|\phi_{\Delta t}^j\|_0^2 \leq \frac{1}{\gamma} \|\sqrt{l} \phi_{\Delta t}^j\|_0^2,$$

for $j = n, n+1$, and inequality (3.90) follows.

For the boundary integral term, I_6 , we use some properties of the inner product in the space $L^2(\Gamma_R)$ and the inequality

$$(1 + c_1 \Delta t)^2 \leq 1 + 3c_1 \Delta t$$

to get the estimate

$$\|\psi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 \leq (1 + c_1 \Delta t)^2 \|\psi\|_{0,\Gamma_R}^2 \leq (1 + 3c_1 \Delta t) \|\psi\|_{0,\Gamma_R}^2,$$

for $\psi \in L^2(\Gamma_R)$. Thus, we obtain,

$$\begin{aligned} I_6 &\geq D_{\Delta t}^n \left(\frac{\alpha \Delta t}{4} \|\widehat{\phi_{\Delta t}}\|_{0,\Gamma_R}^2 \right) + \frac{\alpha}{4} \|\phi_{\Delta t}^{n+1} + \phi_{\Delta t}^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 \\ &\quad - \frac{3}{4} c_1 \alpha \Delta t \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2. \end{aligned} \tag{3.92}$$

Finally, by summing up (3.79), (3.80), (3.88), (3.89), (3.90) and (3.92), we have proved inequality (3.78). $\square$

Lemma 3.7 *Let us assume Hypotheses 3.2, 3.6, 3.7 and that $c_1 \Delta t < 1$. Then we have*

$$\begin{aligned} & \left\langle \frac{f^{n+1} + f^n \circ X_E^n}{2}, \psi \right\rangle + \left\langle \frac{g^{n+1} + g^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2}, \psi \right\rangle_{\Gamma_R} \\ &\leq \frac{1}{2} (\|f^{n+1}\|_0^2 + \|f^n\|_0^2) + \frac{1}{2} \|\psi\|_0^2 + \frac{1}{2} \langle g^{n+1}, \psi \rangle_{\Gamma_R} - \frac{1}{2} \langle g^n, \varphi \rangle_{\Gamma_R} \\ &\quad + \frac{3}{\alpha} \|g^n\|_{0,\Gamma_R}^2 + \frac{3\alpha c_1 \Delta t}{4} \|\varphi\|_{0,\Gamma_R}^2 + \frac{\alpha}{8} \|\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2, \end{aligned} \tag{3.93}$$

for all functions $\varphi, \psi \in H^1(\Omega)$, and $n = 0, \dots, N-1$.

PROOF. Firstly, we divide the term to be bounded into two parts, I_1 and I_2 , with

$$I_1 = \left\langle \frac{f^{n+1} + f^n \circ X_E^n}{2}, \psi \right\rangle, \quad I_2 = \left\langle \frac{g^{n+1} + g^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2}, \psi \right\rangle_{\Gamma_R}.$$

The bound for I_1 is achieved by using (3.77) and Lemma 3.3. Indeed,

$$I_1 \leq \frac{\|f^{n+1}\|_0^2}{4} + \frac{\|\psi\|_0^2}{4} + \frac{(1 + c_1 \Delta t) \|f^n\|_0^2}{4} + \frac{\|\psi\|_0^2}{4} \leq \frac{1}{2} (\|f^{n+1}\|_0^2 + \|f^n\|_0^2 + \|\psi\|_0^2).$$

Then I_2 is decomposed into three terms, $I_2 = I_2^1 + I_2^2 + I_2^3$, where

$$\begin{aligned} I_2^1 &= \frac{1}{2} \langle g^{n+1}, \psi \rangle_{\Gamma_R} - \frac{1}{2} \langle g^n, \varphi \rangle_{\Gamma_R}, \\ I_2^2 &= \frac{1}{2} \langle g^n, \varphi \rangle_{\Gamma_R} - \frac{1}{2} \langle g^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \rangle_{\Gamma_R}, \\ I_2^3 &= \frac{1}{2} \langle g^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \rangle_{\Gamma_R} + \frac{1}{2} \langle g^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \psi \rangle_{\Gamma_R}. \end{aligned}$$

Next, we compute an upper bound for I_2^2 and I_2^3 . For the first one, since $c_1 \Delta t < 1$, we have

$$\begin{aligned} & \int_{\Gamma_R} (-(\Delta t^2 \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}))^2 - 2 \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x})) \varphi(\mathbf{x}) g^n(\mathbf{x}) dA_{\mathbf{x}} \\ & \leq 3c_1 \Delta t \|\varphi\|_{0,\Gamma_R} \|g^n\|_{0,\Gamma_R}, \end{aligned} \quad (3.94)$$

and then

$$\begin{aligned} I_2^2 &\leq \frac{3}{2} c_1 \Delta t \|\varphi\|_{0,\Gamma_R} \|g^n\|_{0,\Gamma_R} \leq \frac{3c_1 \Delta t}{4} \left(\alpha \|\varphi\|_{0,\Gamma_R}^2 + \frac{1}{\alpha} \|g^n\|_{0,\Gamma_R}^2 \right) \\ &\leq \frac{3c_1 \Delta t \alpha}{4} \|\varphi\|_{0,\Gamma_R}^2 + \frac{3}{4\alpha} \|g^n\|_{0,\Gamma_R}^2. \end{aligned}$$

where we have used inequality (3.77).

For term I_2^3 , inequality (3.77) and $(1 + c_1 \Delta t)^2 \leq 4$, yield

$$\begin{aligned} I_2^3 &= \left\langle g^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \frac{\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2} \right\rangle_{\Gamma_R} \\ &\leq \frac{1}{2\alpha} \|g^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 + \frac{\alpha}{8} \|\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 \\ &\leq \frac{4}{2\alpha} \|g^n\|_{0,\Gamma_R}^2 + \frac{\alpha}{8} \|\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2. \end{aligned}$$

Finally, by jointly considering the above inequalities we get (3.93). $\square$

Theorem 3.1 (Stability) *Let us assume Hypotheses 3.2, 3.3, 3.5, 3.6 and 3.7 and let $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ be the solution of (3.69) subject to initial value $\phi_{\Delta t}^0 \in H^1(\Omega)$. Then, there exist two positive constants, c and d , such that, for $\Delta t < d$, we have*

$$\begin{aligned} & \frac{1}{\sqrt{2}} \|\widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \operatorname{grad} \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} \\ & + \sqrt{\frac{\alpha \Delta t}{8}} \|\widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Gamma_R))} \leq c \left(\frac{1}{2} \|\phi_{\Delta t}^0\|_0 + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^0\|_0 + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \phi_{\Delta t}^0\|_0 \right. \\ & \quad \left. + \sqrt{\frac{\alpha \Delta t}{8}} \|\phi_{\Delta t}^0\|_{0,\Gamma_R} + \|\widehat{f}\|_{l^2(L^2(\Omega))} + \|\widehat{g}\|_{l^2(L^2(\Gamma_R))} \right), \end{aligned}$$

where $\mathbf{B} \operatorname{grad} \widehat{\phi_{\Delta t}} := \{\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\}$.

PROOF. The sequence $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ satisfies

$$\left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \phi_{\Delta t}^{n+1} \right\rangle = \left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \phi_{\Delta t}^{n+1} \right\rangle.$$

By using Lemma 3.6 and Lemma 3.7 for $\psi = \phi_{\Delta t}^{n+1}$ and $\varphi = \phi_{\Delta t}^n$, we get

$$\begin{aligned}
& D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{\phi_{\Delta t}}\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}}\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \widehat{\phi_{\Delta t}}\|_0^2 + \frac{\alpha \Delta t}{4} \|\widehat{\phi_{\Delta t}}\|_{0,\Gamma_R}^2 \right) \\
& + \frac{\alpha}{8} \|\phi_{\Delta t}^{n+1} + \phi_{\Delta t}^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 \leq \frac{1}{2} (\|f^{n+1}\|_0^2 + \|f^n\|_0^2) + \frac{3}{\alpha} \|g^n\|_{0,\Gamma_R}^2 \\
& + \frac{1}{2} \langle g^{n+1}, \phi_{\Delta t}^{n+1} \rangle_{\Gamma_R} - \frac{1}{2} \langle g^n, \phi_{\Delta t}^n \rangle_{\Gamma_R} + c (\|\phi_{\Delta t}^n\|_0^2 + \|\phi_{\Delta t}^{n+1}\|_0^2) \\
& + c \delta \Delta t (\|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1}\|_0^2) + c \Delta t (\|\sqrt{l} \phi_{\Delta t}^n\|_0^2 + \|\sqrt{l} \phi_{\Delta t}^{n+1}\|_0^2) \\
& + 2c\alpha \Delta t \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2,
\end{aligned} \tag{3.95}$$

with $c = \max \{1, c_1, c_2, (2c_1c_2 + c_1c_2^2)/\delta, c_1c_3/\gamma\}$. Now, for fixed integer $q \geq 1$, let us sum inequality (3.95) multiplied by Δt from $n = 0$ to $n = q - 1$. We obtain

$$\begin{aligned}
& \frac{1}{2} \|\phi_{\Delta t}^q\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \phi_{\Delta t}^q\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^q\|_0^2 + \frac{\alpha \Delta t}{4} \|\phi_{\Delta t}^q\|_{0,\Gamma_R}^2 \\
& - \frac{1}{2} \|\phi_{\Delta t}^0\|_0^2 - \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \phi_{\Delta t}^0\|_0^2 - \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^0\|_0^2 - \frac{\alpha \Delta t}{4} \|\phi_{\Delta t}^0\|_{0,\Gamma_R}^2 \\
& \leq \Delta t \sum_{n=1}^q \|f^n\|_0^2 + \frac{3\Delta t}{\alpha} \sum_{n=1}^{q-1} \|g^n\|_{0,\Gamma_R}^2 + \frac{\Delta t}{2} \langle g^q, \phi_{\Delta t}^q \rangle_{\Gamma_R} - \frac{\Delta t}{2} \langle g^0, \phi_{\Delta t}^0 \rangle_{\Gamma_R} \\
& + 2c \left(\Delta t \sum_{n=0}^q \|\phi_{\Delta t}^n\|_0^2 + \Delta t^2 \sum_{n=0}^q \delta \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 \right. \\
& \quad \left. + \Delta t^2 \sum_{n=0}^q \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 + \Delta t^2 \sum_{n=0}^{q-1} \alpha \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2 \right).
\end{aligned} \tag{3.96}$$

Let us introduce the following notations:

$$\begin{aligned}
\theta_n &:= \frac{1}{2} \|\phi_{\Delta t}^n\|_0^2 + \frac{\delta \Delta t}{4} \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^n\|_0^2, \\
\bar{\theta}_n &:= \frac{\alpha \Delta t}{8} \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2,
\end{aligned}$$

for $n = 0, \dots, N$. We use Hypothesis 3.3 (or, more precisely, equation (3.70)) together with the estimate

$$\frac{\Delta t}{2} \langle g^n, \phi_{\Delta t}^n \rangle_{\Gamma_R} \leq \frac{\Delta t}{2\alpha} \|g^n\|_{0,\Gamma_R}^2 + \frac{\alpha \Delta t}{8} \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2,$$

for $n = 0$ and $n = q$. Thus, inequality (3.96) implies

$$\begin{aligned}
(1 - 8c\Delta t)\theta_q + \bar{\theta}_q &\leq 8c\Delta t \sum_{n=0}^{q-1} \theta_n + 32c\Delta t \sum_{n=0}^{q-1} \bar{\theta}_n \\
&\quad + \tilde{c} \left(\theta_0 + \bar{\theta}_0 + \|\widehat{f}\|_{l^2(L^2(\Omega))}^2 + \|\widehat{g}\|_{l^2(L^2(\Gamma_R))}^2 \right),
\end{aligned}$$

with $c = \max \{1, c_1, c_2, (2c_1c_2 + c_1c_2^2)/\delta, c_1c_3/\gamma\}$ and $\tilde{c}$ a positive constant. For Δt small enough, we can apply the discrete Gronwall inequality (see for instance [95]), and take the maximum in $q \in \{1, \dots, N\}$. Thus, estimate (3.95) follows. $\square$

Remark 3.16 Notice that the above stability result is independent of the approximation used to compute the characteristics.

3.6.2 Another stability result of the semidiscretized scheme

In order to obtain error estimates, it will be useful to study the stability of a scheme with a more general right hand side. More precisely, let us consider the following problem:

$$\left\{ \begin{array}{l} \text{Given } \phi_{\Delta t}^0, \text{ find } \widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N \in [H_{\Gamma_D}^1(\Omega)]^N \text{ such that} \\ \langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \psi \rangle = \langle \mathcal{H}_{\Delta t}^{n+\frac{1}{2}}, \psi \rangle \quad \forall \psi \in H_{\Gamma_D}^1(\Omega) \text{ and } n = 0, \dots, N-1, \end{array} \right. \quad (3.97)$$

with

$$\langle \mathcal{H}_{\Delta t}^{n+\frac{1}{2}}, \psi \rangle = \langle F^{n+1}, \psi \rangle + \langle G^{n+1}, \psi \rangle_{\Gamma_R}. \quad (3.98)$$

for $F^{n+1} \in L^2(\Omega)$ and $G^{n+1} \in L^2(\Gamma_R)$.

We will require the following hypothesis:

Hypothesis 3.8 $F^{n+1} \in L^2(\Omega)$ and $G^{n+1} \in L^2(\Gamma_R)$, for $0 \leq n \leq N-1$.

Constant $\alpha > 0$ of the Robin boundary condition (3.3) appears explicitly in the following lemmas, however, they remain valid for any positive constant.

Lemma 3.8 *Let us assume Hypothesis 3.2 and Hypothesis 3.8. If $c_1 \Delta t < 1$, then*

$$\begin{aligned} \langle F^{n+1}, \psi \rangle + \langle G^{n+1}, \psi \rangle_{\Gamma_R} &\leq \frac{1}{2} (\|F^{n+1}\|_0^2 + \|\psi\|_0^2) + \langle H^{n+1}, \psi - \varphi \rangle_{\Gamma_R} \\ &+ \frac{21}{\alpha} \|H^{n+1}\|_{0,\Gamma_R}^2 + \frac{\alpha c_1 \Delta t}{2} \|\varphi\|_{0,\Gamma_R}^2 + \frac{\alpha}{16} \|\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2, \end{aligned} \quad (3.99)$$

with $H^{n+1} := G^{n+1}/(2 + \Delta t \operatorname{div} \mathbf{v}^{n+1})$, $\forall \varphi, \psi \in H^1(\Omega)$ and $\alpha > 0$.

PROOF. Let us introduce the notation

$$I_1 = \langle F^{n+1}, \psi \rangle, \quad I_2 = \langle G^{n+1}, \psi \rangle_{\Gamma_R}.$$

While for I_1 we only need to apply Cauchy-Schwartz inequality, for I_2 the computations are similar to those of Lemma 3.7, with slight changes. Let us note first that function H^{n+1} is well defined under hypothesis $c_1 \Delta t < 1$. Then I_2 is decomposed into three terms, namely, $I_2 = I_2^1 + I_2^2 + I_2^3$, where

$$\begin{aligned} I_2^1 &= \langle H^{n+1}, \psi \rangle_{\Gamma_R} - \langle H^{n+1}, \varphi \rangle_{\Gamma_R}, \\ I_2^2 &= \langle H^{n+1}, \varphi \rangle_{\Gamma_R} - \langle H^{n+1}(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \rangle_{\Gamma_R}, \\ I_2^3 &= \langle H^{n+1}(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \rangle_{\Gamma_R} \\ &\quad + \langle H^{n+1}(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \psi \rangle_{\Gamma_R}. \end{aligned}$$

In order to estimate I_2^2 we recall computations (3.94), obtaining

$$\begin{aligned} I_2^2 &\leq 3c_1 \Delta t \|\varphi\|_{0,\Gamma_R} \|H^{n+1}\|_{0,\Gamma_R} \leq \frac{c_1 \Delta t}{2} \left(\alpha \|\varphi\|_{0,\Gamma_R}^2 + \frac{9}{\alpha} \|H^{n+1}\|_{0,\Gamma_R}^2 \right) \\ &\leq \frac{c_1 \alpha \Delta t}{2} \|\varphi\|_{0,\Gamma_R}^2 + \frac{9}{2\alpha} \|H^{n+1}\|_{0,\Gamma_R}^2, \end{aligned}$$

where we have used inequality (3.77) with $a = \|\varphi\|_{0,\Gamma_R}$, $b = 3\|H^{n+1}\|_{0,\Gamma_R}$, and $c = \alpha$.

For I_2^3 , inequality (3.77) and estimate $(1 + c_1\Delta t)^2 \leq 4$ lead to

$$\begin{aligned} I_2^3 &= \langle H^{n+1}(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}), \psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \rangle_{\Gamma_R} \\ &\leq \frac{4}{\alpha} \|H^{n+1}(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 + \frac{\alpha}{16} \|\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 \\ &\leq \frac{16}{\alpha} \|H^{n+1}\|_{0,\Gamma_R}^2 + \frac{\alpha}{16} \|\psi + \varphi(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2. \end{aligned}$$

Finally, by jointly considering the above inequalities we get (3.99). $\square$

The following lemma involves functions defined on Γ_R and velocity field $\mathbf{v}$.

Lemma 3.9 *Let us assume Hypothesis 3.2. Let $\{G^n\}_{n=0}^N \in [L^2(\Gamma_R)]^{N+1}$ and $\{H^n\}_{n=0}^N$ be defined as in Lemma 3.8. If $c_1\Delta t < 1$, then $\{H^n\}_{n=0}^N \in [L^2(\Gamma_R)]^{N+1}$ and*

$$\|H^n\|_{0,\Gamma_R} \leq \|G^n\|_{0,\Gamma_R}, \quad (3.100)$$

where H^n has been introduced in Lemma 3.8. Moreover, for any sequence $\{\psi^n\}_{n=0}^N \in [L^2(\Gamma_R)]^{N+1}$ and any integer $q \in \{0, \dots, N-1\}$, the following inequality holds:

$$\begin{aligned} \left| \sum_{n=0}^{q-1} \langle H^{n+1}, \psi^{n+1} - \psi^n \rangle_{\Gamma_R} \right| &\leq \frac{\alpha}{16} \|\psi^q\|_{0,\Gamma_R}^2 + \frac{4}{\alpha} \|G^q\|_{0,\Gamma_R}^2 + \frac{\alpha}{16} \|\psi^0\|_{0,\Gamma_R}^2 + \frac{4}{\alpha} \|G^1\|_{0,\Gamma_R}^2 \\ &+ 3\alpha\Delta t \sum_{n=1}^{q-1} \|\psi^n\|_{0,\Gamma_R}^2 + \frac{\Delta t}{\alpha} \sum_{n=1}^{q-1} \left\| \frac{G^{n+1} - G^n}{\Delta t} \right\|_{0,\Gamma_R}^2 + \frac{c_1\Delta t}{\alpha} \sum_{n=1}^q \|G^n\|_{0,\Gamma_R}^2, \end{aligned} \quad (3.101)$$

for any $\alpha > 0$.

PROOF. Firstly, since $c_1\Delta t < 1$, then

$$\frac{1}{3} \leq \frac{1}{2 + \Delta t \operatorname{div} \mathbf{v}^n} \leq 1. \quad (3.102)$$

Consequently, $|H^n(\mathbf{x})| \leq |G^n(\mathbf{x})|$ a.e. $\mathbf{x} \in \Gamma_R$, $H^n \in L^2(\Gamma_R)$ and (3.100) holds.

For the second part of we use the equality

$$\sum_{n=0}^{q-1} \langle H^{n+1}, \psi^{n+1} - \psi^n \rangle_{\Gamma_R} = \langle H^q, \psi^q \rangle_{\Gamma_R} - \langle H^1, \psi^0 \rangle_{\Gamma_R} - \Delta t \sum_{n=1}^{q-1} \left\langle \frac{H^{n+1} - H^n}{\Delta t}, \psi^n \right\rangle_{\Gamma_R}. \quad (3.103)$$

The third term in (3.103) can be bounded as follows

$$\begin{aligned} \Delta t \left| \sum_{n=1}^{q-1} \left\langle \frac{H^{n+1} - H^n}{\Delta t}, \psi^n \right\rangle_{\Gamma_R} \right| &\leq \frac{\Delta t}{\alpha} \sum_{n=1}^{q-1} \left\| \frac{G^{n+1} - G^n}{\Delta t} \right\|_{0,\Gamma_R}^2 \\ &+ \frac{c_1\Delta t}{\alpha} \sum_{n=0}^{q-1} \|G^n\|_{0,\Gamma_R}^2 + 3\alpha\Delta t \sum_{n=1}^{q-1} \|\psi^n\|_{0,\Gamma_R}^2, \end{aligned} \quad (3.104)$$

where we have used the equality

$$\frac{H^{n+1} - H^n}{\Delta t} = \frac{2(G^{n+1} - G^n)}{\Delta t (2 + \Delta t \operatorname{div} \mathbf{v}^{n+1})(2 + \Delta t \operatorname{div} \mathbf{v}^n)} + \frac{\operatorname{div} \mathbf{v}^n G^{n+1} - \operatorname{div} \mathbf{v}^{n+1} G^n}{(2 + \Delta t \operatorname{div} \mathbf{v}^{n+1})(2 + \Delta t \operatorname{div} \mathbf{v}^n)},$$

together with (3.77), (3.102) and Hypothesis 3.2. The first two terms in (3.103) can be bounded by using (3.77) again and (3.100), obtaining

$$\left| \langle H^i, \psi^j \rangle_{\Gamma_R} \right| \leq \frac{4}{\alpha} \|G^i\|_{0,\Gamma_R}^2 + \frac{\alpha}{16} \|\psi^j\|_{0,\Gamma_R}^2 \quad \text{for } (i, j) = (1, 0), (q, q). \quad (3.105)$$

Finally, by jointly considering (3.103), (3.104) and (3.105) we get (3.101). $\square$

Theorem 3.2 *Let us assume hypotheses 3.2, 3.3, 3.4, 3.5, 3.8. Let $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ be the solution of (3.97) subject to initial value $\phi_{\Delta t}^0 \in H^1(\Omega)$. Let $\alpha > 0$ be the constant appearing in (3.3). Then, there exist two positive constants c and d , such that, if $\Delta t < d$ then*

$$\begin{aligned} & \frac{1}{\sqrt{2}} \left\| \widehat{\phi_{\Delta t}} \right\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\delta \Delta t}{4}} \left\| \mathbf{B} \operatorname{grad} \widehat{\phi_{\Delta t}} \right\|_{l^\infty(L^2(\Omega))} \\ & + \sqrt{\frac{\Delta t}{4}} \left\| \sqrt{l} \widehat{\phi_{\Delta t}} \right\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\alpha \Delta t}{16}} \left\| \widehat{\phi_{\Delta t}} \right\|_{l^\infty(L^2(\Gamma_R))} \\ & \leq c \left(\frac{1}{2} \|\phi_{\Delta t}^0\|_0 + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^0\|_0 + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \phi_{\Delta t}^0\|_0 + \sqrt{\frac{\alpha \Delta t}{16}} \|\phi_{\Delta t}^0\|_{0,\Gamma_R} \right. \\ & \quad \left. + \left\| \widehat{F} \right\|_{l^2(L^2(\Omega))} + \left\| \widehat{G} \right\|_{l^2(L^2(\Gamma_R))} + \Delta t \left\| \widehat{D_{\Delta t} G} \right\|_{l^2(L^2(\Gamma_R))} \right), \end{aligned} \quad (3.106)$$

where $\mathbf{B} \operatorname{grad} \widehat{\phi_{\Delta t}} := \{\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\}$, $\widehat{F} = \{F^n\}_{n=0}^N$ and $\widehat{G} = \{G^n\}_{n=0}^N$.

PROOF. Sequence $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ satisfies

$$\left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \phi_{\Delta t}^{n+1} \right\rangle = \left\langle \mathcal{H}_{\Delta t}^{n+\frac{1}{2}}, \phi_{\Delta t}^{n+1} \right\rangle.$$

We use first Lemma 3.6 in order to obtain a lower bound of the above expression, namely,

$$\begin{aligned} & \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \phi_{\Delta t}^{n+1} \right\rangle \\ & \geq D_{\Delta t}^n \left(\frac{1}{2} \left\| \widehat{\phi_{\Delta t}} \right\|_0^2 + \frac{\Delta t}{4} \left\| \mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}} \right\|_0^2 + \frac{\Delta t}{4} \left\| \sqrt{l} \widehat{\phi_{\Delta t}} \right\|_0^2 + \frac{\alpha \Delta t}{4} \left\| \widehat{\phi_{\Delta t}} \right\|_{0,\Gamma_R}^2 \right) \\ & \quad + \frac{\alpha}{4} \|\phi_{\Delta t}^{n+1} + \phi_{\Delta t}^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 - c \left(\|\phi_{\Delta t}^n\|_0^2 + \|\phi_{\Delta t}^{n+1}\|_0^2 \right) \\ & \quad - c \Delta t \left(\delta \left(\|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1}\|_0^2 \right) \right. \\ & \quad \left. + \left\| \sqrt{l} \phi_{\Delta t}^n \right\|_0^2 + \left\| \sqrt{l} \phi_{\Delta t}^{n+1} \right\|_0^2 + \alpha \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2 \right). \end{aligned}$$

Secondly, we use Lemma 3.8 for $\psi = \phi_{\Delta t}^{n+1}$ and $\varphi = \phi_{\Delta t}^n$, to obtain the upper bound

$$\begin{aligned} \left\langle \mathcal{H}_{\Delta t}^{n+\frac{1}{2}}, \phi_{\Delta t}^{n+1} \right\rangle & \leq \frac{1}{2} \left(\|F^{n+1}\|_0^2 + \|\phi^{n+1}\|_0^2 \right) + \langle H^{n+1}, \phi^{n+1} - \phi^n \rangle_{\Gamma_R} \\ & \quad + \frac{21}{\alpha} \|H^{n+1}\|_{0,\Gamma_R}^2 + \frac{\alpha c_1 \Delta t}{2} \|\phi^n\|_{0,\Gamma_R}^2 \\ & \quad + \frac{\alpha}{16} \|\phi^{n+1} + \phi^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2, \end{aligned}$$

where H^n has been defined in Lemma 3.8. Next, by jointly considering both estimates, regrouping and simplifying terms we get

$$\begin{aligned} & D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{\phi_{\Delta t}}\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \widehat{\phi_{\Delta t}}\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \widehat{\phi_{\Delta t}}\|_0^2 + \frac{\alpha \Delta t}{4} \|\widehat{\phi_{\Delta t}}\|_{0,\Gamma_R}^2 \right) \\ & \leq \frac{1}{2} \|F^{n+1}\|_0^2 + \frac{21}{\alpha} \|H^{n+1}\|_{0,\Gamma_R}^2 + \langle H^{n+1}, \phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \rangle_{\Gamma_R} + c \left(\|\phi_{\Delta t}^n\|_0^2 + \|\phi_{\Delta t}^{n+1}\|_0^2 \right) \\ & \quad + c \Delta t \left(\delta \left(\|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^{n+1}\|_0^2 \right) \right. \\ & \quad \left. + \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 + \|\sqrt{l} \phi_{\Delta t}^{n+1}\|_0^2 + 2\alpha \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2 \right), \end{aligned}$$

with $c = \max \{1, c_1, c_2, (2c_1c_2 + c_1c_2^2)/\delta, c_1c_3/\gamma\}$. Now, for fixed integer $q \geq 1$, let us sum inequality (3.107) multiplied by Δt from $n = 0$ to $n = q - 1$. We obtain

$$\begin{aligned} & \frac{1}{2} \|\phi_{\Delta t}^q\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \phi_{\Delta t}^q\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^q\|_0^2 + \frac{\alpha \Delta t}{4} \|\phi_{\Delta t}^q\|_{0,\Gamma_R}^2 \\ & - \frac{1}{2} \|\phi_{\Delta t}^0\|_0^2 - \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \phi_{\Delta t}^0\|_0^2 - \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^0\|_0^2 - \frac{\alpha \Delta t}{4} \|\phi_{\Delta t}^0\|_{0,\Gamma_R}^2 \\ & \leq \frac{\Delta t}{2} \sum_{n=1}^q \|F^n\|_0^2 + \frac{21\Delta t}{\alpha} \sum_{n=1}^q \|H^n\|_{0,\Gamma_R}^2 + \Delta t \sum_{n=0}^{q-1} \langle H^{n+1}, \phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \rangle_{\Gamma_R} \\ & \quad + 2c\Delta t \sum_{n=0}^q \|\phi_{\Delta t}^n\|_0^2 + 2c\Delta t^2 \left(\sum_{n=0}^q \delta \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \sum_{n=0}^q \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 + \sum_{n=0}^{q-1} \alpha \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2 \right). \end{aligned}$$

Now, the estimates given in (3.70) are used, and Lemma 3.9, for $\psi^n = \phi^n$, is applied to terms with H^n leading to

$$\begin{aligned} & \frac{1}{2} \|\phi_{\Delta t}^q\|_0^2 + \frac{\delta \Delta t}{4} \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^q\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^q\|_0^2 + \frac{3\alpha \Delta t}{16} \|\phi_{\Delta t}^q\|_{0,\Gamma_R}^2 \\ & \leq \frac{1}{2} \|\phi_{\Delta t}^0\|_0^2 + \frac{c_2 \Delta t}{4} \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^0\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^0\|_0^2 + \frac{5\alpha \Delta t}{16} \|\phi_{\Delta t}^0\|_{0,\Gamma_R}^2 \\ & \quad + \frac{\Delta t}{2} \sum_{n=1}^q \|F^n\|_0^2 + \frac{25\Delta t}{\alpha} \sum_{n=1}^q \|G^n\|_{0,\Gamma_R}^2 + \frac{c_1 \Delta t^2}{\alpha} \sum_{n=1}^q \|G^n\|_{0,\Gamma_R}^2 + \frac{\Delta t^2}{\alpha} \sum_{n=1}^{q-1} \left\| \frac{G^{n+1} - G^n}{\Delta t} \right\|_{0,\Gamma_R}^2 \\ & \quad + 2c\Delta t \sum_{n=0}^q \|\phi_{\Delta t}^n\|_0^2 + 2c\Delta t^2 \left(\sum_{n=0}^q \delta \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \sum_{n=0}^q \|\sqrt{l} \phi_{\Delta t}^n\|_0^2 + \frac{5}{2} \sum_{n=1}^{q-1} \alpha \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2 \right). \end{aligned}$$

Let us introduce, for $n = 0, \dots, N$, the notation:

$$\begin{aligned} \theta_n &:= \frac{1}{2} \|\phi_{\Delta t}^n\|_0^2 + \frac{\delta \Delta t}{4} \|\mathbf{B} \operatorname{grad} \phi_{\Delta t}^n\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \phi_{\Delta t}^n\|_0^2, \\ \bar{\theta}_n &:= \frac{\alpha \Delta t}{16} \|\phi_{\Delta t}^n\|_{0,\Gamma_R}^2. \end{aligned}$$

With the above notation we have

$$\begin{aligned} (1 - 8c\Delta t)\theta_q + \bar{\theta}_q &\leq 8c\Delta t \sum_{n=0}^{q-1} \theta_n + 80c\Delta t \sum_{n=0}^{q-1} \bar{\theta}_n \\ &\quad + \tilde{c} \left(\theta_0 + \bar{\theta}_0 + \|\widehat{F}\|_{l^2(L^2(\Omega))}^2 + \|\widehat{G}\|_{l^2(L^2(\Gamma_R))}^2 + \Delta t \|\widehat{D_{\Delta t} G}\|_{l^2(L^2(\Gamma_R))}^2 \right), \end{aligned}$$

with $c = \max \{1, c_1, c_2, (2c_1c_2 + c_1^2c_2^2)/\delta, c_1c_3/\gamma\}$, and $\tilde{c}$ a positive constant. For Δt small enough, we can apply the discrete Gronwall inequality (see for instance [95]), and take the maximum in $q \in \{1, \dots, N\}$. Thus, estimate (3.106) follows. $\square$

Remark 3.17 *If we assume that $G^n = G(t_n)$, $n = 0, \dots, N$ for a function $G \in C^1(L^2(\Gamma_R))$, we can use Barrow's rule and Holder inequality to obtain a bound for $\|\widehat{D_{\Delta t} G}\|_{l^2(L^2(\Gamma_R))}$. More precisely,*

$$\begin{aligned} \|\widehat{D_{\Delta t} G}\|_{l^2(L^2(\Gamma_R))}^2 &= \Delta t \sum_{n=0}^{N-1} \left\| \frac{G^{n+1} - G^n}{\Delta t} \right\|_{0, \Gamma_R}^2 \\ &\leq \Delta t \sum_{n=0}^{N-1} \frac{1}{\Delta t} \left\| \frac{\partial G}{\partial t} \right\|_{L^2((t_n, t_{n+1}); L^2(\Gamma_R))}^2 = \left\| \frac{\partial G}{\partial t} \right\|_{L^2(L^2(\Gamma_R))}^2. \end{aligned}$$

We can also obtain a stability result for problem (3.69) by using Theorem 3.2, which has one more term than Theorem 3.1.

Corollary 3.3 (Stability) *Let us assume Hypotheses 3.2 to 3.7. Let $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ be the solution of (3.69) subject to initial value $\phi_{\Delta t}^0$. Then, there exist two positive constants, c and d , such that, for $\Delta t < d$, we have*

$$\begin{aligned} &\frac{1}{\sqrt{2}} \|\widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \text{ grad } \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} \\ &+ \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\alpha \Delta t}{16}} \|\widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Gamma_R))} \\ &\leq c \left(\frac{1}{2} \|\phi_{\Delta t}^0\|_0 + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \text{ grad } \phi_{\Delta t}^0\|_0 + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \phi_{\Delta t}^0\|_0 \right. \\ &\quad \left. + \sqrt{\frac{\alpha \Delta t}{16}} \|\phi_{\Delta t}^0\|_{0, \Gamma_R} + \|\widehat{f}\|_{l^2(L^2(\Omega))} + \|\widehat{g}\|_{l^2(L^2(\Gamma_R))} + \Delta t \|\widehat{D_{\Delta t} g}\|_{l^2(L^2(\Gamma_R))} \right), \end{aligned}$$

where $\mathbf{B} \text{ grad } \widehat{\phi_{\Delta t}} := \{\mathbf{B} \text{ grad } \phi_{\Delta t}^n\}$.

PROOF. The result straightly follows by replacing F^{n+1} with $f^{n+1} + f^n \circ X_E^n$ and G^{n+1} with $g^{n+1} + g^n(1 + \Delta t \text{ div } \mathbf{v}^{n+1})$ in (3.106). $\square$

3.6.3 Error estimate of the semidiscretized scheme

The aim of the present section is to estimate the difference between the *discrete solution* of (3.69), $\widehat{\phi_{\Delta t}} := \{\phi_{\Delta t}^n\}_{n=0}^N$, and $\widehat{\phi} = \{\phi^n\}_{n=0}^N$, the exact solution of the continuous problem. This estimate is proved in Theorem 3.3 and requires additional regularity to the solution of the continuous problem and to its data.

Firstly, for $n = 0, 1, 2, \dots$, we extend the definition of X_e^n and $\mathbf{F}_e^n$ to intermediate times, namely,

$$X_e^{n+\frac{1}{2}}(\mathbf{x}) := X_e(\mathbf{x}, t_{n+1}; t_{n+\frac{1}{2}}), \quad \mathbf{F}_e^{n+\frac{1}{2}}(\mathbf{x}) := \mathbf{F}_e(\mathbf{x}, t_{n+1}; t_{n+\frac{1}{2}}).$$

According to (3.46) for $t = t_{n+1}$ and $\tau = t_{n+\frac{1}{2}}$, the exact solution solves the problem

$$\langle \mathcal{L}^{n+\frac{1}{2}} \widehat{\phi}, \psi \rangle = \langle \mathcal{F}^{n+\frac{1}{2}}, \psi \rangle \quad \forall \psi \in H_{\Gamma_D}^1(\Omega), \quad (3.107)$$

where $\mathcal{L}^{n+\frac{1}{2}} \widehat{\phi} \in (H^1(\Omega))'$ and $\mathcal{F}^{n+\frac{1}{2}} \in (H^1(\Omega))'$ are defined by

$$\begin{aligned} \langle \mathcal{L}^{n+\frac{1}{2}} \widehat{\phi}, \psi \rangle &:= \left\langle \left(\dot{\phi} \right)^{n+\frac{1}{2}} \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle + \left\langle \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} \left(\mathbf{A} \operatorname{grad} \phi^{n+\frac{1}{2}} \right) \circ X_e^{n+\frac{1}{2}}, \operatorname{grad} \psi \right\rangle \\ &\quad + \left\langle \operatorname{Div} \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-T} \cdot \left(\mathbf{A} \operatorname{grad} \phi^{n+\frac{1}{2}} \right) \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle \\ &\quad + \left\langle \left(l \phi^{n+\frac{1}{2}} \right) \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle + \alpha \left\langle \det \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} \phi^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R}, \\ \langle \mathcal{F}^{n+\frac{1}{2}}, \psi \rangle &:= \left\langle f^{n+\frac{1}{2}} \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle + \left\langle \det \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} g^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R}, \end{aligned}$$

for $\psi \in H^1(\Omega)$.

The error estimate result to be stated in Theorem 3.3, is proved by means of Theorem 3.2 and the forthcoming Lemmas 3.18 and 3.19. In what follows, for the sake of clarity, we prove some results which will be used in Lemmas 3.18 and 3.19. Some auxiliary functions will be introduced. They will be denoted by ξ, ϑ and Ψ if the function is scalar, vectorial and a matrix, respectively. Moreover, $\tilde{c}_1$ denotes a generic positive constant, related to the norm of the velocity field $\mathbf{v}$, not necessarily the same at each occurrence.

Lemma 3.10 *Let us assume that $\mathbf{v} \in C^2(L^\infty(\Omega)) \cap C^1(W^{1,\infty}(\Omega)) \cap C^0(W^{2,\infty}(\Omega))$ and vanishes on the boundary, $\Delta t \|\mathbf{v}\|_{C^0(W^{1,\infty}(\Omega))} < 1/2$, and $\varphi \in Z^3$. Let us define function $\xi^{n+\frac{1}{2}}$ by*

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) := \dot{\varphi}^{n+\frac{1}{2}} \left(X_e^{n+\frac{1}{2}}(\mathbf{x}) \right) - \frac{\varphi^{n+1}(\mathbf{x}) - \varphi^n(X_{RK}^n(\mathbf{x}))}{\Delta t} \quad \forall \mathbf{x} \in \Omega.$$

Then $\xi^{n+\frac{1}{2}} \in L^2(\Omega)$ and the following estimate holds:

$$\|\xi^{n+\frac{1}{2}}\|_0 \leq \tilde{c}_1 \Delta t^2 \|\varphi\|_{Z^3}, \quad n = 0, \dots, N-1.$$

PROOF. Let us first divide function $\xi^{n+\frac{1}{2}}$ into two parts $\xi^{n+\frac{1}{2}}(\mathbf{x}) = \xi_1^{n+\frac{1}{2}}(\mathbf{x}) + \xi_2^{n+\frac{1}{2}}(\mathbf{x})$ with

$$\begin{aligned} \xi_1^{n+\frac{1}{2}}(\mathbf{x}) &:= \dot{\varphi}^{n+\frac{1}{2}} \left(X_e^{n+\frac{1}{2}}(\mathbf{x}) \right) - \frac{\varphi^{n+1}(\mathbf{x}) - \varphi^n(X_e^n(\mathbf{x}))}{\Delta t} \quad \forall \mathbf{x} \in \Omega, \\ \xi_2^{n+\frac{1}{2}}(\mathbf{x}) &:= \frac{\varphi^n(X_e^n(\mathbf{x})) - \varphi^n(X_{RK}^n(\mathbf{x}))}{\Delta t} \quad \forall \mathbf{x} \in \Omega. \end{aligned}$$

Firstly, by using the Taylor expansion of degree three of the function

$$G(\tau) := \varphi(X_e(\mathbf{x}, t_{n+1}; \tau), \tau),$$

around $\tau = t_{n+\frac{1}{2}}$, easily we get

$$\frac{G(t_{n+1}) - G(t_n)}{\Delta t} = G'(t_{n+\frac{1}{2}}) + \frac{1}{\Delta t} \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} (t_{n+1} - s)^2 G'''(s) ds - \frac{1}{\Delta t} \int_{t_{n+\frac{1}{2}}}^{t_n} (t_n - s)^2 G'''(s) ds.$$

Now, we rewrite this expression in terms of φ , obtaining

$$\begin{aligned} \frac{\varphi^{n+1}(\mathbf{x}) - \varphi^n(X_e^n(\mathbf{x}))}{\Delta t} &= \dot{\varphi}^{n+\frac{1}{2}}\left(X_e^{n+\frac{1}{2}}(\mathbf{x})\right) + \frac{1}{\Delta t} \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} (t_{n+1} - s)^2 \ddot{\varphi}(X_e(\mathbf{x}, t_{n+1}; s), s) ds \\ &\quad - \frac{1}{\Delta t} \int_{t_{n+\frac{1}{2}}}^{t_n} (t_n - s)^2 \ddot{\varphi}(X_e(\mathbf{x}, t_{n+1}; s), s) ds. \end{aligned}$$

Moreover, by computing $\ddot{\varphi}(X_e(\mathbf{x}, t_{n+1}; s), s)$ it is easy to show that $\xi_1^{n+\frac{1}{2}} \in L^2(\Omega)$ and

$$\left\| \xi_1^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \|\varphi\|_{Z^3}.$$

Next, we use the Taylor expansion of degree one of the function $G(\mathbf{y}) := \varphi(\mathbf{y}, t_n)$ around point $X_{RK}^n(\mathbf{x})$, obtaining

$$\begin{aligned} \varphi(X_e^n(\mathbf{x}), t_n) - \varphi(X_{RK}^n(\mathbf{x}), t_n) \\ = \text{grad } \varphi(X_{RK}^n(\mathbf{x}) + \theta(X_e^n(\mathbf{x}) - X_{RK}^n(\mathbf{x}))) \cdot (X_e^n(\mathbf{x}) - X_{RK}^n(\mathbf{x})), \end{aligned}$$

for some number $\theta \in (0, 1)$. By noticing that $X_{RK}^n(\mathbf{x})$ is a second order approximation of $X_e^n(\mathbf{x})$, namely,

$$|X_e^n(\mathbf{x}) - X_{RK}^n(\mathbf{x})| \leq \tilde{c}_1 \Delta t^3,$$

we deduce that

$$\left\| \xi_2^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \|\varphi^n\|_{H^1(\Omega)},$$

and the result is concluded. $\square$

Lemma 3.11 *Let us assume that $\mathbf{v} \in C^1(L^\infty(\Omega)) \cap C^0(W^{1,\infty}(\Omega))$ and vanishes on the boundary. Let $\varphi : \Omega \times [0, T] \rightarrow \mathbb{R}$, $\varphi \in Z^2$ be a given function and $\xi^{n+\frac{1}{2}}$ be defined by*

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) := \varphi(X_e^{n+\frac{1}{2}}(\mathbf{x}), t_{n+\frac{1}{2}}) - \frac{\varphi(\mathbf{x}, t_{n+1}) + \varphi(X_e^n(\mathbf{x}), t_n)}{2} \quad \forall \mathbf{x} \in \Omega.$$

Then, $\xi^{n+\frac{1}{2}} \in L^2(\Omega)$ and we have

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) = -\frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} (t_{n+1} - s) \ddot{\varphi}(X_e(\mathbf{x}, t_{n+1}; s), s) ds - \frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_n} (t_n - s) \ddot{\varphi}(X_e(\mathbf{x}, t_{n+1}; s), s) ds,$$

a.e. $\mathbf{x} \in \Omega$, $n = 0, \dots, N-1$.

PROOF. For $\tau \in (0, T)$ let us introduce the auxiliary function $G(\tau) := \varphi(X_e(\mathbf{x}, t_{n+1}; \tau), \tau)$. We have

$$G(t) = G(t_{n+\frac{1}{2}}) + (t - t_{n+\frac{1}{2}})G'(t_{n+\frac{1}{2}}) + \int_{t_{n+\frac{1}{2}}}^t (t - s)G''(s)ds.$$

The result is obtained by replacing successively t by t_n and t_{n+1} and summing both expressions. $\square$

Lemma 3.12 *Let us assume that $\mathbf{v} \in C^1(W^{1,\infty}(\Omega)) \cap C^0(W^{2,\infty}(\Omega))$ and vanishes on the boundary. Let $\mathbf{w} : \Omega \times [0, T] \longrightarrow \mathbb{R}^m$, $\mathbf{w} \in Z^2$ be a given function and $\vartheta^{n+\frac{1}{2}}$ be defined by*

$$\vartheta^{n+\frac{1}{2}}(\mathbf{x}) := \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; t_{n+\frac{1}{2}}) \mathbf{w}(X_e^{n+\frac{1}{2}}(\mathbf{x}), t_{n+\frac{1}{2}}) - \frac{\mathbf{w}(\mathbf{x}, t_{n+1}) + (\mathbf{F}_e^n)^{-1}(\mathbf{x}) \mathbf{w}(X_e^n(\mathbf{x}), t_n)}{2}$$

for all $\mathbf{x} \in \Omega$. Then, $\vartheta^{n+\frac{1}{2}} \in L^2(\Omega)$ and the following estimate holds

$$\|\vartheta^{n+\frac{1}{2}}\|_0 \leq \tilde{c}_1 \Delta t^2 \|\mathbf{w}\|_{Z^2}, \quad n = 0, \dots, N-1. \quad (3.108)$$

Moreover, if $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$ and $\mathbf{w} \in C^i(H^{3-i}(\Omega))$, $i = 0, 1, 2$ then $\vartheta^{n+\frac{1}{2}} \in H^1(\Omega)$ and

$$\|\text{Div } \vartheta^{n+\frac{1}{2}}\|_0 \leq \tilde{c}_1 \Delta t^2 \|\text{Div } \mathbf{w}\|_{Z^2}, \quad n = 0, \dots, N-1. \quad (3.109)$$

PROOF. We introduce an auxiliary vector function

$$G(\tau) := \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; \tau) \mathbf{w}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau),$$

and apply the Taylor formula, obtaining

$$G(t_{n+\frac{1}{2}}) - \frac{G(t_{n+1}) + G(t_n)}{2} = -\frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} (t_{n+1} - s) G''(s) ds - \frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_n} (t_n - s) G''(s) ds.$$

If we compute the second time derivative appearing in the above integrals we obtain

$$\begin{aligned} G''_i(s) &= \sum_j \frac{\partial^2}{\partial \tau^2} [\mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; \tau)]_{ij} \mathbf{w}_j(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) \\ &\quad + 2 \sum_j \frac{\partial}{\partial \tau} [\mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; \tau)]_{ij} \dot{\mathbf{w}}_j(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) \\ &\quad + \sum_j [\mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; \tau)]_{ij} \ddot{\mathbf{w}}_j(X_e(\mathbf{x}, t_{n+1}; \tau), \tau), \quad \text{for } i = 1, \dots, m. \end{aligned}$$

In view of the above expression we can discuss the regularity of $\vartheta^{n+\frac{1}{2}}$:

- If $\mathbf{w} \in Z^2$ and $\mathbf{v} \in C^0(W^{2,\infty}(\Omega)) \cap C^1(W^{1,\infty}(\Omega))$ then $\vartheta^{n+\frac{1}{2}} \in L^2(\Omega)$ and estimate (3.108) holds (we use Proposition 3.6).
- If $\mathbf{w} \in C^i(H^{3-i}(\Omega))$ for $i = 0, 1, 2$ and $\mathbf{v} \in C^0(W^{3,\infty}(\Omega)) \cap C^1(W^{2,\infty}(\Omega))$ then $\vartheta^{n+\frac{1}{2}} \in H^1(\Omega)$ and estimate (3.109) holds (we use Remark 3.4).

□

Lemma 3.13 *Let us assume that $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$ and vanishes on the boundary. Let $\mathbf{w} : \Omega \times [0, T] \longrightarrow \mathbb{R}^m$, $\mathbf{w} \in Z^2$ be a given function and $\xi^{n+\frac{1}{2}}$ be defined by*

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) := \text{Div } \mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; t_{n+\frac{1}{2}}) \cdot \mathbf{w}(X_e^{n+\frac{1}{2}}(\mathbf{x}), t_{n+\frac{1}{2}}) - \frac{\text{Div } (\mathbf{F}_e^n)^{-T}(\mathbf{x}) \mathbf{w}(X_e^n(\mathbf{x}), t_n)}{2}$$

for all $\mathbf{x} \in \Omega$. Then, $\xi^{n+\frac{1}{2}} \in L^2(\Omega)$ and the following estimate holds

$$\|\xi^{n+\frac{1}{2}}\|_0 \leq \tilde{c}_1 \Delta t^2 \|\mathbf{w}\|_{Z^2}, \quad n = 0, \dots, N-1. \quad (3.110)$$

PROOF. We introduce an auxiliary scalar function

$$G(\tau) := \text{Div } \mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; \tau) \cdot \mathbf{w}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau),$$

and apply Taylor formula obtaining

$$G(t_{n+\frac{1}{2}}) - \frac{G(t_{n+1}) + G(t_n)}{2} = -\frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} (t_{n+1} - s) G''(s) ds - \frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_n} (t_n - s) G''(s) ds.$$

Notice that, since $\mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; t_{n+1}) = \mathbf{I}$, its divergence is zero. By developing the second time derivative appearing in the above integrals we obtain

$$\begin{aligned} G''_i(s) &= \frac{\partial^2}{\partial \tau^2} \text{Div } \mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; \tau) \cdot \mathbf{w}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) \\ &\quad + 2 \frac{\partial}{\partial \tau} \text{Div } \mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; \tau) \cdot \dot{\mathbf{w}}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) \\ &\quad + \text{Div } \mathbf{F}_e^{-T}(\mathbf{x}, t_{n+1}; \tau) \cdot \ddot{\mathbf{w}}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau). \end{aligned}$$

In order to have $\xi^{n+\frac{1}{2}} \in L^2(\Omega)$ and estimate (3.110) we need $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$ (see Remark 3.4) and $\mathbf{w} \in Z^2$. $\square$

Lemma 3.14 *Assume that $\mathbf{v} \in C^1(L^\infty(\Omega)) \cap C^0(W^{1,\infty}(\Omega))$ vanishes on the boundary and $\Delta t \|\mathbf{v}\|_{C^0(W^{1,\infty}(\Omega))} < 1$. Let $\varphi \in H^1(\Omega)$ and ξ^n be defined by*

$$\xi^n(\mathbf{x}) := \varphi(X_e^n(\mathbf{x})) - \varphi(X_E^n(\mathbf{x})) \quad \text{a.e. } \mathbf{x} \in \Omega, \quad n = 0, \dots, N-1.$$

Then $\xi^n \in L^2(\Omega)$ and

$$\xi^n(\mathbf{x}) = \text{grad } \varphi(X_E^n(\mathbf{x}) + \theta \mathbf{y}) \cdot \mathbf{y} \quad \text{a.e. } \mathbf{x} \in \Omega, \quad (3.111)$$

with $\theta \in (0, 1)$, and

$$\mathbf{y} = \int_{t_n}^{t_{n+1}} (s - t_n) \dot{\mathbf{v}}(X_e(\mathbf{x}, t_{n+1}; s), s) ds. \quad (3.112)$$

Moreover, if $\mathbf{v} \in C^1(W^{1,\infty}(\Omega)) \cap C^0(W^{2,\infty}(\Omega))$ and $\varphi \in H^2(\Omega)$ then $\xi^{n+\frac{1}{2}} \in H^1(\Omega)$.

PROOF. We use the Taylor expansion of degree one of function $\varphi \in H^1(\Omega)$ around point $X_E^n(\mathbf{x})$ evaluated at point $X_e^n(\mathbf{x})$, obtaining

$$\varphi(X_e^n(\mathbf{x}), t_n) = \varphi(X_E^n(\mathbf{x}), t_n) + \text{grad } \varphi(X_E^n(\mathbf{x}) + \theta(X_e^n(\mathbf{x}) - X_E^n(\mathbf{x}))) \cdot (X_e^n(\mathbf{x}) - X_E^n(\mathbf{x})),$$

for some number $\theta \in (0, 1)$.

