Università Commerciale Luigi Bocconi
Dottorato di Ricerca in Statistica: XX Ciclo
Coordinatore: Chiar.mo Prof. Pietro Muliere

Perturbation Methods for The Approximation of The Transition Density With Applications in Finance

Alberto Antonio Mietto
Matricola: P1003672

Relatore: Prof. Francesco Corielli

31 Gennaio 2008

Contents

Abstract

In this thesis we describe, design and analyse some perturbation techniques for the analytical approximation of solutions of Cauchy problems for linear parabolic partial differential equations (PDEs), with applications to continuous-time Finance.

In the context of continuous-time Finance parabolic PDEs arise in at least two fundamental problems: i) The computation of the transition density when the statistical model is described by a diffusion process solution of a stochastic differential equation (SDE). In particular, the transition density corresponds to the fundamental solution associated to each of the two Kolmogorov equations. ii) The risk-neutral valuation of the price of many derivative securities, amongst which the so-called T-contingent claims. Actually, the option price can also be obtained as the conditional expectation of the payoff of the contract (under the risk-neutral measure). Therefore, when the transition density associated with the solution of the SDE-type model is known, in principle the valuation problem can always be solved at least numerically, e.g., using Monte Carlo methods.

However, in practice, the transition density is known in closed-form only for a few diffusions. A way around this problem is to look for numerical solutions of the PDEs considered. Another approach, typically exploited in many branches of applied Mathematics, but that has only recently found application in the financial literature, is to look for a perturbation of a solvable, auxiliary problem which we consider ”close” to a true statistical, but unsolvable model. A perturbation is an expansion in series of the solution of a complex and unsolvable PDE, obtained with a suitable modification (i.e., perturbation) of the considered auxiliary model. The auxiliary model can be interpreted as the leading term of the expansion, while the truncation of the series gives an approximation. But the convergence of the perturbation series is a delicate issue and is often excluded.

In particular, we consider three methods, recently proposed for the approximation of option prices, and study their theoretical properties, their generalizations and also possible applications to the parametric estimation of discretely observed diffusion process, highlighting the relation of the proposed methods with well established techniques.

Acknowledgments

I wish to thank my advisor, Professor F. Corielli, and Prof. A. Mele for their discussions, scientific support and constant encouragement. I would like to thank all the professors of the Ph.D., in particular Prof. D.M. Cifarelli, Prof. S. Fortini and Prof. P. Muliere for their critical and helpful comments. I gratefully acknowledge Prof. R. Anderson and Prof. D. Webb for the opportunity to join for a whole year the prestigious Financial Markets Group. My special thanks to my colleagues and also to all the members of the Financial Markets Group not only for the good atmosphere, but also for the critical discussions. Finally, I wish to thank very much my family and, in particular, Federica.

Chapter 1

Introduction: Motivation and Results

This thesis analyses, designs and compares some classes of approximation techniques for the analytical approximation to the solution of parabolic Cauchy problems encountered in many important financial problems. These methods can be grouped under the name "perturbation methods".

Perturbation techniques have found proficient applications in virtually all branches of applied Mathematics (see [Bel64], [Hin91], [Nay73]). It is a fact that even easy, or at least apparently uncomplicated, equations need not possess easy solutions. Mathematical models to describe physical phenomenons are often written in terms of complicated differential equations and so, typically, to extract meaning from those analytical formulations, one must resort to some kind of approximation. A good approximation could be defined as an approximation with an error which is understood and controllable. There are two methods to obtain such precise approximations: analytical or numerical methods. An analytical approximation is useful to gain qualitative and quantitative insight into the problem. However, analytical and numerical methods are not in strict competition because it is often useful to use a combination of both.

Applied perturbation theory is a way to devise precise analytical approximations: given a difficult and not readily solvable problem, the aim is to reduce its study to that of the solution of a more tractable problem. Therefore, a perturbation approach is natural when the equation to be solved is "close" to a simpler equation that can be solved exactly, called the auxiliary model. This auxiliary model is "perturbed" in the sense that the full problem is sometimes interpretable as a disturbed version of the simple problem, but this is just a possible interpretation and the word perturbation is used in many other contexts in the literature. However, the basic idea in any perturbation method is the application of some systematic transformations to get a series expansion relating the solvable to the unsolvable problem. The truncation of the series, typically the first few terms for a good perturbation, should give an accurate approximation.