Next, by applying a Taylor expansion to function $X_e(\mathbf{x}, t_{n+1}; \tau)$ with respect to τ at time t_{n+1} evaluated at time t_n , it follows that

$$X_e^n(\mathbf{x}) - X_E^n(\mathbf{x}) = \int_{t_{n+1}}^{t_n} (t_n - s) \dot{\mathbf{v}}(X_e(\mathbf{x}, t_{n+1}; s), s) ds.$$

By jointly considering both Taylor expansions then (3.111) and (3.112) follow, and the regularity of ξ^n is consequence of the regularity of φ and $\mathbf{v}$. $\square$

Corollary 3.4 *Let us assume that $\mathbf{v} \in C^1(L^\infty(\Omega)) \cap C^0(W^{1,\infty}(\Omega))$ and vanishes on the boundary and $\Delta t \|\mathbf{v}\|_{C^0(W^{1,\infty}(\Omega))} < 1$. Let $\varphi : \Omega \times [0, T] \rightarrow \mathbb{R}$, $\varphi \in Z^2$ be a given function. Let $\xi^{n+\frac{1}{2}}$ be the function defined by*

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) := \varphi(X_e^{n+\frac{1}{2}}(\mathbf{x}), t_{n+\frac{1}{2}}) - \frac{\varphi(\mathbf{x}, t_{n+1}) + \varphi(X_E^n(\mathbf{x}), t_n)}{2} \quad \forall \mathbf{x} \in \Omega.$$

Then $\xi^{n+\frac{1}{2}} \in L^2(\Omega)$ and

$$\|\xi^{n+\frac{1}{2}}\|_0 \leq \tilde{c}_1 \Delta t^2 \|\varphi\|_{Z^2}, \quad n = 0, \dots, N-1,$$

where $\tilde{c}_1$ is independent of Δt .

PROOF. It directly follows from writing $\xi^{n+\frac{1}{2}}$ as

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) := \varphi(X_e^{n+\frac{1}{2}}(\mathbf{x}), t_{n+\frac{1}{2}}) - \frac{\varphi(\mathbf{x}, t_{n+1}) + \varphi(X_e^n(\mathbf{x}), t_n)}{2} - \frac{\varphi(X_e^n(\mathbf{x}), t_n) - \varphi(X_E^n(\mathbf{x}), t_n)}{2}$$

and applying Lemma 3.11 and Lemma 3.14. $\square$

Remark 3.18 *Notice that result of Corollary 3.4 holds true for any approximation of the characteristic lines of, at least, first order.*

Lemma 3.15 *Let us assume that $\mathbf{v} \in C^1(W^{1,\infty}(\Omega)) \cap C^0(W^{2,\infty}(\Omega))$ and vanishes on the boundary. Let Ψ^n be defined by*

$$\Psi^n(\mathbf{x}) := (\mathbf{F}_e^n)^{-1}(X_e^n(\mathbf{x})) - (\mathbf{I} + \Delta t \mathbf{L}^n)(X_e^n(\mathbf{x})) \quad \forall \mathbf{x} \in \Omega.$$

Then $\Psi^n \in \mathbf{L}^2(\Omega)$ for $n = 0, \dots, N-1$ and

$$\Psi^n(\mathbf{x}) = - \int_{t_n}^{t_{n+1}} (\tau - t_{n+1}) \operatorname{grad} \dot{\mathbf{v}}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) \mathbf{F}_e(X_e^n(\mathbf{x}), t_n; \tau) d\tau \quad a.e. \mathbf{x} \in \Omega. \quad (3.113)$$

Moreover, if $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$ then $\Psi^n \in \mathbf{H}^1(\Omega)$.

PROOF. The result directly follows from replacing $t = t_{n+1}$ and $s = t_n$ in (3.16). $\square$

Lemma 3.16 *Let us assume that $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$ and vanishes on the boundary. Let ϑ^n be defined by*

$$\vartheta^n(\mathbf{x}) := \operatorname{Div} (\mathbf{F}_e^n)^{-T}(X_e^n(\mathbf{x})) - \Delta t \operatorname{grad} \operatorname{div} \mathbf{v}^n(X_e^n(\mathbf{x})) \quad \forall \mathbf{x} \in \Omega.$$

Then $\vartheta^n \in L^2(\Omega)$ for $n = 0, \dots, N-1$ and

$$\vartheta^n(\mathbf{x}) = - \int_{t_n}^{t_{n+1}} (\tau - t_{n+1}) \operatorname{Div} ((\mathbf{F}_e)^T (\operatorname{grad} \dot{\mathbf{v}})^T) (X_e(\mathbf{x}, t_{n+1}; \tau)) d\tau \quad a.e. \mathbf{x} \in \Omega.$$

PROOF. The regularity assumed for $\mathbf{v}$ allows us to apply Lemma 3.15 and define a function $\Psi^n \in \mathbf{H}^1(\Omega)$ by (3.113). The result directly follows from applying the Div operator to this function and by taking into account that

$$\operatorname{div} (\mathbf{I} + \Delta t (\mathbf{L}^n)^T) (\mathbf{y}) = \Delta t \operatorname{grad} \operatorname{div} \mathbf{v}^n(\mathbf{y}).$$

$\square$

Whereas the previous lemmas concern the approximation of functions defined on the whole domain Ω , the following lemma gives a second order approximation of a function defined on the boundary Γ_R . Notice that no characteristic lines are involved since we have assumed a velocity field vanishing on the boundary.

Lemma 3.17 *Let $\mathbf{v} \in C^0(W^{2,\infty}(\Omega)) \cap C^1(W^{1,\infty}(\Omega))$ vanishing on the boundary. Let $\varphi : \Gamma_R \times [0, T] \longrightarrow \mathbb{R}$, $\varphi \in C^2(L^2(\Gamma_R))$ and let $\xi^{n+\frac{1}{2}}$ be a function defined on the boundary Γ_R by*

$$\xi^{n+\frac{1}{2}}(\mathbf{x}) := \det \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1}(\mathbf{x}) \varphi(\mathbf{x}, t_{n+\frac{1}{2}}) - \frac{\varphi(\mathbf{x}, t_{n+1}) + (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x})) \varphi(\mathbf{x}, t_n)}{2}$$

a.e. $\mathbf{x} \in \Gamma_R$. Then $\xi^{n+\frac{1}{2}} \in L^2(\Gamma_R)$ and

$$\|\xi^{n+\frac{1}{2}}\|_{0,\Gamma_R} \leq \tilde{c}_1 \Delta t^2 \|\varphi\|_{C^2(L^2(\Gamma_R))} \quad n = 0, \dots, N-1$$

and with $\tilde{c}_1$ is independent of Δt .

PROOF. Let us first divide function $\xi^{n+\frac{1}{2}}$ into two parts $\xi^{n+\frac{1}{2}}(\mathbf{x}) = \xi_1^{n+\frac{1}{2}}(\mathbf{x}) - \xi_2^n(\mathbf{x})$ with

$$\begin{aligned} \xi_1^{n+\frac{1}{2}}(\mathbf{x}) &:= \det \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1}(\mathbf{x}) \varphi(\mathbf{x}, t_{n+\frac{1}{2}}) - \frac{\varphi(\mathbf{x}, t_{n+1}) + \det(\mathbf{F}_e^n)^{-1}(\mathbf{x}) \varphi(\mathbf{x}, t_n)}{2}, \\ \xi_2^n(\mathbf{x}) &:= \frac{\det(\mathbf{F}_e^n)^{-1}(\mathbf{x}) \varphi(\mathbf{x}, t_n) - (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x})) \varphi(\mathbf{x}, t_n)}{2}. \end{aligned}$$

Firstly we consider Taylor expansion of function

$$G(\tau) = \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; \tau) \varphi(\mathbf{x}, \tau)$$

at point $\tau = t_{n+\frac{1}{2}}$ evaluated at t_n and t_{n+1} and sum both expressions. We obtain

$$\xi_1^{n+\frac{1}{2}}(\mathbf{x}) = -\frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} (t_{n+1} - s) G''(s) ds - \frac{1}{2} \int_{t_{n+\frac{1}{2}}}^{t_n} (t_n - s) G''(s) ds \quad a.e. \mathbf{x} \in \Gamma_R \quad (3.114)$$

with

$$\begin{aligned} G''(s) &= \frac{\partial^2}{\partial s^2} \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; s) \varphi(\mathbf{x}, s) + 2 \frac{\partial}{\partial s} \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; s) \frac{\partial \varphi}{\partial s}(\mathbf{x}, s) \\ &\quad + \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; s) \frac{\partial^2 \varphi}{\partial s^2}(\mathbf{x}, s). \end{aligned} \quad (3.115)$$

In order to establish a bound for the terms in $\det \mathbf{F}_e^{-1}$ in (3.115) we replace in (3.17) and (3.20) $t = t_{n+1}$ obtaining

$$\left| \frac{\partial^i}{\partial s^i} \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; s) \right| \leq c |\det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; s)| \quad \text{for } i = 1, 2,$$

where constant c depends on the number

$$\max \{ \|\mathbf{v}\|_{C^0(W^{2,\infty}(\Omega))}, \|\mathbf{v}\|_{C^1(W^{1,\infty}(\Omega))} \}. \quad (3.116)$$

Moreover, by replacing in estimate (3.18) $t = t_{n+1}$ and $\tau \in [t_n, t_{n+1}]$ we obtain

$$|\det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; \tau)| \leq e^{\|\mathbf{v}\|_{C^0(W^{1,\infty}(\Omega))} \Delta t}.$$

Thus, we have proved that $\xi_1^{n+\frac{1}{2}} \in L^2(\Gamma_R)$ and that it can be bounded by

$$\left\| \xi_1^{n+\frac{1}{2}} \right\|_{0,\Gamma_R} \leq \tilde{c}_1 \Delta t^2 \|\varphi\|_{C^2(L^2(\Gamma_R))}.$$

Secondly, the same regularity of $\mathbf{v}$ is required in order to apply Proposition 3.9 for $t = t_{n+1}$ and $s = t_n$, obtaining

$$\begin{aligned} & \left| \det (\mathbf{F}_e^n)^{-1}(\mathbf{x}) - (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x})) \right| \\ &= \left| \int_{t_n}^{t_{n+1}} (s - t_n) \frac{\partial^2}{\partial s^2} \det \mathbf{F}_e^{-1}(\mathbf{x}, t_{n+1}; s) ds \right| \leq c \Delta t^2, \end{aligned}$$

where constant c depends again on the number (3.116). Thus, $\xi_2^n \in L^2(\Gamma_R)$ and the following estimate holds

$$\|\xi_2^n\|_{0,\Gamma_R} \leq \tilde{c}_1 \Delta t^2 \|\varphi\|_{C^0(L^2(\Gamma_R))}.$$

Finally, the result for $\xi^{n+\frac{1}{2}}$ directly follows from the partial results obtained for $\xi_1^{n+\frac{1}{2}}$ and ξ_2^n . $\square$

An estimate of the difference between the continuous and the discrete “left hand side operator,” $\mathcal{L}$, (respectively, “right hand side operator,” $\mathcal{F}$), is given in Lemma 3.18 (respectively, in Lemma 3.19).

Lemma 3.18 *Assume Hypothesis 3.3, 3.4, and 3.5, and that the coefficients of the problem satisfy:*

$$\mathbf{v} \in C^0(W^{3,\infty}(\Omega)) \cap C^1(W^{2,\infty}(\Omega)) \cap C^2(L^\infty(\Omega)) \quad \mathbf{v}|_\Gamma = 0, \quad \mathbf{A} \in \mathbf{W}^{3,\infty}(\Omega), \quad l \in W^{2,\infty}(\Omega),$$

and that $\Delta t \|\mathbf{v}\|_{C^0(W^{1,\infty})} < 1/2$. Let the solution of (3.107) satisfy

$$\phi \in Z^3, \quad \operatorname{grad} \phi \in Z^3, \quad \phi|_{\Gamma_R} \in C^2(L^2(\Gamma_R)).$$

Then, for each $n = 0, 1, \dots, N-1$, there exist two functions

$$\xi_{\mathcal{L}_1}^{n+\frac{1}{2}} : \Omega \longrightarrow \mathbb{R} \qquad \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} : \Gamma_R \longrightarrow \mathbb{R},$$

such that

$$\left\langle \left(\mathcal{L}^{n+\frac{1}{2}} - \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \right) \hat{\phi}, \psi \right\rangle = \left\langle \xi_{\mathcal{L}_1}^{n+\frac{1}{2}}, \psi \right\rangle + \left\langle \xi_{\mathcal{L}_2}^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R}, \quad (3.117)$$

for all $\psi \in H_{\Gamma_D}^1(\Omega)$. Moreover, $\xi_{\mathcal{L}_1}^{n+\frac{1}{2}} \in L^2(\Omega)$ and $\xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \in L^2(\Gamma_R)$ and the following estimates hold

$$\begin{aligned} \left\| \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} \right\|_0 &\leq \tilde{c}_1 \Delta t^2 (\|\phi\|_{Z^3} + \|\mathbf{A} \operatorname{grad} \phi\|_{Z^3} + \|l\phi\|_{Z^2}), \\ \left\| \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R} &\leq \tilde{c}_1 \Delta t^2 \left(\|\mathbf{A} \operatorname{grad} \phi \cdot \mathbf{n}\|_{Z^2,\Gamma_R} + \alpha \|\phi\|_{C^2(L^2(\Gamma_R))} \right), \end{aligned}$$

where $\tilde{c}_1$ denotes a constant independent of Δt and $\alpha > 0$ is the constant appearing in (3.3).

PROOF. The left hand side of (3.117) is equal to $I_1 + I_2 + I_3 + I_4 + I_5$ with

$$\begin{aligned}
I_1 &= \left\langle \left(\dot{\phi} \right)^{n+\frac{1}{2}} \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle - \left\langle \frac{\phi^{n+1} - \phi^n \circ X_{RK}^n}{\Delta t}, \psi \right\rangle, \\
I_2 &= \left\langle \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} \left(\mathbf{A} \operatorname{grad} \phi^{n+\frac{1}{2}} \right) \circ X_e^{n+\frac{1}{2}}, \operatorname{grad} \psi \right\rangle \\
&\quad - \left\langle \frac{\mathbf{A} \operatorname{grad} \phi^{n+1} + ((\mathbf{I} + \Delta t \mathbf{L}^n) \mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n}{2}, \operatorname{grad} \psi \right\rangle \\
I_3 &= \left\langle \operatorname{Div} \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-T} \cdot \left(\mathbf{A} \operatorname{grad} \phi^{n+\frac{1}{2}} \right) \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle \\
&\quad - \left\langle \frac{\Delta t (\operatorname{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n}{2}, \psi \right\rangle, \\
I_4 &= \left\langle (l\phi^{n+\frac{1}{2}}) \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle - \left\langle \frac{l\phi^{n+1} + (l\phi^n) \circ X_E^n}{2}, \psi \right\rangle. \\
I_5 &= \alpha \left\langle \det \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} \phi^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R} - \alpha \left\langle \frac{\phi^{n+1} + \phi^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2}, \psi \right\rangle_{\Gamma_R}.
\end{aligned}$$

The bound for I_1 directly follows from Lemma 3.10 for $\varphi = \phi$, so we can define a function $\xi_{I_1}^{n+\frac{1}{2}} \in L^2(\Omega)$ such that

$$I_1 = \left\langle \xi_{I_1}^{n+\frac{1}{2}}, \psi \right\rangle \quad \text{with} \quad \|\xi_{I_1}^{n+\frac{1}{2}}\|_0 \leq \tilde{c}_1 \Delta t^2 \|\phi\|_{Z^3}. \quad (3.118)$$

Term I_2 is firstly divided into three terms $I_2 = I_2^1 + I_2^2 + I_2^3$, where

$$\begin{aligned}
I_2^1 &= \left\langle \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} \left(\mathbf{A} \operatorname{grad} \phi^{n+\frac{1}{2}} \right) \circ X_e^{n+\frac{1}{2}} \right. \\
&\quad \left. - \frac{\mathbf{A} \operatorname{grad} \phi^{n+1} + (\mathbf{F}_e^n)^{-1} (\mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n}{2}, \operatorname{grad} \psi \right\rangle, \\
I_2^2 &= \left\langle \frac{(\mathbf{F}_e^n)^{-1} (\mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n - ((\mathbf{I} + \Delta t \mathbf{L}^n) \mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n}{2}, \operatorname{grad} \psi \right\rangle, \\
I_2^3 &= \left\langle \frac{((\mathbf{I} + \Delta t \mathbf{L}^n) \mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n - ((\mathbf{I} + \Delta t \mathbf{L}^n) \mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n}{2}, \operatorname{grad} \psi \right\rangle.
\end{aligned}$$

For I_2^1 we apply Lemma 3.12 for $\mathbf{w} = \mathbf{A} \operatorname{grad} \phi \in C^i(\mathbf{H}^{3-i}(\Omega))$ for $i = 0, 1, 2$, so a vector valued function $\vartheta_{I_2^1}^{n+\frac{1}{2}} \in \mathbf{H}^1(\Omega)$ can be defined and Green's formula can be applied. Thus we have

$$I_2^1 = \left\langle \vartheta_{I_2^1}^{n+\frac{1}{2}}, \operatorname{grad} \psi \right\rangle = \left\langle \vartheta_{I_2^1}^{n+\frac{1}{2}} \cdot \mathbf{n}, \psi \right\rangle_{\Gamma_R} - \left\langle \operatorname{Div} \vartheta_{I_2^1}^{n+\frac{1}{2}}, \psi \right\rangle$$

where the functions involved are bounded as follows

$$\left\| \vartheta_{I_2^1}^{n+\frac{1}{2}} \cdot \mathbf{n} \right\|_{0, \Gamma_R} \leq \tilde{c}_1 \Delta t^2 \|\mathbf{A} \operatorname{grad} \phi \cdot \mathbf{n}\|_{Z^2, \Gamma_R}, \quad \left\| \operatorname{Div} \vartheta_{I_2^1}^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \|\operatorname{Div} \mathbf{A} \operatorname{grad} \phi\|_{Z^2}. \quad (3.119)$$

For I_2^2 we apply Lemma 3.15 for $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$, finding a matrix valued function $\Psi_{I_2^2}^n \in \mathbf{H}^1(\Omega)$ satisfying

$$\begin{aligned} I_2^2 &= \left\langle \Psi_{I_2^2}^n (\mathbf{A} \operatorname{grad} \phi) \circ X_e^n, \operatorname{grad} \psi \right\rangle \\ &= \left\langle \Psi_{I_2^2}^n (\mathbf{A} \operatorname{grad} \phi) \circ X_e^n \cdot \mathbf{n}, \psi \right\rangle_{\Gamma_R} - \left\langle \operatorname{Div} \left(\Psi_{I_2^2}^n (\mathbf{A} \operatorname{grad} \phi) \circ X_e^n \right), \psi \right\rangle, \end{aligned} \quad (3.120)$$

where Green's formula has been used for the last equality and with functions on the left hand side of the inner products bounded as in (3.119).

For I_2^3 we apply Lemma 3.14 componentwise for

$$\varphi^n(\mathbf{x}) = [(\mathbf{I} + \Delta t \mathbf{L}^n(\mathbf{x}))\mathbf{A}(\mathbf{x}) \operatorname{grad} \phi(\mathbf{x})]_i \in H^2(\Omega)$$

and we find a function $\vartheta_{I_2^3}^n \in H^1(\Omega)$, to which we can apply Green's formula obtaining bounds analogous to (3.119).

Summing up, we can write I_2 as

$$I_2 = \left\langle \xi_{I_2^A}^{n+\frac{1}{2}}, \psi \right\rangle + \left\langle \xi_{I_2^B}^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R} \quad (3.121)$$

where

$$\left\| \xi_{I_2^A}^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \|\mathbf{A} \operatorname{grad} \phi\|_{Z^3}, \quad \left\| \xi_{I_2^B}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R} \leq \tilde{c}_1 \Delta t^2 \|\mathbf{A} \operatorname{grad} \phi \cdot \mathbf{n}\|_{Z^2,\Gamma_R}.$$

By noticing that $\mathbf{F}_e^{n+1} = \mathbf{I}$, and then $\operatorname{Div} \mathbf{F}_e^{n+1} = 0$, I_3 can be divided into three parts, $I_3 = I_3^1 + I_3^2 + I_3^3$, where

$$\begin{aligned} I_3^1 &= \left\langle \operatorname{Div} \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-T} \cdot (\mathbf{A} \operatorname{grad} \phi^{n+\frac{1}{2}}) \circ X_e^{n+\frac{1}{2}}, \psi \right\rangle \\ &\quad - \left\langle \frac{\operatorname{Div} (\mathbf{F}_e^{n+1})^{-T} \cdot \mathbf{A} \operatorname{grad} \phi^{n+1} + \operatorname{Div} (\mathbf{F}_e^n)^{-T} \cdot (\mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n}{2}, \psi \right\rangle, \\ I_3^2 &= \left\langle \frac{\operatorname{Div} (\mathbf{F}_e^n)^{-T} \cdot (\mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n - \Delta t (\operatorname{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n}{2}, \psi \right\rangle, \\ I_3^3 &= \Delta t \left\langle \frac{(\operatorname{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi^n) \circ X_e^n - (\operatorname{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi^n) \circ X_E^n}{2}, \psi \right\rangle. \end{aligned}$$

For I_3^1 we apply Lemma 3.13 for $\mathbf{w} = \mathbf{A} \operatorname{grad} \phi \in Z^2$, obtaining

$$I_3^1 = \left\langle \xi_{I_3^1}^{n+\frac{1}{2}}, \psi \right\rangle \quad \text{with} \quad \left\| \xi_{I_3^1}^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \|\mathbf{A} \operatorname{grad} \phi\|_{Z^2}.$$

For I_3^2 we apply Lemma 3.16 finding a vector valued function $\vartheta_{I_3^2}^n \in \mathbf{L}^2(\Omega)$ for $\mathbf{v} \in C^1(W^{2,\infty}(\Omega)) \cap C^0(W^{3,\infty}(\Omega))$ satisfying

$$\begin{aligned} I_3^2 &= \left\langle \vartheta_{I_3^2}^n \cdot (\mathbf{A} \operatorname{grad} \phi) \circ X_e^n, \psi \right\rangle \\ &\quad \text{with} \quad \left\| \vartheta_{I_3^2}^n \cdot (\mathbf{A} \operatorname{grad} \phi) \circ X_e^n \right\|_0 \leq \tilde{c}_1 \Delta t^2 \|\mathbf{A} \operatorname{grad} \phi\|_{C^0(\mathbf{L}^2(\Omega))}. \end{aligned}$$

Notice that I_3^3 already has a Δt , thus if we apply Lemma 3.14 for

$$\varphi(\mathbf{x}) = \mathbf{grad} \operatorname{div} \mathbf{v}^n \cdot \mathbf{A} \operatorname{grad} \phi^n \in H^1(\Omega)$$

we find a function $\xi_{I_3^3}^n \in L^2(\Omega)$ satisfying

$$I_3^3 = \left\langle \vartheta_{I_3^3}^n, \psi \right\rangle \quad \text{with} \quad \left\| \vartheta_{I_3^3}^n \right\|_0 \leq \tilde{c}_1 \Delta t^3 \left\| \mathbf{A} \operatorname{grad} \phi \right\|_{C^0(H^1(\Omega))}.$$

Summing up, we get

$$I_3 = \left\langle \xi_{I_3}^{n+\frac{1}{2}}, \psi \right\rangle, \quad \text{with} \quad \left\| \xi_{I_3}^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \left\| \mathbf{A} \operatorname{grad} \phi \right\|_{Z^2}. \quad (3.122)$$

For $\xi = l\phi \in Z^2$ we can apply Corollary 3.4 to I_4 obtaining

$$I_4 = \left\langle \xi_{I_4}^{n+\frac{1}{2}}, \psi \right\rangle, \quad \text{with} \quad \left\| \xi_{I_4}^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \left\| l\phi \right\|_{Z^2}. \quad (3.123)$$

The estimate for the last term, I_5 , is a direct application of Lemma 3.17 for $\xi = \alpha\phi|_{\Gamma_R} \in C^2(L^2(\Gamma_R))$. Indeed,

$$I_5 = \left\langle \xi_{I_5}^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R}, \quad \text{with} \quad \left\| \xi_{I_5}^{n+\frac{1}{2}} \right\|_{0, \Gamma_R} \leq \tilde{c}_1 \Delta t^2 \left\| \phi \right\|_{C^2(L^2(\Gamma_R))}. \quad (3.124)$$

Finally, partial results (3.118), (3.121), (3.122), (3.123) and (3.124) imply (3.117). $\square$

Lemma 3.19 *Assume that $\mathbf{v} \in C^0(W^{2,\infty}(\Omega)) \cap C^1(W^{1,\infty}(\Omega))$, vanishes on the boundary and $\Delta t \|\mathbf{v}\|_{C^0(W^{1,\infty})} < 1/2$. Let $f \in Z^2$ and $g \in C^2(L^2(\Gamma_R))$. Then, for each $n = 0, 1, \dots, N-1$, there exist two functions*

$$\xi_f^{n+\frac{1}{2}} : \Omega \longrightarrow \mathbb{R}, \quad \xi_g^{n+\frac{1}{2}} : \Gamma_R \longrightarrow \mathbb{R},$$

satisfying

$$\left\langle \mathcal{F}^{n+\frac{1}{2}} - \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \psi \right\rangle = \left\langle \xi_f^{n+\frac{1}{2}}, \psi \right\rangle + \left\langle \xi_g^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R} \quad \forall \psi \in H^1(\Omega). \quad (3.125)$$

Moreover, $\xi_f \in L^2(\Omega)$, $\xi_g \in L^2(\Gamma_R)$ and the following estimates hold

$$\left\| \xi_f^{n+\frac{1}{2}} \right\|_0 \leq \tilde{c}_1 \Delta t^2 \left\| f \right\|_{Z^2}, \quad \left\| \xi_g^{n+\frac{1}{2}}(\mathbf{x}) \right\|_{0, \Gamma_R} \leq \tilde{c}_1 \Delta t^2 \left\| g \right\|_{C^2(L^2(\Gamma_R))},$$

with constant $\tilde{c}_1$ independent of Δt .

PROOF. Result (3.125) straightly follows by applying Corollary 3.4 with $\xi = f$ and Lemma 3.17 with $\xi = g$ to expression

$$\begin{aligned} \left\langle \mathcal{F}^{n+\frac{1}{2}} - \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \psi \right\rangle &= \frac{1}{2} \left\langle f^{n+\frac{1}{2}} \circ X_e^{n+\frac{1}{2}} - \frac{f^{n+1} + f^n \circ X_E^n}{2}, \psi \right\rangle \\ &+ \left\langle \det \left(\mathbf{F}_e^{n+\frac{1}{2}} \right)^{-1} g^{n+\frac{1}{2}} \circ X_e^{n+\frac{1}{2}} - \frac{g^{n+1} + g^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2}, \psi \right\rangle_{\Gamma_R}. \end{aligned}$$

$\square$

Lemmas stated in the present section hold under Hypothesis 3.3, 3.4 and 3.5, and the following ones:

Hypothesis 3.9 Functions appearing in problem (3.1)-(3.4) satisfy:

- $\mathbf{A} \in \mathbf{W}^{3,\infty}(\Omega)$, $l \in W^{2,\infty}(\Omega)$,
- $\mathbf{v} \in C^0(\mathbf{W}^{3,\infty}(\Omega)) \cap C^1(W^{2,\infty}(\Omega)) \cap C^2(L^\infty(\Omega))$ and $\mathbf{v}|_\Gamma = 0$,
- $f \in Z^2, g \in Z^3(\Gamma_R)$ and $\alpha > 0$.

Remark 3.19 Although in Lemma 3.4 only $g \in C^2(L^2(\Gamma_R))$ was required, more smoothness will be necessary in the following theorem.

Theorem 3.3 (Error estimate) Assume Hypothesis 3.3, 3.4, 3.5, and 3.9. Let

$$\phi \in Z^3, \text{ grad } \phi \in Z^3, \quad \phi|_{\Gamma_R} \in Z^3(\Gamma_R),$$

be the solution of (3.107) and let $\phi_{\Delta t} = \{\phi_{\Delta t}^n\}$ be the solution of (3.69) subject to initial value $\phi_{\Delta t}^0 = \phi^0$. Then, there exist two positive constants c and d such that, if $\Delta t < d$ we have

$$\begin{aligned} & \sqrt{\frac{1}{2}} \|\widehat{\phi} - \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t \delta}{4}} \|\mathbf{B} \text{ grad } \widehat{\phi} - \mathbf{B} \text{ grad } \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} \\ & + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l}\widehat{\phi} - \sqrt{l}\widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t \alpha}{16}} \|\widehat{\phi} - \widehat{\phi_{\Delta t}}\|_{l^\infty(L^2(\Gamma_R))} \\ & \leq c \Delta t^2 (\|\phi\|_{Z^3} + \|\mathbf{A} \text{ grad } \phi\|_{Z^3} + \|l\phi\|_{Z^2} + \|\phi\|_{Z^2, \Gamma_R} + \|f\|_{Z^2} + \|g\|_{Z^2, \Gamma_R}) \\ & \quad + c \Delta t^3 (\|\phi\|_{Z^3, \Gamma_R} + \|g\|_{Z^3, \Gamma_R}), \end{aligned} \quad (3.126)$$

where $\mathbf{B} \text{ grad } \widehat{\phi_{\Delta t}} := \{\mathbf{B} \text{ grad } \phi_{\Delta t}^n\}$.

PROOF. Let us denote by $e_{\Delta t}$ the difference between the continuous and the discrete solutions, i.e., $e_{\Delta t} = \{e_{\Delta t}^n\}$, with $e_{\Delta t}^n = \phi^n - \phi_{\Delta t}^n$. Then, by using (3.107) and (3.69) we have

$$\begin{aligned} \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} e_{\Delta t}, \psi \right\rangle &= \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi}, \psi \right\rangle - \left\langle \mathcal{L}^{n+\frac{1}{2}} \widehat{\phi}, \psi \right\rangle + \left\langle \mathcal{L}^{n+\frac{1}{2}} \widehat{\phi}, \psi \right\rangle + \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi_{\Delta t}}, \psi \right\rangle \\ &= \left\langle \left(\mathcal{L}_{\Delta t}^{n+\frac{1}{2}} - \mathcal{L}^{n+\frac{1}{2}} \right) \widehat{\phi}, \psi \right\rangle + \left\langle \mathcal{F}^{n+\frac{1}{2}} - \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \psi \right\rangle. \end{aligned} \quad (3.127)$$

then, as a consequence of Lemmas 3.18 and 3.19, we are led to the following scheme

$$\left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} e_{\Delta t}, \psi \right\rangle = \left\langle \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} - \xi_f^{n+\frac{1}{2}}, \psi \right\rangle + \left\langle \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} - \xi_g^{n+\frac{1}{2}}, \psi \right\rangle_{\Gamma_R} \quad \forall \psi \in H_{\Gamma_D}^1(\Omega). \quad (3.128)$$

Next, we apply Theorem 3.2 to (3.128), noting that $e_{\Delta t}^0 = 0$. We obtain

$$\begin{aligned} & \frac{1}{\sqrt{2}} \|\widehat{e_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \text{ grad } \widehat{e_{\Delta t}}\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l}\widehat{e_{\Delta t}}\|_{l^\infty(L^2(\Omega))} \\ & + \sqrt{\frac{\alpha \Delta t}{16}} \|\widehat{e_{\Delta t}}\|_{l^\infty(L^2(\Gamma_R))} \leq c \Delta t \left(\|\widehat{D_{\Delta t} \xi_{\mathcal{L}_2}}\|_{l^2(L^2(\Gamma_R))} + \|\widehat{D_{\Delta t} \xi_g}\|_{l^2(L^2(\Gamma_R))} \right) \\ & + c \left(\|\widehat{\xi_{\mathcal{L}_1}}\|_{l^2(L^2(\Omega))} + \|\widehat{\xi_f}\|_{l^2(L^2(\Omega))} + \|\widehat{\xi_{\mathcal{L}_2}}\|_{l^2(L^2(\Gamma_R))} + \|\widehat{\xi_g}\|_{l^2(L^2(\Gamma_R))} \right). \end{aligned} \quad (3.129)$$

Thus, (3.126) follows by using the upper bounds for $\xi_{\mathcal{L}_1}, \xi_f, \xi_{\mathcal{L}_2}$ and ξ_g given in Lemmas 3.18 and 3.19, with the following two remarks:

- By using the Robin boundary condition of (3.107), the estimate of $\xi_{\mathcal{L}_2}$ given in Lemma 3.18 is replaced by

$$\|\xi_{\mathcal{L}_2}\|_{l^2(L^2(\Gamma_R))} \leq \tilde{c}_1 \Delta t^2 (\alpha \|\phi\|_{Z^2, \Gamma_R} + \|g\|_{Z^2, \Gamma_R}).$$

- Taking into account Remark 3.17, terms multiplied by Δt in (3.129) can be bounded if $\xi_{\mathcal{L}_2}^{n+\frac{1}{2}}$ and $\xi_g^{n+\frac{1}{2}}$ have their respective time derivatives in $L^2(\Gamma_R)$, given rise to the terms multiplied by Δt^3 in (3.126).

□

3.7 Clasical semi-Lagrangian method

The classical characteristics method (3.61) has proved to be first order accurate in time and unconditionally stable for convection-diffusion equations (see [45, 89, 14]), independently of the diffusion coefficient. In fact, this method presents similar properties when applied to the linear transport equation (see, for instance, [88, 82]). Notice that, in this last case, classical and Crank-Nicholson semidiscretized schemes coincide, except for the approximation of the characteristic lines.

In the following we indicate how to obtain stability and consistency results for the classical characteristics method (3.61) under similar hypothesis to those assumed in Section 3.6. Let us introduce

$$\mathcal{L}_{\Delta t}^{n+1} \phi \in (H^1(\Omega))' \text{ and } \mathcal{F}_{\Delta t}^{n+1} \in (H^1(\Omega))' :$$

defined, for $\phi \in C^0(H^1(\Omega))$, by

$$\langle \mathcal{L}_{\Delta t}^{n+1} \phi, \psi \rangle := \frac{1}{\Delta t} \langle \mathcal{D}_E^{n+1}[\phi], \psi \rangle + \langle \mathcal{M}_{\Delta t}^{n,1}[\phi], \psi \rangle \quad \forall \psi \in H^1(\Omega),$$

and

$$\langle \mathcal{F}_{\Delta t}^{n+1}, \psi \rangle := \langle \mathcal{N}_{\Delta t}^{n,1}, \psi \rangle \quad \forall \psi \in H^1(\Omega).$$

Then, the semidiscretized time scheme can be written as follows:

$$\left\{ \begin{array}{l} \text{Given } \phi_{\Delta t}^0, \text{ find } \widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N \in [H_{\Gamma_D}^1(\Omega)]^N \text{ such that} \\ \langle \mathcal{L}_{\Delta t}^{n+1} \widehat{\phi_{\Delta t}}, \psi \rangle = \langle \mathcal{F}_{\Delta t}^{n+1}, \psi \rangle \quad \forall \psi \in H_{\Gamma_D}^1(\Omega), \text{ for } n = 0, \dots, N-1. \end{array} \right. \quad (3.130)$$

In the following two lemmas, a lower bound of the left hand side, and an upper bound of the right hand side are given. Using them, the stability result follows.

Lemma 3.20 *Let us assume Hypotheses 3.2, 3.3 3.5, $\alpha > 0$ and $c_1 \Delta t < 1$. If $\widehat{\phi_{\Delta t}} = \{\phi_{\Delta t}^n\}_{n=1}^N$ denotes the solution of (3.130), then*

$$\begin{aligned} \langle \mathcal{L}_{\Delta t}^{n+1} \widehat{\phi_{\Delta t}}, \phi_{\Delta t}^{n+1} \rangle &\geq D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{\phi_{\Delta t}}\|_0^2 \right) + \|\mathbf{C} \text{ grad } \phi_{\Delta t}^{n+1}\|_0^2 + \|\sqrt{l} \phi_{\Delta t}^{n+1}\|_0^2 + \alpha \|\phi_{\Delta t}^{n+1}\|_{0, \Gamma_R}^2 \\ &\quad + \frac{1}{2\Delta t} \|\phi_{\Delta t}^{n+1} - \phi_{\Delta t}^n \circ X_E^n\|_0^2 - \frac{c_1}{2} \|\phi_{\Delta t}^n\|_0^2. \end{aligned} \quad (3.131)$$

PROOF. The proof is equal to the one of Lemma 3.6 for the material derivative term. For the other terms the equality holds. $\square$

Lemma 3.21 *Let us assume Hypothesis 3.2, 3.6 and 3.7, for $0 \leq n \leq N-1$. If $c_1 \Delta t < 1$, then*

$$\langle \mathcal{F}^{n+1}, \psi \rangle \leq \frac{1}{2} (\|f^{n+1}\|_0^2 + \|\psi\|_0^2) + \frac{1}{2\alpha} \|g^{n+1}\|_{0,\Gamma_R}^2 + \frac{\alpha}{2} \|\psi\|_{0,\Gamma_R}^2, \quad (3.132)$$

for any function $\psi \in H^1(\Omega)$, and $\alpha > 0$ the constant in (3.3).

PROOF. The result straightly follows by applying inequality (3.77). $\square$

Theorem 3.4 (Stability) *Under Hypotheses 3.2, 3.3, 3.5, 3.6 and 3.7. Let $\widehat{\phi}_{\Delta t} = \{\phi_{\Delta t}^n\}_{n=1}^N$ be the solution of (3.130) subject to initial value $\phi_{\Delta t}^0 \in L^2(\Omega)$. Then, there exist two positive constants c and d , such that, if $\Delta t < d$ then*

$$\begin{aligned} & \frac{1}{\sqrt{2}} \|\widehat{\phi}_{\Delta t}\|_{l^\infty(L^2(\Omega))} + \sqrt{\Delta t} \|\mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t}\|_{l^2(L^2(\Omega))} + \sqrt{\Delta t} \|\sqrt{l} \widehat{\phi}_{\Delta t}\|_{l^2(L^2(\Omega))} \\ & + \sqrt{\frac{\alpha \Delta t}{2}} \|\widehat{\phi}_{\Delta t}\|_{l^2(L^2(\Gamma_R))} \leq c \left(\frac{1}{2} \|\phi_{\Delta t}^0\|_0^2 + \|\widehat{f}\|_{l^2(L^2(\Omega))} + \|\widehat{g}\|_{l^2(L^2(\Gamma_R))} \right), \end{aligned} \quad (3.133)$$

where $\mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} := \{\mathbf{C} \text{ grad } \phi_{\Delta t}^n\}$.

PROOF. By using Lemma 3.20 and Lemma 3.21 with $\psi = \phi^{n+1}$ and rearranging terms, we have

$$\begin{aligned} & \frac{1}{2\Delta t} \|\phi^{n+1}\|_0^2 - \frac{1}{2\Delta t} \|\phi^n\|_0^2 + \|\mathbf{C} \text{ grad } \phi^{n+1}\|_0^2 + \|\sqrt{l} \phi^{n+1}\|_0^2 + \frac{\alpha}{2} \|\phi^{n+1}\|_{0,\Gamma_R}^2 \\ & \leq \frac{c_1}{2} \|\phi^n\|_0^2 + \frac{1}{2} \|\phi^{n+1}\|_0^2 + \frac{1}{2} \|f^{n+1}\|_0^2 + \frac{1}{2\alpha} \|g^{n+1}\|_{0,\Gamma_R}^2. \end{aligned}$$

Summing the above inequality from $n = 0$ to $q-1$ we lead to

$$\begin{aligned} & \frac{1}{2} (1 - \Delta t) \|\phi^q\|_0^2 + \Delta t \sum_0^{q-1} \|\mathbf{C} \text{ grad } \phi^{n+1}\|_0^2 + \Delta t \sum_0^{q-1} \|\sqrt{l} \phi^{n+1}\|_0^2 + \Delta t \sum_0^{q-1} \frac{\alpha}{2} \|\phi^{n+1}\|_{0,\Gamma_R}^2 \\ & \leq \frac{1}{2} (1 + c_1) \Delta t \sum_0^{q-1} \|\phi^n\|_0^2 + \frac{1}{2} \Delta t \sum_0^{q-1} \|f^{n+1}\|_0^2 + \frac{1}{2} \Delta t \sum_0^{q-1} \|g^{n+1}\|_{0,\Gamma_R}^2, \end{aligned}$$

and the result directly follows by the discrete Gronwall's inequality. $\square$

According to (3.46) for $t = t_{n+1}$ and $\tau = t_{n+1}$, the continuous solution solves the problem

$$\langle \mathcal{L}^{n+1} \widehat{\phi}, \psi \rangle = \langle \mathcal{F}^{n+1}, \psi \rangle, \forall \psi \in H_{\Gamma_D}^1(\Omega), \quad (3.134)$$

where operators $\mathcal{L}^{n+1} \widehat{\phi} \in (H^1(\Omega))'$ and $\mathcal{F}^{n+1} \in (H^1(\Omega))'$ are defined by

$$\begin{aligned} \langle \mathcal{L}^{n+1} \widehat{\phi}, \psi \rangle & := \left\langle \left(\dot{\phi} \right)^{n+1}, \psi \right\rangle + \langle \mathbf{A} \text{ grad } \phi^{n+1}, \text{ grad } \psi \rangle + \langle l \phi^{n+1}, \psi \rangle + \alpha \langle \phi^{n+1}, \psi \rangle_{\Gamma_R}, \\ \langle \mathcal{F}^{n+1}, \psi \rangle & := \langle f^{n+1}, \psi \rangle + \langle g^{n+1}, \psi \rangle_{\Gamma_R}, \end{aligned}$$

for $\psi \in H^1(\Omega)$.

Clearly, the error comes only from the approximation of the material derivative, and then, the method is of order $O(\Delta t)$. More precisely, a result similar to Lemma 3.10 holds replacing the Runge-Kutta approximation by the Euler approximation.

Lemma 3.22 *If $\mathbf{v} \in C^1(L^\infty(\Omega)) \cap C^0(W^{1,\infty}(\Omega))$, $\mathbf{v}|_\Gamma = 0$, $c_1 \Delta t < 1$, and $\varphi \in Z^2$. Let ξ^{n+1} be a function defined by*

$$\xi^{n+1}(\mathbf{x}) := \dot{\varphi}^{n+1}(\mathbf{x}) - \frac{\varphi^{n+1}(\mathbf{x}) - \varphi^n(X_E^n(\mathbf{x}))}{\Delta t} \quad \forall \mathbf{x} \in \Omega.$$

Then, $\xi^{n+1} \in L^2(\Omega)$ and

$$\|\xi^{n+1}\|_0 \leq \Delta t \tilde{c}_1 \|\varphi\|_{Z^2}.$$

PROOF. Firstly notice that function ξ can be also written as

$$\xi^{n+1}(\mathbf{x}) = \dot{\varphi}^{n+1}(\mathbf{x}) - \frac{\varphi^{n+1}(\mathbf{x}) - \varphi^n(X_e^n(\mathbf{x}))}{\Delta t} + \frac{\varphi^n(X_E^n(\mathbf{x})) - \varphi^n(X_e^n(\mathbf{x}))}{\Delta t} \quad (3.135)$$

The first two terms above are bounded by applying Taylor approximation to function $G(\tau) = \varphi(X_e(\mathbf{x}, t_{n+1}; \tau), \tau)$ up to second order. Namely,

$$\varphi(X_e^n(\mathbf{x}), t_n) = \varphi(\mathbf{x}, t_{n+1}) + (t_n - t_{n+1})\dot{\varphi}(\mathbf{x}, t_{n+1}) + \int_{t_{n+1}}^{t_n} (t_n - s) \ddot{\varphi}(X_e(\mathbf{x}, t_{n+1}; s), s) ds.$$

Next, by using Lemma 3.14, the last term in (3.135) is suitably bounded and the result is concluded. $\square$

The error sequence, $\{\widehat{e}_{\Delta t}\}, e_{\Delta t}^n := \phi^n - \phi_{\Delta t}^n$, satisfies the discrete equation (3.130) with a right hand side given by the previous Lemma, and thus we have the following consistency result.

Theorem 3.5 (Consistency error) *Assume Hypotheses 3.2, 3.3, 3.5, 3.6, 3.7 and that $\mathbf{v} \in C^1(L^\infty(\Omega)) \cap C^0(W^{1,\infty}(\Omega))$. Let $\phi \in Z^2$ be the solution of (3.134) and $\widehat{\phi}_{\Delta t} = \{\phi_{\Delta t}^n\}_{n=1}^N$ be the solution of (3.130) subject to initial value $\phi_{\Delta t}^0 = \phi^0$. Then, there exist two positive constants c and d , such that, if $\Delta t < d$ then*

$$\begin{aligned} & \frac{1}{\sqrt{2}} \left\| \widehat{\phi} - \widehat{\phi}_{\Delta t} \right\|_{l^\infty(L^2(\Omega))} + \sqrt{\Delta t} \left\| \mathbf{C} \text{ grad } \widehat{\phi} - \mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Omega))} \\ & + \sqrt{\Delta t} \left\| \sqrt{l} \widehat{\phi} - \sqrt{l} \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Omega))} + \sqrt{\frac{\alpha \Delta t}{2}} \left\| \widehat{\phi} - \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Gamma_R))} \leq c \Delta t \|\phi\|_{Z^2}, \end{aligned} \quad (3.136)$$

where $\mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} := \{\mathbf{C} \text{ grad } \phi_{\Delta t}^n\}$.

Remark 3.20 *Regarding now the two step Lagrange scheme, a multistep scheme like (3.66) is proposed and analyzed in [46] when applied to a one-dimensional convection diffusion equation, and time-independent velocity. In that case, first order Euler scheme is enough to approximate the characteristic lines maintaining the second order accuracy in time. In [30] a similar multistep scheme (for $N_{step} = 1, 2, 3$) is proposed for the Navier-Stokes equations, and an analysis of that method for $N_{step} = 2$ is developed in [31]. In both papers, second order approximations of the characteristic lines are used.*

Second order accuracy in time is proved in both [46] and [31], but it seems that it is needed for stability a diffusion coefficient bounded from below by a strictly positive constant. In fact, in [30] authors pointed out: “The stability constants depend on the viscosity coefficient. Furthermore, these constants blow up exponentially when the viscosity coefficient tends to zero.”

Chapter 4

Lagrange-Galerkin approximation of convection-diffusion-reaction equations

4.1 Introduction

As we have mentioned in Chapter 3, the time semidiscretization given by the characteristics method can be combined with space discretizations of finite differences [45], spectral finite elements [4], discontinuous finite elements [6, 5, 7], etc. We will consider in the present chapter the combination of the characteristics method with a spatial discretization using conforming Lagrange finite elements satisfying an interpolation property.

There exists an extensive literature studying the classical first order characteristic method combined with (linear) finite elements applied to convection-diffusion equations. If Δt denotes the time step and h the spatial step, and k the degree of the finite element space, error estimates of the form $O(h^k + \Delta t)$ in $l^\infty(L^2(\mathbb{R}^m))$ norm have been stated in [103] (m denotes the dimension of the spatial domain). In [88] estimates of the form $O(h^k + \Delta t + h^{k+1}/\Delta t)$ in $l^\infty(L^2(\Omega))$ norm have been shown under the assumption that the normal component of the velocity field vanishes on the boundary of the spatial domain Ω and for an approximate discrete velocity field. In all of the above estimates the constant depend on norms on the solutions. More recently, for linear finite elements and for a velocity field vanishing on the boundary, a convergence of order $O(h^2 + \min(h, h^2/\Delta t) + \Delta t)$ in $l^\infty(L^2(\Omega))$ norm has been shown in [14], where the constants in the estimate only depends on the data.

In the first part of the present chapter, we continue the analysis of Chapter 3 by proving an error estimate $O(\Delta t^2 + h^k)$ for the second order Crank-Nicholson scheme when combined with finite elements “of order k ”. Similar results are indicated for the classical scheme.

In practice, we will consider linear ($k = 1$) and quadratic ($k = 2$) finite elements over triangular and quadrangular meshes. It is known that, in this latter case, the result may exhibit an oscillatory behaviour. A limiting procedure as the proposed in [18, 17] could be considered to get a non-oscillatory scheme.

Moreover, the unconditional stability of the characteristics method is established only under the assumption that the inner products in the Galerkin formulation are calculated exactly. This is rarely possible in practice so numerical quadrature has to be used instead, producing, in some cases, the loss of the unconditional stability. This aspect was studied for the first order scheme

in [82, 103, 94]. In the second part of this chapter, an analysis of the influence of quadrature formulas is developed. We restrict our study to Lagrange finite element spaces over “triangular” and “quadrangular” meshes with $k = 1, 2$, for which adequate quadrature formulas are proposed. The stability of the classical first order and the second order Crank-Nicholson Lagrange-Galerkin scheme when combined with some of these quadrature formulas is rigourously studied by using Fourier analysis. In this aspect, previous studies about the influence of quadratures in the case of the classical first order Lagrange-Galerkin method applied to transport [82] and convection-diffusion [103] equations, are here extended to the second order Lagrange-Galerkin one.

Finally, we illustrate the theoretical results by two numerical examples of parabolic convection-diffusion-reaction problems in dimension $m = 2$.

4.2 Space discretization: finite element method

We propose a spatial discretization by using finite element spaces V_h^k , where h denotes the mesh parameter and the positive integer $k \geq 1$ is the “approximation degree” in the following sense:

Hypothesis 4.1 *There exists an interpolation operator $\pi_h : C^0(\overline{\Omega}) \longrightarrow V_h^k$, with $k \geq 1$, satisfying*

$$\|\pi_h \psi - \psi\|_s \leq K h^{r-s} \|\psi\|_r \quad \forall \psi \in C^0(\overline{\Omega}) \cap H^r(\Omega) \quad 0 \leq s \leq r \leq k+1, \quad (4.1)$$

for a positive constant K independent of h .

We refer to [35] for a complete textbook on the finite element method.

By dropping subscript Δt , for simplicity, the fully discrete counterparts of (3.61), (3.62) and (3.66) read, respectively,

$$\left\{ \begin{array}{l} \text{Given } \phi_h^0 \in V_h^k, \text{ find } \widehat{\phi}_h = \{\phi_h^n\}_{n=1}^N \in [V_h^k]^N \text{ such that} \\ \frac{1}{\Delta t} \langle \mathcal{D}_E^{n+1}[\phi_h], \psi_h \rangle + \langle \mathcal{M}_{\Delta t}^{n,1}[\phi], \psi \rangle = \langle \mathcal{N}_{\Delta t}^{n,1}, \psi_h \rangle \\ \text{for all } \psi_h \in V_h^k \text{ and } n = 0, \dots, N-1. \end{array} \right. \quad (4.2)$$

$$\left\{ \begin{array}{l} \text{Given } \phi_h^0 \in V_h^k, \text{ find } \widehat{\phi}_h = \{\phi_h^n\}_{n=1}^N \in [V_h^k]^N \text{ such that} \\ \frac{1}{\Delta t} \langle \mathcal{D}_{RK}^{n+1}[\phi_h], \psi_h \rangle + \langle \mathcal{M}_{\Delta t}^{n, \frac{1}{2}}[\phi_h], \psi \rangle = \langle \mathcal{N}_{\Delta t}^{n, \frac{1}{2}}, \psi_h \rangle \\ \text{for all } \psi \in V_h^k \text{ and } n = 0, \dots, N-1. \end{array} \right. \quad (4.3)$$

$$\left\{ \begin{array}{l} \text{Given } \phi_h^0, \phi_h^1 \in V_h^k \text{ find } \widehat{\phi}_h = \{\phi_h^n\}_{n=2}^N \in [V_h^k]^{N-1} \text{ such that} \\ \frac{1}{2\Delta t} \langle \mathcal{D}_{TS}^{n+1}[\phi_h], \psi_h \rangle + \langle \mathcal{M}_{\Delta t}^{n,1}[\phi_h], \psi \rangle = \langle \mathcal{N}_{\Delta t}^{n,1}, \psi_h \rangle \\ \text{for all } \psi \in V_h^k \text{ and } n = 0, \dots, N-1. \end{array} \right. \quad (4.4)$$

4.3 Analysis of the Crank-Nicholson Lagrange-Galerkin method

We study the fully discretized scheme (4.3), or, equivalently,

$$\left\{ \begin{array}{l} \text{Given } \phi_h^0 \in V_h^k, \text{ find } \widehat{\phi}_h := \{\phi_h^n\}_{n=1}^N \in [V_h^k]^N \text{ such that} \\ \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi}_h, \psi_h \right\rangle = \left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, \psi_h \right\rangle \quad \forall \psi_h \in V_h^k \text{ for } n = 0, \dots, N-1, \end{array} \right. \quad (4.5)$$

where the operators have been introduced in Section 3.6.

4.3.1 Stability for the fully discretized scheme

We give the stability result. The proof is the same to the one of Theorem 3.1 for the stability of the semidiscretized scheme.

Theorem 4.1 (Stability) *Let us assume Hypotheses 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 and let $\widehat{\phi}_h = \{\phi_h^n\}_{n=1}^N$ be the solution of (4.5) subject to initial value ϕ_h^0 . Then, there exist two positive constants, c and d , such that for $\Delta t < d$ we have*

$$\begin{aligned} & \frac{1}{\sqrt{2}} \|\widehat{\phi}_h\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \text{ grad } \widehat{\phi}_h\|_{l^\infty(L^2(\Omega))} \\ & + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \widehat{\phi}_h\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\alpha \Delta t}{8}} \|\widehat{\phi}_h\|_{l^\infty(L^2(\Gamma_R))} \\ & \leq c \left(\frac{1}{2} \|\phi_h^0\|_0 + \sqrt{\frac{\delta \Delta t}{4}} \|\mathbf{B} \text{ grad } \phi_h^0\|_0 + \sqrt{\frac{\Delta t}{4}} \|\sqrt{l} \phi_h^0\|_0 \right. \\ & \quad \left. + \sqrt{\frac{\alpha \Delta t}{8}} \|\phi_h^0\|_{0,\Gamma_R} + \|f\|_{l^2(L^2(\Omega))} + \|g\|_{l^2(L^2(\Gamma_R))} \right), \end{aligned}$$

where $\mathbf{B} \text{ grad } \widehat{\phi}_h := \{\mathbf{B} \text{ grad } \phi_h^n\}_{n=1}^N$.

4.3.2 Error estimate of the fully discretized scheme

We study consistency errors of the fully discretized scheme (4.5), i.e., we establish a bound for the difference $\widehat{\phi} - \widehat{\phi}_h$. Let us introduce the notations

$$\widehat{e}_h := \widehat{\phi}_h - \widehat{\pi}_h \widehat{\phi}, \quad \widehat{\eta}_h := \widehat{\phi} - \widehat{\pi}_h \widehat{\phi} \quad (4.6)$$

then $\widehat{\phi} - \widehat{\phi}_h = \widehat{\eta}_h - \widehat{e}_h$, and, since $\widehat{\eta}_h$ can be estimated by Hypothesis 4.1, the problem is reduced to establish a bound for $\widehat{e}_h$. In order to prove the main result given in Theorem 4.2 we will use Hypotheses 3.2 to 3.5, notations and some results of Section 3.6.1, Hypothesis 4.1 and the following lemma.

Lemma 4.1 *Under Hypotheses 3.2, 3.3, 3.4, 3.5 and 4.1, if $\phi \in C^0(C^0(\overline{\Omega})) \cap C^0(H^{k+1}(\Omega)) \cap H^1(H^k(\Omega))$, and $c_1 \Delta t < 1/2$, the following inequality holds:*

$$\begin{aligned}
& \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\eta}_h, e_h^{n+1} \right\rangle \\
& \leq \frac{1}{8} \left\| \mathbf{C} \text{ grad } e_h^{n+1} + (\mathbf{C} \text{ grad } e_h^n) \circ X_E^n \right\|_0^2 + D_{\Delta t}^n \left(\frac{\Delta t}{2} \langle \mathbf{C} \text{ grad } \widehat{\eta}_h, \mathbf{C} \text{ grad } \widehat{e}_h \rangle \right) \\
& \quad + \frac{1}{8} \left\| \sqrt{l} e_h^{n+1} + (\sqrt{l} e_h^n) \circ X_E^n \right\|_0^2 + D_{\Delta t}^n \left(\frac{\Delta t}{2} \langle \sqrt{l} \widehat{\eta}_h, \sqrt{l} e_h^n \rangle \right) \\
& \quad + \frac{\alpha}{8} \left\| e_h^{n+1} + e_h^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1}) \right\|_{0, \Gamma_R}^2 + D_{\Delta t}^n \left(\frac{\alpha \Delta t}{2} \langle \widehat{\eta}_h, \widehat{e}_h \rangle_{\Gamma_R} \right) + c \|e_h^{n+1}\|_0^2 \\
& \quad + c \Delta t \left(\delta \left(\|\mathbf{B} \text{ grad } e_h^n\|_0^2 + \|\mathbf{B} \text{ grad } e_h^{n+1}\|_0^2 \right) + \left\| \sqrt{l} e_h^n \right\|_0^2 + \left\| \sqrt{l} e_h^{n+1} \right\|_0^2 + \alpha \|e_h^n\|_{0, \Gamma_R}^2 \right) \\
& \quad + \widetilde{c} K^2 h^{2k} \left(\frac{1}{\Delta t} \|\phi'\|_{L^2((t_n, t_{n+1}), H^k(\Omega))}^2 + \frac{1}{\Delta t} \|\phi\|_{L^2((t_n, t_{n+1}), H^{k+1}(\Omega))}^2 + \Delta t \|\phi^n\|_{k+1}^2 \right)
\end{aligned} \tag{4.7}$$

with $c = \max \{1, (c_1 c_2 + 1)/4\delta, c_1 c_3/4, c_1\}$, $\widetilde{c}$ a positive constant, and $\delta > 0$ and $\alpha > 0$ being the constants appearing, respectively, in Hypothesis 3.3 and equation (3.3).

PROOF. For the sake of clarity, the left hand side in (4.7) is firstly decomposed as a sum of the following terms:

$$\begin{aligned}
I_1 &= \left\langle \frac{\eta_h^{n+1} - \eta_h^n \circ X_{RK}^n}{\Delta t}, e_h^{n+1} \right\rangle, \\
I_2 &= \left\langle \frac{\mathbf{A} \text{ grad } \eta_h^{n+1} + (\mathbf{A} \text{ grad } \eta_h^n) \circ X_E^n}{2}, \text{ grad } e_h^{n+1} \right\rangle, \\
I_3 &= \frac{\Delta t}{2} \langle (\mathbf{L}^n \mathbf{A} \text{ grad } \eta_h^n) \circ X_E^n, \text{ grad } e_h^{n+1} \rangle, \\
I_4 &= \frac{\Delta t}{2} \langle (\text{grad div } \mathbf{v}^n \cdot \mathbf{A} \text{ grad } \eta_h^n) \circ X_E^n, e_h^{n+1} \rangle, \\
I_5 &= \left\langle \frac{l \eta_h^{n+1} + (l \eta_h^n) \circ X_E^n}{2}, e_h^{n+1} \right\rangle, \\
I_6 &= \alpha \left\langle \frac{\eta_h^{n+1} + \eta_h^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})}{2}, e_h^{n+1} \right\rangle_{\Gamma_R}.
\end{aligned}$$

In order to work with I_1 let us introduce the function

$$\begin{aligned}
Y_{RK}^n(\mathbf{y}, \cdot) : [t_n, t_{n+1}] &\longrightarrow \overline{\Omega} \\
s &\longrightarrow Y_{RK}^n(\mathbf{y}, s) := \mathbf{y} - \mathbf{v} \left(\mathbf{y} - \mathbf{v}^{n+1}(\mathbf{y}) \frac{t_{n+1} - s}{2}, \frac{t_{n+1} + s}{2} \right) (t_{n+1} - s),
\end{aligned}$$

for which we compute its partial derivative with respect to s ,

$$\begin{aligned}
\frac{\partial Y_{RK}^n}{\partial s}(\mathbf{y}, s) &= \mathbf{v} \left(\mathbf{y} - \mathbf{v}^{n+1}(\mathbf{x}) \frac{t_{n+1} - s}{2}, \frac{t_{n+1} + s}{2} \right) \\
&\quad - \frac{t_{n+1} - s}{2} \mathbf{v}' \left(\mathbf{y} - \mathbf{v}^{n+1}(\mathbf{y}) \frac{t_{n+1} - s}{2}, \frac{t_{n+1} + s}{2} \right) \\
&\quad - \frac{t_{n+1} - s}{2} \text{ grad } \mathbf{v} \left(\mathbf{y} - \mathbf{v}^{n+1}(\mathbf{y}) \frac{t_{n+1} - s}{2}, \frac{t_{n+1} + s}{2} \right) \mathbf{v}^{n+1}(\mathbf{y}),
\end{aligned} \tag{4.8}$$

and its gradient with respect to $\mathbf{y}$,

$$\begin{aligned} \nabla Y_{RK}^n(\mathbf{y}, s) &= \mathbf{I} - (t_{n+1} - s) \mathbf{L} \left(\mathbf{y} - \mathbf{v}^{n+1}(\mathbf{y}) \frac{t_{n+1} - s}{2}, \frac{t_{n+1} + s}{2} \right) \\ &\quad + \frac{(t_{n+1} - s)^2}{2} \mathbf{L} \left(\mathbf{y} - \mathbf{v}^{n+1}(\mathbf{y}) \frac{t_{n+1} - s}{2}, \frac{t_{n+1} + s}{2} \right) \mathbf{L}^{n+1}(\mathbf{y}). \end{aligned} \quad (4.9)$$

Thus, by using that $c_1 \Delta t < 1$ we have

$$\left| \frac{\partial Y_{RK}^n}{\partial s}(\mathbf{y}, s) \right| \leq c_1 + \frac{\Delta t}{2} c_1 + \frac{\Delta t}{2} c_1^2 \leq 1 + 2c_1 \quad \forall (\mathbf{y}, s) \in \Omega \times [t_n, t_{n+1}], \quad (4.10)$$

and analogous computations to those developed in Lemma 3.2 and Corollary 3.2 lead to

$$\left| \det (\nabla Y_{RK}^n)^{-1}(\mathbf{y}, s) \right| \leq 1 + c_1 \Delta t, \quad \forall (\mathbf{y}, s) \in \Omega \times [t_n, t_{n+1}]. \quad (4.11)$$

Moreover, noting that $Y_{RK}^n(\mathbf{y}, t_{n+1}) = \mathbf{y}$ and $Y_{RK}^n(\mathbf{y}, t_n) = X_{RK}^n(\mathbf{y})$, and by using Barrow's rule we have

$$\begin{aligned} &\frac{\eta_h^{n+1}(\mathbf{y}) - \eta_h^n(X_{RK}^n(\mathbf{y}))}{\Delta t} \\ &= \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \frac{d\eta_h}{ds} (Y_{RK}^n(\mathbf{y}, s), s) ds \\ &= \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \eta_h'(Y_{RK}^n(\mathbf{y}, s), s) ds + \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \text{grad } \eta_h(Y_{RK}^n(\mathbf{y}, s), s) \cdot \frac{\partial Y_{RK}^n}{\partial s}(\mathbf{y}, s) ds, \end{aligned} \quad (4.12)$$

where the chain rule has been used for the last equality. Next, by applying Holder inequality we get

$$\frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \eta_h'(Y_{RK}^n(\mathbf{y}, s), s) ds \leq \frac{1}{\sqrt{\Delta t}} \left(\int_{t_n}^{t_{n+1}} (\eta_h'(Y_{RK}^n(\mathbf{y}, s), s))^2 ds \right)^{\frac{1}{2}},$$

and then

$$\begin{aligned} &\int_{\Omega} \left(\frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \eta_h'(Y_{RK}^n(\mathbf{y}, s), s) ds \right)^2 d\mathbf{y} \leq \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \int_{\Omega} \eta_h'(Y_{RK}^n(\mathbf{y}, s), s)^2 d\mathbf{y} ds \\ &= \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \int_{\Omega} \eta_h'(\mathbf{z}, s)^2 \det (\nabla Y_{RK}^n)^{-1}(\mathbf{z}, s) d\mathbf{z} ds \leq \frac{(1 + c_1 \Delta t)}{\Delta t} \|\eta_h'\|_{L^2((t_n, t_{n+1}), L^2(\Omega))}^2, \end{aligned}$$

where the change of variable $\mathbf{z} = Y_{RK}^n(\mathbf{y}, s)$ and estimate (4.11) have been used. Analogously, for the last term of (4.12) we get

$$\begin{aligned} &\int_{\Omega} \left(\frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \text{grad } \eta_h(Y_{RK}^n(\mathbf{y}, s), s) \cdot \frac{\partial Y_{RK}^n}{\partial s}(\mathbf{y}, s) ds \right)^2 d\mathbf{y} \\ &\leq \frac{(1 + c_1 \Delta t)(1 + 2c_1)^2}{\Delta t} \|\text{grad } \eta_h\|_{L^2((t_n, t_{n+1}), L^2(\Omega))}^2, \end{aligned}$$

where we have considered also estimate (4.10). Finally, by applying (3.77) to term I_1 and taking into account the above results and Hypothesis 4.1 for $s = 0$ and $r = k$ we obtain

$$\begin{aligned} I_1 &\leq \frac{(1 + c_1 \Delta t)(1 + 2c_1)^2 K^2 h^{2k}}{2\Delta t} \left(\|\phi'\|_{L^2((t_n, t_{n+1}), H^k(\Omega))}^2 + \|\phi\|_{L^2((t_n, t_{n+1}), H^{k+1}(\Omega))}^2 \right) \\ &\quad + \frac{1}{2} \|e_h^{n+1}\|_0^2. \end{aligned} \quad (4.13)$$

Next, we decompose term I_2 into $I_2 = I_2^1 + I_2^2 + I_2^3$, with

$$\begin{aligned} I_2^1 &= \frac{1}{2} \langle \mathbf{C} \operatorname{grad} \eta_h^{n+1}, \mathbf{C} \operatorname{grad} e_h^{n+1} \rangle - \frac{1}{2} \langle \mathbf{C} \operatorname{grad} \eta_h^n, \mathbf{C} \operatorname{grad} e_h^n \rangle, \\ I_2^2 &= \frac{1}{2} \langle \mathbf{C} \operatorname{grad} \eta_h^n, \mathbf{C} \operatorname{grad} e_h^n \rangle - \frac{1}{2} \langle (\mathbf{C} \operatorname{grad} \eta_h^n) \circ X_E^n, (\mathbf{C} \operatorname{grad} e_h^n) \circ X_E^n \rangle, \\ I_2^3 &= \frac{1}{2} \langle (\mathbf{C} \operatorname{grad} \eta_h^n) \circ X_E^n, (\mathbf{C} \circ X_E^n)(\operatorname{grad} e_h^{n+1} + \operatorname{grad} e_h^n \circ X_E^n) \rangle. \end{aligned}$$

Notice that I_2^1 explicitly appears in (4.7). For I_2^2 we apply first the change of variable $\mathbf{y} = X_E^n(\mathbf{x})$ in the integral and Lemma 3.1. We obtain

$$\begin{aligned} I_2^2 &= \frac{1}{2} \int_{\Omega} (\mathbf{C} \operatorname{grad} \eta_h^n)(\mathbf{x}) \cdot (\mathbf{C} \operatorname{grad} e_h^n)(\mathbf{x}) \, d\mathbf{x} - \frac{1}{2} \int_{\Omega} (\mathbf{C} \operatorname{grad} \eta_h^n \cdot \mathbf{C} \operatorname{grad} e_h^n) \circ X_E^n(\mathbf{x}) \, d\mathbf{x} \\ &= \frac{1}{2} \int_{\Omega} (1 - \det((\mathbf{F}_E^n)^{-1}(\mathbf{x}))) (\mathbf{C} \operatorname{grad} \eta_h^n)(\mathbf{x}) \cdot (\mathbf{C} \operatorname{grad} e_h^n)(\mathbf{x}) \, d\mathbf{x} \\ &\leq \frac{c_1 \Delta t}{2} \left| \int_{\Omega} (\mathbf{C} \operatorname{grad} \eta_h^n)(\mathbf{x}) \cdot (\mathbf{C} \operatorname{grad} e_h^n)(\mathbf{x}) \, d\mathbf{x} \right| \\ &\leq \frac{c_1 \Delta t}{4} \left(\|\mathbf{C} \operatorname{grad} \eta_h^n\|_0^2 + \|\mathbf{C} \operatorname{grad} e_h^n\|_0^2 \right) \\ &\leq \frac{c_1 c_2 \Delta t}{4} \left(K^2 h^{2k} \|\phi^n\|_{k+1}^2 + \|\mathbf{B} \operatorname{grad} e_h^n\|_0^2 \right), \end{aligned}$$

where, in the last inequalities we have used, respectively, Corollary 3.2, inequality (3.77) and Hypotheses 3.3 and 4.1.

Now, by replacing in I_2^3 equality

$$\mathbf{C}(X_E^n(\mathbf{x})) = \mathbf{C}(\mathbf{x}) - \mathbf{D}^n(\mathbf{x}),$$

justified in (3.82)-(3.83), with $\mathbf{D}$ introduced in (3.84), we are led to

$$I_2^3 = \frac{1}{2} \langle (\mathbf{C} \operatorname{grad} \eta_h^n) \circ X_E^n, \mathbf{C} \operatorname{grad} e_h^{n+1} - \mathbf{D} \operatorname{grad} e_h^{n+1} + (\mathbf{C} \operatorname{grad} e_h^n) \circ X_E^n \rangle,$$

and, by using again (3.77),

$$I_2^3 \leq \|(\mathbf{C} \operatorname{grad} \eta_h^n) \circ X_E^n\|_0^2 + \frac{1}{8} \|\mathbf{C} \operatorname{grad} e_h^{n+1} + (\mathbf{C} \operatorname{grad} e_h^n) \circ X_E^n\|_0^2 + \frac{1}{8} \|\mathbf{D} \operatorname{grad} e_h^{n+1}\|_0^2.$$

Moreover, we have

$$\|(\mathbf{C} \operatorname{grad} \eta_h^n) \circ X_E^n\|_0^2 \leq (1 + c_1 \Delta t) c_2 K^2 h^{2k} \|\phi^n\|_{k+1}^2, \quad (4.14)$$

$$\|\mathbf{D} \operatorname{grad} e_h^{n+1}\|_0^2 \leq c_1^2 c_2 \Delta t^2 \|\mathbf{B} \operatorname{grad} e_h^{n+1}\|_0^2, \quad (4.15)$$

where, for (4.14) Lemma 3.3, Hypothesis 3.3 and Hypothesis 4.1 with $r = k + 1$ and $s = 1$ have been required, and for (4.15) estimate (3.85) has been applied.

Now, by jointly considering estimates for I_2^2 and I_2^3 and the lower bound δ of Hypothesis 3.3 we can state

$$\begin{aligned} I_2 &\leq I_2^1 + \left(\frac{c_1 c_2 \Delta t}{4} + (1 + c_1 \Delta t) c_2 \right) K^2 h^{2k} \|\phi^n\|_{k+1}^2 \\ &\quad + \frac{(2c_1 + c_1^2 \Delta t) c_2 \Delta t}{8\delta} \delta \left(\|\mathbf{B} \operatorname{grad} e_h^n\|_0^2 + \|\mathbf{B} \operatorname{grad} e_h^{n+1}\|_0^2 \right) \\ &\quad + \frac{1}{8} \|\mathbf{C} \operatorname{grad} e_h^{n+1} + (\mathbf{C} \operatorname{grad} e_h^n) \circ X_E^n\|_0^2. \end{aligned} \quad (4.16)$$

Similar reasoning, i.e., Lemma 3.3, Hypotheses 3.2, 3.3 and 4.1 lead to

$$\begin{aligned} \|(\mathbf{L}^n \mathbf{A} \nabla \eta_h^n) \circ X_E^n\|_0^2 &\leq (1 + c_1 \Delta t) c_1^2 c_2^2 K^2 h^{2k} \|\phi^n\|_{k+1}^2, \\ \|(\text{grad div } \mathbf{v}^n \cdot \mathbf{A} \nabla \eta_h^n) \circ X_E^n\|_0^2 &\leq (1 + c_1 \Delta t) c_1^2 c_2^2 K^2 h^{2k} \|\phi^n\|_{k+1}^2. \end{aligned}$$

Using these inequalities, I_3 and I_4 can be bounded as follows

$$I_3 \leq \frac{(1 + c_1 \Delta t) c_1^2 c_2^2 \Delta t K^2 h^{2k}}{4} \|\phi^n\|_{k+1}^2 + \frac{\Delta t}{4\delta} \delta \|\mathbf{B} \text{ grad } e_h^{n+1}\|_0^2, \quad (4.17)$$

$$I_4 \leq \frac{(1 + c_1 \Delta t) c_1^2 c_2^2 \Delta t^2 K^2 h^{2k}}{4} \|\phi^n\|_{k+1}^2 + \frac{1}{4} \|e_h^{n+1}\|_0^2, \quad (4.18)$$

where, in order to estimate I_3 , we have used Remark 3.15 and Hypothesis 3.3.

Next, term I_5 can be decomposed like term I_2 , namely, $I_5 = I_5^1 + I_5^2 + I_5^3$, where

$$\begin{aligned} I_5^1 &= \frac{1}{2} \left\langle \sqrt{l} \eta_h^{n+1}, \sqrt{l} e_h^{n+1} \right\rangle - \frac{1}{2} \left\langle \sqrt{l} \eta_h^n, \sqrt{l} e_h^n \right\rangle \\ I_5^2 &= \frac{1}{2} \left\langle \sqrt{l} \eta_h^n, \sqrt{l} e_h^n \right\rangle - \frac{1}{2} \left\langle (\sqrt{l} \eta_h^n) \circ X_E^n, (\sqrt{l} e_h^n) \circ X_E^n \right\rangle, \\ I_5^3 &= \frac{1}{2} \left\langle (\sqrt{l} \eta_h^n) \circ X_E^n, \sqrt{l} \circ X_E^n (e_h^{n+1} + e_h^n \circ X_E^n) \right\rangle. \end{aligned}$$

Moreover, by using again the function introduced in (3.82) and expression (3.91) we can rewrite I_5^3 as

$$\frac{1}{2} \left\langle (\sqrt{l} \eta_h^n) \circ X_E^n, \sqrt{l} e_h^{n+1} - \left(\int_{t_n}^{t_{n+1}} \text{grad } \sqrt{l} (Y_E^n(\mathbf{x}, s)) \cdot \mathbf{v}^{n+1}(\mathbf{x}) ds \right) e_h^{n+1} + (\sqrt{l} e_h^n) \circ X_E^n \right\rangle.$$

Thus, the same kind of computations used for I_2 lead to the following estimate:

$$\begin{aligned} I_5 &\leq I_5^1 + \left(\frac{c_1 c_3 \Delta t}{4} + (1 + c_1 \Delta t) c_3 \right) K^2 h^{2k} \|\phi^n\|_k^2 + \\ &\quad + \frac{(2c_1 + c_1^2 \Delta t) c_3 \Delta t}{8} \left(\|\sqrt{l} e_h^n\|_0^2 + \|\sqrt{l} e_h^{n+1}\|_0^2 \right) + \frac{1}{8} \|\sqrt{l} e_h^{n+1} + (\sqrt{l} e_h^n) \circ X_E^n\|_0^2. \end{aligned} \quad (4.19)$$

Finally, the boundary integral term I_6 is decomposed as a sum of the three terms

$$\begin{aligned} I_6^1 &= \frac{\alpha}{2} \langle \eta_h^{n+1}, e_h^{n+1} \rangle_{\Gamma_R} - \frac{\alpha}{2} \langle \eta_h^n, e_h^n \rangle_{\Gamma_R}, \\ I_6^2 &= \frac{\alpha}{2} \langle \eta_h^n, e_h^n \rangle_{\Gamma_R} - \frac{\alpha}{2} \langle \eta_h^n (1 + \Delta t \text{ div } \mathbf{v}^{n+1}), e_h^n (1 + \Delta t \text{ div } \mathbf{v}^{n+1}) \rangle_{\Gamma_R}, \\ I_6^3 &= \frac{\alpha}{2} \langle \eta_h^n (1 + \Delta t \text{ div } \mathbf{v}^{n+1}), e_h^{n+1} + e_h^n (1 + \Delta t \text{ div } \mathbf{v}^{n+1}) \rangle_{\Gamma_R}. \end{aligned}$$

Term I_6^1 appears explicitly in (4.7).

For I_6^2 and I_6^3 we use Hypothesis 3.2 and that $c_1 \Delta t < 1$ to establish the following bounds:

$$|1 - (1 + \Delta t \text{ div } \mathbf{v}^{n+1})^2(\mathbf{x})| \leq c_1 \Delta t (2 + c_1 \Delta t), \quad (1 + \Delta t \text{ div } \mathbf{v}^{n+1})^2(\mathbf{x}) \leq 1 + 2c_1 \Delta t + c_1^2 \Delta t^2,$$

$\forall \mathbf{x} \in \Omega$. Thus we have

$$I_6^2 \leq \frac{\alpha c_1 \Delta t (2 + c_1 \Delta t)}{4} \left(\|\eta_h^n\|_{0,\Gamma_R}^2 + \|e_h^n\|_{0,\Gamma_R}^2 \right)$$

and

$$I_6^3 \leq \frac{\alpha(1 + 2c_1\Delta t + c_1^2\Delta t^2)}{2} \|\eta_h^n\|_{0,\Gamma_R}^2 + \frac{\alpha}{8} \|e_h^{n+1} + e_h^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2.$$

Next, by using the continuity of the trace mapping, there exists a positive constant c_Ω such that

$$\|\eta_h^n\|_{0,\Gamma_R}^2 \leq c_\Omega \|\eta_h^n\|_1^2.$$

We deduce, by using also Hypothesis 4.1, that

$$\begin{aligned} I_6 &\leq I_6^1 + \alpha \left(\frac{c_1\Delta t(2 + c_1\Delta t)}{4} + \frac{(1 + 2c_1\Delta t + c_1^2\Delta t^2)}{2} \right) c_\Omega K^2 h^{2k} \|\phi^n\|_{k+1}^2 \\ &\quad + \frac{\alpha}{8} \|e_h^{n+1} + e_h^n(1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 + \frac{\alpha c_1\Delta t(2 + c_1\Delta t)}{4} \|e_h^n\|_{0,\Gamma_R}^2. \end{aligned} \quad (4.20)$$

Finally, by jointly considering (4.13), (4.16), (4.17), (4.18), (4.19) and (4.20), and taking into account that $c_1\Delta t < 1$ result (4.7) follows. $\square$

Theorem 4.2 (Error estimate) *Let us assume Hypotheses 3.3, 3.4, 3.5, 3.9 and 4.1. Let*

$$\phi \in Z^3 \cap C^0(H^{k+1}(\Omega)) \cap H^1(H^k(\Omega)), \quad \operatorname{grad} \phi \in Z^3, \quad \phi|_{\Gamma_R} \in Z^3(\Gamma_R),$$

be the solution of (3.107) and $\widehat{\phi}_h$ be the solution of (4.5) subject to the initial value $\phi_h^0 = \pi_h \phi^0$. Then, there exist two positive constants c and d , independent of h and Δt , such that, if $\Delta t < d$ we have

$$\begin{aligned} &\sqrt{\frac{1}{2}} \|\widehat{\phi} - \widehat{\phi}_h\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t \delta}{8}} \|\mathbf{B} \operatorname{grad} \widehat{\phi} - \mathbf{B} \operatorname{grad} \widehat{\phi}_h\|_{l^\infty(L^2(\Omega))} \\ &+ \sqrt{\frac{\Delta t}{8}} \|\sqrt{l}\widehat{\phi} - \sqrt{l}\widehat{\phi}_h\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t \alpha}{16}} \|\widehat{\phi} - \widehat{\phi}_h\|_{l^\infty(L^2(\Gamma_R))} \\ &\leq ch^k \left(\|\phi\|_{H^1(H^k(\Omega))} + \|\phi\|_{C^0(H^{k+1}(\Omega))} \right) \\ &\quad + c\Delta t^2 \left(\|\phi\|_{Z^3} + \|\mathbf{A} \operatorname{grad} \phi\|_{Z^3} + \|l\phi\|_{Z^2} + \|\phi\|_{Z^2,\Gamma_R} + \|f\|_{Z^2} + \|g\|_{Z^2,\Gamma_R} \right) \\ &\quad + c\Delta t^{\frac{5}{2}} \left(\|\phi\|_{Z^3,\Gamma_R} + \|g\|_{Z^3,\Gamma_R} \right), \end{aligned} \quad (4.21)$$

where $\mathbf{B} \operatorname{grad} (\widehat{\phi} - \widehat{\phi}_h) := \{\mathbf{B} \operatorname{grad} (\phi^n - \phi_h^n)\}_{n=1}^N$.

PROOF.

Firstly, recall that, according to (4.6) we have that $\widehat{e}_h = \widehat{\eta}_h - \widehat{\phi} + \widehat{\phi}_h$. By using also the definitions of schemes (4.5) and (3.107) the following identity holds

$$\begin{aligned} \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{e}_h, e_h^{n+1} \right\rangle &= \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} (\widehat{\eta}_h - \widehat{\phi} + \widehat{\phi}_h), e_h^{n+1} \right\rangle \\ &= \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\eta}_h, e_h^{n+1} \right\rangle - \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\phi}, e_h^{n+1} \right\rangle + \left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}}, e_h^{n+1} \right\rangle \\ &= \left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{\eta}_h, e_h^{n+1} \right\rangle + \left\langle \left(\mathcal{L}^{n+\frac{1}{2}} - \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \right) \widehat{\phi}, e_h^{n+1} \right\rangle + \left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}} - \mathcal{F}^{n+\frac{1}{2}}, e_h^{n+1} \right\rangle. \end{aligned} \quad (4.22)$$

A lower bound for (4.22) is given by Lemma 3.6, namely,

$$\begin{aligned}
\left\langle \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \widehat{e}_h, e_h^{n+1} \right\rangle &\geq D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{e}_h\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \widehat{e}_h\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \widehat{e}_h\|_0^2 + \frac{\alpha \Delta t}{4} \|\widehat{e}_h\|_{0,\Gamma_R}^2 \right) \\
&+ \frac{1}{2\Delta t} \|e_h^{n+1} - e_h^n \circ X_{RK}^n\|_0^2 + \frac{1}{4} \|\mathbf{C} \operatorname{grad} e_h^{n+1} + (\mathbf{C} \operatorname{grad} e_h^n) \circ X_E^n\|_0^2 \\
&+ \frac{1}{4} \|\sqrt{l} e_h^{n+1} + (\sqrt{l} e_h^n) \circ X_E^n\|_0^2 + \frac{\alpha}{4} \|\phi^{n+1} + \phi^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2 \\
&- \frac{c}{2} \left(\|e_h^n\|_0^2 + \|e_h^{n+1}\|_0^2 \right) - c\alpha \Delta t \|e_h^n\|_{0,\Gamma_R}^2 \\
&- c\Delta t \left(\|\sqrt{l} e_h^n\|_0^2 + \|\sqrt{l} e_h^{n+1}\|_0^2 \right) - c\delta \Delta t \left(\|\mathbf{B} \operatorname{grad} e_h^n\|_0^2 + \|\mathbf{B} \operatorname{grad} e_h^{n+1}\|_0^2 \right),
\end{aligned} \tag{4.23}$$

with $c = \max \{1, c_1, c_2, (2c_1c_2 + c_1c_2^2)/\delta, c_1c_3/\gamma\}$.

Now, by using Lemmas 3.18 and 3.19 we have

$$\begin{aligned}
&\left\langle \left(\mathcal{L}^{n+\frac{1}{2}} - \mathcal{L}_{\Delta t}^{n+\frac{1}{2}} \right) \widehat{\phi}, e_h^{n+1} \right\rangle + \left\langle \mathcal{F}_{\Delta t}^{n+\frac{1}{2}} - \mathcal{F}^{n+\frac{1}{2}}, e_h^{n+1} \right\rangle \\
&= \left\langle \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} - \xi_f^{n+\frac{1}{2}}, e_h^{n+1} \right\rangle + \left\langle \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} - \xi_g^{n+\frac{1}{2}}, e_h^{n+1} \right\rangle_{\Gamma_R} \\
&\leq \left(\left\| \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_f^{n+\frac{1}{2}} \right\|_0^2 \right) + \left\langle H^{n+\frac{1}{2}}, e_h^{n+1} - e_h^n \right\rangle + \frac{21}{\alpha} \left(\left\| \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 + \left\| \xi_g^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 \right) \\
&+ \|e_h^{n+1}\|_0^2 + \alpha c_1 \Delta t \|e_h^n\|_{0,\Gamma_R}^2 + \frac{\alpha}{8} \|e_h^{n+1} + e_h^n (1 + \Delta t \operatorname{div} \mathbf{v}^{n+1})\|_{0,\Gamma_R}^2,
\end{aligned} \tag{4.24}$$

with $H^{n+\frac{1}{2}}(\mathbf{x}) := (\xi_{\mathcal{L}_2}^{n+\frac{1}{2}}(\mathbf{x}) - \xi_g^{n+\frac{1}{2}}(\mathbf{x})) / (2 + \Delta t \operatorname{div} \mathbf{v}^{n+1}(\mathbf{x}))$ a.e. $\mathbf{x} \in \Omega$. Lemma 3.8 has been applied in the last inequality for the choices $\psi = e_h^{n+1}$ and $\varphi = e_h^n$, firstly for $F^{n+1} = \xi_{\mathcal{L}_1}^{n+\frac{1}{2}}$, $G^{n+1} = \xi_{\mathcal{L}_2}^{n+\frac{1}{2}}$ and secondly for $F^{n+1} = -\xi_f^{n+\frac{1}{2}}$, $G^{n+1} = -\xi_g^{n+\frac{1}{2}}$.

By jointly considering the lower bound of (4.22) given in (4.23) and the upper bound given in (4.24) and Lemma 4.1 we deduce

$$\begin{aligned}
&D_{\Delta t}^n \left(\frac{1}{2} \|\widehat{e}_h\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \operatorname{grad} \widehat{e}_h\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l} \widehat{e}_h\|_0^2 + \frac{\alpha \Delta t}{4} \|\widehat{e}_h\|_{0,\Gamma_R}^2 \right) \\
&\leq \widetilde{c} K^2 h^{2k} \left(\frac{1}{\Delta t} \|\phi'\|_{L^2((t_n, t_{n+1}), H^k(\Omega))}^2 + \frac{1}{\Delta t} \|\phi\|_{L^2((t_n, t_{n+1}), H^{k+1}(\Omega))}^2 + \Delta t \|\phi^n\|_{k+1}^2 \right) \\
&+ \widetilde{c} \left(\left\| \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_f^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 + \left\| \xi_g^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 \right) + \left\langle H^{n+\frac{1}{2}}, e_h^{n+1} - e_h^n \right\rangle_{\Gamma_R} \\
&+ \frac{\Delta t}{2} D_{\Delta t}^n \left(\langle \mathbf{C} \operatorname{grad} \widehat{\eta}_h, \mathbf{C} \operatorname{grad} \widehat{e}_h \rangle + \langle \sqrt{l} \widehat{\eta}_h, \sqrt{l} \widehat{e}_h \rangle + \alpha \langle \widehat{\eta}_h, \widehat{e}_h \rangle_{\Gamma_R} \right) \\
&+ c \left(\|e_h^n\|_0^2 + \|e_h^{n+1}\|_0^2 \right) + c\Delta t \left(\|\sqrt{l} e_h^n\|_0^2 + \|\sqrt{l} e_h^{n+1}\|_0^2 \right) \\
&+ c\Delta t \left(\delta \left(\|\mathbf{B} \operatorname{grad} e_h^n\|_0^2 + \|\mathbf{B} \operatorname{grad} e_h^{n+1}\|_0^2 \right) + \alpha \|e_h^n\|_{0,\Gamma_R}^2 \right),
\end{aligned} \tag{4.25}$$

where we have omitted some positive terms in the left hand side. In (4.25), $\widetilde{c}$ is a positive constant and $c = \max \{2, (3c_1c_2 + c_1c_2^2 + 1)/\delta, c_1c_3(1/\gamma + 1/4), c_1, c_2\}$. Now, for fixed integer $q \geq 1$, let

us multiply inequality (4.25) by Δt and sum it from $n = 0$ to $n = q - 1$. We get

$$\begin{aligned}
& \frac{1}{2} \|e_h^q\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \text{ grad } e_h^q\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l}e_h^q\|_0^2 + \frac{\alpha\Delta t}{4} \|e_h^q\|_{0,\Gamma_R}^2 \\
& \leq \frac{1}{2} \|e_h^0\|_0^2 + \frac{\Delta t}{4} \|\mathbf{C} \text{ grad } e_h^0\|_0^2 + \frac{\Delta t}{4} \|\sqrt{l}e_h^0\|_0^2 + \frac{\alpha\Delta t}{4} \|e_h^0\|_{0,\Gamma_R}^2 \\
& \quad + \tilde{c}K^2h^{2k} \left(\sum_{n=0}^{q-1} \|\phi'\|_{L^2((t_n,t_{n+1}),H^k(\Omega))}^2 + \sum_{n=0}^{q-1} \|\phi\|_{L^2((t_n,t_{n+1}),H^{k+1}(\Omega))}^2 + \sum_{n=0}^{q-1} \Delta t^2 \|\phi^n\|_{k+1}^2 \right) \\
& \quad + \tilde{c} \sum_{n=0}^{q-1} \Delta t \left(\left\| \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_f^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 + \left\| \xi_g^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 \right) \\
& \quad + \sum_{n=0}^{q-1} \Delta t \left\langle H^{n+\frac{1}{2}}, e_h^{n+1} - e_h^n \right\rangle_{\Gamma_R} + \frac{\Delta t}{2} (\langle \mathbf{C} \text{ grad } \eta_h^q, \mathbf{C} \text{ grad } e_h^q \rangle - \langle \mathbf{C} \text{ grad } \eta_h^0, \mathbf{C} \text{ grad } e_h^0 \rangle) \\
& \quad + \frac{\Delta t}{2} (\langle \sqrt{l}\eta_h^q, \sqrt{l}e_h^q \rangle - \langle \sqrt{l}\eta_h^0, \sqrt{l}e_h^0 \rangle) + \frac{\alpha\Delta t}{2} (\langle \eta_h^q, e_h^q \rangle_{\Gamma_R} - \langle \eta_h^0, e_h^0 \rangle_{\Gamma_R}) \\
& \quad + 2c\Delta t \sum_{n=0}^q \|e_h^n\|_0^2 + 2c\Delta t^2 \left(\sum_{n=0}^q \|\mathbf{B} \text{ grad } e_h^n\|_0^2 + \sum_{n=0}^q \|\sqrt{l}e_h^n\|_0^2 + \sum_{n=0}^{q-1} \|e_h^n\|_{0,\Gamma_R}^2 \right).
\end{aligned} \tag{4.26}$$

Some of the terms on the right hand side of (4.26) can also be bounded. Firstly, we have

$$\begin{aligned}
& \sum_{n=0}^{q-1} \|\phi'\|_{L^2((t_n,t_{n+1}),H^k(\Omega))}^2 + \sum_{n=0}^{q-1} \|\phi\|_{L^2((t_n,t_{n+1}),H^{k+1}(\Omega))}^2 + \sum_{n=0}^{q-1} \Delta t^2 \|\phi^n\|_{k+1}^2 \\
& \leq \|\phi'\|_{L^2(H^k(\Omega))}^2 + \|\phi\|_{L^2(H^{k+1}(\Omega))}^2 + \Delta t T \|\phi\|_{C^0(H^{k+1}(\Omega))}^2 \\
& \leq \|\phi'\|_{L^2(H^k(\Omega))}^2 + (1 + \Delta t)T \|\phi\|_{C^0(H^{k+1}(\Omega))}^2,
\end{aligned}$$

where T is the measure of the time interval. Secondly, by using also Lemmas 3.18 and 3.19 and the fact that $\mathbf{A} \text{ grad } \phi \cdot \mathbf{n} = g - \alpha\phi$ on the boundary Γ_R we are led to

$$\begin{aligned}
& \sum_{n=0}^{q-1} \Delta t \left(\left\| \xi_{\mathcal{L}_1}^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_f^{n+\frac{1}{2}} \right\|_0^2 + \left\| \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 + \left\| \xi_g^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 \right) \\
& \leq \Delta t^4 \tilde{c}_1 T (\|\phi\|_{Z^3} + \|\mathbf{A} \text{ grad } \phi\|_{Z^3} + \|l\phi\|_{Z^2} + \|\alpha\phi\|_{Z^2,\Gamma_R} + \|f\|_{Z^2} + \|g\|_{Z^2,\Gamma_R}).
\end{aligned}$$

Thirdly, by applying result (3.101) of Lemma 3.9, we obtain the estimate

$$\begin{aligned}
& \left| \sum_{n=0}^{q-1} \Delta t \left\langle H^{n+\frac{1}{2}}, e_h^{n+1} - e_h^n \right\rangle_{\Gamma_R} \right| \leq \frac{\alpha\Delta t}{8} \|e_h^q\|_{0,\Gamma_R}^2 + \frac{\alpha\Delta t}{8} \|e_h^0\|_{0,\Gamma_R}^2 + 6\alpha\Delta t^2 \sum_{n=0}^{q-1} \|e_h^n\|_{0,\Gamma_R}^2 \\
& \quad + \frac{4 + c_1\Delta t}{\alpha} \Delta t \left(\sum_{n=0}^q \left\| \xi_{\mathcal{L}_2}^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 + \sum_{n=0}^q \left\| \xi_g^{n+\frac{1}{2}} \right\|_{0,\Gamma_R}^2 \right) \\
& \quad + \frac{\Delta t^2}{\alpha} \left(\sum_{n=0}^{q-1} \left\| \frac{\xi_{\mathcal{L}_2}^{n+\frac{1}{2}} - \xi_{\mathcal{L}_2}^{n-\frac{1}{2}}}{\Delta t} \right\|_{0,\Gamma_R}^2 + \sum_{n=0}^{q-1} \left\| \frac{h_g^{n+\frac{1}{2}} - h_g^{n-\frac{1}{2}}}{\Delta t} \right\|_{0,\Gamma_R}^2 \right).
\end{aligned}$$

Thus, by using again the bounds given in Lemmas 3.18 and 3.19, and taking into account Remark 3.17, we get

$$\begin{aligned} \left| \sum_{n=0}^{q-1} \left\langle H^{n+\frac{1}{2}}, e_h^{n+1} - e_h^n \right\rangle_{\Gamma_R} \right| &\leq \frac{\alpha \Delta t}{8} \|e_h^q\|_{0,\Gamma_R}^2 + \frac{\alpha \Delta t}{8} \|e_h^0\|_{0,\Gamma_R}^2 + 6\alpha \Delta t^2 \sum_{n=0}^{q-1} \|e_h^n\|_{0,\Gamma_R}^2 \\ &+ \frac{5}{\alpha} \Delta t^4 T \tilde{c}_1 (\|\phi\|_{Z^3} + \|\mathbf{A} \operatorname{grad} \phi\|_{Z^3} + \|l\phi\|_{Z^2} + \|\alpha \phi\|_{Z^2,\Gamma_R} + \|f\|_{Z^2} + \|g\|_{Z^2,\Gamma_R}) \\ &+ \frac{\Delta t^5 T}{\alpha} (\|\alpha \phi\|_{Z^3,\Gamma_R} + \|g\|_{Z^3,\Gamma_R}). \end{aligned}$$

Fourthly, analogous computations to those developed in Lemma 4.1 give

$$\begin{aligned} \frac{\Delta t}{2} \left\langle \mathbf{C} \operatorname{grad} \eta_h^j, \mathbf{C} \operatorname{grad} e_h^j \right\rangle &\leq \frac{K^2 c_2 h^{2k} \Delta t}{2} \|\phi^j\|_{k+1}^2 + \frac{\Delta t}{8} \|\mathbf{C} \operatorname{grad} e_h^j\|_0^2, \\ \frac{\Delta t}{2} \left\langle \sqrt{l} \eta_h^j, \sqrt{l} e_h^j \right\rangle &\leq \frac{K^2 c_3 h^{2k} \Delta t}{2} \|\phi^j\|_{k+1}^2 + \frac{\Delta t}{8} \|\sqrt{l} e_h^j\|_0^2, \\ \frac{\alpha \Delta t}{2} \left\langle \eta_h^j, e_h^j \right\rangle_{\Gamma_R} &\leq \alpha K^2 c_\Omega h^{2k} \Delta t \|\phi^j\|_{k+1}^2 + \frac{\alpha \Delta t}{16} \|e_h^j\|_{0,\Gamma_R}^2 \quad \text{for } j = 0, q. \end{aligned}$$

Let us introduce, for $n = 0, \dots, N$, the notation:

$$\begin{aligned} \theta_n &:= \frac{1}{2} \|e_h^n\|_0^2 + \frac{\delta \Delta t}{8} \|\mathbf{B} \operatorname{grad} e_h^n\|_0^2 + \frac{\Delta t}{8} \|\sqrt{l} e_h^n\|_0^2, \\ \bar{\theta}_n &:= \frac{\alpha \Delta t}{16} \|e_h^n\|_{0,\Gamma_R}^2. \end{aligned}$$

With the above notation, previous estimates lead to

$$(1 - 16c\Delta t)\theta_q + \bar{\theta}_q \leq 16c\Delta t \sum_{n=0}^{q-1} \theta_n + 128c\Delta t \sum_{n=0}^{q-1} \bar{\theta}_n + \tilde{c}(\theta_0 + \bar{\theta}_0 + C),$$

where C contains the constant terms multiplied by h^{2k} , by Δt^4 and by Δt^5 .

Finally, taking into account that $e_h^0 = 0$, the result is concluded by discrete Gronwall's inequality (see for instance [95]). $\square$

4.4 Classical Lagrange-Galerkin scheme

The classical characteristics method (4.2) can be equivalently formulated as

$$\begin{cases} \text{Given } \phi_h^0 \in V_h^k, \text{ find } \widehat{\phi}_h := \{\phi_h^n\}_{n=1}^N \in [V_h^k]^N \text{ such that} \\ \left\langle \mathcal{L}_{\Delta t}^{n+1} \widehat{\phi}_h, \psi_h \right\rangle = \left\langle \mathcal{F}_{\Delta t}^{n+1}, \psi_h \right\rangle \quad \forall \psi_h \in V_h^k \text{ for } n = 0, \dots, N-1, \end{cases} \quad (4.27)$$

where the operators have been introduced in Section (3.7). The previous scheme has proved to be first order accurate in time and space (when combined with linear finite elements) for convection-diffusion equations (see [45, 89, 14]).

Similarly to Section 3.7, we establish analogous stability and error estimates results to those given in Section 4.3.

Theorem 4.3 (Stability) *Under Hypotheses 3.2, 3.3, 3.5, 3.6 and 3.7. Let $\widehat{\phi}_h = \{\phi_h^n\}_{n=1}^N$ be the solution of (4.27) subject to initial value $\phi_h^0 \in L^2(\Omega)$. Then, there exist two positive constants c and d , such that, if $\Delta t < d$ then*

$$\begin{aligned} & \frac{1}{\sqrt{2}} \left\| \widehat{\phi}_h \right\|_{l^\infty(L^2(\Omega))} + \sqrt{\Delta t} \left\| \mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Omega))} + \sqrt{\Delta t} \left\| \sqrt{l} \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Omega))} \\ & + \sqrt{\frac{\alpha \Delta t}{2}} \left\| \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Gamma_R))} \leq c \left(\frac{1}{2} \left\| \phi_{\Delta t}^0 \right\|_0 + \left\| \widehat{f} \right\|_{l^2(L^2(\Omega))} + \left\| \widehat{g} \right\|_{l^2(L^2(\Gamma_R))} \right), \end{aligned} \quad (4.28)$$

where $\mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} := \{\mathbf{C} \text{ grad } \phi_{\Delta t}^n\}$.

PROOF. It is analogous to Theorem 3.4 for the stability of the semidiscretized scheme. $\square$

Theorem 4.4 (Consistency error) *Assume Hypotheses 3.2, 3.3, 3.5, 3.6, 3.7 and that $\mathbf{v} \in C^1(L^\infty(\Omega)) \cap C^0(W^{1,\infty}(\Omega))$. Let*

$$\phi \in Z^2 \cap C^0(H^{k+1}(\Omega)) \cap H^1(H^k(\Omega))$$

be the solution of (3.134) and $\widehat{\phi}_{\Delta t} = \{\phi_{\Delta t}^n\}_{n=1}^N$ be the solution of (3.130) subject to initial value $\phi_{\Delta t}^0 = \phi^0$. Then, there exist two positive constants c and d , such that, if $\Delta t < d$ then

$$\begin{aligned} & \frac{1}{\sqrt{2}} \left\| \widehat{\phi} - \widehat{\phi}_{\Delta t} \right\|_{l^\infty(L^2(\Omega))} + \sqrt{\frac{\Delta t}{4}} \left\| \mathbf{C} \text{ grad } \widehat{\phi} - \mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Omega))} \\ & + \sqrt{\frac{\Delta t}{2}} \left\| \sqrt{l} \widehat{\phi} - \sqrt{l} \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Omega))} + \sqrt{\frac{\alpha \Delta t}{2}} \left\| \widehat{\phi} - \widehat{\phi}_{\Delta t} \right\|_{l^2(L^2(\Gamma_R))} \\ & \leq c \Delta t \left\| \phi \right\|_{Z^2} + c h^k \left(\left\| \phi \right\|_{H^1(H^k(\Omega))} + \left\| \phi \right\|_{C^0(H^{k+1}(\Omega))} \right), \end{aligned} \quad (4.29)$$

where $\mathbf{C} \text{ grad } \widehat{\phi}_{\Delta t} := \{\mathbf{C} \text{ grad } \phi_{\Delta t}^n\}$.

PROOF. We only indicate the steps in the proof, which are similar to those of Theorem 4.2.

Firstly, the following identity holds

$$\langle \mathcal{L}_{\Delta t}^{n+1} \widehat{e}_h, e_h^{n+1} \rangle = \langle \mathcal{L}_{\Delta t}^{n+1} \widehat{\eta}_h, e_h^{n+1} \rangle + \langle (\mathcal{L}^{n+1} - \mathcal{L}_{\Delta t}^{n+1}) \widehat{\phi}, e_h^{n+1} \rangle + \langle \mathcal{F}_{\Delta t}^{n+1} - \mathcal{F}^{n+1}, e_h^{n+1} \rangle. \quad (4.30)$$

Now, from Lemma 3.20, we deduce

$$\begin{aligned} \langle \mathcal{L}_{\Delta t}^{n+1} \widehat{e}_h, e_h^{n+1} \rangle & \geq D_{\Delta t}^n \left(\frac{1}{2} \left\| \widehat{e}_h \right\|_0^2 \right) + \left\| \mathbf{C} \text{ grad } e_h^{n+1} \right\|_0^2 + \left\| \sqrt{l} e_h^{n+1} \right\|_0^2 + \alpha \left\| e_h^{n+1} \right\|_{0,\Gamma_R}^2 \\ & + \frac{1}{2\Delta t} \left\| e_h^{n+1} - e_h^n \circ X_E^n \right\|_0^2 - \frac{c_1}{2} \left\| e_h^n \right\|_0^2. \end{aligned}$$

As we have explained in Section 3.7, the time error only comes from the approximation of the material derivative term. Thus, we have,

$$\langle \mathcal{F}_{\Delta t}^{n+1} - \mathcal{F}^{n+1}, e_h^{n+1} \rangle = 0$$

and

$$\langle (\mathcal{L}^{n+1} - \mathcal{L}_{\Delta t}^{n+1}) \widehat{\phi}, e_h^{n+1} \rangle = \langle \xi^{n+1}, e_h^{n+1} \rangle \leq \widetilde{c}_1 \Delta t^2 \widetilde{c}_1 \left\| \varphi \right\|_{Z^2}^2 + \frac{1}{4} \left\| e_h^{n+1} \right\|_0^2,$$

where Lemma 3.22 and inequality (3.77) have been used.