As already stressed, there is no unique perturbation method, but just a vast set of techniques, each specific for a particular problem. Clearly, a key common issue is given by the convergence of the perturbation series. Depending on the specific problem, it is sometimes possible to prove that the perturbation expansion converges, and in this case one speaks of regular perturbations. The real advantage of regular methods is that the perturbation series defines a genuine solution of the full problem, but, actually, most of the perturbation series are divergent. Although divergent, provided the perturbation series possesses a minimal asymptotic property, these expansion can be very useful (e.g., see [Geo95]). In fact, when one uses an asymptotic (i.e., non-regular) perturbation, typically no other analytical approximations are practically available and an ad-hoc justification must be given.

Perturbation are also studied in pure Mathematics (e.g., see [Kat80]), where there is a special interest in regular perturbations. To be precise, the interest is in some properties of a complex problem (rather than on the solution itself), such as the existence of a solution for a differential equation, and the problem is often solved by means of a convergent series expansion in terms of a simpler, well understood problem. Therefore, when we could supply also a detailed study of the truncation error, these theoretical results can be used as regular perturbations. Even though the conditions are typically restrictive, the convergence of the perturbation series is a rare and appealing feature for complex differential problems, so that we are particularly interested in these results.

The specific problem that we face is the approximation of the solution of linear parabolic Cauchy problems. The analytical solution of parabolic partial differential equations (PDEs) is generally formidable and, departing from a few special cases, one necessarily has to rely on some approximations. In fact, there are a lot of perturbation techniques, often extremely powerful, to cope with Cauchy problems. In this literature (see [Geo95], [KeCo96]), the approach is typically on a case-by-case analysis, being studied very specific and complex equations. As a consequence, asymptotic perturbations are the standard. On the other hand, there are many numerical methods, either deterministic (see [RiMo67], [Smi85], [QuVa97]) or Monte Carlo based (see, e.g., [Gla04], [JoPo04]), which provide accurate approximations.

The object of the thesis is the study of certain particular perturbation techniques for the approximation of the fundamental solution of parabolic PDEs. We consider three techniques, and their extensions, that we believe are particular useful in continuous-time Finance. In this context, Cauchy problems arise when the statistical model is given by a diffusion process, for which the transition density is given by the fundamental solution of a PDE, as prescribed by any of the two Kolmogorov equations. As a matter of fact, the transition density is known in closed-form only for a rather small class of models. Not surprisingly, the modeling in Finance has been mostly restricted to these solvable models and a few of their extensions (e.g., [BlSc73], [CIR85], [Hes93]), due to the manifest advantage to have a closed-form solution in terms of computational time and understanding of the phenomenon described.

Those solvable models, of course, cannot have good statistical properties, and are often an oversimplification. So we propose to use quite simple perturbation techniques, possibly regular and readily interpretable, to extend the class of those standard diffusion models used. This approach, tuned on the most used financial models, allow the preservation of the fundamental feature of analytic tractability and fast evaluation of the chosen auxiliary model. The alternative is to define a much more complex and realistic model, which could be approximated only with either powerful ad hoc asymptotic techniques or numerical approximation, though with considerable time and computational costs.

Our approach has natural applications in at least two relevant financial problems where one faces the problem of solve parabolic PDEs, but very often ends up with the inconsistent choice of using a solvable, but clearly misspecified model. The first problem is the pricing and hedging of derivative securities. The second is the estimation of the parameters of a discretely observed diffusion. In both cases the knowledge of the transition density would, in principle, solve the problem. We briefly discuss these two issues and the related literature below.

1.1 Pricing of Derivative Securities

The techniques considered in this thesis can be used for the approximation of option prices. An option contract is a derivative security, i.e., an investment instrument whose value depends on some other financial quantity, such as stock prices, interest rates, exchange rates, etc., called underlying. When the value of the underlying is given by the solution of a stochastic differential equation (SDE), typically a diffusion, under some fundamental conditions, the option price of any European-style derivative (also called T-claim) can be computed as the solution of a Cauchy problem, as shown in the celebrated models introduced by [BlSc73], [Bla76], [Vas77], [CIR85] (see also [Bjö03] and [Duf01]).