Finally, the spatial error derives from the term

$$\langle \mathcal{L}_{\Delta t}^{n+1} \widehat{\eta}_h, e_h^{n+1} \rangle = I_1 + I_2 + I_3 + I_4,$$

with

$$\begin{aligned} I_1 &= \left\langle \frac{\eta_h^{n+1} - \eta_h^n \circ X_E^n}{\Delta t}, e_h^{n+1} \right\rangle, \\ I_2 &= \langle \mathbf{A} \operatorname{grad} \eta_h^{n+1}, \operatorname{grad} e_h^{n+1} \rangle, \\ I_3 &= \langle l \eta_h^{n+1}, e_h^{n+1} \rangle, \\ I_4 &= \alpha \langle \eta_h^{n+1}, e_h^{n+1} \rangle_{\Gamma_R}. \end{aligned}$$

Similar computations to those given in Lemma 4.1 for I_1 yield

$$I_1 \leq \frac{cK^2 h^{2k}}{\Delta t} \left(\|\phi'\|_{L^2((t_n, t_{n+1}, H^k(\Omega)))}^2 + \|\phi\|_{L^2((t_n, t_{n+1}, H^{k+1}(\Omega)))}^2 \right) + \frac{1}{4} \|e_h^{n+1}\|_0^2.$$

For the other terms, inequality (3.77) and Hypothesis 4.1 give

$$\begin{aligned} I_2 &\leq cK^2 h^{2k} \|\phi^n\|_{k+1}^2 + \frac{1}{2} \|\mathbf{C} \operatorname{grad} e_h^{n+1}\|_0^2, \\ I_3 &\leq cK^2 h^{2k} \|\phi^n\|_k^2 + \frac{1}{2} \|\sqrt{l} e_h^{n+1}\|_0^2, \\ I_4 &\leq cK^2 h^{2k} \|\phi^n\|_{k+1}^2 + \frac{\alpha}{2} \|e_h^{n+1}\|_{0, \Gamma_R}^2. \end{aligned}$$

In the above, c denotes a positive constant.

Finally, by applying the previous estimates to equation (4.30), we have (4.29) follows by the discrete Gronwall's inequality. $\square$

4.5 Finite element spaces and quadrature formulas

Results concerning stability and consistency properties of scheme (4.5) have been proved for a wide class of finite element spaces (we only require the interpolation property given in Hypothesis 4.1), under the assumption that the inner products are exactly integrated. Similar results have been stated in the literature for the classical or the two-step scheme. Nevertheless, numerical integration has to be used in practice to approximate the involved integrals. It is well known that, for the classical first order in time Lagrange-Galerkin method, numerical quadrature can lead to conditionally stable schemes [82, 103, 94]. Moreover, new terms may appear in the error estimates (see [88, 82] and Section 4.6).

In the present section we analyze the stability of (4.5) when combined with some finite element spaces and quadrature formulas, extending the studies in the literature regarding the classical scheme. In Section 4.6 we present some numerical tests showing the influence of quadrature formulas in both stability and consistency errors.

Most of the papers in the literature study the classical Lagrange-Galerkin method for piecewise linear finite elements. In particular, conditional instability is shown in [82] for Gauss-Legendre, Gauss-Lobatto (with more than three points), Radau and Newton-Cotes formula,

when applied to the linear convection equation. This work was extended to the linear convection-diffusion equation in [103] and to a wider class of quadrature formulas in [94]. For both convection and convection-diffusion equations, Gauss-Lobatto quadrature formulas lead to the most stable schemes. However, only the trapezoidal rule (or two-point Gauss-Lobatto) preserves unconditional stability. We have not found any positive statement concerning classical Lagrange-Galerkin method when using quadrature formulas for quadratic elements. In [52] a Fourier analysis is developed for both the classical and the two-step methods, applied to the one dimensional linear advection equation and combined with Gauss-Legendre quadrature formulas. The regions of instability are determined numerically, and they coincide with those predicted in [82] for the classical scheme. Regarding the results, the two-step method seems to be more unstable than the classical one. In [105] authors show experimentally that the Crank-Nicholson scheme is more robust than the first order scheme with respect to numerical integration error and produce better numerical results. More precisely, they consider linear finite elements over triangular meshes and divide each triangle into j^2 small triangles, for $j = 1, 2, 3$. A three-point Gauss-Lobatto quadrature formula is developed in each of the small triangles and it is only applied to the integral terms where the characteristic lines appear. A theoretical analysis supporting the observed results is announced for a forthcoming paper.

The theoretical results of [82, 103, 94] are established by using Fourier analysis for constant coefficients equations and only in the one dimensional case. In [82] the analysis has been generalized to m dimensions for the linear convection equation under some particular conditions. Throughout this section, we develop an analogous approach, extending their results to the Crank-Nicholson Lagrange-Galerkin method, applied to the linear convection-diffusion equation with constant coefficients, namely,

$$\frac{\partial \phi}{\partial t} - \nu \Delta \phi + \mathbf{v} \cdot \nabla \phi = 0, \quad (\mathbf{x}, t) \in \Omega \times [0, T], \quad (4.31)$$

with $\nu, \mathbf{v}$ constant and $\nu \geq 0$. For this purpose, we establish Lemma 4.2, which allows us generalize to m dimensions the analysis developed in one dimension. In posterior sections, the one-dimensional analysis for different Lagrange-Galerkin schemes, finite element spaces and quadrature formulas is carried out.

Remark 4.1 *Although we have stated the analysis of the previous and the present chapter for $m = 2, 3$, it is easy to consider the corresponding scheme for $m = 1$ and to prove similar results. In fact, the only difference appears when writing boundary conditions, but they are not considered in Fourier analysis.*

Let us first introduce the notation

$$T_x[\psi] := \psi(x - v\Delta t),$$

for $x \in \Omega \subset \mathbb{R}$ and $\psi : \Omega \rightarrow \mathbb{R}$. Thus, scheme (4.5) applied to the one dimensional version of equation (4.31), without considering boundary conditions, has the form

$$\frac{1}{\Delta t} \int_{\Omega} \phi_h^{n+1} \psi_j dx + \frac{\nu}{2} \int_{\Omega} \frac{d\phi_h^{n+1}}{dx} \frac{d\psi_j}{dx} dx = \frac{1}{\Delta t} \int_{\Omega} T_x[\phi_h^n] \psi_j dx - \frac{\nu}{2} \int_{\Omega} T_x \left[\frac{d\phi_h^n}{dx} \right] \frac{d\psi_j}{dx} dx, \quad (4.32)$$

where ψ_j is the j -th basis function of the chosen one-dimensional finite element space. Similarly, for the two-dimensional case,

$$\begin{aligned} T_{(x_1, x_2)}[\psi] &:= \psi(x_1 - v_1 \Delta t, x_2 - v_2 \Delta t), \\ T_{x_1}[\psi] &:= \psi(x_1 - v_1 \Delta t, x_2), \\ T_{x_2}[\psi] &:= \psi(x_1, x_2 - v_2 \Delta t), \end{aligned}$$

for $\psi : (x_1, x_2) \in \Omega_1 \times \Omega_2 \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, and $\mathbf{v} = (v_1, v_2)$. In the previous definitions, Ω_1 and Ω_2 are intervals of $\mathbb{R}$.

Finally, scheme (4.5) applied to equation (4.31) assuming that $\Omega := \Omega_1 \times \dots \times \Omega_m$, with Ω_i intervals of $\mathbb{R}$, and with similar notation to the introduced above, has the form

$$\begin{aligned} & \frac{1}{\Delta t} \int_{\Omega} \phi_h^{n+1} \psi_j \, d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} \text{grad } \phi_h^{n+1} \cdot \text{grad } \psi_j \, d\mathbf{x} \\ &= \frac{1}{\Delta t} \int_{\Omega} T_{\mathbf{x}} [\phi_h^n] \psi_j \, d\mathbf{x} - \frac{\nu}{2} \int_{\Omega} T_{\mathbf{x}} [\text{grad } \phi_h^n] \cdot \text{grad } \psi_j \, d\mathbf{x}, \end{aligned} \quad (4.33)$$

Remark 4.2 Notice that, for $\mathbf{v}$ constant, Euler and Runge Kutta approximations lead to the same scheme.

Lemma 4.2 For linear convection-diffusion equation in m dimensions with constant coefficients, the second order Lagrange-Galerkin method (4.33) is just a tensor product of one-dimensional second order Lagrange-Galerkin methods assuming that the basis functions themselves are tensor products of one-dimensional basis functions on a grid which is uniform in each coordinate direction.

PROOF. We follow the proof in [82] for the first order Lagrange-Galerkin method in the case $m = 2$, adding the new diffusion terms. The general case can be solved by induction on m . For the sake of simplicity, the notation used for the two-dimensional case is similar to that for the one-dimensional case.

By hypothesis, the two-dimensional basis functions are tensor products of one dimensional basis functions, namely,

$$\psi_j(x_1, x_2) = \psi_{j_1}(x_1) \psi_{j_2}(x_2), \quad (4.34)$$

for indices j_1, j_2 , with $j = (j_1, j_2)$. Moreover, since $\mathbf{v} = (v_1, v_2)$ is constant and the mesh is uniform, we have

$$T_{(x_1, x_2)} = T_{x_1} \circ T_{x_2}, \quad (4.35)$$

$$\int_{\Omega_1 \times \Omega_2} \psi(x_1, x_2) \, dx_1 dx_2 = \int_{\Omega_2} \left(\int_{\Omega_1} \psi(x_1, x_2) \, dx_1 \right) dx_2. \quad (4.36)$$

Next, we analyze the different terms appearing in the discretized problem.

- The two-dimensional mass matrix is the tensor product of one dimensional mass matrices:

$$\begin{aligned} & \int_{\Omega_1 \times \Omega_2} \psi_i(x_1, x_2) \psi_j(x_1, x_2) \, dx_1 dx_2 \\ &= \int_{\Omega_2} \int_{\Omega_1} \psi_{i_1}(x_1) \psi_{i_2}(x_2) \psi_{j_1}(x_1) \psi_{j_2}(x_2) \, dx_1 dx_2 \\ &= \left(\int_{\Omega_1} \psi_{i_1}(x_1) \psi_{j_1}(x_1) \, dx_1 \right) \left(\int_{\Omega_2} \psi_{i_2}(x_2) \psi_{j_2}(x_2) \, dx_2 \right). \end{aligned}$$

- The two-dimensional stiffness matrix is the tensor product of one dimensional mass and stiffness matrices. More precisely, the stiffness matrix is computed as follows:

$$\sum_{r=1}^2 \int_{\Omega_1 \times \Omega_2} \frac{\partial \psi_i}{\partial x_m}(x_1, x_2) \frac{\partial \psi_j}{\partial x_l}(x_1, x_2) \, dx_1 dx_2.$$

Applying the same arguments as for the mass matrix and the product rule we obtain

$$\left(\int_{\Omega_1} \frac{d\psi_{i_1}}{dx_1}(x_1) \frac{d\psi_{j_1}}{dx_1}(x_1) dx_1 \right) \left(\int_{\Omega_2} \psi_{i_2}(x_2) \psi_{j_2}(x_2) dx_2 \right),$$

for $l = 1$, and

$$\left(\int_{\Omega_1} \psi_{i_1}(x_1) \psi_{j_1}(x_1) dx_1 \right) \left(\int_{\Omega_2} \frac{d\psi_{i_2}}{dx_2}(x_2) \frac{d\psi_{j_2}}{dx_2}(x_2) dx_2 \right),$$

for $l = 2$.

- For the second member associated to the first order characteristics method we use properties (4.34)-(4.36) and take into account that T_{x_i} and I_{x_i} have no effect on a function depending only on x_j with $i \neq j$. We get

$$\begin{aligned} & \int_{\Omega_1 \times \Omega_2} \psi_i(x_1, x_2) T_{(x_1, x_2)} [\phi^n(x_1, x_2)] dx_1 dx_2 \\ &= \int_{\Omega_1} \psi_{i_1}(x_1) T_{x_1} \left[\int_{\Omega_2} \psi_{i_2}(x_2) T_{x_2} \phi_h^n(x_1, x_2) dx_2 \right] dx_1. \end{aligned} \quad (4.37)$$

- For the second member associated to the second order characteristics method, the term containing the derivative with respect to x_1 is given by

$$\begin{aligned} & \int_{\Omega_2} \int_{\Omega_1} \frac{d\psi_{i_1}}{dx_1}(x_1) \psi_{i_2}(x_2) T_{x_1} \left[T_{x_2} \left[\frac{\partial \phi_h^n}{\partial x_1}(x_1, x_2) \right] \right] dx_1 dx_2 \\ &= \int_{\Omega_1} \frac{d\psi_{i_1}}{dx_1}(x_1) T_{x_1} \left[\int_{\Omega_2} \psi_{i_2}(x_2) T_{x_2} \left[\frac{\partial \phi_h^n}{\partial x_1}(x_1, x_2) \right] \right] dx_2 dx_1. \end{aligned} \quad (4.38)$$

In the same way, we can deal with the term containing the derivative with respect to x_2 .

Finally, for both (4.37) and (4.38), the result follows from the fact that

$$\phi_h^n(x_1, x_2) = \sum_j (\phi_h)_j^n \psi_{j_1}(x_1) \psi_{j_2}(x_2),$$

where we have used the notation $(\phi_h)_j^n := \phi_h(x_j, t_n)$ for a meshpoint (x_j, t_n) . $\square$

Now, for a family of quadrangular meshes of parameter h , $\mathcal{T}_h$, let us introduce the finite element spaces

$$\mathcal{Q}_h^k := \{f \in C^0(\overline{\Omega}), f|_K \in \mathcal{Q}^k, \forall K \in \mathcal{T}_h\},$$

with $\mathcal{Q}^k$ being the space of polynomials of degree less than or equal to k in each variable. Analogously, for a family of triangular meshes of parameter h , $\mathcal{T}_h$, let us introduce the finite element spaces

$$\mathcal{P}_h^k := \{f \in C^0(\overline{\Omega}), f|_K \in \mathcal{P}^k, \forall K \in \mathcal{T}_h\},$$

with $\mathcal{P}^k$ the space of polynomials of degree less than or equal to k .

Remark 4.3 Recall that, in order to preserve the finite element error $O(h^k)$ in a problem with second order spatial derivatives (as the one considered here) it is enough to use a quadrature that integrates exactly polynomials of the space $\mathcal{Q}^{2(k-1)}$ (respectively, $P^{2(k-1)}$). (See for instance [115]).

Remark 4.4 Although the analysis of stability performed in this section does not apply directly to problems with variable coefficients, “the stability conditions obtained for constant coefficients schemes can be used to give stability conditions for the same scheme applied to equations with variable coefficients” (see [101]). In general, if each of the problems obtained by fixing the coefficients for the points of the domain is stable, then the variable coefficient problem is also stable.

Remark 4.5 In the Fourier or von Neumann analysis we are going to perform below, “lower order terms” (in our case, reaction terms) do not modify stability properties of the resulting scheme (see [101]).

4.5.1 Study of the one dimensional problem

In the following, we develop Fourier analysis to study the stability of scheme (4.32) applied to the one-dimensional version of equation (4.31). We recall the definition of the Courant number, $\mu := v\Delta t/h$, and the Peclet number, $\rho := \nu\Delta t/h^2$. Moreover, in Table 4.1 we write some difference operators with their corresponding Fourier transforms.

Operator	Fourier transform
$\Upsilon^2[U_j] := U_{j+1} - 2U_j + U_{j-1}$	$-4s^2$
$\Delta_0[U_j] = (U_{j+1} - U_{j-1})/2$	$2isc$
$\Delta_-[U_j] = U_j - U_{j-1}$	$2(is + s^2)$
$E_r[U_i] = U_{i-r}$	$e^{-ir\theta}$

Table 4.1: Operators and corresponding Fourier transforms, with $s = \sin(\theta/2)$ and $c = \cos(\theta/2)$.

Once we have stated Lemma 4.2, we consider the one dimensional Lagrange finite elements of degree k (notice that $\mathcal{Q}_h^k = \mathcal{P}_h^k$). In particular, we study cases $k = 1$ and $k = 2$ combined with Gauss-Lobatto quadrature formulas.

- Trapezoidal rule: it integrates exactly polynomials of degree 1.

$$\int_{x_1}^{x_2} \psi(x) dx \approx \frac{x_2 - x_1}{2} (\psi(x_1) + \psi(x_2)). \quad (4.39)$$

- Simpson rule: it integrates exactly polynomials of degree 3.

$$\int_{x_1}^{x_2} \psi(x) dx \approx \frac{x_2 - x_1}{6} \left(\psi(x_1) + 4\psi\left(\frac{x_1 + x_2}{2}\right) + \psi(x_2) \right). \quad (4.40)$$

Proposition 4.1 If scheme (4.5) with $\mathcal{P}_h^1$ finite elements on a uniform mesh is applied to the one-dimensional equation (4.31) combined with two point Gauss Lobatto quadrature (4.39) in all the terms, the method is unconditionally stable.

PROOF. Firstly, let us compute the terms appearing in the j -th equation. We use the notation $(\phi_h)_j^n := \phi_h(x_j, t_n)$ for meshpoint (x_j, t_n) .

- The mass term is

$$\frac{1}{\Delta t} \int_{\Omega} \phi_h^{n+1}(x) \psi_j(x) dx \approx \frac{h}{\Delta t} E_0 [(\phi_h)_j^{n+1}].$$

- The stiffness term is exactly integrated giving rise to

$$\frac{\nu}{2} \int_{\Omega} \frac{d\phi_h^{n+1}}{dx}(x) \frac{d\psi_j}{dx}(x) dx = -\frac{\nu}{2} \frac{1}{h} \Upsilon^2 [(\phi_h)_j^{n+1}].$$

- The integral of the second member term associated to the first order characteristics method, when it is approximate by two point Gauss-Lobatto quadrature, depends on μ in the form

$$\frac{1}{\Delta t} \int_{\Omega} T_x[\phi_h^n(x)] \psi_j(x) dx \approx \frac{h}{\Delta t} (E_0 + (n - \mu)\Delta - E_{m-1}) [(\phi_h)_j^n],$$

for a positive integer n such that $n - 1 < |\mu| < n$.

- The integral of the second member term associated to the second order characteristics method is

$$-\frac{\nu}{2} \int_{\Omega} \frac{dT_x[\phi_h^n]}{dx}(x) \frac{d\psi_j}{dx}(x) dx \approx \frac{\nu}{2} \frac{1}{h} \Delta_0 \Delta - E_{n-1} [(\phi_h)_j^n],$$

for a positive integer n such that $n - 1 < |\mu| < n$.

In order to apply von Neumann analysis the amplification factor is needed. Considering the approximated integrals computed above and replacing $(\phi_h)_j^n$ with $g^n e^{i\theta j}$, the following expression for the amplification factor is obtained when $|\mu| < 1$:

$$g_{\mu,\rho}(\theta) = \frac{1 - 2\mu(s^2 + isc) + 2\rho(is^3c - s^2c^2)}{1 + 2\rho s^2},$$

with $s = \sin(\theta/2)$ and $c = \cos(\theta/2)$. Now, since

$$|g_{\mu,\rho}(\theta)|^2 = \frac{1 + 4s^2\mu(\mu - 1) + 4c^2(\rho^2s^4 - \rho s^2)}{1 + 4\rho^2s^4 + 4\rho s^4},$$

easy computations lead to $|g_{\mu,\rho}(\theta)| \leq 1$ for all $\theta \in [-\pi/2, \pi/2]$. We can proceed analogously for arbitrary μ . Thus, unconditional stability is stated. $\square$

Remark 4.6 Note that Proposition 4.1 generalizes results given in [82] and [103] for the classical first order Lagrange-Galerkin (totally implicit time discretization), for which we would have obtained

$$g_{\mu,\rho}(\theta) = \frac{1 - 2\mu(s^2 + isc)}{1 + 4\rho s^2}.$$

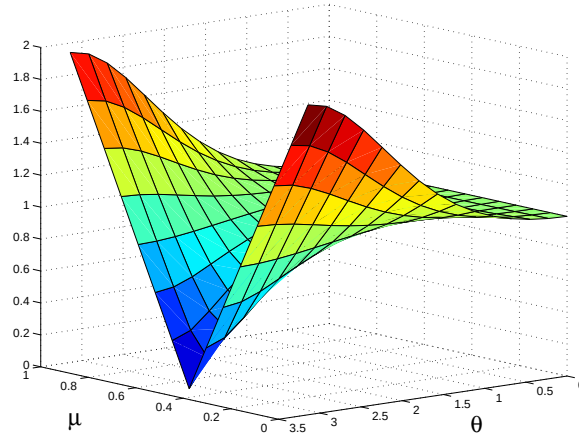


Figure 4.1: Modulus of the amplification factor for second order Lagrange-Galerkin method with $\mathcal{P}_h^1$ when the exact mass matrix is used.

Remark 4.7 *An exact integration of the mass matrix yields*

$$I_{m1}^j = \frac{h}{\Delta t} \left(\frac{1}{6} \Upsilon^2 [(\phi_h)_j^{n+1}] - \frac{2}{3} E_0 [(\phi_h)_j^{n+1}] \right),$$

which, combined with the other terms computed in Proposition 4.1, leads to regions of instability. This fact is illustrated in Figure 4.1 by plotting the modulus of the amplification factor as a function of μ and θ for fixed $\rho = 0.19$.

Next, we study quadratic elements for which we have only found the following negative result in the literature:

Lemma 4.3 *The classical Lagrange-Galerkin method applied to the one dimensional version of (4.31) with $\nu = 0$ and using piecewise quadratic elements has regions of instability if the mass matrix is exactly computed and the right hand side is evaluated by using a quadrature formula with only interior nodes.*

PROOF. See [82]. □

In the following lemma we prove that the combination of the classical Lagrange-Galerkin method with quadratic elements and Simpson quadrature preserves the unconditional stability of the scheme when applied to the linear convection equation. It is straightforward to prove that a similar result holds for the linear convection diffusion equation, since the diffusion term is evaluated implicitly.

Proposition 4.2 *If the classical Lagrange-Galerkin scheme with $\mathcal{P}_h^2$ finite elements on a uniform mesh is applied to the one-dimensional convection equation ((4.31) with $\nu = 0$) combined with Simpson quadrature (4.40) in all the terms, the method is unconditionally stable.*

PROOF. Firstly, let us introduce the basis functions on the reference element,

$$p_1(x) = 1 - 3x + 2x^2, \quad p_2(x) = 4(x - x^2), \quad p_3(x) = -x + 2x^2,$$

with the corresponding derivatives

$$dp_1(x) = -3 + 4x, \quad dp_2(x) = 4 - 8x, \quad dp_3(x) = -1 + 4x.$$

In this case we must take into account that the integrals depend on whether the basis function, ψ_j , corresponds to an interior or to a vertex node. Moreover, for the sake of simplicity, we only consider the case $|\mu| < 1$.

- The mass term provides the approximation

$$\frac{1}{\Delta t} \frac{2h}{3} E_0 \left[(\phi_h)_j^{n+1} \right],$$

when j is an interior node, and

$$\frac{1}{\Delta t} \frac{h}{3} E_0 \left[(\phi_h)_j^{n+1} \right],$$

otherwise.

- For the second member associated to the first order characteristics method we distinguish two cases. Firstly, when j is an interior node and $|\mu| < \frac{1}{2}$ we have

$$\frac{1}{\Delta t} \frac{2h}{3} \left((\phi_h)_{j-1}^n p_1 \left(\frac{1}{2} - \mu \right) + (\phi_h)_j^n p_2 \left(\frac{1}{2} - \mu \right) + (\phi_h)_{j+1}^n p_3 \left(\frac{1}{2} - \mu \right) \right);$$

and when $\frac{1}{2} < |\mu| < \frac{3}{2}$,

$$\frac{1}{\Delta t} \frac{2h}{3} \left((\phi_h)_{j-3}^n p_1 \left(\frac{3}{2} - \mu \right) + (\phi_h)_{j-2}^n p_2 \left(\frac{3}{2} - \mu \right) + (\phi_h)_{j-1}^n p_3 \left(\frac{3}{2} - \mu \right) \right).$$

Secondly, when j is a vertex node and $|\mu| < 1$,

$$\frac{1}{\Delta t} \frac{h}{3} \left((\phi_h)_{j-2}^n p_1(1 - \mu) + (\phi_h)_{j-1}^n p_2(1 - \mu) + (\phi_h)_j^n p_3(1 - \mu) \right).$$

The corresponding amplification factors satisfy condition $|g| \leq 1$ (see Figure 4.2), so unconditional stability is achieved. $\square$

Remark 4.8 *Notice that, when applied to the linear convection equation, the classical scheme and the second order Lagrange-Galerkin scheme (4.5) are exactly the same (assuming the same approximation of the characteristic lines).*

In the case of scheme (4.5) combined with quadratic elements and Simpson quadrature formula only conditional stability has been obtained when $\nu > 0$ (equivalently, when $\rho > 0$). Moreover, the region of instability grows up with the Peclet number and it is very small for low Peclet numbers (see, in Figure 4.3, the norm of the amplification factor as a function of θ and μ for different ρ). Thus, it could be used when convection-dominated features are present. In fact, as Priestley [94] pointed out, “results concerning stability are largely academic in that, for the schemes using the higher order quadratures it can be very hard to generate signs of instability. We know of no examples where, in a physical situation, the quadrature instability has caused any problems.” We were also unable to observe instabilities for the second order Lagrange-Galerkin with quadratic elements and Simpson quadrature. We will present some numerical results in Section 4.6.

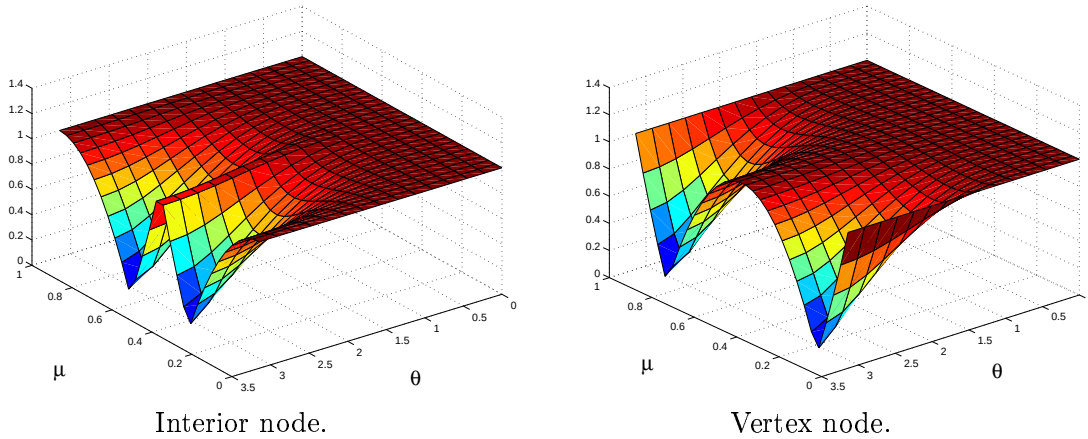


Figure 4.2: Modulus of the amplification factor for Lagrange-Galerkin with P_h^2 applied to linear convection equation when Simpson quadrature is used.

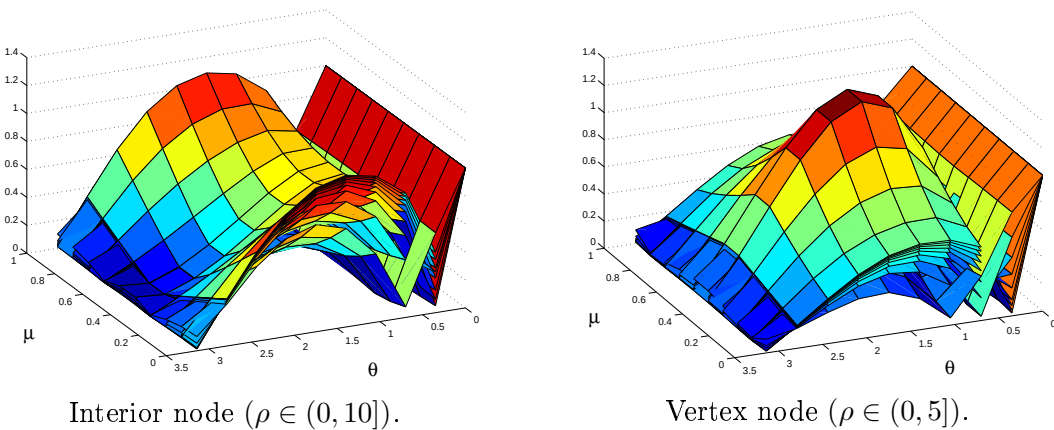


Figure 4.3: Norm of the amplification factor for second order Lagrange-Galerkin with Q_h^2 applied to the linear convection diffusion equation when Simpson quadrature is used and for different Peclet numbers.

4.5.2 Analysis in m dimensions

Since the finite element space $\mathcal{Q}_h^k$ satisfies hypothesis of Lemma 4.2, results given in Lemma 4.1 for $k = 1$ and Lemma 4.2 for $k = 2$ can be extended to m dimensions for the corresponding tensor product finite element space and quadrature formulas. In Figure 4.4 nodes of quadrature corresponding to $k = 1$ (left) and $k = 2$ (right), for the two-dimensional case, are represented. Notice that these quadratures imply mass lumping.

However, spaces $\mathcal{P}_h^k$ do not satisfy the product property, thus, the theoretical Fourier analysis developed in the present section can not be applied. Nonetheless, we propose, for the two dimensional case, vertex quadrature for $\mathcal{P}_h^1$ (see Table 4.2) which integrates exactly polynomials

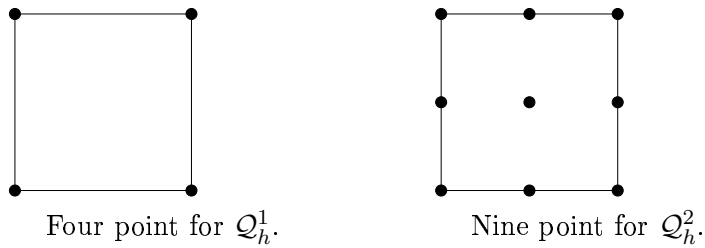


Figure 4.4: Nodes of quadrature for $d = 2$ and $\mathcal{Q}_h^k$ finite elements.

of degree one and produces mass lumping.

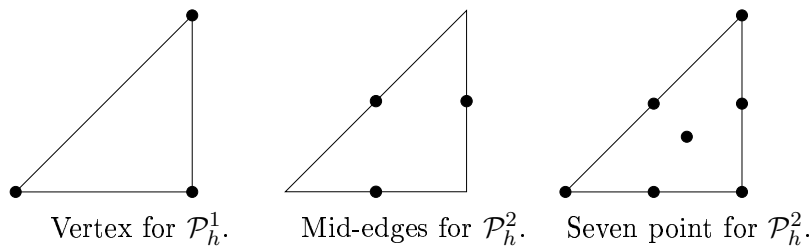


Figure 4.5: Nodes of quadrature for $m = 2$ and $\mathcal{P}_h^k$.

For finite elements $\mathcal{P}_h^2$ we have considered:

- Three-point mid-edges formula (see Table 4.3), which integrates exactly polynomials of space $\mathcal{P}_h^2$.

This is a cheap formula which leads to a diagonal mass matrix. However, the mass matrix is always singular. Depending on the boundary conditions and on the diffusion coefficients, the complete linear system may be singular or not and requires a particular analysis. For instance, the pure transport equation can not be solved with this formula.

- Seven-point formula (see Table 4.4), which integrates exactly polynomials of space $\mathcal{P}_h^3$.

One of the quadrature nodes in this formula is not a node in the finite element space, so the mass matrix is non-diagonal.

With the presented formulas we have obtained satisfactory results, as we will illustrate in the next section.

4.6 Numerical results

We show numerical results for two numerical examples in two space dimensions. We have tested the above properties of the proposed schemes. We point out that we have not found any sign of instability when using scheme (4.5) combined with either $\mathcal{Q}_h^2$ and Simpson rule (only conditionally

Node number	Barycentric coordinates	Weights
1	(1,0,0)	1/3
2	(0,1,0)	1/3
3	(0,0,1)	1/3

Table 4.2: Quadrature formula (vertex formula) used for space $\mathcal{P}_h^1$.

Node number	Barycentric coordinates	Weights
1	(0.5,0.5,0)	1/3
2	(0.5,0,0.5)	1/3
3	(0,0.5,0.5)	1/3

Table 4.3: Quadrature formula (mid-edges formula) used for space $\mathcal{P}_h^2$.

Node number	Barycentric coordinates	Weights
1	(1,0,0)	3/60
2	(0,1,0)	3/60
3	(0,0,1)	3/60
4	(0.5,0.5,0)	8/60
5	(0,0.5,0.5)	8/60
6	(0.5,0,0.5)	8/60
7	(1/3,1/3,1/3)	27/60

Table 4.4: Quadrature formula used for space $\mathcal{P}_h^2$.

stable) or $\mathcal{P}_h^k$ and the proposed quadrature formulas (for which, the developed Fourier analysis does not apply).

We notice that, instead of the theoretical $l^\infty((0,T);L^2(\Omega))$ norm, we use an approximation denoted by $l^\infty((0,T);l^2(\Omega))$. If ϕ denotes the exact solution and ϕ_h the one approximated by the numerical algorithm we want to compute

$$\|\phi - \phi_h\|_{l^\infty((0,T);L^2(\Omega))} = \max_n \|\phi^n - \phi_h^n\|_0.$$

But using triangular inequality we have

$$\|\phi^n - \phi_h^n\|_0 \leq \|\phi^n - \pi_h \phi^n\|_0 + \|\pi_h \phi^n - \phi_h^n\|_0,$$

and the first term is bounded $O(h^k)$ by Hypothesis 4.1. Thus, it is enough to estimate

$$\|\pi_h \phi^n - \phi_h^n\|_0 := \left(\int_{\Omega} (\pi_h \phi^n(\mathbf{x}) - \phi_h^n(\mathbf{x}))^2 d\mathbf{x} \right)^{\frac{1}{2}}. \quad (4.41)$$

Indeed, if this term is also $O(h^k)$, then the error $\|\phi^n - \phi_h^n\|_0$ is $O(h^k)$. Since both functions in the integral (4.41) belong to the finite element space, it can be exactly computed by using an adequate quadrature formula.

Throughout this section, all the spatial quadrangular meshes are structured and uniform and with edges parallel to the axis, so the number of degrees of freedom (d.o.f.) in each spatial direction completely characterizes the mesh. Let us denote by N_{x_1} (respectively, by N_{x_2}) the number of d.o.f. in the x_1 direction (respectively, in the x_2 direction). Moreover, we suppose all the triangular spatial meshes are built by dividing each element of the corresponding quadrangular mesh into two triangles; i.e., for the same number of d.o.f., triangular and quadrangular meshes have the same mesh parameter h .

Example 1

This is an “academic” test aiming to check the order of convergence.

We study a convection-(degenerated) diffusion-reaction problem with variable coefficients. The spatial domain is $\Omega = (0, 1) \times (0, 1)$ and $T = 1$. The only non null coefficient of the diffusion matrix is $\mathbf{A}_{22}(x_1, x_2) = x_1^2 + 0.5$. Moreover $\mathbf{v}(x_1, x_2) = (0, x_2)$, $l(x_1, x_2) = x_2$. Neumann boundary conditions are imposed on $\Gamma_{2,-} := \Gamma \cap \{x_2 = 0\}$ (i.e., Robin condition with $\alpha = 0$) and Dirichlet boundary conditions on $\Gamma \setminus \Gamma_{2,-}$. Functions f and g are computed so that the solution is the function $\phi(x_1, x_2, t) = e^{x_1+x_2+t}$.

For this problem we want not only to show the performance of scheme (4.5) with different finite element spaces, but also to make comparisons with to the more classical characteristics method, when the diffusion, reaction and source terms are totally implicit in time, but by using a second order approximation of the characteristics trajectories. More precisely, we use the method (4.27) with Runge-Kutta approximation of the trajectories instead of Euler approximation.

Table 4.5 shows the $l^\infty((0, T); l^2(\Omega))$ error obtained for different temporal/spatial meshes and the second order Lagrange-Galerkin with $\mathcal{Q}_h^1 ((\mathcal{LG})_2/\mathcal{Q}_h^1)$. Notice that

$$N_{dof} = N_{x_1} = N_{x_2},$$

and that meshes were built in such a way that

$$h_j = h_{j-1}/\sqrt{2}, \quad (\Delta t)_i = (\Delta t)_{i-1}/\sqrt{2}.$$

For the first order Lagrange-Galerkin method with $\mathcal{Q}_h^1$ finite elements $((\mathcal{LG})_1/\mathcal{Q}_h^1)$, we observe the same error for a fixed spatial mesh and different number of time steps because for all the spatial meshes the errors obtained were similar. This is because the term $O(\Delta t)$ dominates the global error.

Clearly, $(\mathcal{LG})_2/\mathcal{Q}_h^1$ achieves better errors than the corresponding classical first order method. Moreover, we have observed, for fixed h , a term $1/\Delta t$, added by the quadrature formula to the error. This behavior has also been illustrated in figure 4.6 (top). Notice that an analogous term has been already observed in the literature for the $(\mathcal{LG})_1/\mathcal{Q}_h^1$ method (see [88, 82]).

Similar comments also hold when using $\mathcal{Q}_h^2$ finite elements (see Table 4.6). However, in this case, it seems that the quadrature formula does not add any error term or, at least, a very small time step would be needed to observe it (see Figure 4.6 (bottom)).

Whereas in Table 4.6 the $O(h)$ error can be observed (in the lower rows), in Table 4.6 it is difficult to observe the second order spatial error due to the fact that the temporal error seems to dominate the global error. In Figure 4.7 we have fixed a finer temporal mesh, and we show the error versus the d.o.f. in each spatial direction (or, in other words, the error versus $1/h$). A much finer temporal mesh would be required to isolate the $O(h^2)$ term for $(\mathcal{LG})_1/\mathcal{Q}_h^2$.

	N_{dof}					
N_{ts}	25	34	49	67	97	97
10	7.943E-04	9.425E-04	9.484E-04	8.072E-04	8.050E-04	2.541E-02
14	3.788E-04	5.120E-04	3.947E-04	5.545E-04	3.669E-04	1.859E-02
20	3.200E-04	2.221E-04	1.962E-04	1.569E-04	1.493E-04	1.326E-02
28	3.021E-04	2.028E-04	1.151E-04	1.457E-04	1.061E-04	9.605E-03
40	5.080E-04	1.861E-04	1.215E-04	7.740E-05	7.357E-05	6.809E-03
57	9.486E-04	3.957E-04	1.364E-04	8.454E-05	4.481E-05	4.924E-03
80	4.837E-04	6.359E-04	2.510E-04	8.847E-05	5.297E-05	3.507E-03
113	5.081E-04	3.415E-04	4.626E-04	2.044E-04	6.313E-05	2.555E-03
160	7.153E-04	3.886E-04	2.415E-04	3.161E-04	1.250E-04	1.909E-03
226	8.984E-04	5.423E-04	2.632E-04	1.710E-04	2.305E-04	1.456E-03
320	1.038E-03	6.752E-04	3.709E-04	1.995E-04	1.207E-04	1.132E-03
453	1.138E-03	7.741E-04	4.643E-04	2.783E-04	1.339E-04	9.052E-04
640	1.211E-03	8.463E-04	5.351E-04	3.457E-04	1.888E-04	7.432E-04
	$(\mathcal{LG})_2/\mathcal{Q}_h^1$					$(\mathcal{LG})_1/\mathcal{Q}_h^1$

Table 4.5: Computed $l^\infty((0, T); l^2(\Omega))$ error norm for *Example 1*. Uniform temporal meshes with N_{ts} time steps and uniform spatial meshes with N_{dof} d.o.f. in each direction have been used.

Notice that, at the boundary where the Neumann boundary condition is imposed, the velocity field vanishes, $\text{div } \mathbf{v} = 1$, and the term

$$\mathbf{A}(\mathbf{x}) \nabla \phi(\mathbf{x}, t) \cdot \mathbf{n}|_{\Gamma_{2,-}} = -(x_1^2 + 0.5)e^{x_1+t}$$

is not null. The necessity of including the $(1 + \Delta t \text{ div } \mathbf{v})$ term at the boundary condition is illustrated in Figure 4.8, where, instead of the boundary term

$$\int_{\Gamma_{2,-}} g^n(\mathbf{x})(1 + \Delta t \text{ div } \mathbf{v}^n(\mathbf{x}))\psi(\mathbf{x})dA_{\mathbf{x}},$$

we have taken

$$\int_{\Gamma_{2,-}} g^n(\mathbf{x})\psi(\mathbf{x})dA_{\mathbf{x}}.$$

If Figure 4.8 this latter method is referred as “Bad $(\mathcal{LG})_2/\mathcal{Q}_h^2$ ”.

With respect to the computational cost of the algorithms, we remark that first order and second order Lagrange-Galerkin methods take approximately the same time for the same meshes. Moreover, for the same number of degrees of freedom, the $(\mathcal{LG})_2/\mathcal{Q}_h^2$ is quicker than $(\mathcal{LG})_2/\mathcal{Q}_h^1$ due to the different amount of mesh elements (for the same number of nodes, a mesh of linear elements have four times the number of elements of a mesh of quadratic elements).

The conclusion after the first test is that we have obtained better results with $(\mathcal{LG})_2$ than with $(\mathcal{LG})_1$ for similar computation times. Moreover, we have obtained better (accuracy) and quicker (computation time) results with $(\mathcal{LG})_2/\mathcal{Q}_h^2$ than with $(\mathcal{LG})_2/\mathcal{Q}_h^1$ for the same number of degrees of freedom.

	N_{dof}					
N_{ts}	25	34	49	67	97	97
10	7.745E-04	7.732E-04	7.739E-04	7.752E-04	7.753E-04	2.545E-02
14	3.847E-04	3.871E-04	3.871E-04	3.869E-04	3.870E-04	1.861E-02
20	1.869E-04	1.851E-04	1.862E-04	1.862E-04	1.861E-04	1.326E-02
28	9.249E-05	9.406E-05	9.330E-05	9.375E-05	9.371E-05	9.593E-03
40	4.018E-05	4.454E-05	4.565E-05	4.520E-05	4.544E-05	6.779E-03
57	2.219E-05	2.021E-05	2.271E-05	2.311E-05	2.291E-05	4.873E-03
80	1.344E-05	1.100E-05	9.877E-06	1.098E-05	1.127E-05	3.428E-03
113	8.676E-06	6.831E-06	5.384E-06	4.886E-06	5.514E-06	2.435E-03
160	5.799E-06	4.425E-06	3.333E-06	2.717E-06	2.445E-06	1.724E-03
226	3.999E-06	2.978E-06	2.164E-06	1.698E-06	1.334E-06	1.222E-03
320	2.811E-06	2.050E-06	1.451E-06	1.105E-06	8.297E-07	8.643E-04
453	2.016E-06	1.442E-06	9.988E-07	7.448E-07	5.405E-07	6.124E-04
640	1.468E-06	1.028E-06	6.978E-07	5.120E-07	3.627E-07	4.328E-04
	$(\mathcal{LG})_2/\mathcal{Q}_h^1$					$(\mathcal{LG})_1/\mathcal{Q}_h^1$

Table 4.6: Computed $l^\infty((0, T); l^2(\Omega))$ error norm for *Example 1*. Uniform temporal meshes with N_{ts} time steps and uniform spatial meshes with N_{dof} d.o.f. in each direction were used.

Finally, we have observed an analogous behavior for $\mathcal{P}_h$ finite elements. In Figure 4.9 we show the error versus the number of time steps for fixed spatial meshes.

Example 2

We solve the problem of the rotating Gaussian hill. More precisely, we choose $\mathbf{A} = \sigma_1 I$, $\mathbf{v} = (-y, x)$, $l = 0$ and $f = 0$ in the domain $\Omega = (-1., 1.) \times (-1., 1.)$, and $T = 2\pi$. We also impose appropriate Dirichlet boundary conditions and initial condition such that the solution of the problem is

$$\phi(x, y, t) = \frac{\sigma_2}{\sigma_2 + 4\sigma_1 t} \exp \left\{ -\frac{(\bar{x}(t) - x_c)^2 + (\bar{y}(t) - y_c)^2}{\sigma_2 + 4\sigma_1 t} \right\}, \quad (4.42)$$

where

$$\begin{aligned} \bar{x} &= x \cos t + y \sin t, \quad \bar{y} = -x \sin t + y \cos t, \\ (x_c, y_c) &= (0.25, 0), \quad \sigma_1 = 0.001, \quad \sigma_2 = 0.01. \end{aligned} \quad (4.43)$$

Moreover, the velocity field does not vanish on the boundary, thus we have artificially imposed $\phi = 0$ and $\nabla \phi = \mathbf{0}$ wherever the characteristic trajectories lead outside the computational domain.

In [99], results corresponding to $\mathcal{P}_h^1$ finite elements and a quadrature formula with 10 quadrature nodes per triangle are shown.

If Figure 4.10 and Figure 4.11 we represent the computed $l^\infty((0, T); l^2(\Omega))$ error obtained versus the number of time steps for three uniform spatial meshes with $N_{dof} = N_{x_1} = N_{x_2}$ degrees of freedom in each direction. We denote by $(\mathcal{LG})_1$ the classical Lagrange-Galerkin scheme (3.61) and by $(\mathcal{LG})_2$ the second order one (4.5). In view of the results, we have the following comments:

- The second order Lagrange-Galerkin method proposed $((\mathcal{LG})_2)$ reduces the time error and, allows for a less number of time steps.
- In both $(\mathcal{LG})_2 \mathcal{Q}_h^1$ and $(\mathcal{LG})_2 \mathcal{Q}_h^2$ a $O(1/\Delta t)$ term is observed for fixed h .
- Quadratic finite elements lead to a smaller $O(h)$ term than linear ones. Notice that we have used meshes with the same number of degrees of freedom, or, equivalently, the linear mesh has four times the number of elements of the quadratic mesh. For this reason, and for the same mesh, quadratic elements lead to quicker algorithms than linear ones.

Remark 4.9 *We have also observed that the classical Lagrange-Galerkin method with Runge-Kutta approximation scheme for the characteristic trajectories gives analogous results to the second order Lagrange-Galerkin (4.5). Recall that, when applied to the transport equations, the classical Lagrange-Galerkin and the second order Lagrange-Galerkin are the same if the characteristic trajectories are computed with the same scheme. In this example, the diffusion term (the only term that is discretized in a different way) seems to be smaller than the discretization error, so both lead to similar results.*

We can see the exact solution compared with the computed solutions in 4.12 and 4.13, with finite elements $\mathcal{Q}_h^1$ and $\mathcal{Q}_h^2$, respectively, and the uniform spatial mesh with $N_{dof} = 133$. In each of the cases, we have chosen the number of time steps that minimizes the error (see Figure 4.10 down). We want to point out that, whereas the $(\mathcal{LG})_2/\mathcal{Q}_h^2$ takes approximately 30 seconds for these parameters, the $(\mathcal{LG})_2/\mathcal{Q}_h^1$ takes 125 seconds.

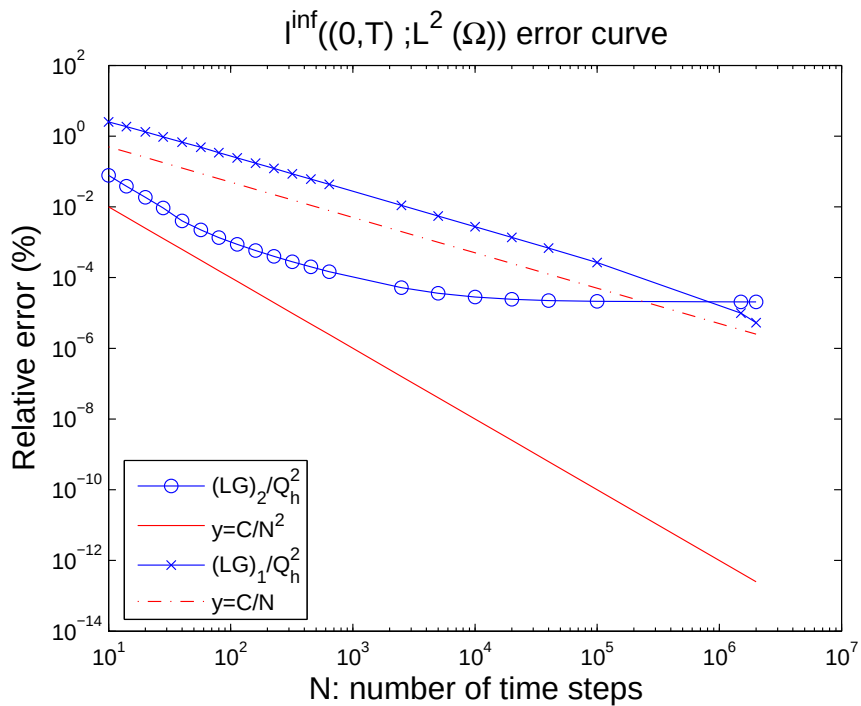
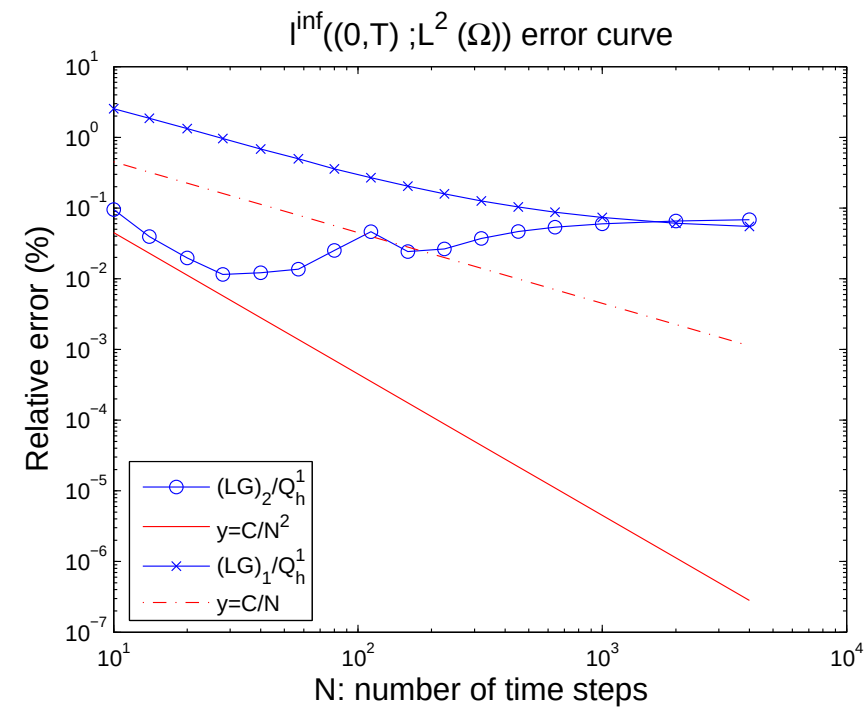


Figure 4.6: Computed $l^\infty((0,T); l^2(\Omega))$ errors, in log-log scale, for *Example 1* versus the number of time steps for a fixed spatial mesh. On the top Q_h^1 FE with 49 d.o.f. in each spatial direction. On the bottom Q_h^2 FE with 25 d.o.f. in each spatial direction.

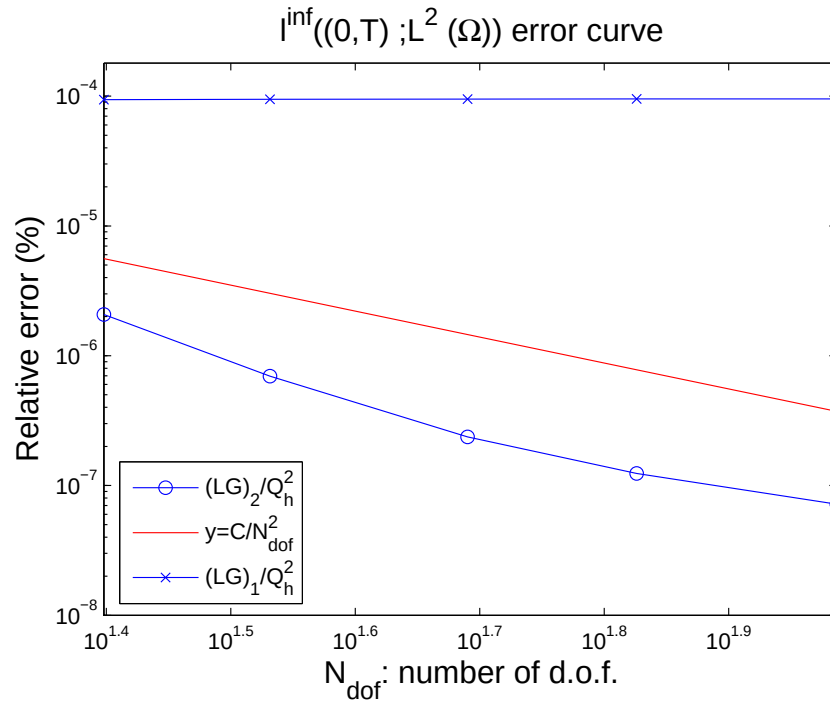


Figure 4.7: Computed $l^\infty((0,T); l^2(\Omega))$ errors, in log-log scale, for **Example 1** by using Q_h^2 finite elements, and a fixed temporal mesh with $\Delta t = 6.67E - 05$.

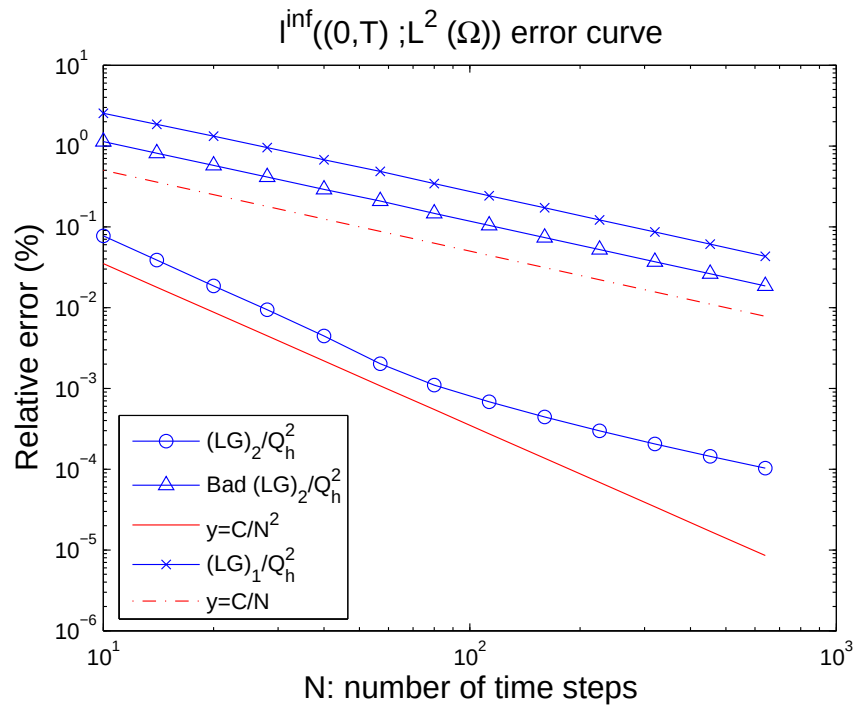


Figure 4.8: Computed $l^\infty((0,T); l^2(\Omega))$ errors, in log-log scale, for **Example 1** by using Q_h^2 finite elements and different Lagrangian methods.

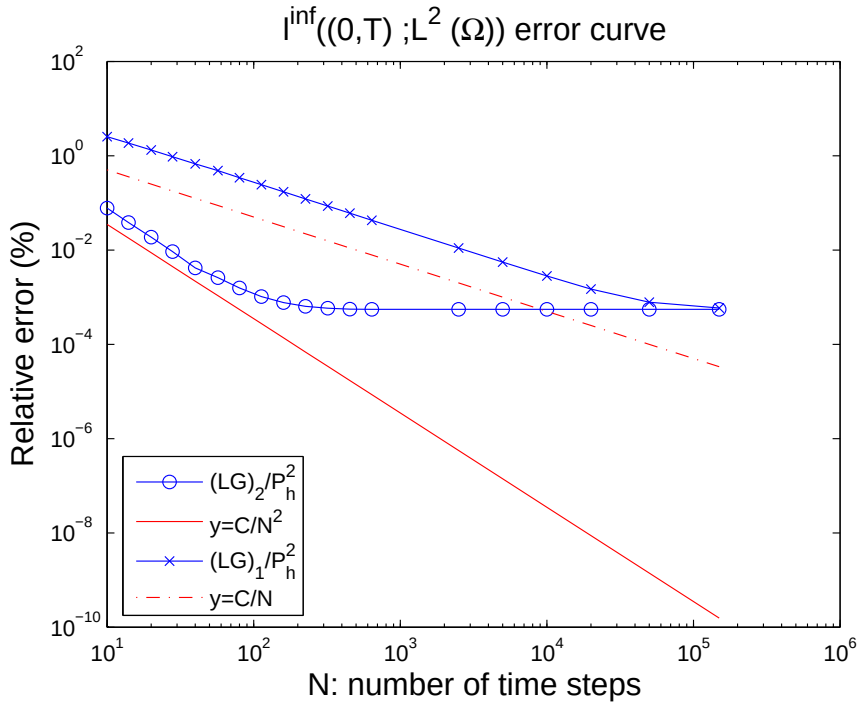
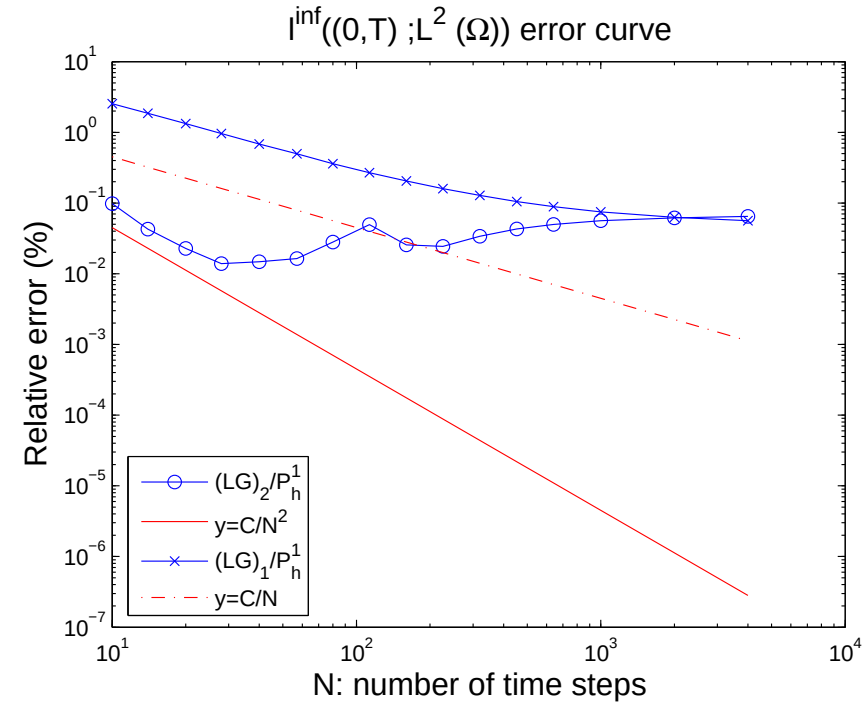


Figure 4.9: Computed $l^{\infty}((0,T);l^2(\Omega))$ errors, in log-log scale, for *Example 1* versus the number of time steps for a fixed spatial mesh. On the top $\mathcal{P}_h^1$ FE with 49 d.o.f. in each spatial direction. On the bottom $\mathcal{P}_h^2$ with 25 d.o.f. in each spatial direction.

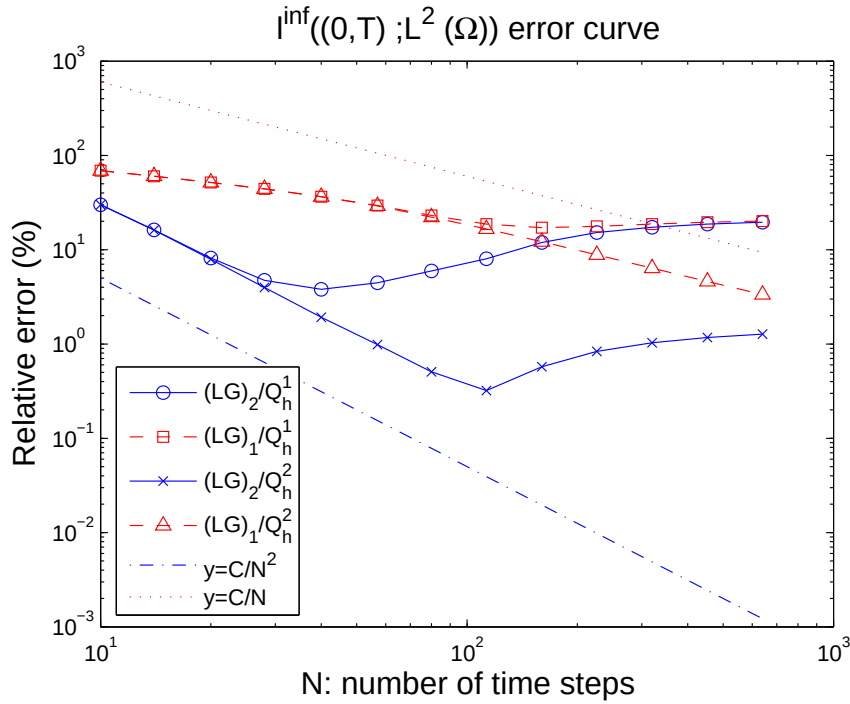
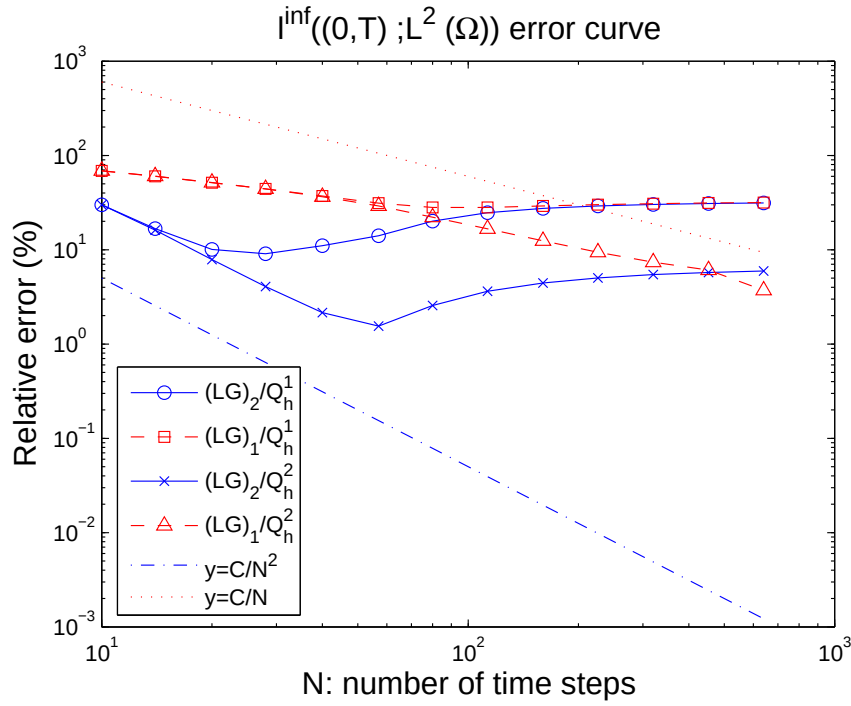


Figure 4.10: Computed $l^\infty((0,T); L^2(\Omega))$ errors, in log-log scale, for **Example 2** versus the number of time steps for two fixed spatial meshes: on the top with $N_{dof} = 67$ and on the bottom with $N_{dof} = 133$.

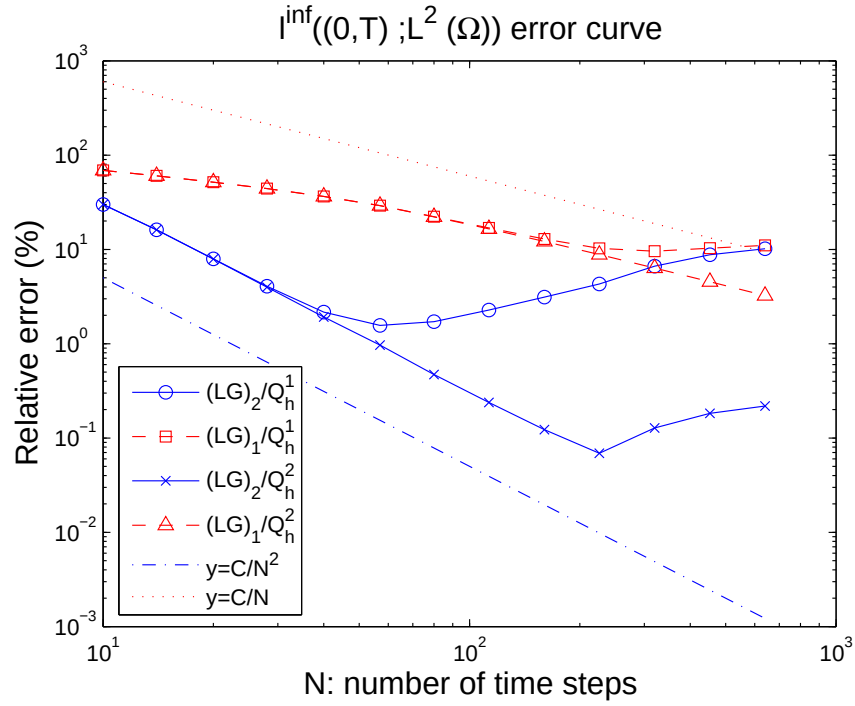
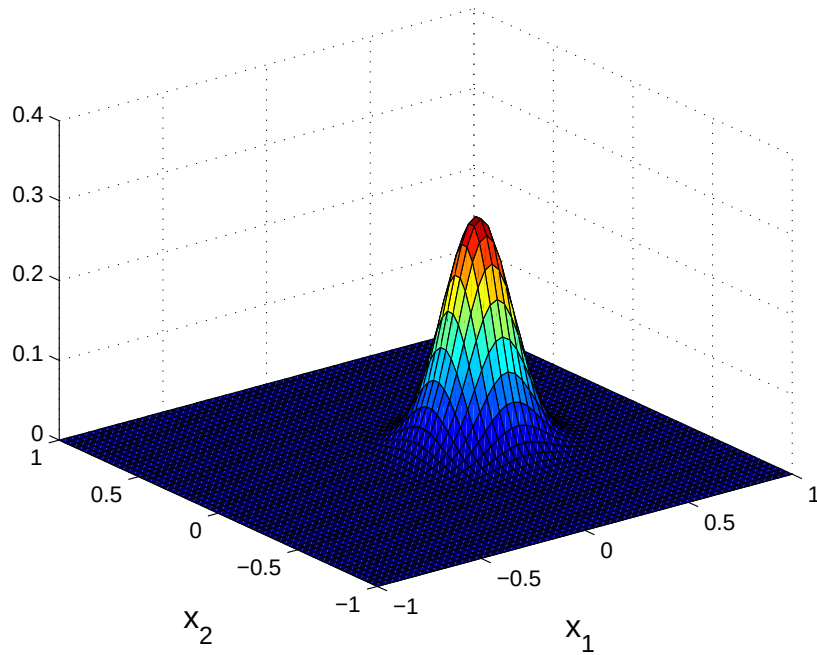


Figure 4.11: Computed $l^\infty((0,T);l^2(\Omega))$ errors, in log-log scale, for **Example 2** versus the number of time steps for a fixed spatial meshes with $N_{dof} = 265$.



Exact solution.

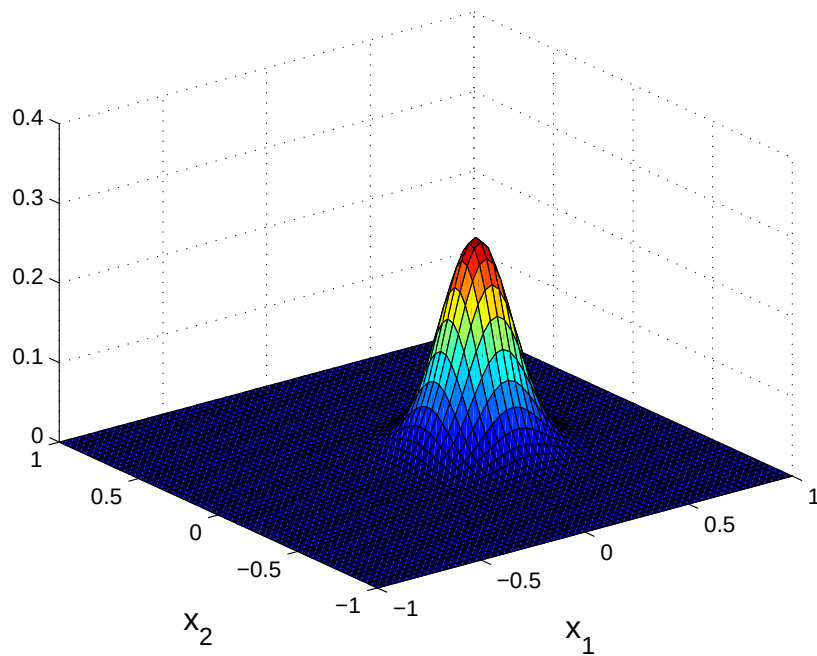
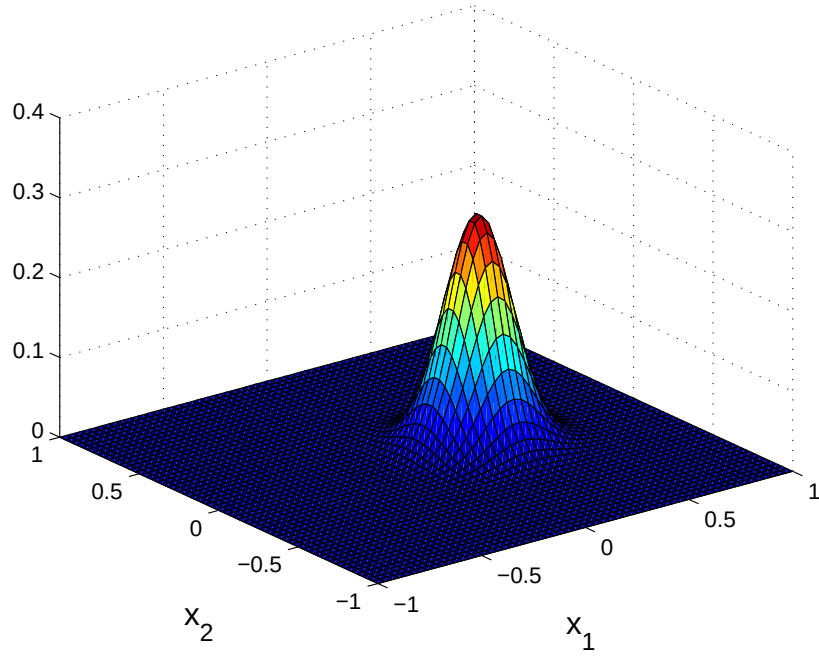
 $(\mathcal{LG})_2/Q_h^1$ computed solution.

Figure 4.12: Exact (top) and computed (bottom) solution of **Example 1** at time $t = 2\pi$ with $(\mathcal{LG})_2/Q_h^1$ and mesh parameters $N_{dof} = 133$ (in each direction) and 40 time steps.



Exact solution.

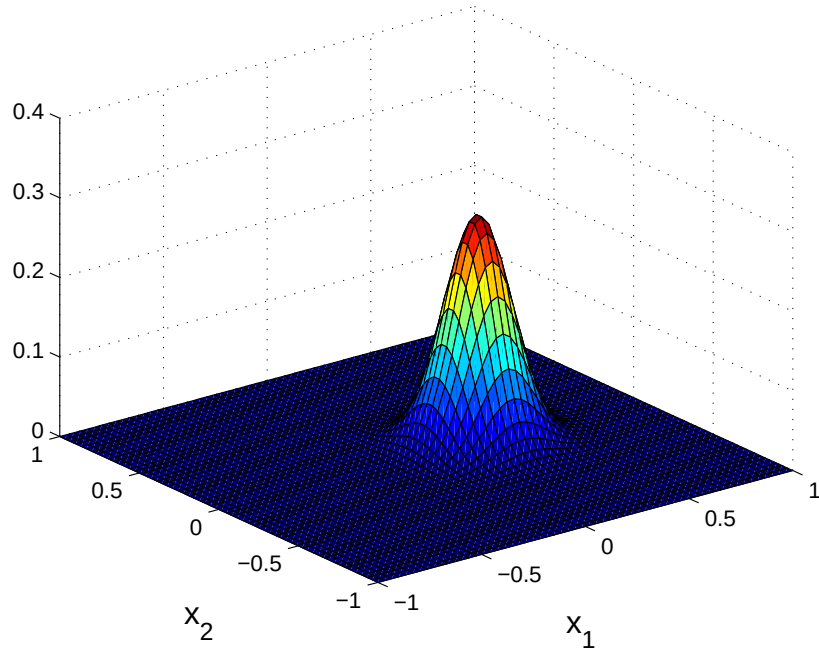
 $(\mathcal{LG})_2/\mathcal{Q}_h^2$ computed solution.

Figure 4.13: Exact (top) and computed (bottom) solution of **Example 1** at time $t = 2\pi$ $(\mathcal{LG})_2/\mathcal{Q}_h^2$ and mesh parameters $N_{dof} = 133$ (in each direction) and 113 time steps.

Chapter 5

Numerical solution of Asian options pricing problems

5.1 Introduction

In this chapter we mainly discuss the numerical solution of the Cauchy problems formulated in Sections 1.3.3 and 1.3.4 for fixed-strike Eurasian and Amerasian options, by using the second order Crank Nicholson characteristics method introduced in Chapter 3 and analyzed in Chapters 3 and 4. Moreover, we compare two algorithms based in the augmented Lagrangian formulation for the unilateral obstacle problem appearing in the Amerasian pricing problems.

As we have already pointed out, in the literature regarding the fixed-strike Eurasian pricing problem, some changes of variable were proposed (see [97] or [108]) reducing the dimension of the problem in one. Nevertheless, these techniques cannot be applied to the American case, or to more general problems as with share-dependent volatility, where it is mandatory to solve the two dimensional problem. Thus, we develop the numerical solution of unilateral-obstacle problems for degenerated diffusion-convection-reaction differential operators, in two spatial dimensions. The algorithm proposed consists of the combination of:

(1) Higher order characteristics method for time discretization.

Very often, in differential equations for pricing financial products, the diffusion is quite small relative to the convection for some regions of the domain and/or for particular values of the parameters. This is reinforced for Asian options, due to the fact that there is no diffusion in one of the spatial dimensions. In such circumstances numerical schemes present difficulties. Many different ideas and approaches have been proposed in widely different contexts to solve these difficulties, and, as we have seen in the last two chapters, a possible upwinding scheme that leads to symmetric and stable approximations of the transport PDE, reducing temporal errors and allowing for large timesteps without loss of accuracy, is the characteristics method for time discretization. An application in finance of the classical characteristics method has been already developed in [107, 10, 23, 43].

(2) Higher order finite elements for space discretization.

While most papers and books on financial derivatives employ finite differences for the numerical solution, the use of finite elements has several advantages. Firstly, unstructured meshes can be convenient to make refinements at some particular parts of the domain as, for instance, near free boundaries or where the initial condition is less smooth (see [90]). Secondly, it provides greater

flexibility in terms of changing final or boundary conditions as well as incorporating inequality constraints, if necessary.

Finite element spaces introduced in Chapter 4 are considered here, i.e., linear and quadratic finite elements over triangular and quadrangular meshes. Moreover, the algebraic structure of the fully discretized scheme for the Asian options pricing problem can be improved if we use meshes with the edges of the finite elements parallel to the axis and with a suitable nodes numbering.

(3) **Two iterative algorithms for the unilateral-obstacle problem.**

The numerical solution of free boundary problems is difficult, especially when it involves two factors. The most common method of handling the early exercise condition in numerical finance is simply to advance the discrete solution over a timestep ignoring the restriction and then to make its projection on the set of constraints (see for example [38]). In the case of a single factor (American vanilla put pricing problem, for instance) the algebraic linear complementary problem is commonly solved using a projected iteration method (PSOR) that captures the unknown exercise boundary at each time step (see Wilmott [110]). In [37] a multigrid method to accelerate convergence of the basic relaxation method is suggested and in [107] a Uzawa's algorithm to better capture the free boundary is used. Moreover, in [51] an implicit penalty method for pricing American options is proposed. Authors show that, when variable timestep is used, quadratic convergence is achieved. The drawbacks of projected relaxation methods are that their rates of convergence depend on the choice of the relaxation parameter, they deteriorate when the meshes are refined and, moreover, they do not take into account the domain decomposition given by the free boundary. In this chapter we study two algorithms based on the mixed formulation of the problem for which the regularization developed does not introduce any further source of error as penalty methods do:

- *Bermúdez-Moreno algorithm (BM)*. This method has been introduced in [22] for solving elliptic variational inequalities of the second kind (see [85] for further analysis). It consists of approximating the solution of the variational inequality by a sequence of solutions of variational equalities. While this method has been extensively applied to solve free boundary problems in computational mechanics, its application to price financial derivatives has been recently proposed [23]. In fact, although the numerical results included in that work and in the present chapter show a good behavior when applied to both degenerated and non degenerated pricing models, theoretical convergence results existing in the literature only cover the latter.
- *Augmented Lagrangian Active Set method (ALAS)*. In this method the basic iteration of the active set consists of two steps. In the first one the domain is decomposed into active and inactive parts (depending on whether the constraint “acts” or not), and then, a reduced linear system associated to the inactive part is solved. We use the algorithm presented in [69] for unilateral and bilateral obstacle problems, which is based on the augmented Lagrangian formulation of these problems. It can be also interpreted as a semi-smooth Newton method (see [61]), so it is supposed to have super-linear convergence rate. A similar Newton-type iterative algorithm has already been applied to option pricing in [80], and to solve contact problems in solids mechanics in [11]. Some known a priori properties of our particular problem are taken into account in order to improve the performance of this method.

Coming back to the Amerasian/Eurasian options pricing problem, an algorithm based on some analytical arguments and probabilistic remarks (the forward shooting algorithm) is proposed in [12]. Moreover, finite volume methods with high order nonlinear flux limiter for the

convective terms combined with a penalty method for the inequality constraints, are applied in [116] to the same problem. In [76] an implicit finite element method combined with a PSOR procedure and operator regularization is proposed. Recently, in [84], solution methods from compressible fluid dynamics, as TVD and WENO discretizations, have been applied to the problem. In [43] authors study the numerical solution of Amerasian option pricing problem under jump diffusion models, i.e., they solve numerically integro-partial differential equations. They apply a splitting technique in the sense that they solve the transport equation in the average direction, by using a Lagrangian scheme, and then they solve a one dimensional Black-Scholes equation in the asset direction using classical second order implicit finite differences.

In [23] a numerical algorithm consisting of combining the iterative algorithm introduced in [22] with first order Lagrange-Galerkin methods to solve general early exercise two factor pricing problems is proposed and applied. The slow convergence observed motivated us to improve the accuracy of the numerical scheme by using higher order Lagrange-Galerkin methods. Eventhough the Asian problem does not satisfy all the hypothesis required in Chapters 3 and 4 for proving second order convergence, this accuracy has been obtained in practice. Notice that the above methodology (1)-(2)-(3) is quite general and can be applied to general two factor products as, for instance, convertible bonds (see [10]). However, we have taken into account the specific features of Asian options. They concern, for instance, the exact computation of the characteristic lines, the fact that the “velocity” is non null at the boundary, the particular boundary conditions, the way of solving the linear system, and the computation of the “active” set for the ALAS iterative algorithm, etc.

In the following, we apply this Lagrange-Galerkin method to the degenerate linear operator of Eurasian options. This algorithm is also used in the American case, where we also have to solve nonlinear problems. In both European and American cases, qualitative and quantitative numerical results are shown and compared with others from the literature.

5.2 Fixed-strike Eurasian options

We are going to study the numerical solution of the Eurasian call option pricing problem, formulated in Section 2.4.3. Given the positive numbers r, σ, d_0, K and $T_i < T_f$, and fixing two strictly positive numbers x_1^*, x_2^* , we are going to apply a Crank-Nicholson Lagrange-Galerkin method to the Cauchy problem

$$\begin{cases} \mathcal{L}[V] = f & \text{in } \Omega^* \times (0, T), \\ V(x_1, x_2, 0) = \Lambda(x_1, x_2) & \text{in } \Omega^*, \\ \frac{\partial V}{\partial x_1}(x_1, x_2, \tau) = g(x_1, x_2, \tau) & \text{on } \Gamma_{1,+}^* \times (0, T), \end{cases} \quad (5.1)$$

where $\Omega^* := (0, x_1^*) \times (0, x_2^*)$ and $T := T_f - T_i$. Linear operator $\mathcal{L}$ takes the form

$$\mathcal{L}[\phi] = \frac{\partial \phi}{\partial \tau} - \text{Div}(\mathbf{A} \nabla \phi) + \mathbf{v} \cdot \nabla \phi + l\phi, \quad (5.2)$$

for ϕ defined in $\Omega^* \times (0, T)$, and

$$\mathbf{A} = \begin{pmatrix} \frac{1}{2}\sigma^2 x_1^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} (\sigma^2 - r + d_0)x_1 \\ \frac{x_2 - x_1}{T - \tau} \end{pmatrix}, \quad l = r. \quad (5.3)$$

Moreover,

$$\Lambda(x_1, x_2) = (x_2 - K)_+, \quad (5.4)$$

$$f(x_1, x_2, \tau) = 0, \quad (5.5)$$

and

$$g(x_1, x_2, \tau) := \begin{cases} \frac{\tau}{T} e^{-r\tau}, & \text{if } 0 < x_2 < K \frac{T}{T-\tau}, \\ \frac{e^{-d_0\tau} - e^{-r\tau}}{T(r-d_0)} & \text{if } K \frac{T}{T-\tau} < x_2 < x_2^*. \end{cases} \quad (5.6)$$

Remark 5.1 We will present the numerical discretization for general l and f functions in order to apply it to the numerical solution of Amerasian options pricing problem in Section 5.3.

Remark 5.2 In the numerical solution of pricing problems by the PDE approach it is usual to make a logarithmic change of variable, namely, $x = \log S$. However, it will not be applied here because a new artificial boundary and a non justified concentrated grid near $S = 0$ arise (see [90], [66], [100]). Moreover, this change of variable does not eliminate the strong degeneracy of the parabolic operator $\mathcal{L}$.

5.2.1 Crank-Nicholson semi-Lagrangian time discretization

Let us introduce the number of time steps, N , the time step, $\Delta t = T/N$ and the time meshpoints $t_n = n\Delta t$, for $n = 0, 1/2, 1, 3/2, \dots, N$. Now, let us write the Crank-Nicholson characteristics semidiscretization of the first equation of (5.1), namely,

$$\begin{aligned} & \frac{V^{n+1}(\mathbf{x}) - V^n(X_e^n(\mathbf{x}))}{\Delta t} - \frac{1}{2} \operatorname{div}(\mathbf{A} \operatorname{grad} V^{n+1})(\mathbf{x}) - \frac{1}{2} \operatorname{div}(\mathbf{A} \operatorname{grad} V^n)(X_e^n(\mathbf{x})) \\ & + \frac{1}{2} (lV^{n+1})(\mathbf{x}) + \frac{1}{2} (lV^n)(X_e^n(\mathbf{x})) = \frac{1}{2} f^{n+1}(\mathbf{x}) + \frac{1}{2} f^n(X_e^n(\mathbf{x})). \end{aligned} \quad (5.7)$$

Let us recall the notations

$$X_e^n(\mathbf{x}) := X_e(\mathbf{x}, t_{n+1}; t_n), \quad \mathbf{F}_e^n(\mathbf{x}) := \nabla X_e(\mathbf{x}, t_{n+1}; t_n), \quad (5.8)$$

where $X_e(\mathbf{x}, t_{n+1}; \tau)$ solves the Cauchy's problem

$$\begin{cases} \frac{dX_e}{d\tau}(\mathbf{x}, t_{n+1}; \tau) = \mathbf{v}(X_e(\mathbf{x}, t_{n+1}; \tau), \tau) & \text{for } (\mathbf{x}, \tau) \in \Omega^* \times (t_n, t_{n+1}), \\ X_e(\mathbf{x}, t_{n+1}; t_{n+1}) = \mathbf{x} & \text{for } \mathbf{x} \in \Omega^*, \end{cases} \quad (5.9)$$

where $\mathbf{v}$ has been defined in (5.3). Since $\mathbf{v} \in C^0((t_n, t_{n+1}); \mathbb{C}^\infty(\Omega^*))$, problem (5.9) has smooth unique solution. In fact, an analytical solution can be easily obtained. Indeed,

- If $\sigma^2 - r + d_0 \neq 0$

$$\begin{aligned} [X_e^n]_1(x_1, x_2) &= x_1 e^{(\sigma^2 - r + d_0) \Delta t}, \\ [X_e^n]_2(x_1, x_2) &= \frac{1 - e^{(\sigma^2 - r + d_0) \Delta t}}{(\sigma^2 - r + d_0)(T - t_n)} x_1 + \frac{T - t_{n+1}}{T - t_n} x_2, \end{aligned} \quad (5.10)$$

- If $\sigma^2 - r + d_0 = 0$

$$\begin{aligned} [X_e^n]_1(x_1, x_2) &= x_1, \\ [X_e^n]_2(x_1, x_2) &= \frac{\Delta t}{(T - t_n)} x_1 + \frac{T - t_{n+1}}{T - t_n} x_2. \end{aligned} \quad (5.11)$$

Consequently, the gradients of the characteristics are also exactly computed :

- If $\sigma^2 - r + d_0 \neq 0$

$$\mathbf{F}_e^n(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ \frac{1 - e^{(\sigma^2 - r + D_0) \Delta t}}{(\sigma^2 - r + D_0)(T - t_n)} & \frac{T - t_{n+1}}{T - t_n} \end{pmatrix}.$$

- If $\sigma^2 - r + d_0 = 0$

$$\mathbf{F}_e^n(x_1, x_2) = \begin{pmatrix} e^{(\sigma^2 - r + D_0) \Delta t} & 0 \\ \frac{\Delta t}{(T - t_n)} & \frac{T - t_{n+1}}{T - t_n} \end{pmatrix}.$$

Notice that, for $n = N - 1$, the above matrices are singular. For $n = 1, 2, \dots, N - 2$ we have

- If $\sigma^2 - r + d_0 \neq 0$

$$(\mathbf{F}_e^n)^{-1}(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ \frac{1 - e^{-(\sigma^2 - r + D_0) \Delta t}}{(\sigma^2 - r + D_0)(T - t_{n+1})} & \frac{T - t_n}{T - t_{n+1}} \end{pmatrix}.$$

- If $\sigma^2 - r + d_0 = 0$

$$(\mathbf{F}_e^n)^{-1}(x_1, x_2) = \begin{pmatrix} e^{-(\sigma^2 - r + D_0) \Delta t} & 0 \\ -\frac{\Delta t}{(T - t_{n+1})} & \frac{T - t_n}{T - t_{n+1}} \end{pmatrix}.$$

We point out that our algorithm can use exact characteristic lines and their gradients, instead of their second order approximations proposed in Chapter 3. In other words, (3.50) will be solved, rather than (3.54).

Next, we are going to study the problems related to the fact that the velocity field does not vanish at the boundary and hence the “spatial domain” may change with time, i.e.,

$$X_e^n(\Omega^*, t_{n+1}; t_n) \neq \Omega^*.$$

From the theoretical point of view, the classical first order scheme has been analyzed in [91] for the linear convection-diffusion equation with time-dependent domains. Under the condition

$$X_i^n(\Omega^*, t_{n+1}; t_n) \subset \Omega^* \tag{5.12}$$

for $i = E$, unconditional stability is proved and, if the “distance” between two consecutive domains is $O(\Delta t^2)$, the first order accuracy of time discretization is preserved. However, the second order Crank-Nicholson scheme has not been analyzed there.

From the practical implementation point of view, it is also important to have condition (5.12) for $i = e$, because in order to compute the solution at time t_{n+1} evaluating $V^n(X_e^n(\mathbf{y}))$ for a quadrature node $\mathbf{y} \in \Omega^*$, is needed. However, this inclusion is not assured in the Asian options pricing problem. More precisely, let us consider the velocity field associated to the Asian options pricing model (see Figures 5.1 and 5.2):

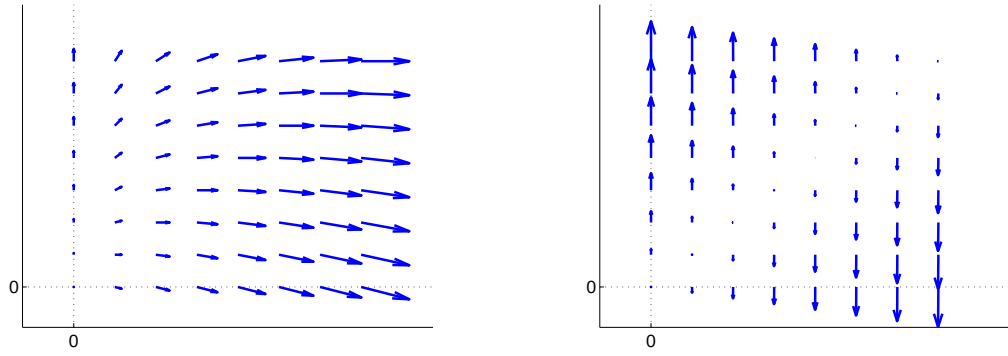


Figure 5.1: Two different velocity fields $\mathbf{v}(x_1, x_2, \tau)$. On the left, for parameters $\sigma = 0.9$, $r = 0.05$, $d_0 = 0.03$, $T = 4.5$, $\tau = 0.6$ and, on the right, for parameters $\sigma = 0.1$, $r = 0.1$, $d_0 = 0.0$, $T = 1.5$, $\tau = 0.5$

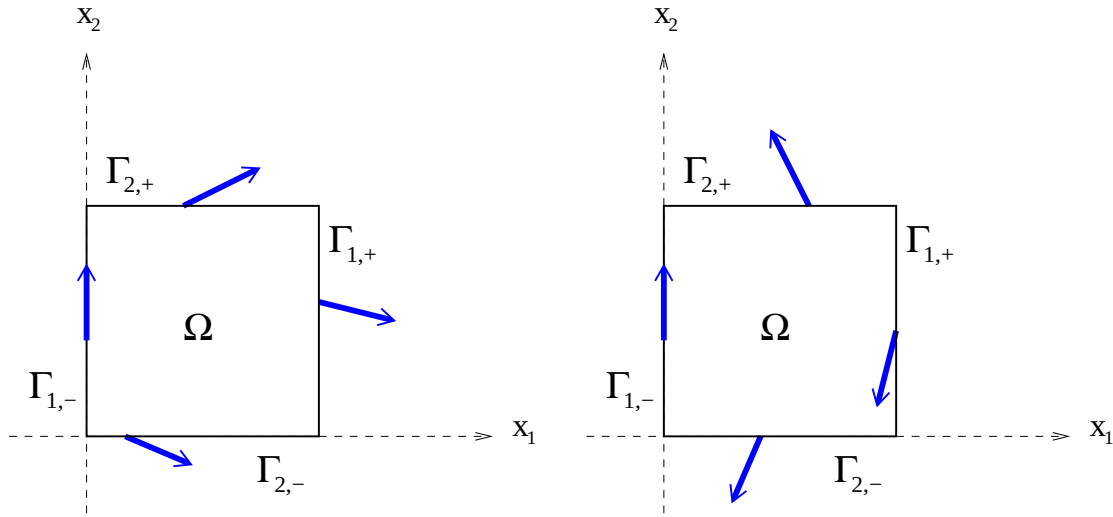


Figure 5.2: Possible qualitative behavior of the velocity field on the boundaries with $\sigma^2 - r + d_0 > 0$, on the left, and $\sigma^2 - r + d_0 < 0$, on the right.

- On boundary $\Gamma_{1,-}^*$, since $x_1 = 0$ then $\mathbf{v} = \left(0, \frac{x_2 - x_1}{T - \tau}\right)^T$ so the velocity field is tangent to the boundary.
- On boundary $\Gamma_{2,-}^*$, since $x_2 = 0$ then $\mathbf{v} = \left((\sigma^2 - r + d_0)x_1, \frac{-x_1}{T - \tau}\right)^T$ has always negative second component.
- On boundary $\Gamma_{2,+}^*$, if we add the condition $x_2^* \geq x_1^*$ then $\mathbf{v} = \left((\sigma^2 - r + d_0)x_1, \frac{x_2^* - x_1}{T - \tau}\right)^T$ has always positive second component.
- On boundary $\Gamma_{1,+}^*$, since $x_1 = x_1^*$ then $\mathbf{v} = \left((\sigma^2 - r + d_0)x_1^*, \frac{x_2 - x_1^*}{T - \tau}\right)^T$ and the sign of its

first component only depends on the coefficients of the problem. If $\sigma^2 - r + d_0 < 0$ we are in the case shown on the right hand side of Figure 5.2 and we develop an approximation on that boundary. More precisely, let us assume that $\mathbf{x} = (x_1, x_2) \in \Omega^*$ (point X in Figure 5.3) but $X_e^n(\mathbf{x}) \notin \Omega^*$ (point Y in Figure 5.3). In that case, we use a Taylor approximation of $V^n(X_e^n(\mathbf{x}))$ as follows:

- If $[X_e^n]_1(\mathbf{x}) > x_1^*$ and $[X_e^n]_2(\mathbf{x}) \leq x_2^*$, the Taylor approximation is developed around point $(x_1^*, [X_e^n]_2(\mathbf{x}))$ (point Z on the left of Figure 5.3),

$$V^n(X_e^n(\mathbf{x})) \approx V^n(x_1^*, [X_e^n]_2(\mathbf{x})) + g^n(x_1^*, [X_e^n]_2(\mathbf{x})) ([X_e^n(\mathbf{x})]_1 - x_1^*), \quad (5.13)$$

where g is the function appearing in (5.6).

- If $[X_e^n]_1(\mathbf{x}) > x_1^*$ and $[X_e^n]_2(\mathbf{x}) > x_2^*$, the Taylor approximation is developed around point (x_1^*, x_2^*) (point Z on the right of Figure 5.3),

$$\begin{aligned} V^n(X_e^n(\mathbf{x})) &\approx V^n(x_1^*, x_2^*) + g^n(x_1^*, x_2^*) ([X_e^n(\mathbf{x})]_1 - x_1^*) \\ &\quad + \frac{\partial V^n}{\partial x_2}(x_1^*, x_2^*) ([X_e^n(\mathbf{x})]_2 - x_2^*) + \frac{\partial^2 V^n}{\partial x_2^2}(x_1^*, x_2^*) ([X_e^n(\mathbf{x})]_2 - x_2^*)^2. \end{aligned} \quad (5.14)$$

Notice that, assuming that the $\partial V / \partial x_1$ is given by function g defined in (5.6), which is space-independent, we have

$$\frac{\partial^2 V}{\partial x_1^2} = 0, \quad \frac{\partial^2 V}{\partial x_1 \partial x_2} = 0,$$

so (5.13) and (5.14) are second order approximations.

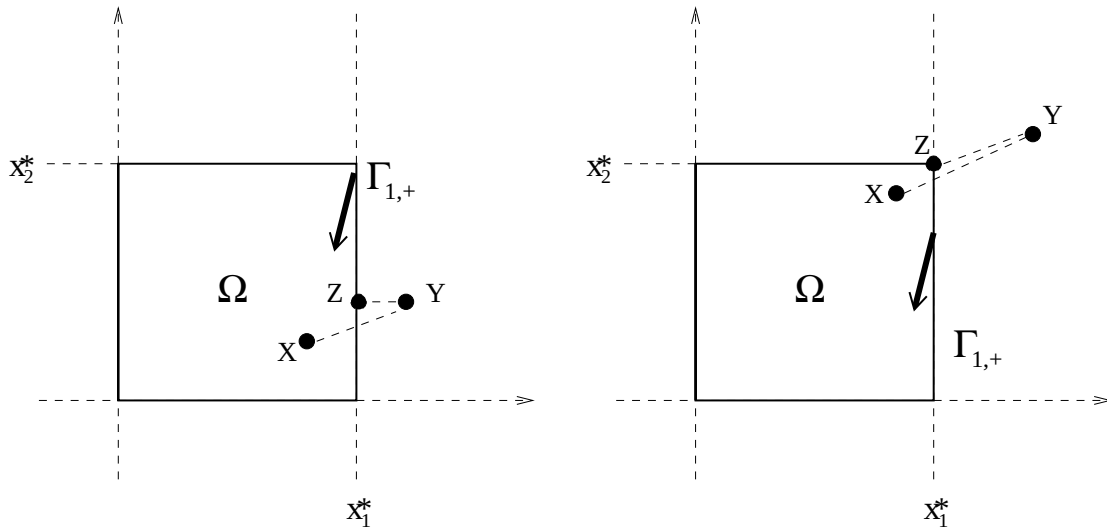


Figure 5.3: Taylor approximation on $\Gamma_{1,+}^*$.

Remark 5.3 *The discussion on whether $X_e^n(\Omega^*, t_{n+1}; t_n) \subset \Omega^*$ or not is clearly connected with the discussion of Section 2.4.3 on the well-posedness of the Dirichlet problem in bounded domains.*

Another question associated to the non vanishing velocity field on the boundaries is that Lemma 3.5 does not hold. Nevertheless we can state the following appropriate version of Lemma 3.4 applied to our particular Asian case:

Lemma 5.1 *For $n = 0, 1, \dots, N - 2$, we have $X_e^n \in C^2(\Omega^*)$ and $(\mathbf{F}_e^n)^{-1} \in C^1(\Omega^*)$. Moreover, if $\mathbf{A} \operatorname{grad} V \in H^1(\Omega^*)$, then*

$$\begin{aligned} & \int_{\Omega^*} \operatorname{div} (\mathbf{A} \operatorname{grad} V^n)(X_e^n(\mathbf{x})) \psi(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Gamma^*} (\mathbf{F}_e^n)^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot (\mathbf{A} \operatorname{grad} V^n)(X_e^n(\mathbf{x})) \psi(\mathbf{x}) dA_{\mathbf{x}} \\ & \quad - \int_{\Omega^*} (\mathbf{F}_e^n)^{-1}(\mathbf{x}) (\mathbf{A} \operatorname{grad} V^n)(X_e^n(\mathbf{x})) \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x}, \end{aligned} \quad (5.15)$$

for all $\psi \in H^1(\Omega^*)$.

PROOF. The regularity of the involved functions is straightforward in view of the analytical expressions of X_e^n and $\mathbf{F}_e^n$. Thus, we can apply Lemma 3.4 and use that $\operatorname{Div} (\mathbf{F}_e^n)^{-T}(\mathbf{x}) = 0$. $\square$

In order to clarify the expression, let us develop the terms in the boundary integral of (5.15),

$$\begin{aligned} & (\mathbf{F}_e^n)^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot (\mathbf{A} \operatorname{grad} V^n)(X_e^n(\mathbf{x})) \psi(\mathbf{x}) = \\ & \quad ([(\mathbf{F}_e^n)^{-T}]_{11}(\mathbf{x}) n_1(\mathbf{x}) + [(\mathbf{F}_e^n)^{-T}]_{12}(\mathbf{x}) n_2(\mathbf{x})) a_{11}(X_e^n(\mathbf{x})) \frac{\partial V}{\partial x_1}(X_e^n(\mathbf{x})). \end{aligned} \quad (5.16)$$

Now, regarding each of the boundaries:

- On $\Gamma_{1,-}^*$ expression (5.16) is zero because $[X_e^n]_1(0, x_2) = 0$ and $a_{11}(0, x_2) = 0$.
- On $\Gamma_{2,-}^*$ we get

$$- [(\mathbf{F}_e^n)^{-T}]_{12}(\mathbf{x}) a_{11}(X_e^n(\mathbf{x})) \frac{\partial V}{\partial x_1}(X_e^n(\mathbf{x})).$$

- On $\Gamma_{2,+}^*$ we get

$$[(\mathbf{F}_e^n)^{-T}]_{12}(\mathbf{x}) a_{11}(X_e^n(\mathbf{x})) \frac{\partial V}{\partial x_1}(X_e^n(\mathbf{x})).$$

- On $\Gamma_{1,+}^*$ we get

$$[(\mathbf{F}_e^n)^{-T}]_{11}(\mathbf{x}) a_{11}(X_e^n(\mathbf{x})) \frac{\partial V}{\partial x_1}(X_e^n(\mathbf{x})) \approx [(\mathbf{F}_e^n)^{-T}]_{11}(\mathbf{x}) a_{11}(X_e^n(\mathbf{x})) g^n(X_e^n(\mathbf{x})).$$

Thus, instead of the boundary term (5.15), in

$$\int_{\Gamma^*} (\mathbf{F}_e^n)^{-T}(\mathbf{x}) \mathbf{n}(\mathbf{x}) \cdot (\mathbf{A} \nabla V^n)(X_e^n(\mathbf{x})) \psi(\mathbf{x}) dA_{\mathbf{x}},$$

we compute

$$\int_{\Gamma^*} \widetilde{g}^n(\mathbf{x}) \psi(\mathbf{x}) dA_{\mathbf{x}},$$

where

$$\widetilde{g}^n(\mathbf{x}) := \begin{cases} 0 & \text{on } \Gamma_{1,-}^*, \\ -[(\mathbf{F}_e^n)^{-T}]_{12}(\mathbf{x})a_{11}(X_e^n(\mathbf{x}))\frac{\partial V}{\partial x_1}(X_e^n(\mathbf{x})) & \text{on } \Gamma_{2,-}^*, \\ -[(\mathbf{F}_e^n)^{-T}]_{12}(\mathbf{x})a_{11}(X_e^n(\mathbf{x}))\frac{\partial V}{\partial x_1}(X_e^n(\mathbf{x})) & \text{on } \Gamma_{2,+}^*, \\ [(\mathbf{F}_e^n)^{-T}]_{11}(\mathbf{x})a_{11}(X_e^n(\mathbf{x}))g^n(X_e^n(\mathbf{x})) & \text{on } \Gamma_{1,+}^*. \end{cases} \quad (5.17)$$

Notice that, for $\mathbf{x} \in \Gamma_{2,-}^*$ or $\mathbf{x} \in \Gamma_{2,+}^*$, we have the inclusion $X_e^n(\mathbf{x}) \in \Omega^*$.

For the implicit part of the diffusion term of scheme (5.7), the classical Green's formula can be applied for $n = 0, \dots, N-1$ to our particular model, it reads

$$\begin{aligned} & \int_{\Omega^*} \operatorname{div}(\mathbf{A} \operatorname{grad} V^{n+1})(\mathbf{x}) \psi(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Gamma_{1,+}^*} \bar{g}(\mathbf{x}) \psi(\mathbf{x}) dA_{\mathbf{x}} - \int_{\Omega^*} (\mathbf{A} \operatorname{grad} V^{n+1})(\mathbf{x}) \cdot \operatorname{grad} \psi(\mathbf{x}) d\mathbf{x}, \end{aligned} \quad (5.18)$$

where $\bar{g}(\mathbf{x}) := g(\mathbf{x})a_{11}(\mathbf{x})$.