Each valuation model is characterized by a particular choice of the underlying diffusion and a payoff function defining the nature of the derivative. E.g., the transition density of the diffusions used in [BlSc73], [Vas77] and [CIR85] are, respectively, lognormal, Gaussian and gamma. When, as in these famous examples, the transition density is known in closed form, it turns out that the specific type of payoff function does not affect the complexity of the problem, since any (European) option price can be obtained integrating its payoff w.r.t. the transition density. In the language of Finance (see [Duf01]), this fact is described by saying that the set of Arrow-Debreu securities, each paying one unit of the account in a specific state of nature and nothing otherwise, are the building blocks of Asset pricing theory. The price of an Arrow-Debreu security, called state-price density, is given by the transition density of the Markov process (Arrow-Debreu securities satisfy the pricing PDE with payoff given by the terminal condition of the Kolmogorov equations for the transition density, i.e., the delta function). Modulo this remark, we can directly consider the perturbation of state-price densities.

Even when the Markov process for the underlying is given by a diffusion process, the state-price density is known in closed-form pretty much only in the famous exampled cited above. More generally, the option price is analytically available only for some linear and affine SDEs (see [Duf01] and [DFS03]) and very few other examples (see [Epp00], [Hul06]). In any other case, one must resort to some kind of approximation of the state-price density. The standard approach in Finance is twofold. On one hand, sometimes it is defined a possibly very complex SDE-type model that is solved numerically (see [Duf06] and [Gla04]). Otherwise, it is considered an oversimplified, but analytically tractable model (such as [BlSc73] or [Bla76]), even though the statistical properties (i.e., the fitting of the observed data) of solvable models are generally quite unsatisfactory, proving that solvable models are often very rough approximations. We briefly explain these attitudes.

The market practice is to use tractable models in the daily analysis. In fact, T-claims are often liquid traded assets, so that their prices are observed. The pricing of the derivative is necessary for the financial institutions that issue these contracts in order to assess and hedge the so-called market risk, i.e., the exposure to market fluctuations. Remarkably, Markov Valuation theory explains also how to hedge those risks (see [Bjö03]): market risk can be managed computing some sensitivities (or Greeks) of the option price, i.e., derivatives w.r.t. parameters of the derivative price. These can of course be proficiently computed numerically (since they will solve suitable PDEs), but with considerable computational and time costs, that cannot be borne when a real-time position is managed.

A special role in this practice is played by the Black-Scholes (BS) model [BlSc73]. The BS model has been the standard market model for a long time and it is still important for quoted prices. Most of the intuition about derivative prices dynamics has been gained by means of the closed-form formulas from this model. And it has also proven to be quite robust for practical purposes (see [Hul06], [Reb04]). The practitioner's way out from the poor statistical properties of the BS model is represented by inconsistent but effective procedure, such parameters and initial conditions recalibration. This market practice implies the necessity to offset a second important kind of risk[1]: model risk, i.e., the consequences of model misspecification. In this case numerical methods are widely applied, despite the intrinsic difficulty in the comparative assessment of non-tractable models. A possible alternative, or better a complement of the numerical approach, is represented by some analytical approximations.

While well-know and extensively used in other fields of applied Mathematics, perturbation techniques have had only a limited used in Finance, which appears surprising (see the discussion in [CaCo06], [Cor06b]), as the problem faced are formally quite similar. In fact, some singular perturbations has been applied to specific problems, with both modeling and computational ends,

[1]A financial institution has to face many risks, such as market risk, credit or counterpart risk, operational risk. Model risk, in fact, surfaces in all these categories.

and are effectively used in the market practice, such as [HKLW02]. Other possible examples are [FPSS03], [WDAN05] and some methods discussed in [FPS00]. Our approach is, instead, to propose simple perturbations of the basic solvable models, with the broader possible applicability. The goal is to retain the tractability of simple models, but improving the empirical performances. Therefore, their application to the management of the model error is straightforward. In this direction, it makes sense that, throughout the thesis, we mainly discuss the perturbation of the BS example.