Summarizing, in order to write a weak formulation of the time semidiscretized problem, we multiply equation (5.7) by a test function, integrate in Ω^* and use (5.18) and Lemma 5.1. Recall that matrix $\mathbf{F}_e^n$ becomes singular for $n = N-1$, so we use the second order algorithm up to step $n = N-2$ and the first order one in the last time step. Thus, we are led to solve the problem: *For $n = 0, \dots, N-2$, find $V^{n+1} \in H^1(\Omega^*)$ such that*

$$\begin{aligned} & \int_{\Omega^*} \frac{V^{n+1} - V^n \circ X_e^n}{\Delta t} W d\mathbf{x} + \frac{1}{2} \int_{\Omega^*} \mathbf{A} \operatorname{grad} V^{n+1} \cdot \operatorname{grad} W d\mathbf{x} \\ &+ \frac{1}{2} \int_{\Omega^*} (\mathbf{F}_e^n)^{-1} ((\mathbf{A} \operatorname{grad} V^n) \circ X_e^n) \cdot \operatorname{grad} W d\mathbf{x} \\ &+ \frac{1}{2} \int_{\Omega^*} lV^{n+1} W d\mathbf{x} + \frac{1}{2} \int_{\Omega^*} (lV^n) \circ X_e^n W d\mathbf{x} \\ &= \frac{1}{2} \int_{\Gamma_{1,+}^*} \bar{g}^{n+1} W dA_{\mathbf{x}} + \int_{\Gamma^*} \bar{g}^n W dA_{\mathbf{x}} + \frac{1}{2} \int_{\Omega^*} (f^{n+1} + f^n \circ X_e^n) W d\mathbf{x} \quad \forall W \in H^1(\Omega^*), \end{aligned} \quad (5.19)$$

and find $V^N \in H^1(\Omega^*)$ such that

$$\begin{aligned} & \int_{\Omega^*} \frac{V^N - V^{N-1} \circ X_e^{N-1}}{\Delta t} W d\mathbf{x} + \int_{\Omega^*} \mathbf{A} \operatorname{grad} V^N \cdot \operatorname{grad} W d\mathbf{x} + \int_{\Omega^*} lV^N W d\mathbf{x} \\ &= \int_{\Gamma_{1,+}^*} \bar{g}^N W dA_{\mathbf{x}} + \int_{\Omega^*} f^N W d\mathbf{x}, \quad \forall W \in H^1(\Omega^*). \end{aligned} \quad (5.20)$$

The corresponding first order classical scheme can be formulated as:

For $n = 0, \dots, N-1$, find $V^{n+1} \in H^1(\Omega^*)$ such that

$$\begin{aligned} & \int_{\Omega^*} \frac{V^{n+1} - V^n \circ X_e^n}{\Delta t} W d\mathbf{x} + \int_{\Omega^*} \mathbf{A} \operatorname{grad} V^{n+1} \cdot \operatorname{grad} W d\mathbf{x} + \int_{\Omega^*} lV^{n+1} W d\mathbf{x} \\ &= \int_{\Gamma_{1,+}^*} \bar{g}^{n+1} W dA_{\mathbf{x}} + \int_{\Omega^*} f^{n+1} W d\mathbf{x}, \quad \forall W \in H^1(\Omega^*). \end{aligned} \quad (5.21)$$

5.2.2 Galerkin projection, quadrature formulas and computational issues

With respect to the spatial discretization, we consider the Galerkin approximations introduced in Chapter 4, i.e., $\mathcal{Q}_h$ and $\mathcal{P}_h$ finite elements spaces.

Firstly, notice that the velocity field is time-dependent so that the searching process along the characteristic trajectories needs to be developed at each time step and could be computationally costly. If meshes satisfy the condition that their edges are parallel to the axis, this process can be highly optimized for both $\mathcal{Q}_h$ and $\mathcal{P}_h$ choices.

Moreover, if we consider a finite element space of dimension N_{dof} , we have to solve a $N_{dof} \times N_{dof}$ linear system

$$\mathbf{M}_h V_h^n = b_h^{n-1}, \quad (5.22)$$

at each time step, where matrix $\mathbf{M}_h$ does not depend on time. This fact allows us to compute its Choleski factorization only once. Moreover, in some cases, this matrix is block diagonal as we will explain below.

$\mathcal{Q}_h^k$ Finite elements

In the case of quadrangular finite elements of degrees $k = 1$ and $k = 2$, we use the quadrature formulas shown in Figure 4.4. Notice that these quadrature formulas imply mass lumping (the discrete mass matrix is diagonal). In this way, the use of an spatial mesh with the edges of the elements parallel to the axis and an adequate mesh-node numbering leads to:

- a tridiagonal symmetric linear system when $k = 1$,
- a pentadiagonal symmetric linear system when $k = 2$.

Both of them allow us a practical solution by blocks, equivalent to handle a set of 1-d problems (this is similar to the efficient technique proposed in [86] or [43], where nonsymmetric problems are solved). More precisely, if the mesh has N_{x_1} nodes in the x_1 direction and N_{x_2} nodes in the x_2 one, we solve N_{x_2} linear systems of dimension N_{x_1} . As an example, in (5.23) and (5.24) the form of the matrices, when $N_{x_2} = 2$, are shown for $k = 1$ and 2, respectively.

$$\left(\begin{array}{cccc|cccc} x & x & & & & & & \\ x & x & x & & & & & \\ & x & x & x & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & x & x & x & & \\ \hline & & & & x & x & & \\ & & & & x & x & x & \\ & & & & & x & x & x \\ & & & & & \ddots & \ddots & \ddots \\ & & & & & & x & x & x \end{array} \right) \quad (5.23)$$

which can be easily deduced from the put-call parity relation (1.39) and taking into account that the value of a zero-strike put option is identically zero. We have solved problem (5.1) for this case, by using the exact Neumann condition on $\Gamma_{1,+}^*$.

- *Academic test 2:*

We have considered problem (5.1) with the exact solution

$$V(x_1, x_2, \tau) = e^{x_1 + x_2 + \tau},$$

and appropriate functions f , g and Λ .

Firstly, we have compared the classical Lagrange-Galerkin method (5.21), denoted by $(\mathcal{LG})_1$, with the second order method (5.19)-(5.20), denoted by $(\mathcal{LG})_2$, for finite elements $\mathcal{Q}_h^1$ and $\mathcal{Q}_h^2$. In Figures 5.4 (corresponding to *Academic test 1*) and 5.5 (corresponding to *Academic test 2*) we show the computed errors versus the number of time steps for a uniform spatial mesh with 133 degrees of freedom in each spatial direction.

Figure 5.4 (top) and 5.5 (top) show the $l^\infty((0, T); l^2(\Omega))$ error versus the number of time steps in log-log scale. As predicted, this curve has slope -1 for $(\mathcal{LG})_1$ and -2 for $(\mathcal{LG})_2$. With respect to the $(\mathcal{LG})_2/\mathcal{Q}_h^1$ method, in Figure 5.4 (corresponding to *Academic test 1*) it seems that the $O(h)$ term dominates the global error soon, whereas in Figure 5.5 (corresponding to *Academic test 2*) the error curve has initially slope -2 and later, it has slope 1 due to the $1/\Delta t$ error term. Then, the conclusion is that the Lagrange-Galerkin methods applied to Asian options, with the slight modifications and approximations explained in this chapter, exhibit analogous behavior to the general methods studied in the previous chapters.

Although our analysis does not predict the behavior of pointwise errors, since for the real pricing case we are interested in pointwise values, these errors are also shown, for a fixed mesh, in Figures 5.4 (bottom) and 5.5 (bottom). For *Academic test 1* (see Figure 5.4 (bottom)) we have clearly obtained a second order pointwise error in time for $(\mathcal{LG})_2$ and a first order for $(\mathcal{LG})_1$. Moreover, similar results are obtained when using linear and quadratic finite elements. It may be due to the fact that the solution is very simple (linear in each variable). However, for *Academic test 2* (see figure 5.5 (bottom)), a similar behavior for the errors in the $l^\infty((0, T); l^2(\Omega))$ norm can be observed. These facts have been tested for different volatilities and times to maturity.

Finally, we point out that method $(\mathcal{LG})_i/\mathcal{Q}_h^1$ takes approximately the same time for $i = 1$ and $i = 2$ (similarly for $(\mathcal{LG})_i/\mathcal{Q}_h^2$), whereas methods with $k = 2$ are quicker than methods with $k = 1$ for the same mesh. In Table 5.1 we present the computing time in seconds corresponding to *Academic test 1* for different number of time steps and a fixed spatial mesh for each finite element type. Both meshes have the same number of degrees of freedom, i.e., for $\mathcal{Q}_h^2$ we have used a mesh with 4356 total number of elements and with 133 d.o.f. in each direction, and for $\mathcal{Q}_h^1$ we have used a mesh with 17424 total number of elements and with 133 d.o.f. in each direction. Notice that these computation times correspond to the error shown in Figure 5.4.

Real life Eurasian options

In Table 5.2 we show pointwise prices for Asian call options with different parameters. These results were obtained with the second-order characteristics method combined with either $\mathcal{Q}_h^1$ or $\mathcal{Q}_h^2$ finite elements, i.e. $(\mathcal{LG})_2/\mathcal{Q}_h^1$ and $(\mathcal{LG})_2/\mathcal{Q}_h^2$, and with the classical one $(\mathcal{LG})_1/\mathcal{Q}_h^1$. Moreover, we also show other results appearing in the literature. More precisely, we denote by BP the Barraquand and Pudet [12] results, where the forward shooting algorithm is used; and by

N_{ts}	10	20	40	80	160	320
$(\mathcal{LG})_2/\mathcal{Q}_h^1$	550	638	810	1150	1843	3233
$(\mathcal{LG})_2/\mathcal{Q}_h^2$	54	79	120	213	403	769

Table 5.1: Computing time in seconds corresponding to the $(\mathcal{LG})_2$ methods applied to *Academic test 1* for a fixed spatial mesh with 133 d.o.f. in each direction and N_{ts} number of time steps.

FVZ the Forsyth, Vetzal and Zvan [116] ones, where finite volumes with a high order nonlinear flux limiter for the convective terms are applied.

We have fixed $x_1^* = x_2^* = 3K$ and we have used a uniform discretization in time. Moreover, non-uniform meshes in space (see Figure 5.6), more refined near the most interesting region from the financial point of view (i.e., nearby the lines $x_1 = K$ and $x_2 = K$) are handled. From the numerical point of view, it seems also to be suitable to have a more refined mesh near the region where the initial condition is only continuous, i.e, nearby the line $x_2 = K$. We denote by N_{dof} the number of degrees of freedom in each spatial direction and by N_{ts} the number of time steps.

Remark 5.4 *The lack of smoothness of the payoff is common in financial pricing problems, where we can find even discontinuous payoffs (digital options, for example). It may cause the loss of the expected convergence rates. In [93], this subject is analyzed and some smoothing techniques combined with a non-uniform timestepping are proposed. The use of mesh adaptivity could be another way to alleviate the problem (see [90] for an application of the mesh adaptativity by the metric-Voronoi approach, to the American vanilla options pricing problem).*

Table 5.2 illustrates the slow convergence of the classical method $(\mathcal{LG})_1/\mathcal{Q}_h^1$ and the improvement when using the Crank-Nicholson Lagrange-Galerkin one, $(\mathcal{LG})_2/\mathcal{Q}_h^1$. However, the method which provides dramatically improved results is the Crank-Nicholson Lagrange-Galerkin method with quadratic finite elements, $(\mathcal{LG})_2/\mathcal{Q}_h^2$. Moreover, as we have said, this method is more efficient in terms of computing time.

To be noticed is that, with spatially uniform meshes, the convergence results were no so good.

5.3 Fixed-strike Amerasian options

In this section we extend the numerical solution to the American case, i.e., we are going to discretize the unilateral obstacle problem formulated in Section 2.5.2 for pricing Amerasian options. Notice that the obstacle function does not depend on time; in fact, it is equal to the payoff function defined in (5.4), i.e., $\underline{\Lambda} = \Lambda$. The Cauchy problem in this case is the following:

Find $V : (x_1, x_2, \tau) \in \Omega^* \times [0, T] \longrightarrow \mathbb{R}$, satisfying

$$\mathcal{L}[V] \leq 0, \quad V \geq \Lambda, \quad \mathcal{L}[V](V - \Lambda) = 0 \quad \text{in } \Omega^* \times (0, T), \quad (5.25)$$

subjected to the initial condition

$$V(S, M, 0) = \Lambda(S, M) \quad \text{for } (S, M) \in \Omega^*, \quad (5.26)$$

and to the boundary condition

$$\frac{\partial V}{\partial x_1}(x_1, x_2, \tau) = g(x_1, x_2, \tau) \quad \text{on } \Gamma_{1,+}^* \times (0, T). \quad (5.27)$$

The differential operator has been defined in (5.2), and involved functions have been introduced in (5.3), (5.4), (5.5) and (5.6).

The iterative algorithms we propose for the numerical solution of (5.25)-(5.26)-(5.27) are based on the Lagrange formulation. It refers to the fact that the inequality involving the operator $\mathcal{L}$ is replaced by an equality by means of an appropriate Lagrange variable or multiplier unknown P . Thus, problem (5.25)-(5.26)-(5.27) is equivalent to the following mixed formulation:

Find V and $P : \Omega^* \times [0, T] \longrightarrow \mathbb{R}$ satisfying the partial differential equation

$$\frac{\partial V}{\partial \tau} - \text{Div}(\mathbf{A} \nabla V) + \mathbf{v} \cdot \nabla V + rV + P = 0 \quad \text{in } \Omega^* \times (0, T), \quad (5.28)$$

the complementarity conditions

$$V \geq \Lambda, \quad P \leq 0, \quad (V - \Lambda)P = 0 \quad \text{in } \Omega^* \times (0, T), \quad (5.29)$$

and initial and boundary conditions (5.26)-(5.27). The functions $\mathbf{A}$ and $\mathbf{v}$ have been introduced in (5.3) and Λ in (5.4).

Remark 5.5 *The new unknown P plays the role of a Lagrange multiplier associated to the unilateral constraint $V \geq \Lambda$. This mixed formulation appears when dealing with duality methods for solving variational inequalities associated to obstacle problems (see [49], for example). In this setting, V and P are termed the primal and dual variables, respectively.*

5.3.1 The Bermúdez-Moreno iterative algorithm (BM)

In order to apply the duality method proposed in [22], we introduce a new Lagrange multiplier, Q , in terms of a parameter $\omega > 0$, by

$$Q := P - \omega V. \quad (5.30)$$

Then, condition (5.29) can be equivalently formulated as

$$Q(\mathbf{x}, t) \in G^\omega(V(\mathbf{x}, t)) \quad \text{a.e. in } \Omega^* \times (0, T),$$

where $G^\omega := G - \omega I$, G denotes the following multi-valued maximal monotone operator (see [33]):

$$G(Y) = \begin{cases} \emptyset & \text{if } Y < \Lambda, \\ (-\infty, 0] & \text{if } Y = \Lambda, \\ 0 & \text{if } \Lambda < Y, \end{cases} \quad (5.31)$$

and I is the identity function.

We recall that, if B is a maximal monotone operator in a Hilbert space then its *resolvent* operator is the single-valued contraction $J_\lambda = (I + \lambda B)^{-1}$ and its *Yosida regularization* is the Lipschitz-continuous mapping $G_\lambda = \lambda^{-1}(I - J_\lambda)$, where λ is any positive real number (see for instance [33]).

The following equivalences are straightforward:

$$\begin{aligned} p \in B(v) &\Leftrightarrow v + \lambda p \in (I + \lambda B)(v) \Leftrightarrow v = J_\lambda(v + \lambda p) \\ &\Leftrightarrow \lambda p = v + \lambda p - J_\lambda(v + \lambda p) \Leftrightarrow p = B_\lambda(v + \lambda p). \end{aligned} \quad (5.32)$$

In fact, one can easily show that similar equivalences hold by replacing B with $B^\omega := B - \omega I$, for any $\lambda < 1/\omega$.

In the particular case of $B = G$ given by (5.31), the Yosida regularization of G^ω is

$$G_\lambda^\omega(Y) = \begin{cases} \frac{Y - \Lambda}{\lambda} & \text{if } Y < (1 - \omega\lambda)\Lambda, \\ -\frac{\omega}{1 - \omega\lambda}Y & \text{if } Y \geq (1 - \omega\lambda)\Lambda, \end{cases} \quad (5.33)$$

and equivalence (5.32) yields

$$Q = G_\lambda^\omega(V + \lambda Q). \quad (5.34)$$

The above developments lead to consider the following algorithm, first introduced in [22] for elliptic variational inequalities in an abstract setting:

1. *Initialization*: Q_0 is arbitrarily given.

2. *Iteration m* : Q_m is known.

(a) Compute V_{m+1} by solving

$$\left\{ \begin{aligned} \frac{\partial V_{m+1}}{\partial \tau} - \text{Div}(\mathbf{A} \nabla V_{m+1}) + \mathbf{v} \cdot \nabla V_{m+1} \\ \quad + (r + \omega)V_{m+1} + Q_m &= 0 && \text{in } \Omega^* \times (0, T), \\ V_{m+1}(x_1, x_2, 0) &= \Lambda(x_1, x_2) && \text{in } \Omega^*, \\ \frac{\partial V_{m+1}}{\partial x_1}(x_1, x_2, \tau) &= g(x_1, x_2, \tau) && \text{on } \Gamma_{1,+}^* \times (0, T). \end{aligned} \right. \quad (5.35)$$

(b) Update Lagrange multiplier Q by

$$Q_{m+1} = \mu G_\lambda^\omega[V_{m+1} + \lambda Q_m] + (1 - \mu)Q_m \quad \text{in } \Omega^* \times (0, T), \quad (5.36)$$

where μ is a relaxation parameter, $\mu \in (0, 1]$.

Remark 5.6 *We emphasize that, since (5.34) is completely equivalent to (5.30), this algorithm does not introduce any further source of error as penalty methods do. In fact, parameter λ does not need to be small.*

Actually, the implementation of the algorithm is slightly different from the one described above. Indeed, we first discretize the problem in time and space by using the characteristics/finite element method explained in Section 5.2. Notice that problem (5.35) is exactly problem (5.1) for the particular choices

$$l = r + \omega \quad \text{and} \quad f = -Q_m.$$

Thus, at each time step, we use the iterative algorithm to solve the corresponding stationary variational inequality. Thus, by using the notation introduced in (5.22) instead of (5.35) we solve the following linear system:

$$\mathbf{M}_h V_{h,m+1}^n = -Q_{h,m}^n + b_h^{n-1}, \quad (5.37)$$

where, for appropriate matrix M_h and vectors b_h^{n-1} and $Q_{h,m}^n$, vector $V_{h,m+1}^n$ gives the discrete solution at time t_n and iteration m .

In this way, convergence results in [22] can be easily adapted to our degenerate case. In particular, one can show convergence as far as ω and λ are chosen such that $\omega\lambda = 1/2$ but, unfortunately, the speed of convergence depends on these parameters.

5.3.2 Augmented Lagrangian Active set method (ALAS)

The ALAS algorithm proposed in [69] applies to the fully discretized in time and space mixed formulation (5.28)-(5.29).

Firstly, we introduce some notation: let N_{dof} be the number of degrees of freedom of the finite element space and $N := \{1, 2, \dots, N_{dof}\}$ the set of indices in $\mathbb{R}^{N_{dof}}$. Moreover, for any decomposition $N = I \cup J$, let us denote by $[M_h]_{II}$ the principal minor of matrix M_h and by $[M_h]_{IJ}$ the codiagonal block indexed by I and J .

Thus, for each mesh time t_n , the ALAS algorithm consists of finding not only V_h^n and P_h^n but also a decomposition $N = J^n \cup I^n$ such that

$$\begin{aligned} M_h V_h^n + P_h^n &= b_h^{n-1}, \\ [P_h^n]_j + \beta [V_h^n - \Lambda]_j &\leq 0 \quad \forall j \in J^n, \\ [P_h^n]_i &= 0 \quad \forall i \in I^n, \end{aligned} \tag{5.38}$$

for any positive constant β . In the above, I^n and J^n are, respectively, the *inactive* and the *active* sets at time t_n .

At each time step, the iterative algorithm builds sequences $\{V_{h,m}^n\}_m$, $\{P_{h,m}^n\}_m$, $\{I_m^n\}_m$, and $\{J_m^n\}_m$, converging to V_h^n , P_h^n , I^n and J^n , respectively. More precisely, it reads:

1. Initialize $V_{h,0}^n$ and $P_{h,0}^n \leq 0$. Choose $\beta > 0$. Set $m = 0$.

2. Compute

$$\begin{aligned} Q_{h,m}^n &= \min \left\{ 0, P_{h,m}^n + \beta (V_{h,m}^n - \Lambda) \right\}, \\ J_m^n &= \left\{ j \in N, [Q_{h,m}^n]_j < 0 \right\}, \\ I_m^n &= \{ i \in N, [Q_{h,m}^n]_i = 0 \}. \end{aligned}$$

3. If $m \geq 1$ and $J_m^n = J_{m-1}^n$ then convergence and stop.

4. Let V and P be the solution of the linear system

$$\begin{aligned} M_h V + P &= b^{n-1}, \\ P &= 0 \text{ on } I_m^n \text{ and } V = \Lambda \text{ on } J_m^n. \end{aligned} \tag{5.39}$$

Set $V_{h,m+1}^n = V$, $P_{h,m+1}^n = \min\{0, P\}$, $m = m + 1$ and go to 2.

It is important to notice that, instead of solving the full linear system in (5.39), the following reduced one on the inactive set is solved:

$$\begin{aligned} [\mathbf{M}_h]_{\text{II}} [V]_{\text{I}} &= [b^{n-1}]_{\text{I}} - [\mathbf{M}_h]_{\text{IJ}} [\Lambda]_{\text{J}}, \\ [V]_{\text{J}} &= [\Lambda]_{\text{J}}, \\ P &= b^{n-1} - \mathbf{M}_h V, \end{aligned} \tag{5.40}$$

for $\text{I} = \text{I}_m^n$ and $\text{J} = \text{J}_m^n$.

Remark 5.7 *In an unilateral obstacle problem, the parameter β only influences the first iteration (as it is explained in [69], the role of this parameter results more evident for bilateral obstacle problems).*

In [69], authors proved convergence of the algorithm in finitely many steps under the condition that the matrix of the system is a Stieltjes matrix and if the algorithm is suitably initialized. They also proved that $\text{I}_m \subset \text{I}_{m+1}$. A Stieltjes matrix is a real symmetric positive definite matrix with all the off-diagonal entries negative (see for instance [106]). This condition can be only satisfied by the linear elements but never by “our” quadratic elements because we have some positive off-diagonal entries coming from the stiffness matrix (recall that we have a lumped mass matrix). However, we have obtained good results by using ALAS algorithm with quadratic finite elements and some particular additional features. More precisely, at each time step n :

- The initialization of the algorithm is developed as proposed in [69], namely,

$$V_{h,0}^n = \Lambda \quad \text{and} \quad P_{h,0}^n = b^n - \mathbf{M}_h V_{h,0}^n.$$

- We compute the set

$$\text{I}_*^n := \{i, \in \text{N}, \mathbf{x}^i = (x_1^i, x_2^i) \text{ is a mesh node such that, } x_2^i < K, x_1^i > (1 + r(t_n - T_i)x_2^i)\},$$

and impose that $\text{I}_*^n \subset \text{I}_m^n$ for every m . Notice that Propositions 1.6 and Proposition 1.8 have been used.

- We do not assume monotonicity with respect to m for the sets $\{\text{I}_m^n\}$.

Special care has to be taken for an efficient solution of the linear system when using the ALAS algorithm. In Section 5.2.2 we have explained how to take advantage of the special algebraic structure of the Asian options pricing problem, by using meshes with edges parallel to the axis and with suitable mesh numbering. This efficient implementation has been used for the Eurasian option pricing problem and for the Amerasian option pricing problem when discretized with the BM algorithm. The fact that in the ALAS algorithm only an incomplete linear system is solved requires a deeper study. More precisely, by ordering the nodes from right to left and from bottom to top, we are led to a matrix with N_{x_2} blocks of dimension N_{x_1} . In other words, each set of nodes with the same x_2 coordinate gives rise to a block in the matrix. Thus, for each block either we have all the nodes inside the inactive set (the case of Block “r” in Figure 5.7) or only the first $n(x_2)$ nodes (with $n(x_2) \leq N_{x_1}$) lie inside the inactive set (the case of Block “s” in Figure 5.7). The main point is that also for the ALAS algorithm we develop the factorization of the (complete) matrix only once outside the time loop and the iterative algorithm loop, and, at each iteration, we solve N_{x_2} systems of variable dimension (less or equal than N_{x_1}).

A comparison between the two iterative algorithms is out of our scope; moreover, it is not practical because the performance of this second algorithm is very problem-dependent. For

example, the larger the active set, the more efficient the second algorithm is. Nevertheless, we can establish some a priori comments related to the comparison of the two algorithms when applied to our particular problem. The following comments will be completed when showing the numerical results in the next section:

- Linear systems arising in the ALAS algorithm are smaller than linear systems in the BM algorithm.
- The ALAS algorithm uses some a priori known data about the situation of the inactive set.
- BM algorithm is strongly parameter (ω) dependent, whereas the (β) parameter appearing in ALAS algorithm only influences the first iteration.
- ALAS algorithm can be interpreted as a semi-smooth Newton method, and thus it exhibits a super-linear convergence rate.

5.3.3 Numerical results

Firstly, some results obtained with the BM algorithm combined with $(\mathcal{LG})_2/Q_h^2$ are shown and compared with other results appearing in the literature. Secondly, we compare the two proposed iterative algorithms.

BM algorithm performance

In Table 5.3 we present pointwise values obtained for different real life Amerasian call options by using the $(\mathcal{LG})_2/Q_h^2$ method and the BM algorithm. They are compared with the results in the papers by Barraquand and Pudet [12] (denoted by BP) and Zvan, Forsyth and Vetzal [116] (denoted by ZFV). Recall that the forward shooting algorithm is used in [12] whereas finite volumes with high order nonlinear flux limiter for the convective terms combined with a penalty method for the inequality constraints, are applied in [116].

For this, we have fixed $x_1^* = x_2^* = 3K$ and used uniform meshes in time, but non-uniform meshes in space (more refined near the most interesting region from the financial point of view, i.e. near the lines $x_1 = K$ and $x_2 = K$). By N_{dof} we denote the number of degrees of freedom in each spatial direction and by N_{ts} the number of time steps. We denote by N_{it} the mean number of iterations in the iterative algorithm.

Clearly, Table 5.3 illustrates pointwise convergence and accordance with the literature results.

Comparison between BM algorithm and ALAS algorithm

Next, for three sets of parameters we compare the performance of the BM and the ALAS algorithms. In Table 5.4 we show the results corresponding to the computational domain $\Omega^* = (0, x_1^*) \times (0, x_2^*)$, with $x_1^* = x_2^* = 3K$. Uniform meshes in time and non-uniform meshes in space have been used. If N_{ts} denotes the number of time steps and $N_{dof} = N_{x_1} = N_{x_2}$ denotes the number of d.o.f. in each spatial direction, we introduce the following notation for the meshes:

	Mesh-1	Mesh-2	Mesh-3	Mesh-4
N_{ts}	34	67	133	265
N_{dof}	67	133	265	529

Regarding Table 5.4 we can conclude that the ALAS algorithm is a little more efficient than the BM one because the computational time is smaller for the same meshes and leads to analogous results. Notice that the main difference between these algorithms is the solution of the linear systems and the updating of the Lagrange multiplier (or active/inactive sets), whereas the computation and factorization of the full matrix, the computations related to the characteristic lines, etc ... are common to both algorithms. Thus, since the algebraic part of the program (in particular, the resolution of the already factorized linear systems) has been optimized in both cases, the difference in time is not so much.

On the other hand, the mean number of iterations is clearly less in the ALAS algorithm than in the BM one. With respect to this subject, it is interesting to note that in the ALAS algorithm the parameter β seems not to influence the number of iterations (as it is claimed in [69]). However, the mean number of iterations in the BM algorithm crucially depends on the choice of the parameter ω . Furthermore, we have observed that the number of iterations as a function of the parameter ω is a convex function, and that its minimum depends not only on the test data (i.e. volatility, time to maturity, etc) but also on the mesh parameters. However, we have not found an explicit formula which allows us to deduce the optimum ω for an specific mesh or data parameters. For this reason, we have first searched the value of ω giving the minimum number of iterations for each test data and mesh. This optimal choice is the one used in Table 5.4. In Figure 5.8 we show the mean number of iterations in the BM algorithm as a function of the parameter ω for the four mesh refinements and the data set $\sigma = 0.2, T_f - T_i = 1, r = 0.1, d_0 = 0, K = 100$.

5.4 Qualitative results and the Greeks

In this section we show some graphical results obtained for an Eurasian call option and an Amerasian call option. All results correspond to:

- *Computational domain*: $\Omega^* = (0, x_1^*) \times (0, x_2^*)$, with $x_1^* = x_2^* = 3K$.
- *Meshes*: uniform in time ($N_{ts} = 67$) and non-uniform in space ($N_{dof} = N_{x_1} = N_{x_2} = 133$).
- *Method*: $(\mathcal{LG})_2/Q_h^2$ for Eurasian options and $(\mathcal{LG})_2/Q_h^2$ combined with BM iterative algorithm for Amerasian options.

We also show some computed “Greeks”, which are derivatives of option the price with respect to some variables (as S and t) and with respect to some of the parameters (as σ and r). We have already used the derivative of the option price with respect to S , called “Delta”, in Chapter 1 (see also Appendix B), to derive the pde for pricing path-dependent options. All of the Greeks are important in practice for the hedging of an option position (see, for instance, [110]). The Greeks have been computed by interpolation in the FE space in the case of derivatives with respect to one of the spatial variables, and with appropriate second order numerical differentiation techniques otherwise.

Firstly, we show graphical representations (the option value and the Greeks) of an Eurasian option characterized by parameters $\sigma = 0.4, r = 0.1, d_0 = 0, K = 100, T_0 = 0, T_i = 0, T_f = 1$. In Figures 5.9 to 5.14 we have fixed two different times $t = T_i$ and $t = T_i + (T_i + T_f)/2$ and shown the values with respect to spatial variables.

Secondly we show graphical representations of one Amerasian option (the option value and the Lagrange multiplier) characterized by parameters $\sigma = 0.1, r = 0.1, d_0 = 0, K = 100, T_0 = 0, T_i = 0., T_f = 0.25$. In Figures 5.15 to 5.16 we have fixed two different times $t = T_i$ and $t = T_i + (T_i + T_f)/2$ and shown the values with respect to spatial variables $\mathbf{x} = (x_1, x_2)$.

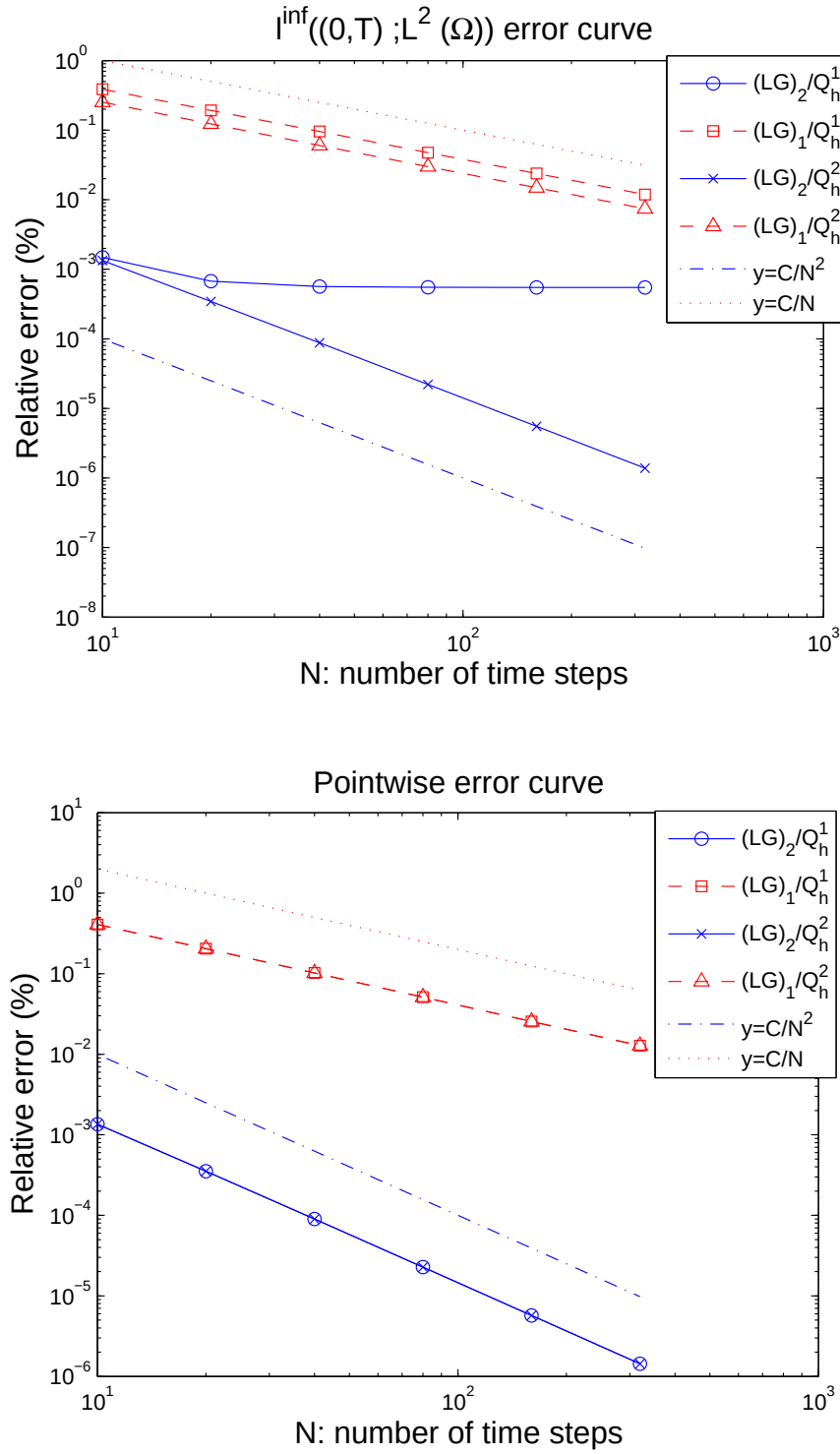


Figure 5.4: Error behavior for *Academic test 1* and parameters $\sigma = 0.4, r = 0.1, d_0 = 0, T_f - T_i = 0.5$. On the top, the relative error in $l^\infty((0, 0.5); l^2(\Omega))$ -norm *versus* the number of time steps is presented, for a uniform spatial mesh with 133 d.o.f. in each direction (i.e., with 4356 total number of elements for $\mathcal{Q}_h^2$ and with 17424 total number of elements for $\mathcal{Q}_h^1$). On the bottom, pointwise errors at $(x_1, x_2, \tau) = (1, 1, 0.5)$ for the same mesh are presented.

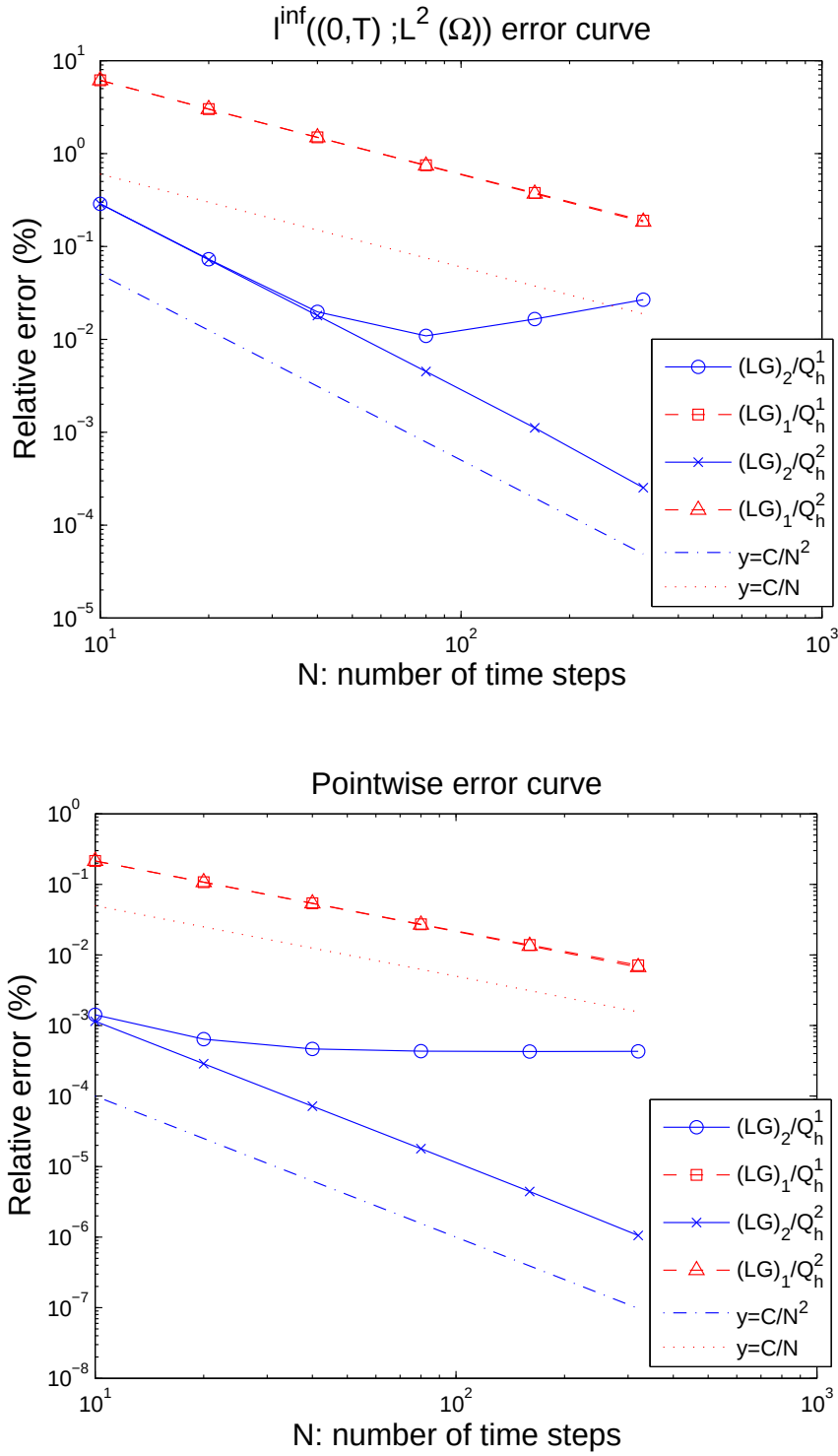


Figure 5.5: Error behavior for *Academic test 2* and parameters $\sigma = 0.1, r = 0.1, d_0 = 0, T_f - T_i = 0.45$. On the top, the relative error in $l^\infty((0, 0.25); l^2(\Omega))$ norm *versus* the number of time steps is presented, for a uniform spatial mesh with 133 d.o.f. in each direction (i.e., with 4356 total number of elements for Q_h^2 and with 17424 total number of elements for Q_h^1). On the bottom, pointwise errors at $(x_1, x_2, \tau) = (1, 1, 0.25)$ for the same mesh are presented.

		$\sigma = 0.4, T = 1$			$\sigma = 0.4, T = 0.25$		
N_{dof}	N_{ts}	$(\mathcal{LG})_1/Q_h^1$	$(\mathcal{LG})_2/Q_h^1$	$(\mathcal{LG})_2/Q_h^2$	$(\mathcal{LG})_1/Q^1$	$(\mathcal{LG})_2/Q_h^1$	$(\mathcal{LG})_2/Q_h^2$
49	25	11.56076	11.17693	11.11681	5.377461	5.238728	5.145673
67	34	11.44995	11.16534	11.12776	5.326913	5.223867	5.173030
97	49	11.35114	11.15191	11.12590	5.277259	5.205231	5.169899
133	67	1.129102	11.14443	11.12453	5.248473	5.195584	5.167598
193	97	11.23934	11.13754	11.12392	5.222775	5.186118	5.167117
265	133	11.20774	11.13325	11.12322	5.222775	5.186118	5.167117
385	193			11.12322			5.166238
529	265			11.12333			5.166680
BP		11.213			5.218		
ZFV		11.115			5.175		

		$\sigma = 0.2, T = 1$			$\sigma = 0.2, T = 0.25$		
N_{dof}	N_{ts}	$(\mathcal{LG})_1/Q_h^1$	$(\mathcal{LG})_2/Q_h^1$	$(\mathcal{LG})_2/Q_h^2$	$(\mathcal{LG})_1/Q_h^1$	$(\mathcal{LG})_2/Q_h^1$	$(\mathcal{LG})_2/Q_h^2$
49	25	7.263293	7.129032	7.022569	3.094146	3.027484	2.886315
67	34	7.208916	7.109001	7.047066	3.059829	3.011628	2.944738
97	49	7.156962	7.087036	7.044338	3.018175	2.984850	2.938239
133	67	7.126622	7.075271	7.042183	2.996152	2.971821	2.933466
193	97	7.099834	7.064247	7.041640	2.975125	2.958318	2.932641
265	133	7.083288	7.057278	7.040539	2.961778	2.949515	2.930225
385	193			7.040820			2.931068
529	265			7.041198			2.932069
BP		7.079			2.960		
ZFV		7.042			2.928		

		$\sigma = 0.1, T = 1$			$\sigma = 0.1, T = 0.25$		
N_{dof}	N_{ts}	$(\mathcal{LG})_1/Q_h^1$	$(\mathcal{LG})_2/Q_h^1$	$(\mathcal{LG})_2/Q_h^2$	$(\mathcal{LG})_1/Q^1$	$(\mathcal{LG})_2/Q_h^1$	$(\mathcal{LG})_2/Q_h^2$
49	25	5.453204	5.387069	5.219144	2.026422	1.973995	1.761066
67	34	5.406129	5.356978	5.261479	1.998952	1.963917	1.874352
97	49	5.357967	5.323282	5.258044	1.950819	1.926939	1.862189
133	67	5.330935	5.305318	5.255204	1.926878	1.909593	1.853855
193	97	5.306530	5.288660	5.254821	1.889831	1.889831	1.852559
265	133	5.291355	5.278237	5.253318	1.885329	1.876545	1.848239
385	193			5.253921			1.849970
529	265			5.254609			1.851962
BP		5.279			1.869		
ZFV		5.261			1.793		

Table 5.2: Computed real life Eurasian options prices for $(S, M, t) = (100, 100, T_i)$, with data $r = 0.1, d_0 = 0, K = 100, T = T_f - T_i$ and uniform temporal meshes with N_{ts} time steps and non-uniform spatial meshes with N_{dof} degrees of freedom in each direction.

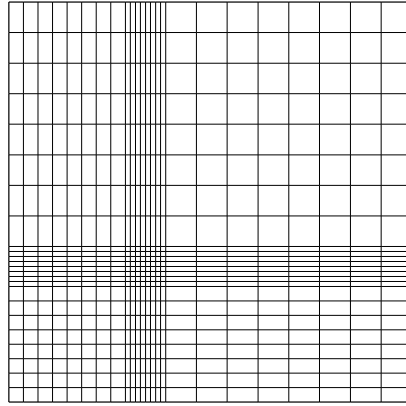


Figure 5.6: Non-uniform spatial mesh with 24 mesh elements in each spatial direction, corresponding to $N_{dof} = 49$, in each spatial direction, for $\mathcal{Q}_h^2$ finite elements.

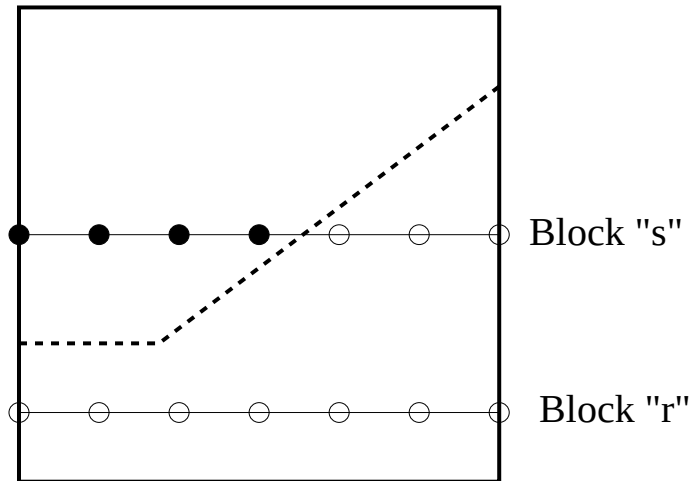


Figure 5.7: Spatial domain of solution for the Amerasian call options pricing problem, separating the active and inactive set. Two sets of FE nodes with the same x_2 coordinate are represented, and the nodes inside the active set are filled.

N_{dof}	N_{ts}	$\sigma = 0.4, T = 1.$	$\sigma = 0.4, T = 0.5$	$\sigma = 0.4, T = 0.25$
49	25	12.56541	8.547581	5.803121
67	34	12.54951	8.552752	5.852107
97	49	12.53221	8.538125	5.840765
133	67	12.52252	8.531369	5.836706
193	97	12.51601	8.526534	5.833239
265	133	12.51295	8.524181	5.831462
385	193	12.51119	8.523064	5.831010
529	265	12.51048	8.522680	5.831012
N_{it}		29	26	22
BP		12.439	8.519	5.826
ZFV		12.497	8.527	5.846

N_{dof}	N_{ts}	$\sigma = 0.2, T = 1.$	$\sigma = 0.2, T = 0.5$	$\sigma = 0.2, T = 0.25$
49	25	7.502504	4.823648	3.110095
67	34	7.547609	4.889710	3.223165
97	49	7.546300	4.885601	3.211567
133	67	7.544279	4.884481	3.213565
193	97	7.542679	4.883438	3.213082
265	133	7.541638	4.882581	3.212758
385	193	7.541593	4.883146	3.214360
529	265	7.541744	4.883727	3.215686
N_{it}		22	17	15
BP		7.521	4.893	3.219
ZFV		7.528	4.901	3.224

N_{dof}	N_{ts}	$\sigma = 0.1, T = 1.$	$\sigma = 0.1, T = 0.5$	$\sigma = 0.1, T = 0.25$
49	25	5.319105	3.085532	1.812080
67	34	5.386284	3.194547	1.975918
97	49	5.385620	3.190993	1.941031
133	67	5.385239	3.192280	1.952915
193	97	5.385290	3.192530	1.951097
265	133	5.384598	3.191569	1.951080
385	193	5.385470	3.193714	1.955484
529	265	5.386255	3.195429	1.958986
N_{it}		16	13	11
BP		5.394	3.212	1.967
ZFV		5.398	3.180	1.903

Table 5.3: Computed Amerasian options prices by $(\mathcal{LG})_2/Q_h^2$ and Bermudez-Moreno algorithm for $(S, M, t) = (100, 100, T_i)$, with data $r = 0.1, d_0 = 0, K = 100, T = T_f - T_i$ and non-uniform spatial meshes.

$\sigma = 0.4, T_f - T_i = 0.5$								
	Mesh-1		Mesh-2		Mesh-3		Mesh-4	
	BM	ALAS	BM	ALAS	BM	ALAS	BM	ALAS
Parameter	1.5E2	1.E5	4.E2	1.E5	1.15E3	1.E5	1.9E3	1.E5
N_{it}	23	4	26	5	31	6	33	9
Op. Value	8.5528	8.5503	8.5310	8.5299	8.5220	8.5232	8.5222	8.5223
Time	4	3	38	24	411	271	4689	3560

$\sigma = 0.1, T_f - T_i = 0.25$								
	Mesh-1		Mesh-2		Mesh-3		Mesh-4	
	BM	ALAS	BM	ALAS	BM	ALAS	BM	ALAS
Parameter	1.5E2	1.E5	3.25E2	1.E5	7.E2	1.E5	1.15E3	1.E5
N_{it}	10	3	11	3	12	3	13	3
Op. Value	1.9759	1.9560	1.9529	1.9308	1.9511	1.9348	1.9590	1.9590
Time	4	3	38	30	413	349	5259	4685

$\sigma = 0.2, T_f - T_i = 1.$								
	Mesh-1		Mesh-2		Mesh-3		Mesh-4	
	BM	ALAS	BM	ALAS	BM	ALAS	BM	ALAS
Parameter	7.5E1	1.E5	1.5E2	1.E5	4.5E2	1.E5	1.05E3	1.E5
N_{it}	19	3	20	4	25	5	24	7
Op. Value	7.5475E	7.5445E	7.5441	7.5425E	7.5405	7.5406	7.5386	7.5417
Time	5	3	43	31	474	355	5589	4737

Table 5.4: Results comparing, for successive mesh refinements, the BM and the ALAS algorithms applied to the Amerasian option pricing problem with data $r = 0.1, d_0 = 0, K = 100$. The parameters of these algorithms (Parameter) are λ for BM and β for ALAS. N_{it} denotes the mean number of iterations and option value (Op. Value) corresponds to $(S, M, t) = (100, 100, T_i)$. The computing time (Time) is measured in seconds.

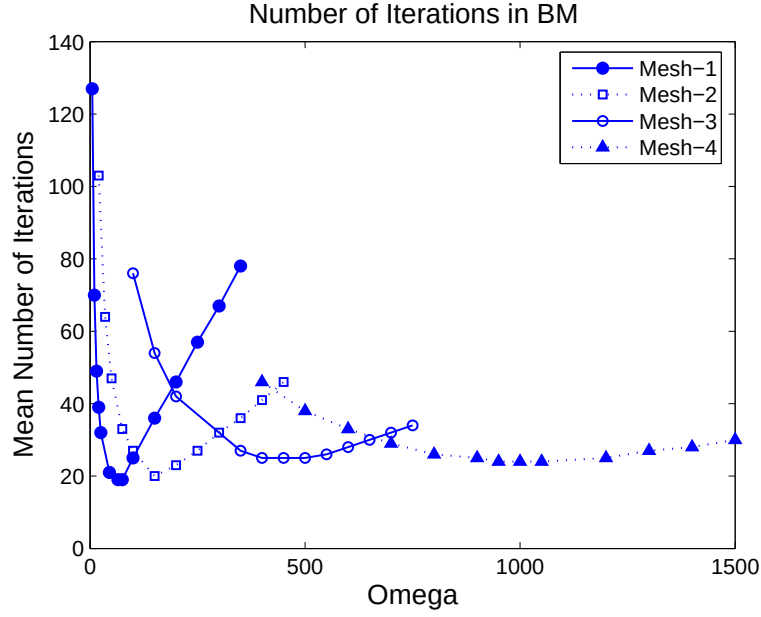


Figure 5.8: Mean number of iterations taken by the BM algorithm as a function of parameter ω for different meshes and for data $\sigma = 0.2, T_f - T_i = 1., r = 0.1, d_0 = 0, K = 100$.

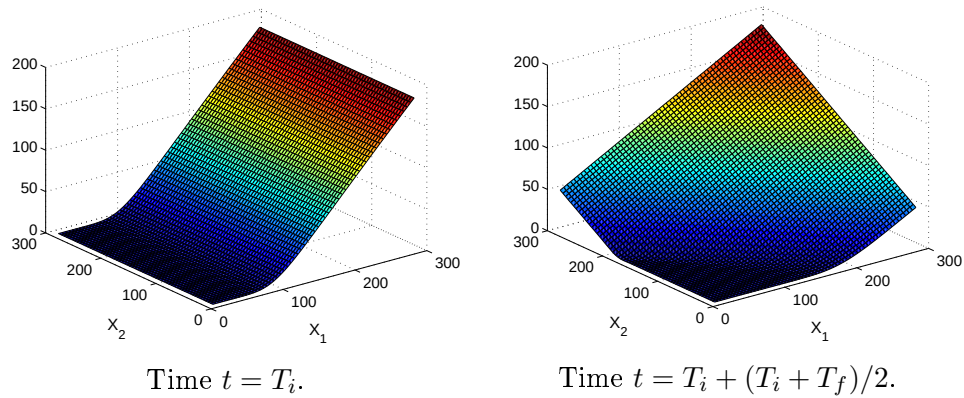
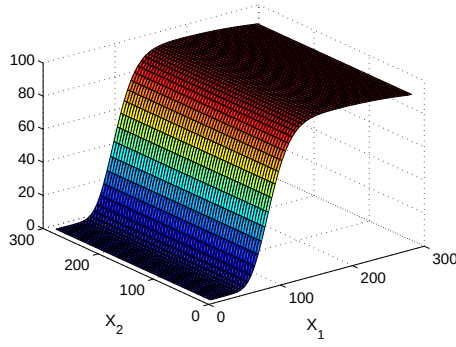
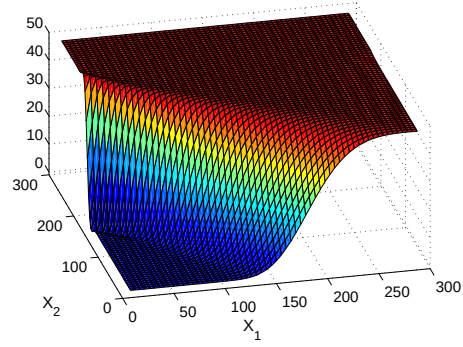


Figure 5.9: Eurasian call option value.