1.2 Parametric Estimation For Diffusions

Diffusion processes are the standard statistical models in Finance, not only because of the elegant and effective Valuation theory[2] briefly discussed in the previous section, but also because they are consistent with the fact that financial data is observed most often at irregular times, and sometimes even randomly (see, e.g., [AïMy03] and [DuGl01]). From this, any discrete time formulation of the problem would contain a clear element of arbitrariness. However, essentially all financial data is recorded at discrete time points, and the continuous time modeling complicates the statistical analysis considerably. The extension to other Markov processes is possible (e.g., see [DuGl01]), and sometimes more natural, but much more complicated and so not considered in the thesis. Actually, we are interested in the perturbation techniques themselves and, therefore, we choose the simpler setting. This remark about the natural possibility of the extension to general Markov processes applies throughout the thesis.

We consider the following parametric model. Let $S(t;\theta)$ be the diffusion solution of the SDE $dS(t;\theta) = \mu(x,t;\theta)dt + \sigma(x,t;\theta)dW(t)$, where $\theta \in \Theta \subset \mathbb{R}^m$ is a vector of parameters (which often has to be taken in the form $\theta = (\theta_1, \theta_2)$, where θ_1 enters μ only and θ_2 enters σ only) and μ, σ are known functions. For estimation ends the existence of a weak solution for any $\theta \in \Theta$ is often enough (and maybe a more natural concept, see [Aït02] for a discussion), but we directly suppose that the existence of a strong solution; this requires asking for more regularity for the coefficients of $S(t;\theta)$. We always suppose that there is a true parameter value $\theta_0 \in \Theta$, which in turn is asked to be a convex and compact set.

Suppose that we have $N+1$ discrete (exactly measured) observations $S(t_i;\theta)$, $i = 0, \ldots, N$, $0 = t_0 < t_1 < \cdots < t_N = T$, and let $p(x,t;y,s;\theta)$, $0 \le t \le s \le T$, be the transition density associated to $S(t;\theta)$, for θ fixed. We recall the technical points that, in this context, the conditional expectations are taken under θ_0 and the conditioning is w.r.t. the σ-field generated by the observations. Given the Markov structure of $S(t;\theta)$, the likelihood function is given by

$$L_N(\theta) = \prod_{i=1}^{N} p(x_{i-1}, t_{i-1}; x_i, t_i; \theta), \tag{1.1}$$

where $x_i = S(t_i;\theta)$ and the first observation $S(t_0) = S(0)$ is typically discarded, as natural if $S(0)$ is non-random, and anyway an innocuous assumption if N is large. The maximum likelihood estimator (MLE) is found equating the score function to 0, i.e., $\widehat{\theta}_{MLE} : \nabla_\theta \ln L_N(\theta) = 0$. The existence of $\widehat{\theta}_{MLE}$ is guarantee by the assumptions on Θ. However, is $\widehat{\theta}_{MLE}$ a feasible solution in this situation?

First notice that the considered diffusion $S(t;\theta)$ is not necessarily time-homogeneous. When this is true, the statistical analysis, per se already complex, becomes extremely involved, when meaningful. So we distinguish two fundamental cases: A) $S(t;\theta)$ is ergodic (hence necessarily

[2] This is of course the main reason of the success of diffusion models, after [BlSc73], as can be explained with the following citation (see the introduction of [ReSi80], Vol.1) about a widespread issue: "When a successful mathematical model is created for a physical phenomenon, that is, a model which can be used for accurate computations and predictions, the mathematical structure itself provides a new way of thinking about a phenomenon."

time-homogeneous), with invariant measure $q(x;\theta)$, for all $\theta \in \Theta$[3]. Then, under weak conditions, $\widehat{\theta}_{MLE}$ has the usual optimal properties (see [DaFl86]). B) $S(t;\theta)$ is non-ergodic, possibly time-inhomogeneous (as most financial series are; furthermore it is very difficult that a multivariate diffusion is ergodic). Then, under strong identification assumption, $\widehat{\theta}_{MLE}$ still has the optimal properties, despite to be suitably interpreted (see [BaSo94]). In particular, the ratio between the observed and expected (now time-dependent) information matrices, as $T = T_N \to \infty$, $N \to \infty$, is random, and the usual LAN property has to be replaced with the so called LAMN property (see [BaSc83], [Pra99a] and [Pra99b], for the details as well as the identification assumptions for specific examples). When the conditions are violated, (w.l.o.g., we can suppose here that $t_i - t_{i-1} = \Delta$) for Δ fixed, it could be that even $\widehat{\theta}_{MLE}$ is not consistent, which is the key property in Finance since the observed time-series are easily quite long.