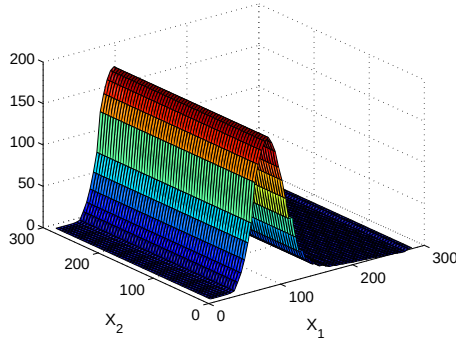


Time $t = T_i$.

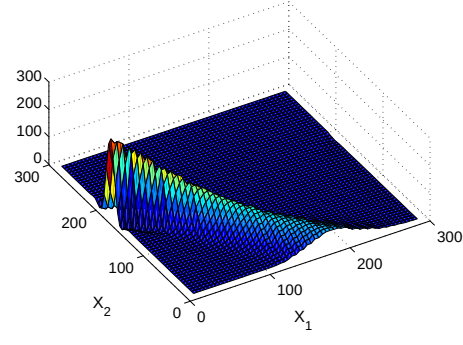


Time $t = T_i + (T_i + T_f)/2$.

Figure 5.10: Eurasian call option “Delta” ($\Delta = \frac{\partial V}{\partial S}$).

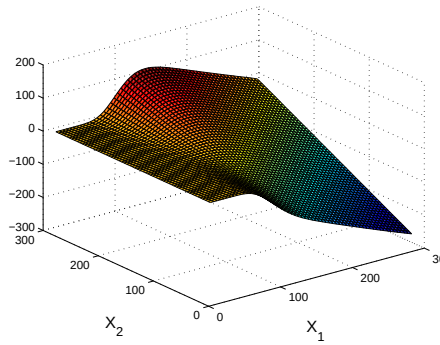


Time $t = T_i$.

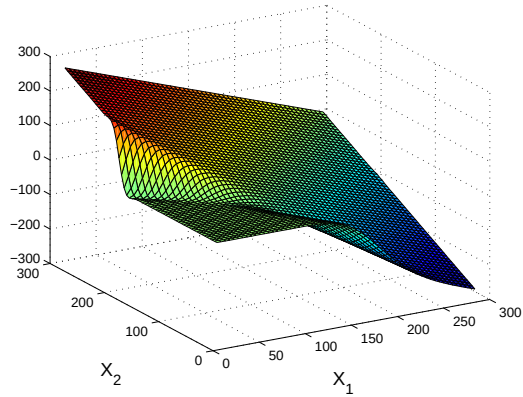


Time $t = T_i + (T_i + T_f)/2$.

Figure 5.11: Eurasian call option “Gamma” ($\Delta = \frac{\partial^2 V}{\partial S^2}$).



Time $t = T_i$.



Time $t = T_i + (T_i + T_f)/2$.

Figure 5.12: Eurasian call option “Theta” ($\Delta = \frac{\partial V}{\partial t}$).

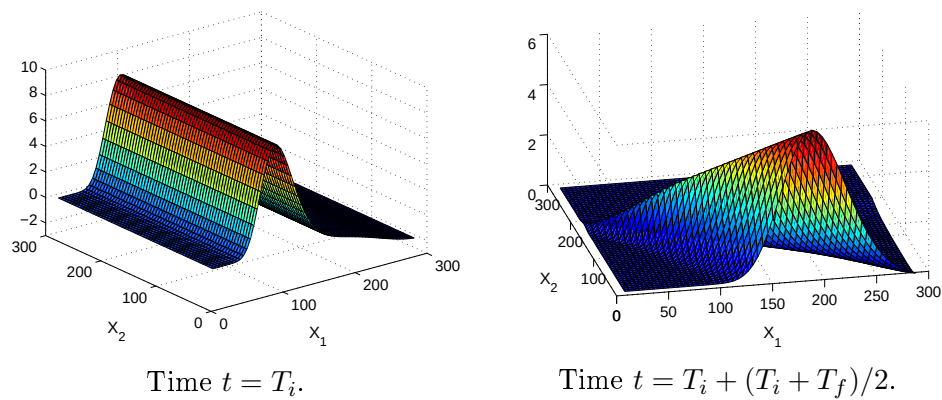
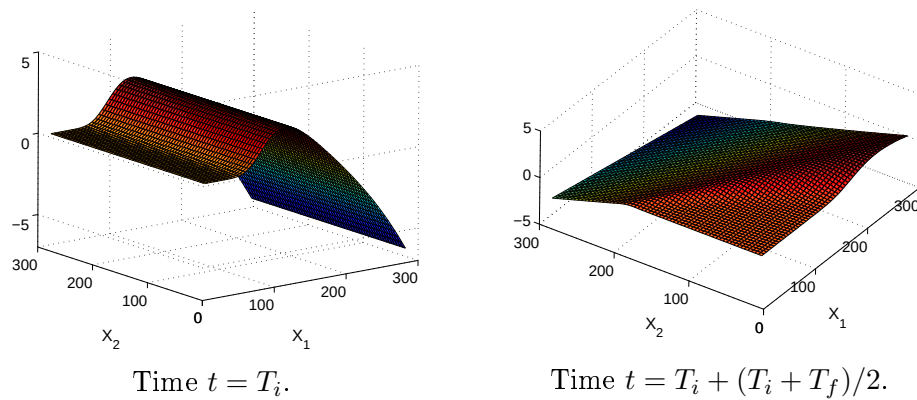
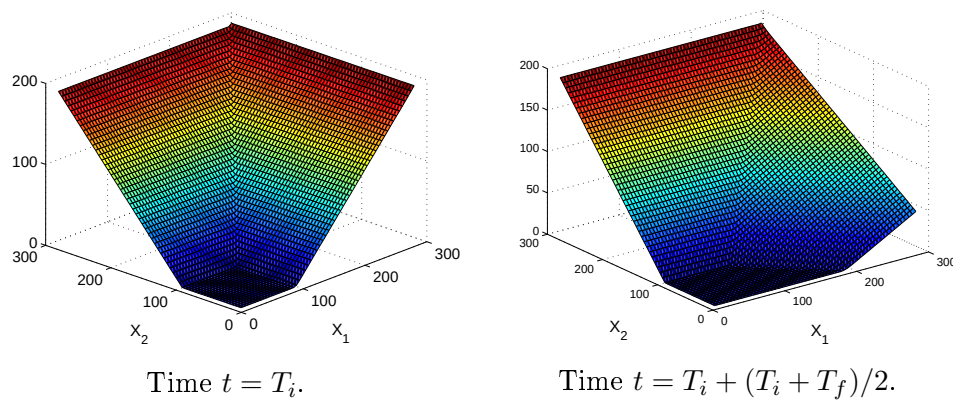
Figure 5.13: Eurasian call option “Vega” ($\Delta = \frac{\partial V}{\partial \sigma}$).Figure 5.14: Eurasian call option “Rho” ($\Delta = \frac{\partial V}{\partial r}$).

Figure 5.15: Amerasian call option value.

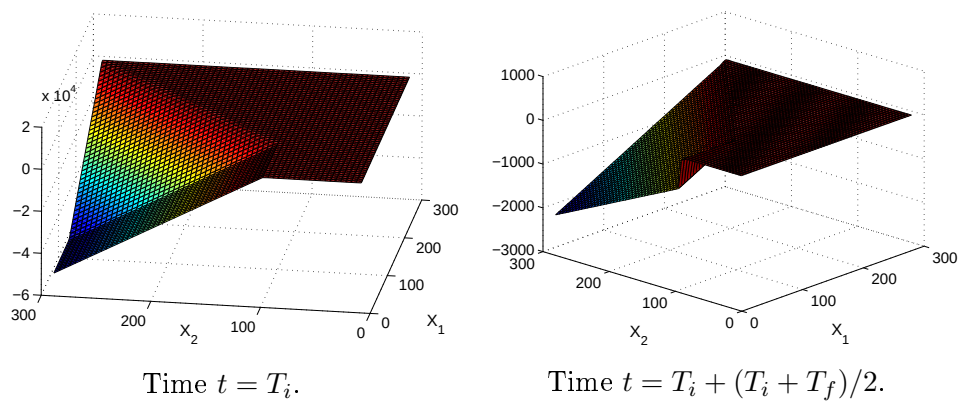


Figure 5.16: Amerasian call option Lagrange multiplier.

Conclusions

Our objective in this work has been the numerical solution of some pricing problems appearing in financial mathematics which can be formulated in terms of parabolic convection-diffusion-reaction partial differential equations (pdes), and eventually involving inequality constraints. As a particular application we have considered Asian options pricing problems, both in European and American styles. In order to provide accuracy and efficiency in the computation of price values we propose second order Lagrange-Galerkin methods.

So, we have performed the numerical analysis of second order Lagrange-Galerkin method for solving linear convection-diffusion-reaction equations. These methods have been completely analyzed in the following sense:

- The method has been introduced by using the formalism of continuum mechanics and weak formulations.
- For the characteristics time semidiscretizations, stability and second order error estimates have been obtained when smooth enough data and solutions are available.
- Lagrange-Galerkin schemes are analyzed to establish stability and second order accuracy error estimates for the fully discretized problem.
- In the light of the practical implementation, the use of quadrature formulas is taken into account providing a Fourier analysis of some quadrature choices. FORTRAN code programs have been developed and several appropriate test examples illustrate the theoretical results.

Therefore, an extensive numerical analysis of the proposed Lagrange-Galerkin second order schemes has been performed to deduce the improvement of these methods with respect to the first order classical methods. Although the analysis is only given for velocity fields null on the boundary, the stated Green's formula allow us to write the weak formulation of problems for which this assumption is not satisfied and, thereby, to write second order schemes for their numerical solution.

Regarding the application to Eurasian options pricing problems, we first study the pde model which presents a degenerated diffusive term, in spite of which the existence, uniqueness and C^∞ regularity of the solution can be established. In this way, the use of higher order Lagrange-Galerkin scheme is justified. Moreover, it has been applied paying attention to specific features.

In view of the numerical results obtained herein and those in the literature, second order schemes are necessary to achieve accuracy and efficiency in Eurasian pricing problems, and they improve previous results which use first order schemes.

Next, we have addressed the valuation of Amerasian options, which is governed by a complementarity problem associated with the same pde. The nonlinearity of the model is given by a

unilateral constraint on price (greater or equal to the exercise value). The iterative Bermúdez-Moreno algorithm has been successfully applied to deal with this obstacle-type problem. We have compared it with a more recently proposed active set method, and have obtained improvements in efficiency.

Appendix A

Background in Stochastic Calculus

In order to obtain a self-contained document, in this section we recall some definitions and results, without proofs, of stochastic calculus. All of them are rigorously stated in general textbooks as [67] or [56]-[55], but they can also be found in financial-oriented books as [74] and [72]. For a more basic overview on stochastic calculus applied to finance see [79].

Definition A.1 *A continuous time process (or stochastic process) taking its values in a space $\mathcal{E}$ endowed with a σ -algebra ε is a family $(X(t), t \geq 0)$ of random variables defined in a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with values in $(\mathcal{E}, \varepsilon)$.*

A stochastic process can also be considered as a map from $\mathbb{R}^+ \times \Omega$ in $\mathcal{E}$. In this case, we use the notation $(X(t, \omega), t \geq 0, \omega \in \Omega)$. We shall always consider that this map is measurable from $(\mathbb{R}^+ \times \Omega, \mathcal{B}(\mathbb{R}^+) \times \mathcal{A})$ to $(\mathcal{E}, \varepsilon)$, where $\mathcal{B}$ denotes the Borel measure. For a fixed sample point $\omega \in \Omega$, the function $t \mapsto X(t, \omega)$, $t \geq 0$ is called *sample path, realization* or *trajectory*.

Given a stochastic process $(X(t), t \geq 0)$ and an open subset of $\mathcal{E}$, $\mathcal{E}_2$, the first exit time of $\mathcal{E}_2$ of the process $(X(t), t \geq 0)$ is defined as:

$$\tau := \inf\{t : t \geq 0, X(t) \notin \mathcal{E}_2\}.$$

Definition A.2 *Consider the probability space $(\Omega, \mathcal{A}, \mathbf{P})$, a filtration $(\mathcal{F}_t, t \geq 0)$ is an increasing family of σ -algebras included in $\mathcal{A}$.*

Definition A.3 *We say that a process $(X(t), t \geq 0)$ is adapted with respect to $(\mathcal{F}_t, t \geq 0)$ if for each t , $X(t)$ is $\mathcal{F}_t$ -measurable.*

The natural filtration of a stochastic process $(X(t), t \geq 0)$, is defined as the “completion” of the sets $\mathcal{F}_t = \sigma(X(s), s \leq t)$, i.e., the completion of the smallest σ -field for which $X(s)$ is measurable for every $s \in [0, T]$ (see [74]). Any process is adapted to its natural filtration.

Definition A.4 *A random variable τ with values in $\mathbb{R}_+ \cup \{+\infty\}$ is an stopping time with respect to the filtration $(\mathcal{F}_t, t \geq 0)$ if the event $\{\tau \leq t\}$ belongs to the σ -field $\mathcal{F}_t$ for every $t \geq 0$.*

What $\{\tau \leq t\} \in \mathcal{F}_t$ means is that, for every instant t , taking into account the available information (i.e., $\mathcal{F}_t$), we know whether $\tau \leq t$ or not.

Some properties of stopping times can be deduced from the definition. For instance, given τ_1 and τ_2 two stopping times, then the random variables $\min\{\tau_1, \tau_2\}$, $\max\{\tau_1, \tau_2\}$ and $\tau_1 + \tau_2$ are also stopping times.

There is a σ -algebra associated to a stopping time τ defined by

$$\mathcal{F}_\tau = \{A : A \in \mathcal{A}, A \cap \{\tau \leq t\} \in \mathcal{F}_t \forall t \geq 0\}.$$

It represents all available information until the random time τ .

Before introducing the concept of martingale, let us recall that, given a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, we say that the sequence $\{A_n\}$ converges almost surely in probability $\mathbf{P}$ ($\mathbf{P}$ a.s.) if

$$\mathbf{P}(\{\omega : \omega \in \Omega, A_n(\omega) \longrightarrow A(\omega)\}) = 1.$$

Definition A.5 A stochastic process $(M(t), t \geq 0)$ is a $\mathcal{F}_t$ -martingale if $\mathbf{E}(|M(t)|) < \infty$ for all $t \geq 0$, $M(t)$ is $\mathcal{F}_t$ -adapted and furthermore

$$\mathbf{E}(M(t)|\mathcal{F}_s) = M(s) \quad a.s. \quad \forall 0 \leq s \leq t.$$

Martingales are related to models of fair gambling: for instance, if $M(s)$ is a martingale process representing the amount of money a player possesses at time s , the martingale property means that the expected amount of the player in a posterior time $t > s$, given $M(s)$, is equal to $M(s)$, regardless the past history of fortune.

An example of martingale is a Brownian motion. Actually, the Brownian motion is the *fundamental martingale with continuous sample paths*. Its name comes from the botanist Robert Brown, who observed in 1828 the irregular movement of pollen suspended in water. After that, the range of application of Brownian motion has been extended to other physical, biological, economical and management systems.

Definition A.6 A $\mathcal{F}_t$ -Brownian motion is a real-valued, continuous stochastic process $(W(t), t \geq 0)$ with $W(0) = 0$ a.s., adapted to a given filtration $(\mathcal{F}_t, t \geq 0)$ and with independent and stationary increments.

The above definitions means that Brownian motion satisfies:

- For each $\omega \in \Omega$, the map $s \longrightarrow X(s, \omega)$ is continuous $\mathbf{P}$ a.s..
- If $s \leq t$, then $W(t) - W(s)$ is independent of the σ -algebra $\mathcal{F}_s$.
- If $s \leq t$, then $W(t) - W(s)$ has the same probability distribution as $W(t - s) - W(0)$.

Theorem A.1 If $(W(t), t \geq 0)$ is a Brownian motion, then $W(t) - W(0)$ is a gaussian random variable with mean rt and variance $\sigma^2 t$, where r and σ are constant real numbers.

Definition A.7 A Brownian motion $(W(t), t \geq 0)$ is a standard Brownian motion if

$$W(0) = 0 \quad \mathbf{P} \text{ a.s.}, \quad \mathbf{E}(W(t)) = 0, \quad \mathbf{E}(W^2(t)) = t.$$

Proposition A.1 Let $(W(t), 0 \leq t \leq T)$ a $\mathcal{F}_t$ -standard Brownian motion. Then

- $(W(t), 0 \leq t \leq T)$ is a $\mathcal{F}_t$ -martingale.

- $(W^2(t) - t, 0 \leq t \leq T)$ is a $\mathcal{F}_t$ -martingale.
- $(e^{\sigma W(t) - \frac{\sigma^2 t}{2}}, 0 \leq t \leq T)$ is a $\mathcal{F}_t$ -martingale.

Brownian motions have unbounded quadratic variation, which implies that sample paths of a Brownian motion are nowhere differentiable (see [67] Section 2.9 for a study of the Brownian motion sample paths). This has major consequences for the definition of a stochastic integral with respect to Brownian sample paths. The idea is that, if we try to extend the Riemann-Stieltjes integral, we find out that the Brownian sample paths cannot be integrated with respect to themselves. Thus, another integral definition is considered, the *Ito stochastic integral*, that can be seen as a mean square limit of certain Riemann-Stieltjes sums.

The construction is due to Ito in 1944. It has been extended to square-integrable martingales with continuous sample paths by Kunita and Watanabe in 1967. The rigorous construction can be seen in [67] Chapter 3. We recall here the fundamental ideas and some properties.

Let $(W(t), t \geq 0)$ be a real valued $\mathcal{F}_t$ -Brownian motion with respect to a given filtration $(\mathcal{F}_t, t \geq 0)$. Given an adapted continuous real valued process $(\beta(t), t \geq 0)$, the Ito integral of β with respect to $(W(t), t \geq 0)$ between 0 and T is defined as a limit of the discrete sums

$$I_T^n := \sum_{i=0}^{n-1} \beta(t_i^n) (W(t_{i+1}^n) - W(t_i^n)),$$

where t_n^i is a family of partitions of $[0, T]$ such that the mesh step tends to 0 with n .

Starting from the called simple process (see [67]), the Ito stochastic integral can be extended to the adapted square integrable processes

$$M_{\mathcal{F}}^2(0, T) = \left\{ (\Gamma(t), 0 \leq t \leq T), \mathcal{F}_t - \text{adapted process, } \mathbf{E} \left(\int_0^T \Gamma^2(s) ds \right) < \infty \right\}. \quad (\text{A.1})$$

Proposition A.2 *Let $(H(t), 0 \leq t \leq T) \in M_{\mathcal{F}}^2(0, T)$, then the process*

$$\left(\int_0^t H(s) dW_s, 0 \leq t \leq T \right)$$

is well defined, and it is a martingale.

The Ito stochastic integral can also be extended to the class of adapted almost sure square integrable processes

$$L_{\mathcal{F}}^2(0, T) = \left\{ (\Gamma(t), 0 \leq t \leq T), \mathcal{F}_t - \text{adapted process, } \mathbf{E} \left(\int_0^T \Gamma^2(s) ds \right) < \infty \text{ a.s.} \right\} \quad (\text{A.2})$$

but the martingale property is not preserved. All the Brownian martingales can be represented by a stochastic integral, as we will see in the following theorem.

Theorem A.2 *Let $(M(t), 0 \leq t \leq T) \in M_{\mathcal{F}}^2(0, T)$. Then there exists a $\mathcal{F}_t$ -adapted process $(H(t), 0 \leq t \leq T)$ such that $\mathbf{E}(\int_0^T H^2(s) ds) < \infty$ and*

$$\forall t \in [0, T] \quad M(t) = M(0) + \int_0^t H(s) dW(s) \quad \text{a.s.}$$

The differential calculus corresponding to the Ito stochastic integral is called Ito Calculus, and its fundamental tool is the Ito formula.

Definition A.8 Let $(\Omega, \mathcal{A}, (\mathcal{F}_t, t \geq 0), \mathbf{P})$ be a probability space and $(W(t), t \geq 0)$ an $\mathcal{F}_t$ -Brownian motion on it. A process $(X(t), t \geq 0)$ is a $\mathbb{R}$ -valued Ito process if it can be written as

$$X(t) = X(0) + \int_0^t K(s)ds + \int_0^t \Gamma(s)dW(s), \quad a.s. \quad \forall t \leq T, \quad (\text{A.3})$$

where

- $X(0)$ is $\mathcal{F}_0$ -measurable,
- $(K(t), 0 \leq t \leq T)$ and $(\Gamma(t), 0 \leq t \leq T)$ are $\mathcal{F}_t$ -adapted processes,
- $\int_0^t |K(s)|ds < \infty$ a.s.,
- $\int_0^t |\Gamma(s)|^2 ds < \infty$ a.s..

Theorem A.3 Let $(X(t), t \geq 0)$ be an Ito process as introduced in Definition (A.8) and $f \in C^{2,1}(\mathbb{R} \times [0, T])$ function, then

$$\begin{aligned} f(X(t), t) &= F(X(0), 0) + \int_0^t \frac{\partial f}{\partial t}(X(s), s)ds + \int_0^t \frac{\partial f}{\partial x}(X(s), s)dX(s) \\ &\quad + \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(X(s), s)d\langle X, X \rangle_s, \end{aligned} \quad (\text{A.4})$$

where, by definition

$$\langle X, X \rangle_t := \int_0^t \Gamma^2(s)ds,$$

and

$$\int_0^t \frac{\partial f}{\partial x}(X(s), s)dX(s) := \int_0^t \frac{\partial f}{\partial x}(X(s), s)K(s)ds + \int_0^t \frac{\partial f}{\partial x}(X(s), s)\Gamma(s)dW(s).$$

Definition A.9 Let $(\Omega, \mathcal{A}, (\mathcal{F}_t, t \geq 0), \mathbf{P})$ be a probability space and $(W(t), t \geq 0)$ a p dimensional $\mathcal{F}_t$ -Brownian motion on it, $W(t) = (W^{(1)}(t), \dots, W^{(p)}(t))$. A process $(X(t), t \geq 0)$ is a $\mathbb{R}$ -valued Ito process if it can be written as

$$X(t) = X(0) + \int_0^t K(s)ds + \sum_{i=1}^p \int_0^t \Gamma^{(i)}(s)dW^{(i)}(s), \quad a.s. \quad \forall t \leq T, \quad (\text{A.5})$$

where

- $X(0)$ is $\mathcal{F}_0$ -measurable,
- $(K(t), 0 \leq t \leq T)$ and all the processes $(\Gamma^{(i)}(t), 0 \leq t \leq T)$ are $\mathcal{F}_t$ -adapted processes,
- $\int_0^t |K(s)|ds < \infty$ a.s.,
- $\int_0^t |\Gamma^{(i)}(s)|^2 ds < \infty$ a.s, and for all $1 \leq i \leq p$.

Theorem A.4 Let $(X^{(1)}(t), \dots, X^{(n)}(t))$ be Ito processes

$$X^{(i)}(t) = X^{(i)}(0) + \int_0^t K^{(i)}(s)ds + \sum_{j=1}^p \int_0^t \Gamma^{(i,j)}(s)dW^j(s), \text{ a.s. } \forall t \leq T \quad (\text{A.6})$$

and $f \in C^{2,1}(\mathbb{R}^n \times [0, T])$, then

$$\begin{aligned} f(X^{(1)}(t), \dots, X^{(n)}(t), t) &= f(X^{(1)}(0), \dots, X^{(n)}(0), 0) + \int_0^t \frac{\partial f}{\partial t}(X^{(1)}(s), \dots, X^{(n)}(s), s)ds \\ &\quad + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X^{(1)}(s), \dots, X^{(n)}(s), s)dX^{(i)}(s) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X^{(1)}(s), \dots, X^{(n)}(s), s)d\langle X^{(i)}, X^{(j)} \rangle_s, \end{aligned} \quad (\text{A.7})$$

with

- $dX^{(i)}(s) = K^{(i)}(s)ds + \sum_{j=1}^p \Gamma^{(i,j)}(s)dW^{(j)}(s),$
- $d\langle X^{(i)}, X^{(j)} \rangle_s = \sum_{m=1}^p \Gamma^{(i,m)}(s)\Gamma^{(j,m)}(s)ds.$

Definition A.10 Let σ be a measurable function from $\mathbb{R}^d \times [0, T]$ to the set of $d \times p$ matrices, b be a measurable function from $\mathbb{R}^d \times [0, T]$ to $\mathbb{R}^d$, $(W(t), t \geq 0)$ be a p -dimensional Brownian motion and Z be a random variable. A process $(X(t), t \geq 0)$ belonging to $L^2_{\mathcal{F}}(0, T)$ which satisfies

$$X(t) = Z + \int_0^t b(X(s), s)ds + \int_0^t \sigma(X(s), s)dW(s) \text{ a. s.} \quad (\text{A.8})$$

is called a solution of the stochastic differential equation with coefficients b and σ , initial condition Z and Brownian motion $(W(t), t \geq 0)$. Moreover, $(X_t, t \geq 0)$ is called the diffusion process corresponding to the coefficients b (the drift) and σ (the diffusion matrix).

Notice that if $\sigma \equiv 0$ we have an ordinary differential equation. Formally we write (A.8) as

$$\begin{aligned} dX(t) &= b(X(t), t)dt + \sigma(X(t), t)dW(t), \\ X(0) &= Z. \end{aligned}$$

Also formally, the above equation means that if at time t the system has a known state $X(t)$, then, during the interval $(t, t + \Delta t)$ (for small Δt) the variation of state $\Delta X(t)$ is a Gaussian random variable with mean $b(X(t), t)\Delta t$ and variance $\sigma^2(X(t), t)\Delta t$.

Definition A.11 The family of infinitesimal generators of the diffusion $(X(t), t \geq 0)$ introduced in Definition A.10, $(A_t, t \geq 0)$, is the family of second order differential operators defined for functions $f \in C^2(\mathbb{R}^d)$ by

$$A_t f(x) := \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x, t) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{j=1}^p b_j(x, t) \frac{\partial f}{\partial x_j}(x),$$

where $a(x, t) := \sigma(x, t)\sigma^*(x, t)$.

Theorem A.5 Let $|\cdot|$ denote the Euclidean norm in $\mathbb{R}^d$. Suppose that b and σ are continuous functions and there exist constants K and C such that

$$|b(x, t) - b(y, t)| + |\sigma(x, t) - \sigma(y, t)| \leq K|x - y| \quad \forall x, y \in \mathbb{R}^d, t \geq 0, \quad (\text{A.9})$$

$$|b(x, t)| + |\sigma(x, t)| \leq C(1 + |x|) \quad \forall x \in \mathbb{R}^d, t \geq 0. \quad (\text{A.10})$$

Let us assume that $\mathbf{E}(|Z|^2) \leq \infty$. Then, for all $T > 0$, there exists a unique continuous process

$$(X(t), t \geq 0) \in L^2_{\mathcal{F}}(0, T)$$

satisfying (A.8) and

$$\mathbf{E} \left(\sup_{0 \leq s \leq T} |X(s)|^n \right) < \infty, \quad \forall T > 0, n \geq 1. \quad (\text{A.11})$$

Markov property of a process $(X(t), t \geq 0)$ means that the future behavior of this process after time t only depends on $X(t)$, and not on $X(s)$ for $s < t$. There exist some equivalent statements defining the Markov property. We adopt the following definition.

Definition A.12 A process $(X(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t, t \geq 0)$ satisfies the Markov property with respect to $(\mathcal{F}_t, t \geq 0)$ if

$$\mathbf{E}(f(X(t)) | \mathcal{F}_s) = \mathbf{E}(f(X(t)) | X(s))$$

with f being a Borel-measurable bounded function and $s \leq t$.

Let us introduce the notation $(X(s)^{x,t}, s \geq 0)$ for the solution of equation (A.8) with initial condition x at time t , i.e.,

$$X^{x,t}(\theta) = x + \int_t^\theta b(X^{x,t}(s), s) ds + \int_t^\theta \sigma(X^{x,t}(s), s) dW(s) \text{ a. s.} \quad (\text{A.12})$$

Theorem A.6 Let $(X(t), t \geq 0)$ be a solution of (A.8). Then it satisfies the Markov property with respect to the filtration $(\mathcal{F}_t, t \geq 0)$ of the Brownian motion. More precisely, for all Borel-measurable bounded function f we have

$$\mathbf{E}(f(X(t)) | \mathcal{F}_s) = \phi(X(s))$$

where $\phi(x) = \mathbf{E}(f(X^{x,s}(t)))$.

Girsanov theorem provides a device to solve stochastic differential equations driven by Brownian motion, by changing the underlying probability measure. We only recall here the version which claims that a Brownian motion with drift on a bounded interval can be considered, after a change of probability, a standard Brownian motion.

Definition A.13 Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. A probability $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$ if $\forall A \in \mathcal{A}$ with $\mathbf{P}(A) = 0$ then $\mathbf{Q}(A) = 0$.

Theorem A.7 $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$ if and only if there exist a positive random variable Z such that

$$\forall A \in \mathcal{A}, \quad \mathbf{Q}(A) = \int_A Z(\omega) d\mathbf{P}(\omega).$$

In this case, Z is the density of $\mathbf{Q}$ with respect to $\mathbf{P}$.

Definition A.14 Two probabilities $\mathbf{P}$ and $\mathbf{Q}$ are equivalent if each of them is absolutely continuous with respect to the other, or, equivalently, if $\mathbf{P}(Z > 0) = 1$.

Theorem A.8 Let T be a finite positive number. Let $(W(t), 0 \leq t \leq T)$ be a n -dimensional Brownian motion with respect to a given filtration $(\mathcal{F}_t, 0 \leq t \leq T)$. Let $(h(t), 0 \leq t \leq T)$ be a random process $\mathcal{F}_t$ -adapted, taking its values in $\mathbb{R}^n$ and such that

$$\int_0^T |h(s)|^2 ds < \infty \quad \text{a.s.}$$

Let

$$L_T = \exp \left(- \int_0^T h(s) dW(s) - \frac{1}{2} \int_0^T |h(s)|^2 ds \right), \quad (\text{A.13})$$

and

$$\widetilde{W}(t) = W(t) + \int_0^t h(s) ds. \quad (\text{A.14})$$

Then, if $\mathbf{E}(L_T) = 1$, L_T defines a new probability $\mathbf{Q}$ on $\mathcal{F}_T$ by

$$\mathbf{Q}(A) = \mathbf{E}(L_T 1_A) \quad \forall A \in \mathcal{A}. \quad (\text{A.15})$$

Then the process $(\widetilde{W}(t), 0 \leq t \leq T)$ is a standard Brownian motion under probability $\mathbf{Q}$.

Definition A.15 A Markov transition function, or transition probability $p(s, x, t, A)$ is a non-negative function defined for $0 \leq s < t < \infty$, $x \in \mathbb{R}^n$, $A \in \mathcal{B}^n$ and satisfying

- $p(s, x, t, A)$ is Borel measurable as a function of x .
- $p(s, x, t, A)$ is a probability measure as a function of A .
- p satisfies the Chapman-Kolmogorov equation

$$p(s, x, t, A) = \int_{\mathbb{R}^d} p(s, x, \lambda, dy) p(\lambda, y, t, A), \quad \forall s < \lambda < t.$$

Theorem A.9 Let p be a transition probability function. Then, for any $s \geq 0$ and for any probability distribution $F(dx)$ on $(\mathbb{R}^d, \mathcal{B}^d)$, there exists a n -dimensional stochastic process $(X(t), s \leq t < \infty)$ such that

$$\begin{aligned} \mathbf{P}(X(s) \in A) &= F(A) \\ \mathbf{P}(X(t) \in A | \sigma(X(\bar{s}))) &= p(\bar{s}, X(\bar{s}), t, A) \quad \text{a.s. } s \leq \bar{s} < t. \end{aligned}$$

For instance, the n -dimensional Brownian motion has the stationary transition probability function

$$p(t, x, A) = \int_A \frac{1}{(2\pi t)^{n/2}} \exp \left(-\frac{|x-y|^2}{2t} \right) dy. \quad (\text{A.16})$$

Appendix B

Black-Scholes pricing framework for vanilla options

In this section we recall Black-Scholes [29] and Merton [77] hypotheses and methodology to formulate a model for pricing vanilla options. This framework can be used to price other exotic options, and some of the hypotheses can be weakened (see, for instance, [111], [71], [110]).

In addition to the no arbitrage hypothesis and the geometric Brownian motion model for the underlying asset, the new idea here is the elimination of risk or *delta-hedging* by using a *self-financed strategy* replicating options.

We begin with the European vanilla options, developing two different approach: a pde approach and a more probabilistic one. The same structure is repeated for the American vanilla options.

B.1 European vanilla options

An investment strategy is said to be *self financed* if no extra funds are added or withdrawn from the initial investments. Thus, the cost of acquiring more units of one security in the portfolio is completely financed by the sale of some units of other securities within the same portfolio.

Let us take the hypothesis from the original paper [29]:

1. The short-term interest rate is known and constant through time.
2. The stock price follows a random walk in continuous time with a variance rate proportional to the square of the stock process. Thus, the distribution of possible stock prices at the end of any finite interval is lognormal. The variance rate of return on the stocks is constant.
3. The stock pays no dividends or other distributions.
4. The option is European, that is, it can only be exercised at maturity.
5. There are no transaction costs in buying or selling the stock or the option.
6. It is possible to borrow any fraction of the price of a security to buy it or to hold it, at the short-term interest rate.

7. There are no penalties to short selling. A seller who does not own a security will simply accept the price of the security from a buyer, and will agree to settle with the buyer on some future date by paying him an amount equal to the price of the security on that date.

The first assumption implies that a risk-free asset, B , in such a market satisfies

$$dB(t) = rB(t)dt, \quad B(t_0) = B_0,$$

and then

$$B(t) = B_0 e^{-r(t-t_0)}.$$

The second assumption is equivalent to affirm that the risky asset follows a geometric Brownian motion with constant parameters μ and σ , i.e.,

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad t \geq t_0, \quad S(t_0) = S_0, \quad (\text{B.1})$$

or, equivalently,

$$S(t) = S_0 e^{(\mu - \frac{1}{2}\sigma^2)(t-t_0) + \sigma(W(t) - W(t_0))}, \quad (\text{B.2})$$

PDE approach

Now, our aim is to formulate a model for the price, V , of an European option on the underlying asset S , for which the payoff function is a deterministic function $\Lambda = \Lambda(S)$. Concerning the value of the option we assume:

- Function V is a deterministic function on the underlying asset and on time, i.e., at time t , if the underlying asset price is denoted by S , we have $V = V(S, t)$.
- Function V is smooth, $V \in C^{2,1}(\mathbb{R}_+ \times [0, T])$.

Next, the idea of *hedging* is applied by constructing a risk-free portfolio by linearly combining the random walk for dV and the random walk for dS . More precisely, let us consider a self-financed portfolio Π with one European vanilla call option and with short-selling a quantity Δ (unknown a priori) of the underlying asset, so the portfolio value is

$$\Pi(S, t) = V(S, t) - \Delta S(t).$$

The variation of the self-financed portfolio value between t and $t + dt$ is given by

$$d\Pi(t) = dV(t) - \Delta dS(t). \quad (\text{B.3})$$

Since V is smooth enough to apply Ito's lemma (Theorem A.3), then

$$dV(t) = \left(\frac{\partial V}{\partial t} + \mu S(t) \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2(t) \frac{\partial^2 V}{\partial S^2} \right) dt + \sigma S(t) dW(t). \quad (\text{B.4})$$

Next, by replacing (B.1) and (B.4) in (B.3) we obtain

$$d\Pi(t) = \left(\frac{\partial V}{\partial t} + \mu S(t) \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2(t) \frac{\partial^2 V}{\partial S^2} - \Delta \mu S(t) \right) dt + (\sigma S(t) - \Delta \sigma S(t)) dW(t).$$

By choosing $\Delta = \frac{\partial V}{\partial S}$, the random part of the above equation is eliminated and we have

$$d\Pi(t) = \left(\frac{\partial V}{\partial t} + \mu S(t) \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2(t) \frac{\partial^2 V}{\partial S^2} \right) dt, \quad (\text{B.5})$$

obtaining a deterministic equation for the evolution of Π . If we had invested the value of the portfolio on a risk-free asset, we would have obtained

$$d\Pi(t) = r\Pi(t)dt. \quad (\text{B.6})$$

By a no arbitrage argument, (B.5) and (B.6) are equal, and then we deduce Black-Scholes equation

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0. \quad (\text{B.7})$$

The backward parabolic equation (B.7) is called the Black-Scholes equation. Let us introduce the Black-Scholes differential operator, applied to a smooth function ϕ ,

$$\mathcal{L}_{BS}[\phi] := \frac{\partial \phi}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 \phi}{\partial S^2} + rS \frac{\partial \phi}{\partial S} - r\phi. \quad (\text{B.8})$$

Adding a final condition given by the payoff function, the following final value problem is posed for an European option with value V :

$$\begin{cases} \mathcal{L}_{BS}[V] = 0, & \text{for } (S, t) \in \mathbb{R}_+ \times [0, T], \\ V(S, T) = \Lambda(S), & \text{for } S \in \mathbb{R}_+. \end{cases} \quad (\text{B.9})$$

In the particular case of vanilla call or vanilla put options, the payoff is given by

$$\Lambda(S) = (S - K)_+$$

for the call, and

$$\Lambda(S) = (K - S)_+$$

for the put, and the problem has an explicit solution called the Black-Scholes formula.

Remark B.1 Notice that the value of an option is independent of the growth of rate μ , since it does not appear in the equation nor in the final condition. An implication is that two investors can agree on the fair value of an option regardless of their respective views about the expected performance of the underlying security, i.e., the mathematical relationship between prices of the derivative and the asset is invariant to the risk preferences of the investors. This idea is related to the risk-neutral approach, developed in the following paragraph.

Martingale/risk neutral approach

Black-Scholes model can be formulated by using a more probabilistic approach, called martingale or risk-neutral approach. The martingale pricing theory asserts that a continuous time financial market consisting of trading securities and trading strategies is arbitrage free if and only if there exists and equivalent (*martingale*) measure under which the discounted asset prices are martingales. Moreover, if the market is complete, the equivalent measure is unique. In such a market, arbitrage price of contingent claims is given by the discounted value of the terminal payoff under this martingale measure.

The following results can be seen from [72], see also [68] for a deeper study.

Firstly, let us give a mathematical definition for some already introduced financial concepts.

Definition B.1 A strategy is a $\mathbb{R}^2$ -valued stochastic process $(\beta(t), 0 \leq t \leq T)$ with $\beta(t) = (H^0(t), H(t))$, adapted to the natural filtration $(\mathcal{F}_t, 0 \leq t \leq T)$ of the Brownian motion.

At each time t , a portfolio Π following the strategy β , $\Pi\{\beta\}$, has $H^0(t)$ unities of risk-free asset and $H(t)$ unities of risky asset, i.e.,

$$\Pi\{\beta\}(t) = H^0(t)B(t) + H(t)S(t).$$

As we have said, in self-financed portfolios all the gains obtained are reinvested in the portfolio, no consumption is allowed.

Definition B.2 A self-financed strategy $(\beta(t), 0 \leq t \leq T)$ with $\beta(t) = (H^0(t), H(t))$ is a strategy satisfying

$$\begin{aligned} \int_0^T |H^0(t)|dt + \int_0^T H^2(t)dt &< \infty, \quad a.s. \\ H^0(t)B(t) + H(t)S(t) &= H^0(0)B(0) + H(0)S(0) + \int_0^t H^0(u)dB(u) + \int_0^t H(u)dS(u), \end{aligned}$$

a.s., for all $t \in [0, T]$.

Definition B.3 A strategy $(\beta(t), 0 \leq t \leq T)$ is admissible if it is self-financed and the value of its corresponding actualized portfolio

$$\tilde{\Pi}\{\beta\}(t) = H^0(t) + H(t)\tilde{S}(t)$$

is positive for all t and $\sup_{t \in [0, T]} \tilde{\Pi}(t) \in M_{\mathcal{F}}^2$ under $\mathbf{Q}$ (see (A.1) for the definition of square integrable processes $M_{\mathcal{F}}^2$).

We assume that the value of an European option is given by a random variable F_T -measurable h . In this context we will say that an option is replicable if there exists an admissible strategy which is equal to the option value at maturity, and we will say that the market is complete if all assets are replicable.

Next, let us find the martingale measure. For this, we introduce the discounted asset prices

$$\tilde{S}(t) := e^{-rt}S(t),$$

which solves the stochastic differential equation

$$d\tilde{S}(t) = (\mu - r)\tilde{S}(t)dt + \sigma\tilde{S}(t)dW(t).$$

Secondly, we introduce

$$\tilde{W}(t) = W(t) + \frac{\mu - r}{\sigma}t.$$

Now, since $h(\tau) = \frac{\mu - r}{\sigma}$ clearly satisfies

$$\int_0^T h^2(\tau)d\tau = \left(\frac{\mu - r}{\sigma}\right)^2 T < \infty,$$

Theorem A.8 can be applied, so that there exists a new probability (risk-neutral probability) under which $(\tilde{W}(t), 0 \leq t \leq T)$ is a standard Brownian motion. This new probability is defined by

$$\mathbf{Q}(A) = \mathbf{E}(L_T 1_A), \tag{B.10}$$

with

$$L_T = e^{\left(-\frac{\mu-r}{\sigma}W(T) - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2 T\right)}. \quad (\text{B.11})$$

Since the discounted asset prices satisfy

$$d\tilde{S}(t) = \tilde{S}(t)\sigma d\tilde{W}(t),$$

then

$$\tilde{S}(t) = \tilde{S}(0)e^{\sigma W(t) - \frac{\sigma^2}{2}t},$$

and $(\tilde{S}(t), 0 \leq t \leq T)$ is a martingale (see Proposition A.1).

Theorem B.1 *Under Black-Scholes hypotheses, an European option defined by a positive, F_T -measurable random variable h , with $h \in M_{\mathcal{F}}^2$ under $\mathbf{Q}$, is replicable, and its value at time t of a portfolio simulating h is*

$$V(t) = \mathbf{E}^{\mathbf{Q}} \left(e^{-r(T-t)} h | \mathcal{F}_t \right). \quad (\text{B.12})$$

PROOF. We recall here the idea of the proof, the complete one can be seen from [72].

Firstly, we assume the existence of an admissible strategy, $(\beta(t), 0 \leq t \leq T)$, replicating the option, i.e.,

$$V\{\beta\}(t) = H^0(t)B(t) + H(t)S(t), \forall t < T,$$

and that $V(T) = h$. Let be $\tilde{V}$ its discounted value,

$$\tilde{V}(t) = e^{-rt}V(t) = H^0(t) + H(t)\tilde{S}(t).$$

Using now that the strategy is selffinanced, we have

$$\tilde{V}(t) = V(0) + \int_0^t H(u)d\tilde{S}(u) = V(0) + \int_0^t H(u)\sigma\tilde{S}(u)dW(u).$$

And finally, we deduce, using Proposition (A.2), that $\tilde{V}$ is, under $\mathbf{Q}$, a square integrable martingale, i.e.,

$$\tilde{V}(t) = \mathbf{E}^{\mathbf{Q}} \left(\tilde{V}(T) | \mathcal{F}_t \right),$$

and thus (B.12) is deduced.

Secondly, the admissible strategy replicating the option is found by using the theorem of representation of Brownian martingales (Theorem A.2). $\square$ $\square$

Remark B.2 *The probabilistic representation (B.12) could have been deduced from Cauchy problem (B.9) by applying Feynman-Kac formulas (see [74]).*

Easy extensions of the Black-Scholes framework

A more general pricing framework can be easily considered by weakening some of the Black-Scholes and Merton hypotheses. (See Chapter 8 of [110] for simple generalizations of the Black-Scholes world).

For instance, the same analysis holds true for time-dependent parameters (interest rate, drift and volatility). Moreover, a priori known dividend payments can be allowed, given different models depending on whether they are considered to be discrete (dividend payments) or continuous (dividend yield) in time.

In the following, we will allow for proportional continuously distributed dividend yield with rate d_0 (i.e., the continuous payment is $d_0 S$). In this case, the same reasoning as in the absence of dividends developed, replacing (B.6) by

$$d\Pi(t) = r\Pi(t)dt + d_0\Delta Sdt, \quad (\text{B.13})$$

and thus, Black-Scholes differential operator results

$$\mathcal{L}_{BS}[\phi] := \frac{\partial\phi}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2\phi}{\partial S^2} + (r - d_0)S \frac{\partial\phi}{\partial S} - r\phi. \quad (\text{B.14})$$

B.2 Black-Scholes and American options

American options can be exercised at any time prior to expiry. From a financial point of view we not only look for the fair value of the option, but also if it is optimal, for the holder, to hold the option or to exercise it. Following [41], we first describe in this section how to derive a free boundary problem, where the free boundary separates the regions where it is optimal to hold from the regions where it is optimal to exercise it. Then, we also formulate the pricing problem as an optimal stopping problem.

Free boundary problem approach

Let us assume Black-Scholes and Merton hypothesis on the market and let us consider an American vanilla put option with strike price K and maturity T . By definition of the contract, for each time $t < T$, the holder has the right to exercise the option obtaining $(K - S)_+$.

Similarly to the European case, we denote by V the option value, and assume that it is a deterministic and smooth function on the asset value and on time. By no arbitrage arguments, we have deduced in Section 1.3.1 that

$$V(S, t) \geq (K - S)_+, \quad \forall t \leq T, \quad (\text{B.15})$$

which means that we will have a unilateral obstacle type problem. The free boundary is defined as the set of points $(S_f(t), t)$ at which the option value first meets the payoff, i.e., satisfying

$$V(S_f(t), t) = (K - S_f(t))_+, \quad V(S, t) > (K - S)_+ \text{ for } S > S_f(t).$$

Then, at each time t , the spatial domain is divided into two regions:

- For $S > S_f(t)$ we also have, by definition of the free boundary, that $V(S, t) > (K - S)_+$. In this case, the deduction of the Black-Scholes equation remains valid, because it is possible to take either short and long positions on the option.
- For $S < S_f(t)$ the Black-Scholes analysis fails because such a put option will be immediately exercised against the writer, and no longer long positions on the option are possible. Thus, in this case the Black-Scholes riskless portfolio satisfies $d\Pi \leq r\Pi dt$. Moreover, the option price equates the payoff function.

Summing up, we have

$$\begin{aligned} V(S, t) &= (K - S)_+ & \text{and} & \quad \mathcal{L}_{BS}[\phi] \leq 0, & \text{for } S \leq S_f(t), \\ V(S, t) &> (K - S)_+ & \text{and} & \quad \mathcal{L}_{BS}[\phi] = 0, & \text{for } S > S_f(t). \end{aligned} \quad (\text{B.16})$$

To complete the free boundary problem, the final condition is added

$$V(S, T) = (S - K)_+, \quad (\text{B.17})$$

and two boundary conditions are required on the (a priori) unknown free boundary,

$$V(S_f(t), t) = (K - S_f(t))_+ \quad (\text{B.18})$$

$$\frac{\partial V}{\partial S}(S_f(t), t) = -1. \quad (\text{B.19})$$

The second one can be deduced by arbitrage arguments. See [111] for a discussion on the American put option model, and [41] for the corresponding discussion on American call options with continuous dividend payment.

The above free boundary problem can be formulated as a complementarity problem,

$$\left. \begin{aligned} \mathcal{L}_{BS}[V] (V - \underline{\Lambda}) &= 0, \\ \mathcal{L}_{BS}[V] &\leq 0, \\ V - \underline{\Lambda} &\geq 0, \end{aligned} \right\} \text{ in } \mathbb{R}_+ \times (0, T). \quad (\text{B.20})$$

It has been also formulated as a variational inequality (see [41] for a deduction starting from the free boundary problem, and [62] for a variational inequality formulation using a probabilistic approach).

Optimal stopping problem approach

Similarly to the European case, there exists a link between the PDE approach leading to a complementarity problem and a probabilistic representation of the solution by an optimal stopping problem (see the reference textbook [15]). We require consumption strategies.

Definition B.4 *A consumption strategy is given by an adapted process $(\beta(t), 0 \leq t \leq T)$ with values in $\mathbb{R}^2$, $(\beta(t) = (H^0(t), H(t)))$, satisfying*

$$\begin{aligned} \int_0^T |H^0(t)| dt + \int_0^T H^2(t) dt &< \infty, \quad a.s. \\ H^0(t)B(t) + H(t)S(t) &= H^0(0)B(0) + H(0)S(0) + \int_0^t H^0(u)dB(u) + \int_0^t H(u)dS(u) - C(t), \end{aligned}$$

a.s. for all $t \in [0, T]$ where $(C(t), 0 \leq t \leq T)$ is a continuous increasing process, with $C(0) = 0$.

Under this approach, the American option price is defined by a positive valued adapted process, $(h(t), 0 \leq t \leq T)$. We only consider process with form $h(t) = \underline{\Lambda}(S(t))$, being $\underline{\Lambda}$ defined on $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ a continuous function such that

$$\underline{\Lambda}(x) \leq a_1 + a_2 x, \forall x \in \mathbb{R}_+,$$

with a_1 and a_2 positive constants. For instance, for an American vanilla put we have $\underline{\Lambda}(x) = (K - x)_+$. The consumption strategy $(\beta(t), 0 \leq t \leq T)$ hedge the American option if

$$V\{\beta\}(t) \geq \underline{\Lambda}(S(t)) \quad a.s. \quad \forall t \in [0, T].$$

In this approach the value of the American options is considered the minimal value of the consumption strategies hedging the option. More precisely, the price is given by the function $u(S(t), t)$ defined in the following theorem.

Theorem B.2 *Let $u : \mathbb{R}_+ \times [0, T] \longrightarrow \mathbb{R}$ be*

$$u(x, t) = \sup_{\tau \in \Upsilon_{t, T}} \mathbf{E}^{\mathbf{Q}} \left(e^{-r(\tau-t)} \psi(xe^\alpha) \right), \quad (\text{B.21})$$

with $\alpha = (r - \frac{\sigma^2}{2})(t - \tau) + \sigma(W_\tau - W_t)$ and with $\Upsilon_{t, T}$ is the set of stopping times (Definition A.4) with values in $[t, T]$. There exists a strategy β^* hedging the option such that $V\{\beta^*\}(t) = u(t, S(t))$ for all $t \in [0, T]$. Moreover, for all strategy β hedging the option we have $V\{\beta\}(t) \geq u(t, S(t))$, $\forall t \in [0, T]$.

For further reading see, for instance, [72] or [68].