The point, however, is that the transition density of $S(t_i;\theta)$, as already discussed, is known only in a few cases (by the way, some of them ruled out by the identification assumptions, e.g., see the discussion in [Jac06]). There are essentially two alternatives to $\widehat{\theta}_{MLE}$: either approximate (1.1) or look for a completely different estimator (see the excellent surveys [AHS04], [BJS04], [Sør04]). We are interested in the first alternative. And there are many possible approaches to approximate maximum likelihood estimation (AMLE), which means considering the MLE associated to (1.1) with p somehow approximated. Let us consider the numerical methods. A first way is to solve numerically the PDE for the transition density ([Lo88], [Pou99]). Otherwise, some simulation based techniques have been proposed ([Ped95], [Era01], [ECS01], the last two using MCMC, with direct application to simulate Bayesian estimation). The pro is that these methods (actually all devised for ergodic diffusions) obtain AMLEs asymptotically equivalent to the true unknown MLE for any Δ. The con is that one can numerically solve $p(x,t;y,s;\theta)$ only for fixed values $(x,t;y,s;\theta)$. Therefore, one faces the task of solving N times the same problem for each θ in a suitable mesh. The numerical burden is clearly such that undermines the concrete feasibility of these computer-intensive proposals.

A different way out is to allow other sample schemes. In fact, there are at least three sample schemes considered in the literature to study the consistency of the proposed estimators (in general, not necessarily AMLEs): i) Δ fixed, $T_N \to \infty$ as $N \to \infty$ (e.g., [BiSo95] if S is ergodic, [Aït02] in special ergodic cases). ii) T fixed, $\Delta \to 0$ as $N \to \infty$ (e.g., [GeJa93], [Jac06], both in the general non-ergodic case). iii) $\Delta = \Delta_N = T/N \to 0$, $T = T_N \to \infty$ as $N \to \infty$, such that $N\Delta_N \to \infty$ as $N \to \infty$ (e.g., [Kes97] and [Aït02], [Aït04]). Notice that in ii),iii), the limit correspond to the observation of the whole trajectory, which is completely unrealistic. The real role of $\Delta_N \to 0$ is to allow a Gaussian approximation of the transition density, that we will interpret as a singular perturbation.

We conclude saying that foremost amongst the non-AMLE techniques are the so called martingale estimating functions (see [BJS04]). An estimating function is a function $G_N(\theta)$ of the sample and θ, that gives the estimator $\widehat{\theta} : G_N(\theta) = 0$. The idea is to substitute the unknown score function with some other function. As the score is a martingale, it is convenient to look for $G_N(\theta)$ which are martingales, allowing the study of the asymptotic properties. Estimating functions are simple, but rarely explicit (see [Sør97]). Also in this context, we suggest and explain how to integrate well known results with simple perturbations. We stress again that the main interest of the thesis is a theoretical study of the perturbation techniques, and not in the specific applications (many are currently being studied), which are presented in order to highlight the pros and cons of the perturbation methods and in what direction do further research.

[3] A set of sufficient conditions for this are given in [KaTa81], Sections 15.5-6; other conditions can be found, e.g., in [HaSc95].

1.3 Structure of The Thesis And Statement of Contributions

The thesis is structured in two parts and some appendixes.

First Part
In the first part of the thesis we supply the main concepts, assumptions and tools used in the second part. Most of the material presented is well known, and we will need to extend only a few results. We always give a detailed account of the literature and try to explain in an easy way the results, why we need them and what related problems one has to face. We believe that this review could be quite useful not only for a clearer understanding of the techniques exposed in the second part, but also on its own, since we collect here in a single framework many results, some of which fairly advanced, that to our knowledge can be found only spread across many different references.