B.3 Black-Scholes and m-factor models

The same ideas applied to one asset options can be applied to multi-assets options within Black-Scholes and Merton framework. Let us consider an option with m assets, $S^{(i)}, i = 1, \dots, m$, i.e., the payoff function is a function of m variables

$$\Lambda = \Lambda(S^{(1)}, \dots, S^{(m)}).$$

Firstly, we assume that each underlying follows a lognormal random walk

$$dS^{(i)}(t) = S^{(i)}(t) \left(\mu_i dt + \sum_{j=1}^m \sigma_{ij} dW^{(j)}(t) \right), \quad (\text{B.22})$$

where $(W^{(1)}, \dots, W^{(m)})$ is a m -dimensional Brownian motion. We allow to the stocks the possibility of paying a proportional continuously distributed dividends with rates d_i . Thus, the Black-Scholes type operator is (see [110])

$$\mathcal{L}_{BS}[\phi] := \frac{\partial \phi}{\partial t} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m a_{ij} S^{(i)} S^{(j)} \frac{\partial^2 \phi}{\partial S^{(i)} \partial S^{(j)}} + \sum_{i=1}^m (r - d_i) S^{(i)} \frac{\partial \phi}{\partial S^{(i)}} - r\phi, \quad (\text{B.23})$$

where

$$a_{ij} = \sum_{k=1}^m \sigma_{ik} \sigma_{jk}.$$

To obtain (B.23) it suffices with building a self-financed portfolio,

$$\Pi = V - \sum_{i=1}^m \Delta_i S^{(i)},$$

and applying multi-dimensional Ito's formula (A.6), delta-hedging with $\Delta_i = \frac{\partial V}{\partial S^{(i)}}$ and no arbitrage argument.

Appendix C

Resumo da tese en galego

Nesta memoria estudamos a resolución numérica de ecuacións en derivadas parciais de convección-difusión-reacción lineares mediante métodos de Lagrange-Galerkin de segunda orde. A nosa motivación é unha aplicación na matemática financeira: calcular o prezo de produtos financeiros de tipo derivado usando ecuacións en derivadas parciais (edps). O noso estudo focalízase nos *modelos de dous factores*, isto é, en produtos financeiros cuxo valor depende de dúas variabeis aleatorias. Como un exemplo representativo tratamos o problema de valoración das opcións asiáticas de estilo europeo e de estilo americano.

As ecuacións lineares de convección-difusión-reacción aparecen no modelado matemático na mecánica de fluídos, meteoroloxía, finanzas Moi frecuentemente, o termo convectivo é moito máis grande do que o difusivo, orixinando os problemas de convección dominante. Isto acontece, por exemplo, nas ecuacións de Navier-Stokes cun número de Reynolds grande e nas ecuacións de convección-difusión cun número de Péclet grande. Este termo convectivo (dominante) é fonte de dificultades computacionais na solución numérica de tales ecuacións (vexa, por exemplo, [81, 89]).

Nun sistema de coordenadas “Lagrangianas” o efecto da convección non aparece. Así, xorden na literatura métodos numéricos que usan, para a discretización temporal, aproximacións mediante diferencias finitas da derivada material ao longo das traxectorias características (vexa, por exemplo, [45]). Nos problemas con convección significativa, a solución cambia moito menos rapidamente no sentido das características do que no sentido do tempo. Por iso, unha discretización Lagrangiana permite o uso de pasos grandes de tempo mantendo a estabilidade e o erro de consistencia. Ademais, conduce a sistemas simétricos. En oposición a estes métodos, temos os métodos Eulerianos que desenvolven a discretización temporal no sentido do tempo. Estes métodos non poden simular exactamente todas as interaccións da onda que ocorren se a información se propaga máis dunha cela por etapa de tempo, por razóns de estabilidade, no caso dos métodos explícitos, ou por causa do erro, no caso dos métodos implícitos (vexa o traballo de revisión [47]).

Máis concretamente, neste traballo consideramos unha discretización Lagrangiana que calcula a posición no instante t_n das partículas que alcanzan os puntos dunha grella espacial fixada no instante t_{n+1} . Este método, proposto a principios dos oitenta [45, 88], é coñecido como *método (modificado) das características*, ou *método semi-Lagrangiano*. Foi combinado con discretizacións espaciais de diferencias finitas [45], elementos finitos [88, 16, 21, 82, 103, 102, 91], elementos finitos espectrais [104, 4], elementos finitos discontinuos [6, 5, 7], ...

Neste traballo interesámonos pola combinación do método semi-Lagrangiano cos elementos

finitos. Este método, chamado *método de características elementos finitos* ou *método de Lagrange-Galerkin*, usouse na resolución numérica dunha gran variedade de problemas de edps. Por exemplo, aplicado ás ecuacións de Navier-Stokes, foi analizado en [88, 102], onde se obtiveron estabilidade incondicional e estimacións do erro. A solución numérica da ecuación de convección-difusión con velocidade independente do tempo é estudada en [45, 44], e en [88, 103, 14] para velocidade dependente do tempo. A estabilidade incondicional do método foi obtida nestes traballos. Alén diso, en [103] probáronse estimacións do erro da forma $O(h^k + \Delta t)$ na norma $l^\infty(L^2(\mathbb{R}^m))$. No anterior, Δt denota o paso de tempo, h o paso espacial, k o grao do espazo do elemento finito, e m a dimensión do dominio espacial. En [88] obtéñense estimacións $O(h^k + \Delta t + h^{k+1}/\Delta t)$ na norma $l^\infty(L^2(\Omega))$ baixo a hipótese de que a compoñente normal da velocidade é nula na fronteira do dominio espacial Ω . Todas estas estimacións involucran constantes que dependen de normas da solución. Máis recentemente, para elementos finitos lineares e para un campo da velocidades que se anula na fronteira, unha converxencia de orde $O(h^2 + \Delta t + \min(h, h^2/\Delta t))$ na norma $l^\infty(L^2(\Omega))$ próbase en [14], onde as constantes na estimación dependen só dos datos. En todos os casos anteriores, os métodos son de primeira orde en tempo.

Tendo como obxectivo a segunda orde na discretización temporal usando métodos semi-Lagrangianos, atopamos dúas ideas diferentes na literatura. A primeira consiste en aproximar a derivada material cun esquema de tres puntos, e avaliar implicitamente o resto da ecuación. Este método, ao que nós chamamos *método de dous puntos de Lagrange-Galerkin*, foi proposto e analizado para ecuacións de convección-difusión unidimensionais en [46], e para ecuacións de Navier-Stokes no caso incompresible en [31]. A segunda aproximación consiste en usar unha discretización Lagrangiana de tipo Crank-Nicholson. Máis precisamente, a derivada material aproxímase por unha fórmula de dous puntos que sexa de segunda orde nun punto “características-tempo” intermedio, e o resto da ecuación aproxímase nese punto. Aínda que a idea de Crank-Nicholson xa foi suxerida en [88, 91], o correcto descentramento dos termos da ecuación distintos da derivada material propónse en [99]. Este segundo método é analizado cando se aplica a unha ecuación de convección-difusión con coeficientes constantes, condicións Dirichlet e campo de velocidades con diverxencia nula en [99]. Obtense estabilidade incondicional e estimacións do erro de tipo $O(\Delta t^2 + h^k)$ na norma $l^\infty(L^2(\Omega))$.

No presente traballo propomos unha extensión deste último método de características-elementos finitos, que chamamos *método Crank-Nicholson Lagrange-Galerkin*: permitimos un termo difusivo con coeficientes variabeis e eventualmente dexenerado, no canto do máis clásico Laplaciano; consideramos campos de velocidades con diverxencia non nula; un termo de reacción e condicións de contorno xerais de tipo mixto Dirichlet-Robin. Alén diso, usamos o formalismo matemático da mecánica dos medios continuos (vexa por exemplo [58]) para expresar os resultados e as notacións relacionados coas curvas características. De feito, o esquema é introducido e motivado despois dunha formulación feble do problema en termos das características. Esta formulación podería axudar se quixésemos aplicar un esquema similar con outras condicións de contorno, por exemplo, ou a outras ecuacións. Obtemos resultados da estabilidade, e unha estimación do erro $O(\Delta t^2 + h^k)$ na norma $l^\infty(L^2(\Omega))$. Un resumo destes resultados pode verse en [24, 25].

As propiedades dos método de características-elementos finitos introducidos previamente, tanto do método clásico como dos de segunda orde, establécense baixo a hipótese de que as integrais na formulación de Galerkin son calculadas exactamente. Como isto é raramente posíbel na práctica, úsase cuadratura numérica. Nalgúns casos, isto produce a perda da estabilidade incondicional e engade algúns termos ás estimacións finais do erro (vexa [82, 103, 94, 52, 105, 25]).

Hai algúns traballos na literatura que estudan o efecto da integración numérica sobre o

método de Lagrange-Galerkin clásico con elementos finitos lineares a pedazos. En particular, en [82] próbase a inestabilidade condicional para unha ampla clase de fórmulas de cuadratura e ecuacións de transporte lineares. Este traballo foi extendido a ecuacións de convección-difusión lineares en [103], e a unha máis grande clase de fórmulas da cuadratura en [94]. Nos estudos anteriores conclúese que as fórmulas da cuadratura de Gauss-Lobatto conducen aos esquemas máis estabéis, mais só a fórmula trapezoidal (Gauss-Lobatto de dous puntos) preserva a estabilidade incondicional. En [52] faise unha análise de Fourier do método de Lagrange-Galerkin de dous pasos con elementos finitos lineares, e conclúese que o método de dous pasos parece ser máis inestábel do que o clásico. Con respecto ao método de Lagrange-Galerkin Crank-Nicholson, en [105] os autores mostran experimentalmente que é máis robusto do que o esquema clásico con respecto á integración numérica e produce resultados numéricos mellores. No presente traballo estudamos tamén o efecto da cuadratura numérica para os métodos de Lagrange-Galerkin clásico e Crank-Nicholson con elementos finitos de graos $k = 1, 2$, sobre grellas triangulares e propomos fórmulas de cuadratura adecuadas en cada caso. Unha análise teórica baseada no método de Fourier é desenvolvida para algunhas das fórmulas.

Os métodos dos que falamos foron implementados en computador cun código FORTRAN. Adxuntamos nesta memoria resultados numéricos que ilustran e completan a nosa análise.

O motivo de estudar solucións numéricas de ecuacións de convección-difusión-reacción con convección dominante é que as ecuacións deste tipo modelan a evolución do prezo dalgúns contratos financeiros. De feito, a solución numérica de problemas de valoración de modelos de dous factores constitúe a segunda parte deste traballo.

A negociación dos produtos financeiros de tipo derivado remóntase, nos mercados organizados, ao principio dos anos setenta, e, nos mercados non organizados (OTC), a mediados dos oitenta. Este mercado creceu espectacularmente nas últimas tres décadas e os contratos financeiros tornáronse máis e máis complexos. Do mesmo xeito, a investigación sobre derivados financeiros tamén estoupou desde que, no 1973, Black, Scholes [29] e Merton [77] publicaron os seus traballos. Empregando as súas técnicas, os modelos matemáticos para o prezo das opcións financeiras fórmulanse coma problemas de valor final parabólicos de segunda orde, ás veces dexenerados (vexa, por exemplo, [111, 110, 71]). Estes problemas involucran eventualmente desigualdades de tipo obstáculo relacionadas co exercicio adiantado das opcións. A dimensión do dominio espacial depende do produto financeiro concreto. Un produto con m variabeis espaciais denomínase *modelo* ou *produto de m factores*. Baixo as mesmas hipóteses de modelado, o prezo das opcións tamén ten unha representación probabilística. Máis concretamente o valor dunha opción pode ser expresado como o valor previsto de seu *payoff* (ou valor da opción ao final do contrato) actualizado.

Hai algúns exemplos particulares de opcións, como as *vainilla*, para as que se coñece explicitamente a solución. Mais na maioría dos casos os métodos numéricos teñen que ser empregados para aproximar o valor da opción. Pódense distinguir os métodos que se centran en estimar a esperanza condicional do payoff final, dos métodos que aproximan numericamente a solución da edp. Con respecto a este segundo grupo, os métodos máis empregados nas finanzas involucran discretizacións Eulerianas da derivada temporal combinadas con diferencias finitas para a discretización espacial, e os métodos de proxección e relaxación para tratar as restricións unilaterais (vexa, por exemplo, [111]). A maioría deles son aplicados a problemas dun factor. Con todo, nas últimas dúas décadas empregáronse nas finanzas computacionais métodos máis sofisticados da mecánica de fluídos. Temos, por exemplo, métodos de volumes finitos [116] e métodos de elementos finitos [90, 76] para a discretización espacial; ou métodos das características para a discretización temporal [107, 10, 23, 43]. Para resolver as nonlinearidades de tipo obstáculo

propuxéronse, por exemplo, métodos de relaxación combinados con métodos multigrella en [37], o algoritmo de Uzawa en [107], e penalización implícita con paso de tempo variábel en [51].

Nós estudamos a resolución numérica de ecuacións en derivadas parciais que modelan o prezo de produtos de dous factores. Para elo usamos métodos de Lagrange-Galerkin clásicos e de orde máis elevada para a discretización espacial-temporal (vexa [99, 24, 25]) e dous algoritmos iterativos diferentes baseados na formulación mixta para discretizar o problema unilateral de tipo obstáculo, que foran introducidos, respectivamente, en [22] e [69]. A diferenza dos métodos de penalización, a regularización desenvolvida non introduce ningunha fonte de erro adicional. A combinación de métodos clásicos de Lagrange-Galerkin co algoritmo iterativo proposto en [22] foi empregada con éxito para calcular o prezo de produtos de dous factores, tales como os bonos convertibles e as opcións asiáticas, en [23].

Como exemplo de aplicación, consideramos o problema de valoración das opcións asiáticas. Estas opcións son derivados financeiros *dependentes do camiño* xa que os seus *payoffs* dependen nalgún xeito da media, durante un intervalo de tempo, dos prezos doutro produto financeiro chamado *activo subxacente*. Usando a metodoloxía de Black-Scholes e de Merton, o valor dunha opción asiática de tipo europeo resolve unha ecuación parabólica de convección-difusión-reacción linear bidimensional, fortemente dexenerada porque non hai difusión nunha das direccións espaciais. A primeira formulación deste modelo matemático en forma de edp con dúas variábeis espaciais aparece en [60]. Alén diso, cando se considera a mesma opción, pero de tipo americano, o problema de valoración tórnase nonlineal. Este último foi formulado coma un problema de complementariedade linear en [42]. En ambos os dous casos, o operador diferencial é fortemente de convección dominante así que o uso de métodos de Lagrange-Galerkin está xustificado. A regularidade da solución [13] permite que se apliquen con éxito esquemas dunha orde máis elevada. En [26, 27] preséntanse os métodos de Lagrange-Galerkin de orde dúas aplicados a opcións euroasiáticas. Estes estudos foron extendidos ao caso nonlineal en [28], onde se emprega o algoritmo iterativo de [22].

Os resultados resumidos ata agora foron organizados na presente tese como segue:

No **Capítulo 1** introducimos os conceptos e as notacións requiridas para modelar matematicamente o prezo das opcións financeiras dentro do marco de Black-Scholes [29] e de Merton [77]. Isto é, defínense os produtos derivados e, en particular as opcións financeiras, e explícase en que consiste o problema de valoración de opcións. A formulación matemática deste problema conséguese mediante algunhas hipóteses de modelado, coma o principio da ausencia de arbitraje, o modelado do activo subxacente mediante Browniano xeométrico, e de técnicas coma a da cobertura dinámica. Todos estes resultados xerais poden ser ampliados na bibliografía sobre o tema, como [111, 72, 71].

Depois, centramos o noso estudo nas opcións asiáticas, para as que se deduce o modelo matemático dos seus prezos en forma de edp, tanto no caso de opcións de estilo europeo (o exercicio só se permite ao final do contrato), como de estilo americano (o exercicio permítese en calquera momento antes de que finalice o contrato).

O problema de valor final para o prezo das opcións asiáticas de tipo europeo resulta:

Atopar $V = V(S, M, t)$, tal que

$$\begin{cases} \mathcal{L}_M[V] = 0 & \text{para } (S, M, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \\ V(S, M, T_f) = \Lambda(S, M) & \text{para } (S, M) \in \mathbb{R}_+ \times \mathbb{R}_+. \end{cases} \quad (\text{C.1})$$

onde

$$\mathcal{L}_M[V] := \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - d_0)S \frac{\partial V}{\partial S} + \frac{S - M}{t - T_i} \frac{\partial V}{\partial M} - rV,$$

$$\Lambda(S, M) = (M - K)_+,$$

e S, M e t representan, respectivamente, o valor do subxacente e da media e o tempo corrente, $[T_i, T_f]$ é o intervalo no que se calcula a media e o resto de parámetros son parámetros financeiros.

No caso americano, dedúcese o problema de complementariedade:

Atopar $V = V(S, M, t)$, tal que

$$\left. \begin{array}{l} \mathcal{L}_M[V](V - \Lambda) = 0 \\ \mathcal{L}_M[V] \leq 0 \\ V - \Lambda \geq 0 \end{array} \right\} \quad \text{en } \mathbb{R}_+ \times \mathbb{R}_+ \times (T_i, T_f), \quad (\text{C.2})$$

e

$$V(S, M, T_f) = \Lambda(S, M) \quad \text{para } (S, M) \in \mathbb{R}_+ \times \mathbb{R}_+. \quad (\text{C.3})$$

Tamén deducimos, baixo as mínimas hipóteses de modelado, propiedades dos valores das opcións asiáticas, como por exemplo a fórmula de paridade “put-call” (opción de compra-opción de venda) para o caso europeo, ou conxuntos do dominio no que é óptimo non exercer a opción, no caso americano.

No **Capítulo 2** formulamos nun marco funcional adecuado problemas xerais de valoración de opcións. Pensando na súa posterior resolución numérica, discútese o truncamento do dominio espacial, que é non limitado en xeral.

Xa particularizando ao problema de valoración de opcións asiáticas, (C.1) transfórmase, mediante un cambio de variábel adecuado, en:

Atopar $\phi = \phi(y_1, y_2, \tau)$ tal que

$$\left\{ \begin{array}{ll} \mathcal{L}_1[\phi] = 0 & \text{en } \mathbb{R}_+ \times \mathbb{R} \times (0, T), \\ \phi(y_1, y_2, 0) = \Lambda(y_1, y_2) & \text{en } \mathbb{R}_+ \times \mathbb{R}, \end{array} \right. \quad (\text{C.4})$$

onde

$$\mathcal{L}_1[\phi] := \frac{\partial \phi}{\partial \tau} - y_1^2 \frac{\partial^2 \phi}{\partial y_1^2} - y_1 \frac{\partial \phi}{\partial y_2} = \frac{\partial \phi}{\partial \tau} - \frac{\partial}{\partial y_1} \left(y_1^2 \frac{\partial \phi}{\partial y_1} \right) + 2y_1 \frac{\partial \phi}{\partial y_1} - y_1 \frac{\partial \phi}{\partial y_2},$$

e

$$\Lambda(y_1, y_2) = y_1^m \left(\frac{2y_2}{\sigma^2 T} - K \right)_+ \quad \text{para } (y_1, y_2) \in \Omega = \mathbb{R}_+ \times \mathbb{R}_+.$$

Este último problema foi estudado en [13], obténdose resultados de existencia, unicidade e regularidade de solución.

Tendo como obxectivo a súa resolución numérica, o problema nas variabeis orixinais (C.1) trúncase a distancia espacial finita e dase unha formulación feble do mesmo. Usando a teoría clásica de ecuacións de segunda orde “con forma non negativa das características” (vexa o libro de referencia [83]) discútese a necesidade das condicións de contorno en cada unha das fronteiras do dominio espacial truncado. Formúlase entón o problema aproximado:

Atopar $\phi = \phi(\mathbf{x}, t)$, para $(\mathbf{x}, t) \in \Omega^* \times (0, T : T_f - T_i)$, tal que

$$\begin{aligned} & \int_{\Omega^*} \frac{\partial \phi}{\partial \tau}(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^*} A_{11}(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) \frac{\partial \psi}{\partial x_1}(\mathbf{x}) d\mathbf{x} \\ & + \sum_{j=1}^2 \int_{\Omega^*} v_j(\mathbf{x}, \tau) \frac{\partial \phi}{\partial x_j}(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} + \int_{\Omega^*} l(\mathbf{x}, \tau) \phi(\mathbf{x}, \tau) \psi(\mathbf{x}) d\mathbf{x} \\ & = \int_{\Gamma_{1,+}^*} A_{11}(\mathbf{x}, \tau) g(\mathbf{x}, \tau) \psi(\mathbf{x}) dA_{\mathbf{x}}, \quad \forall \psi \in H^1(\Omega), \end{aligned} \quad (\text{C.5})$$

onde

$$\frac{\partial \phi}{\partial x_1}(\mathbf{x}, \tau) = g(\mathbf{x}, \tau) = \begin{cases} \frac{\tau}{T} e^{-r\tau} & se \quad 0 < x_2 < TK, \\ \frac{e^{-d_0\tau} - e^{-r\tau}}{T(r - d_0)} & se \quad x_2 > TK, \end{cases} \quad (C.6)$$

e

$$\mathbf{A}(\mathbf{x}, t) = \begin{pmatrix} \frac{1}{2}\sigma^2 x_1^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{v}(\mathbf{x}, t) = \begin{pmatrix} (\sigma^2 - r + d_0)x_1 \\ \frac{x_2 - x_1}{T - \tau} \end{pmatrix}, \quad l(\mathbf{x}, t) = r. \quad (C.7)$$

Ademais, para dous números suficientemente grandes fixados $x_1^* > 0$ e $x_2^* > 0$, o dominio espacial truncado defínese como $\Omega^* = (0, x_1^*) \times (0, x_2^*)$ e $\Gamma_{1,+}^*$, é a parte da fronteira de Ω^* caracterizada polo vector normal exterior $\mathbf{n} = (1, 0)$.

Tamén se formula no dominio truncado o problema de valoración de opcións asiáticas de tipo americano, xustificando as condicións de contorno.

No **Capítulo 3** propomos un método de características de segunda orde para a discretization temporal de ecuacións de convección-difusión-reacción lineares nun dominio limitado Ω de $\mathbb{R}^m$ ($m = 2, 3$) con fronteira $\Gamma = \Gamma_D \cup \Gamma_R$. Máis concretamente, o *problema forte* escríbese:

Atopar unha función $\phi : \Omega \times (0, T) \longrightarrow \mathbb{R}$ tal que

$$\frac{\partial \phi}{\partial t}(\mathbf{x}, t) - \text{Div}(\mathbf{A}(\mathbf{x})\nabla \phi(\mathbf{x}, t)) + \mathbf{v}(\mathbf{x}, t) \cdot \nabla \phi(\mathbf{x}, t) + l(\mathbf{x})\phi(\mathbf{x}, t) = f(\mathbf{x}, t), \quad (C.8)$$

para $(\mathbf{x}, t) \in \Omega \times (0, T)$, baixo as condicións de contorno

$$\phi(\mathbf{x}, t) = 0 \text{ en } \Gamma_D \times (0, T), \quad (C.9)$$

$$\alpha \phi(\mathbf{x}, t) + \mathbf{A}(\mathbf{x})\nabla \phi(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}) = g(\mathbf{x}, t) \text{ en } \Gamma_R \times (0, T), \quad (C.10)$$

e condición inicial

$$\phi(\mathbf{x}, 0) = \phi^0(\mathbf{x}) \text{ en } \Omega. \quad (C.11)$$

Primeiro estúdanse as curvas características asociadas ao campo de velocidades $\mathbf{v}$, que se supón, en xeral, nulo na fronteira e suficientemente regular. Por definición, estas curvas características resolven o problema de Cauchy

$$\begin{cases} \frac{\partial X_e}{\partial \tau}(\mathbf{x}, t; \tau) = \mathbf{v}(X_e(\mathbf{x}, t; \tau), \tau), \\ X_e(\mathbf{x}, t; t) = \mathbf{x}, \end{cases} \quad (C.12)$$

para $\mathbf{x} \in \Omega$ e $t, \tau \in (0, T)$. Danse resultados, algúns deles clásicos, que involucran a X_e , e aos gradientes de X_e e de $\mathbf{v}$, denotados, respectivamente, $\mathbf{F}_e$ e $\mathbf{L}$.

Resultados similares próbanse logo para algunhas fórmulas de aproximación das liñas características, como o esquema de Euler (X_E) e o de Runge-Kutta (X_{RK}).

Empregando unha fórmula tipo Green,¹ orixinal, dase unha formulación feble do problema (C.8)-(C.11) en termos das características para, de seguido, propoñer distintas variantes de semidiscretizacións en tempo de segunda orde. Por exemplo, empregando as características e

gradientes exactos, temos

$$\begin{aligned}
& \int_{\Omega} \frac{\phi^{n+1}(\mathbf{x}) - \phi^n(X_e^n(\mathbf{x}))}{\Delta t} \psi(\mathbf{x}) \, d\mathbf{x} + \theta \int_{\Omega} \mathbf{A}(\mathbf{x}) \, \text{grad } \phi^{n+1}(\mathbf{x}) \cdot \text{grad } \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \frac{1}{2} \int_{\Omega} (\mathbf{F}_e^n)^{-1}(\mathbf{x}) \mathbf{A}(X_e^n(\mathbf{x})) \, \text{grad } \phi^n(X_e^n(\mathbf{x})) \cdot \text{grad } \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \frac{1}{2} \int_{\Omega} \text{Div } (\mathbf{F}_e^n)^{-T}(\mathbf{x}) \cdot \mathbf{A}(X_e^n(\mathbf{x})) \, \text{grad } \phi^n(X_e^n(\mathbf{x})) \, \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \frac{1}{2} \int_{\Omega} l(\mathbf{x}) \phi^{n+1}(\mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x} + \frac{1}{2} \int_{\Omega} l(X_e^n(\mathbf{x})) \phi^n(X_e^n(\mathbf{x})) \, \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \int_{\Gamma_R} \alpha \left(\frac{1}{2} \phi^{n+1}(\mathbf{x}) + \frac{1}{2} \phi^n(\mathbf{x}) \det (\mathbf{F}_e^n)^{-1}(\mathbf{x}) \right) \psi(\mathbf{x}) \, dA_{\mathbf{x}} \\
& = \int_{\Omega} \left(\frac{1}{2} f^{n+1}(\mathbf{x}) + \frac{1}{2} f^n(X_e^n(\mathbf{x})) \right) \psi(\mathbf{x}) \, d\mathbf{x} \\
& + \int_{\Gamma_R} \left(\frac{1}{2} g^{n+1}(\mathbf{x}) + \frac{1}{2} g^n(\mathbf{x}) \det (\mathbf{F}_e^n)^{-1}(\mathbf{x}) \right) \psi(\mathbf{x}) \, dA_{\mathbf{x}}.
\end{aligned} \tag{C.13}$$

Para un esquema similar a (C.13), pero usando adecuadas aproximacións das liñas características e os seus gradientes próbase a estabilidade en norma $l^\infty((0, T); L^2(\Omega))$ baixo certas condicións de regularidade sobre os datos do problema. A segunda orde no erro de consistencia é establecida baixo condicións de regularidade máis restrictivas, non só dos datos, senón tamén da solución do problema forte e usando un resultado auxiliar de estabilidade para un método con segundo membro máis xeral.

Próbanse resultados de estabilidade e erro, baixo hipóteses similares, para o método das características clásico.

No **Capítulo 4** analízase un esquema completamente discretizado, que resulta de combinar a semidiscretización de segunda orde do capítulo precedente con elementos finitos que verifican a propiedade de interpolación: *Existe un operador de interpolación $\pi_h : C^0(\overline{\Omega}) \longrightarrow V_h^k$, con $k \geq 1$, satisfacendo*

$$\|\pi_h \psi - \psi\|_s \leq K h^{r-s} \|\psi\|_r \quad \forall \psi \in C^0(\overline{\Omega}) \cap H^r(\Omega) \quad 0 \leq s \leq r \leq k+1, \tag{C.14}$$

para unha constante positiva K independente de h .

Para o esquema totalmente discretizado próbase un erro de discretización da forma $O(\Delta t^2 + h^k)$ na norma $l^\infty((0, T); L^2(\Omega))$.

Recórdanse resultados similares para o esquema clásico de Lagrange-Galerkin.

Despois desenvólvese unha análise de Fourier no caso unidimensional e con coeficientes constantes para fórmulas de cuadratura de tipo Gauss-Lobatto de dous e tres puntos cando se usan, respectivamente, elementos finitos lineares e cuadráticos. A análise pode extenderse a dimensión m para elementos finitos sobre grellas cuadrangulares. No caso triangular e en dimensión dúas tamén se propoñen fórmulas de cuadratura.

Dous exemplos académicos ilustran e completan os resultados anteriores. Na práctica, o método de Crank-Nicholson dá mellores resultados que o clásico para as mesmas grellas, e a eficiencia é moito maior cando se combina con elementos cuadráticos que con elementos lineares.

No **Capítulo 5** presentamos unha metodoloxía para resolver numericamente modelos de dous factores con restriccións unilaterais consistente en:

- Método das características de orde alta para discretización temporal.
- Discretización espacial con elementos finitos lineares e cuadráticos.
- Resolución da nonlinearidade mediante algoritmos iterativos baseados na formulación mixta do problema.

Con todo, as técnicas anteriores serán só aplicadas nesta memoria a problemas de valoración de opcións asiáticas.

Primeiro discutimos a resolución numérica de problemas de valoración de opcións euroasiáticas usando unha versión lixeiramente diferente dos métodos de Lagrange-Galerkin introducidos nos capítulos 3 e 4. En concreto, as características calcúlanse exactamente e o campo de velocidades non se anula na fronteira. Isto último, fai que o dominio espacial sexa dependente do tempo, e que haxa que empregar novas aproximacións e técnicas numéricas. Ademais, discútese como facer a numeración dos nodos dos elementos finitos para optimizar a parte alxebrica dos algoritmos. Engáñense tamén exemplos académicos nos que se observa a orde de converxencia esperada, e exemplos “reais” no que se ilustra a converxencia comparando con outros resultados que aparecen na literatura. Á vista dos resultados, a utilización de esquemas de alta orde, coma o proposto nesta memoria, é necesaria para unha resolución numérica do problema de valoración de xeito eficiente.

En segundo lugar, propoñemos dous algoritmos iterativos que aparecen en [22] e en [69] e que se aplican á formulación mixta do problema nonlinear para a valoración de opcións asiáticas de estilo americano. Faise unha comparación a priori, que é completada con resultados numéricos, concluindo que o segundo algoritmo (un método de conxunto activo, que pode ser interpretado coma un método de Newton “semi-smooth”) é máis eficiente que o primeiro (un método iterativo de punto fixo baseado en resultados de [22]).

Rematamos mostrando algunhas gráficas representando valores das opcións, dos *multiplcadores de Lagrange* e das *gregas* (derivadas do valor das opcións con respecto a distintas variabeis e parámetros).

No **Apéndice A** incluímos algunhas definicións e resultados do cálculo estocástico, e, no **Apéndice B** dedúcese o modelo matemático para o valor das opcións máis básicas, as opcións vainilla. Ambos os dous apéndices poden axudar á comprensión da lectura do Capítulo 1.

Bibliography

- [1] R. A. ADAMS, *Sobolev Spaces*, Academic Press, Academic Press, New York, 1975.
- [2] H. AMANN, *Ordinary Differential Equations. An Introduction to Nonlinear Analysis*, vol. 13 of de Gruyter Studies in Mathematics, Walter de Gruyter, Berlin, 1990.
- [3] L. BACHELIER, *Théorie de la spéculation*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Éditions Jacques Gabay, Sceaux, 1995. Théorie mathématique du jeu. [Mathematical theory of games], Reprint of the 1900 original.
- [4] M. D. BAKER, E. SÜLI, AND A. F. WARE, *Stability and convergence of the spectral Lagrange-Galerkin method for mixed periodic/non-periodic convection-dominated diffusion problems*, IMA J. Numer. Anal., 19 (1999), pp. 637–663.
- [5] J. BARANGER, D. ESSLAOUI, AND A. MACHMOUM, *Error estimate for convection problem with characteristics method*, Numer. Algorithms, 21 (1999), pp. 49–56. Numerical methods for partial differential equations (Marrakech, 1998).
- [6] J. BARANGER AND A. MACHMOUM, *Une norme “naturelle” pour la méthode des caractéristiques en éléments finis discontinus: cas 1-D*, RAIRO Modél. Math. Anal. Numér., 30 (1996), pp. 549–574.
- [7] J. BARANGER AND A. MACHMOUM, *A “natural” norm for the method of characteristics using discontinuous finite elements: 2D and 3D case*, Math. Model. Numer. Anal., 33 (1999), pp. 1223–1240.
- [8] G. BARLES, C. DAHER, AND M. ROMANO, *Convergence of numerical schemes for parabolic equations arising in finance theory*, Math. Models Methods Appl. Sci., 5 (1995), pp. 125–143.
- [9] G. BARLES AND P. E. SOUGANIDIS, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Anal., 4 (1991), pp. 271–283.
- [10] G. BARONE-ADESI, A. BERMÚDEZ, AND J. HATGIOANNIDES, *Two-factor convertible bonds valuation using the method of characteristics/finite elements*, J. Econom. Dynam. Control, 27 (2003), pp. 1801–1831.
- [11] P. BARRAL, C. MORENO, P. QUINTELA, AND T. SÁNCHEZ, *A numerical algorithm for a Signorini problem associated with Maxwell-Norton materials by using generalized Newton’s methods*, Computer Methods in Applied Mechanics and Engineering, (to appear).
- [12] J. BARRAQUAND AND T. PUDET, *Pricing of American path-dependent contingent claims*, Math. Finance, 6 (1996), pp. 17–51.

- [13] E. BARUCCI, S. POLIDORO, AND V. VESPRI, *Some results on partial differential equations and Asian options*, Math. Models Methods Appl. Sci., 11 (2001), pp. 475–497.
- [14] M. BAUSE AND P. KNABNER, *Uniform error analysis for Lagrange-Galerkin approximations of convection-dominated problems*, SIAM J. Numer. Anal., 39 (2002), pp. 1954–1984.
- [15] A. BENSOUSSAN AND J.-L. LIONS, *Applications des inéquations variationnelles en contrôle stochastique*, Dunod, Paris, 1978. Méthodes Mathématiques de l'Informatique, No. 6.
- [16] M. BERCOVIER, O. PIRONNEAU, AND V. SASTRI, *Finite elements and characteristics for some parabolic-hyperbolic problems*, Appl. Math. Modelling, 7 (1983), pp. 89–96.
- [17] R. BERMEJO, *Analysis of a class of quasi-monotone and conservative semi-Lagrangian advection schemes*, Numer. Math., 87 (2001), pp. 597–623.
- [18] R. BERMEJO AND A. STANFORTH, *The conversion of semi-Lagrangian advection schemes to quasi-monotone schemes*, Mon. Wea. Rev., 120 (1992), pp. 2622–2632.
- [19] A. BERMÚDEZ, J. BUSCA, AND M. R. NOGUEIRAS, *Numerical calibration of local volatility surfaces*, in Information processing: Recent Mathematical Advances in Optimization and Control. Mathematical and Computational Sciences, Presses de l'Ecole des Mines de Paris, 2004, pp. 205–217.
- [20] A. BERMÚDEZ AND J. DURANY, *La méthode des caractéristiques pour les problèmes de convection-diffusion stationnaires*, Math. Model. Numer. Anal., 21 (1987), pp. 7–26.
- [21] ———, *Numerical solution of steady-state flow through a porous dam*, Comput. Methods Appl. Mech. Engrg., 68 (1988), pp. 55–65.
- [22] A. BERMÚDEZ AND C. MORENO, *Duality methods for solving variational inequalities*, Comput. Math. Appl., 7 (1981), pp. 43–58.
- [23] A. BERMÚDEZ AND M. R. NOGUEIRAS, *Numerical solution of two-factor models for valuation of financial derivatives*, Math. Models Methods Appl. Sci., 14 (2004), pp. 295–327.
- [24] A. BERMÚDEZ, M. R. NOGUEIRAS, AND C. VÁZQUEZ, *Numerical analysis of degenerated convection-diffusion-reaction problems with higher order characteristics/finite elements. Part I: Time discretization.* (Submitted to SIAM J. Numer. Anal.).
- [25] ———, *Numerical analysis of degenerated convection-diffusion-reaction problems with higher order characteristics/finite elements. Part II: Fully discretized scheme and quadrature formulas.* (Submitted to SIAM J. Numer. Anal.).
- [26] ———, *Métodos de características/elementos finitos de orden dos para la valoración de opciones asiáticas. Actas en CD-Rom del XVIII CEDYA/VIII Congreso de Matemática Aplicada*, 2003.
- [27] ———, *Numerical methods for pricing fixed strike Eurasian options with continuous arithmetic averaging*, in CD-ROM Proceedings of Stochastics and Finance, Lisbon, 2004.
- [28] ———, *Numerical solution of variational inequalities for pricing Asian options by higher order Lagrange-Galerkin methods*, Appl. Numer. Math., (to appear).
- [29] F. BLACK AND M. SCHOLES, *The pricing of options and corporate liabilities*, Journal of Political Economy, 81 (1973), pp. 637–654.

-
- [30] K. BOUKIR, Y. MADAY, AND B. MÉTIVET, *A high order characteristics method for the incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg., 116 (1994), pp. 211–218. ICOSAHOM'92 (Montpellier, 1992).
 - [31] K. BOUKIR, Y. MADAY, B. MÉTIVET, AND E. RAZAFINDRAKOTO, *A high-order characteristics/finite element method for the incompressible Navier-Stokes equations*, Internat. J. Numer. Methods Fluids, 25 (1997), pp. 1421–1454.
 - [32] P. P. BOYLE, *Options: A monte carlo approach*, J. Financial Economics, 4 (1990), pp. 323–339.
 - [33] H. BREZIS, *Analyse fonctionnelle*, Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris, 1983.
 - [34] Z. CHEN, R. E. EWING, Q. JIANG, AND A. M. SPAGNUOLO, *Error analysis for characteristics-based methods for degenerate parabolic problems*, SIAM J. Numer. Anal., 40 (2002), pp. 1491–1515 (electronic).
 - [35] P. G. CIARLET, *The finite element method for elliptic problems*, North-Holland Publishing Co., Amsterdam, 1978. Studies in Mathematics and its Applications, Vol. 4.
 - [36] G. CITTI, A. PASCUCCHI, AND S. POLIDORO, *On the regularity of solutions to a nonlinear ultraparabolic equation arising in mathematical finance*, Differential Integral Equations, 14 (2001), pp. 701–738.
 - [37] N. CLARKE AND K. PARROT, *Multigrid american option pricing with stochastic volatility*, App. Math. Finance, 6 (1994), pp. 177–195. ICOSAHOM'92 (Montpellier, 1992).
 - [38] L. CLELOW AND C. STRICKLAND, *Implementing derivative models*, Jonh Wiley & Sons, Nonsei, 1998. Méthodes Mathématiques de l'Informatique, No. 6.
 - [39] R. CONT AND P. TANKOV, *Financial modelling with jump processes*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.
 - [40] M. G. CRANDALL, H. ISHII, AND P.-L. LIONS, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), 27 (1992), pp. 1–67.
 - [41] J. N. DEWYNNE, S. D. HOWISON, I. RUPF, AND P. WILMOTT, *Some mathematical results in the pricing of American options*, European J. Appl. Math., 4 (1993), pp. 381–398.
 - [42] J. N. DEWYNNE AND P. WILMOTT, *Asian options as linear complementarity problems: analysis and finite difference solutions*, Advances in Futures and Options Research, 8 (1995), pp. 145–173.
 - [43] Y. D'HALLUIN, P. A. FORSYTH, AND G. LABAHN, *A semi-Lagrangian approach for American Asian options under jump diffusion*, to appear in SIAM J. Sci. Comp., (2005).
 - [44] J. DOUGLAS, JR., C.-S. HUANG, AND F. PEREIRA, *The modified method of characteristics with adjusted advection*, Numer. Math., 83 (1999), pp. 353–369.
 - [45] J. DOUGLAS, JR. AND T. F. RUSSELL, *Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures*, SIAM J. Numer. Anal., 19 (1982), pp. 871–885.

- [46] R. E. EWING AND T. F. RUSSEL, *Multistep galerkin methods along characteristics for convection-diffusion problems*, IMACS Publications, (1981), pp. 28–36. Advances in Computer Methods for Partial Differential Equations IV, R. Vichtneveski and R.S. Stepleman Eds.
- [47] R. E. EWING AND H. WANG, *A summary of numerical methods for time-dependent advection-dominated partial differential equations*, J. Comput. Appl. Math., 128 (2001), pp. 423–445. Numerical analysis 2000, Vol. VII, Partial differential equations.
- [48] M. FALCONE AND R. FERRETTI, *Convergence analysis for a class of high-order semi-Lagrangian advection schemes*, SIAM J. Numer. Anal., 35 (1998), pp. 909–940.
- [49] P. FAURRE, *Analyse numérique*, École Polytechnique, Palaiseau, 1988. Notes d’optimisation.
- [50] W. H. FLEMING AND H. M. SONER, *Controlled Markov processes and viscosity solutions*, vol. 25 of Applications of Mathematics (New York), Springer-Verlag, New York, 1993.
- [51] P. A. FORSYTH AND K. R. VETZAL, *Quadratic convergence for valuing American options using a penalty method*, SIAM J. Sci. Comput., 23 (2002), pp. 2095–2122.
- [52] G. FOURESTEY, *Stabilité des méthodes de Lagrange-Galerkin du premier et du second ordre*, tech. report, INRIA, Rapport de recherche, 2002.
- [53] G. FOURESTEY AND S. PIPERNO, *A second-order time-accurate ALE Lagrange-Galerkin method applied to wind engineering and control bridge profiles*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 4117–4137.
- [54] A. FRIEDMAN, *Partial differential equations of parabolic type*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.
- [55] ———, *Stochastic differential equations and applications. Vol. 1*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. Probability and Mathematical Statistics, Vol. 28.
- [56] ———, *Stochastic differential equations and applications. Vol. 2*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. Probability and Mathematical Statistics, Vol. 28.
- [57] R. GLOWINSKI, J.-L. LIONS, AND R. TRÉMOLIÈRES, *Numerical analysis of variational inequalities*, vol. 8 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam, 1981. Translated from the French.
- [58] M. GURTIN, *An Introduction to Continuum Mechanics*, vol. 158 of Mathematics in Science and Engineering, Academic Press, San Diego, 1981.
- [59] L. HÖRMANDER, *Hypoelliptic second order differential equations*, Acta Math., 119 (1967), pp. 147–171.
- [60] J. E. INGERSOLL, *Theory of Financial Decision Making*, Mathematics in Science and Engineering, Rowman & Littlefield Publishers, Inc., Maryland, 1987.
- [61] K. ITO AND K. KUNISCH, *Semi-smooth Newton methods for state-constrained optimal control problems*, Systems Control Lett., 50 (2003), pp. 221–228.

-
- [62] P. JAILLET, D. LAMBERTON, AND B. LAPEYRE, *Variational inequalities and the pricing of American options*, Acta Appl. Math., 21 (1990), pp. 263–289.
 - [63] J. KAČUR, *Solution of degenerate convection-diffusion problems by the method of characteristics*, SIAM J. Numer. Anal., 39 (2001), pp. 858–879 (electronic).
 - [64] J. KAČUR, *Solution of nonlinear and degenerate convection-diffusion problems*, in Proceedings of the Third World Congress of Nonlinear Analysts, Part 1 (Catania, 2000), vol. 47, 2001, pp. 123–134.
 - [65] J. KACUR AND R. VAN KEER, *Solution of degenerate parabolic variational inequalities with convection*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 417–431.
 - [66] R. KANGRO AND R. NICOLAIDES, *Far field boundary conditions for Black-Scholes equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1357–1368.
 - [67] I. KARATZAS AND S. E. SHREVE, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.
 - [68] ———, *Methods of mathematical finance*, vol. 39 of Applications of Mathematics (New York), Springer-Verlag, New York, 1998.
 - [69] T. KÄRKKÄINEN, K. KUNISCH, AND P. TARVAINEN, *Augmented Lagrangian active set methods for obstacle problems*, J. Optim. Theory Appl., 119 (2003), pp. 499–533.
 - [70] A. KEMMA AND T. VORST, *A pricing method for options based on an average asset values*, J. Banking & Finance, 14 (1990), pp. 113–129.
 - [71] Y. K. KWOK, *Mathematical models of financial derivatives*, Springer Finance, Springer-Verlag Singapore, Singapore, second ed., 1998.
 - [72] D. LAMBERTON AND B. LAPEYRE, *Introduction au calcul stochastique appliqué à la finance*, Ellipses Édition Marketing, Paris, second ed., 1997.
 - [73] E. LANCONELLI, A. PASCUCCI, AND S. POLIDORO, *Linear and nonlinear ultraparabolic equations of Kolmogorov type arising in diffusion theory and in finance*, in Nonlinear problems in mathematical physics and related topics, II, vol. 2 of Int. Math. Ser. (N. Y.), Kluwer/Plenum, New York, 2002, pp. 243–265.
 - [74] B. LAPEYRE, A. SULEM, AND D. TALAY, *Understanding Numerical Analysis for Option Pricing*, Cambridge University Press, London, 2004.
 - [75] J.-L. LIONS, *Équations différentielles opérationnelles et problèmes aux limites*, Die Grundlehren der mathematischen Wissenschaften, Bd. 111, Springer-Verlag, Berlin, 1961.
 - [76] M. D. MARCOZZI, *On the valuation of Asian options by variational methods*, SIAM J. Sci. Comput., 24 (2003), pp. 1124–1140.
 - [77] R. C. MERTON, *Theory of rational option pricing*, J. of Economics & Management Sci., 4 (1973), pp. 141–183.
 - [78] G. H. MEYER, *On pricing American and Asian options with PDE methods*, Acta Math. Univ. Comenian. (N.S.), 70 (2000), pp. 153–165.

- [79] T. MIKOSCH, *Elementary stochastic calculus—with finance in view*, vol. 6 of Advanced Series on Statistical Science & Applied Probability, World Scientific Publishing Co. Inc., River Edge, NJ, 1998.
- [80] C. MORENO AND A. SAMARTIN, *Numerical methods in several factors option pricing*. (Working paper).
- [81] K. W. MORTON, *Numerical solution of convection-diffusion problems*, vol. 12 of Applied Mathematics and Mathematical Computation, Chapman & Hall, London, 1996.
- [82] K. W. MORTON, A. PRIESTLEY, AND E. SÜLI, *Stability of the Lagrange-Galerkin method with nonexact integration*, Math. Model. Numer. Anal., 22 (1988), pp. 625–653.
- [83] O. A. OLEĬNIK AND E. V. RADKEVIČ, *Second order equations with nonnegative characteristic form*, Plenum Press, New York, 1973. Translated from the Russian by Paul C. Fife.
- [84] C. W. OOSTERLEE, J. C. FRISCH, AND F. J. GASPAR, *TVD, WENO and blended BDF discretizations for Asian options*, Comput. Vis. Sci., 6 (2004), pp. 131–138.
- [85] C. PARÉS, M. CASTRO, AND J. MACÍAS, *On the convergence of the Bermúdez-Moreno algorithm with constant parameters*, Numer. Math., 92 (2002), pp. 113–128.
- [86] K. PARROTT AND N. CLARKE, *A parallel solution of early exercise asian options with stochastic volatility*, in Proceedings of the 11th Domain Decomposition Conference, 1998, Greenwich.
- [87] A. PAZY, *Semigroups of linear operators and applications to partial differential equations*, vol. 44 of Applied Mathematical Sciences, Springer-Verlag, New York, 1983.
- [88] O. PIRONNEAU, *On the transport-diffusion algorithm and its applications to the Navier-Stokes equations*, Numer. Math., 38 (1981/82), pp. 309–332.
- [89] ———, *Finite Element Methods for Fluids*, John Wiley & Sons, Paris, 1989.
- [90] O. PIRONNEAU AND F. HECHT, *Mesh adaption for the Black and Scholes equations*, East-West J. Numer. Math., 8 (2000), pp. 25–35.
- [91] O. PIRONNEAU, J. LIOU, AND T. TEZDUYAR, *Characteristic-Galerkin and Galerkin/least-squares space-time formulations for the advection-diffusion equations with time-dependent domains*, Comput. Methods Appl. Mech. Engrg., 100 (1992), pp. 117–141.
- [92] S. POLIDORO AND M. A. RAGUSA, *Hölder regularity for solutions of ultraparabolic equations in divergence form*, Potential Anal., 14 (2001), pp. 341–350.
- [93] D. M. POOLEY, K. R. VETZAL, AND P. A. FORSYTH, *Convergence remedies for non-smooth payoffs in option pricing*, J. Comput. Finance, 6 (2003), pp. 1–21.
- [94] A. PRIESTLEY, *Exact projections and the Lagrange-Galerkin method: a realistic alternative to quadrature*, J. Comput. Phys., 112 (1994), pp. 316–333.
- [95] A. QUARTERONI AND A. VALLI, *Numerical approximation of partial differential equations*, vol. 23 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1994.

-
- [96] M. A. RAGUSA, *On weak solutions of ultraparabolic equations*, in Proceedings of the Third World Congress of Nonlinear Analysts, Part 1 (Catania, 2000), vol. 47, 2001, pp. 503–511.
 - [97] L. C. G. ROGERS AND Z. SHI, *The value of an Asian option*, J. Appl. Probab., 32 (1995), pp. 1077–1088.
 - [98] R. Y. RUBINSTEIN, *Simulation and the Monte Carlo method*, John Wiley & Sons Inc., New York, 1981. Wiley Series in Probability and Mathematical Statistics.
 - [99] H. RUI AND M. TABATA, *A second order characteristic finite element scheme for convection-diffusion problems*, Numer. Math., 92 (2002), pp. 161–177.
 - [100] S. SANFELICI, *Galerkin infinite element approximation for barrier options and options with discontinuous payoff*, Decisions in Economics and Finance.
 - [101] J. C. STRIKWERDA, *Finite Difference Schemes and Partial Differential Equations*, Wadsworth & Brooks Cole, Pacific Grove, California, 1989.
 - [102] E. SÜLI, *Convergence and nonlinear stability of the Lagrange-Galerkin method for the Navier-Stokes equations*, Numer. Math., 53 (1988), pp. 459–483.
 - [103] E. SÜLI, *Stability and convergence of the Lagrange-Galerkin method with nonexact integration*, in The mathematics of finite elements and applications, VI (Uxbridge, 1987), Academic Press, London, 1988, pp. 435–442.
 - [104] E. SÜLI AND A. WARE, *A spectral method of characteristics for hyperbolic problems*, SIAM J. Numer. Anal., 28 (1991), pp. 423–445.
 - [105] M. TABATA AND S. FUJIMA, *Robustness of a characteristic finite element scheme of second order in time increment*, tech. report, MHF Preprint Series, 2004.
 - [106] R. S. VARGA, *Matrix iterative analysis*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1962.
 - [107] C. VÁZQUEZ, *An upwind numerical approach for an American and European option pricing model*, Appl. Math. Comput., 97 (1998), pp. 273–286.
 - [108] J. VECER, *A new PDE approach for pricing arithmetic average Asian options*, J. Comp. Finance, (2001), pp. 105–113.
 - [109] J. VEČER AND M. XU, *Pricing Asian options in a semimartingale model*, Quant. Finance, 4 (2004), pp. 170–175.
 - [110] P. WILMOTT, *Derivatives. The Theory and Practice of Financial Engineering*, John Wiley & Sons, Chichester, 1999.
 - [111] P. WILMOTT, J. DEWYNNE, AND S. HOWISON, *Option Pricing. Mathematical models and computation*, Oxford Financial Press, Oxford, 1993.
 - [112] H. WINDCLIFF, P. FORSYTH, AND K. VETZAL, *Analysis of the stability of the linear boundary condition for the Black-Scholes equation*, J. Comp. Finance, 8 (2004).
 - [113] D. XIU AND G. E. KARNIADAKIS, *A semi-Lagrangian high-order method for Navier-Stokes equations*, J. Comput. Phys., 172 (2001), pp. 658–684.
 - [114] P. ZHANG, *Exotic options*, World Scientific, Singapore, 1998.

- [115] O. C. ZIENKIEWICZ AND R. L. TAYLOR, *The finite element method. Vol. 1*, Butterworth-Heinemann, Oxford, fifth ed., 2000. The basis.
- [116] R. ZVAN, P. FORSYTH, AND K. VETZAL, *Robust numerical methods for PDE of Asian options*, J. Comput. Finance, 1 (1998), pp. 39–78.