In the thesis we use some operator theory facts, especially from the semigroup formulation of continuous time Markov process (see [EtKu86], [KaTa81]), trying to follow the route opened in Finance by [AHS04] (see also, e.g., [AHS03]). This will allow us to study the Cauchy problems from an abstract point of view, and hence to supply in the second part of the thesis some general results. We remark also that the only prerequisite of the first part is stochastic calculus, at the level of the lectures notes [CiPe98] used in the Ph.D. lectures of Prof. D.M. Cifarelli, while any other prerequisite of Functional analysis and PDE theory is reviewed in the appendixes.

Briefly, the contents of the chapters of the first part are:

Chapter 2: We introduce the basic auxiliary models that we use and the first assumptions.

Chapter 3: We review some results about the Kolmogorov equations and the Feynman-Kac representation theorem. A main goal is to recall what derivative securities can be priced by means of a PDE and then conclude the set of main assumptions.

Chapter 4: The general semigroup theory for not necessarily time-homogeneous diffusions is presented. We sketch the proof for the results that we extend to the general time-dependent case. Then we also review in detail the connections between diffusion processes and linear parabolic PDEs.

Chapter 5: We introduce the problem of the approximation of solutions of Cauchy problems and the central concept of model error discussing a key result in the literature about how to use PDE a-priori inequalities in financial applications.

Chapter 6: This chapter is a brief and, by and large, partial introduction to perturbations of differential problems. We explain the goals and ideas of the particular perturbation methods that we use in the second part. We also consider some basic well-known results about the perturbation of semigroups of operators.

Second Part
The second part is the core of the thesis and is about the derivation and analysis of the perturbation methods considered. In each chapter we start considering a new proposal of the very recent literature, for which we always give precise and complete references: we discuss in detail the method, why it has been proposed and the suggested applications. Then we supply our own contribution: the study of new theoretical properties of the techniques, as well as of possible extensions and further applications, with special emphasis on the parametric estimation of discretely observed diffusion processes.

The approximation methods considered are: regular small-parameter perturbation method, small-time expansions and the parametrix method. We remark that the techniques themselves are classical mathematical tools, while their application to the specific financial problems here is quite new. As classical tools, mostly developed in the field of physics, the techniques well apply to situations where strong regularities conditions and high precision of the observables are met. Furthermore, with the remarkable exception of the parametrix method, in the standard literature there are very few general convergence results and error bounds for the perturbation series, very important in financial and statistical applications.

This additional difficulty of tuning the techniques to the classical financial assumptions has been of great relevance especially for the small-times expansions, which have been our major object of research in the last year. We manage to supply a mathematical justification to the very promising and effective approximation method of option prices proposed by [MeKr06] using a small-time expansion. However, we can give only a partial justification in terms of asymptotic perturbations. Our justification is only partial since it seems more than a mere conjecture that the perturbation is regular, at least for some interesting cases in the numerical experiments of [MeKr06], but we are not able to prove the convergence of the perturbation series, and we will explain why in an example about the BS case.

All the methods considered, not only the small-time expansions, can be proficiently applied to the approximation of option prices which are solution of linear parabolic PDEs. We discuss this application in great depth, extending the study also to the approximation of the Greeks of the options. Each technique has some comparative advantages, and so we believe that the thorough theoretical comparison carried out is useful not only to supply a unitary, general framework, but also to better investigate further applications. In particular, as remarked in [AHS04], operator methods applied to the parametric estimation of discretely observed diffusion processes overcome many relevant difficulties of this latter complex problem. But we have noticed, and it is object of current research, that the perturbation approach from the parametrix method is even more promising, since could be used to remove also the typical unrealistic assumption (for asymptotic results) that the frequency of observations increases with the number of observation.

Let us summarise in detail the contents of the chapters in the second part of the thesis, highlighting our specific contributions to the study of the problem and stressing also how we are developing our current related research in connection to new open problems discovered:

Chapter 7: We start presenting an example of the simplest regular perturbations with respect to a dimensionless parameter, as recently applied to the BS model. Our new contributions are in Sections 7.1, where we give a description of the method that could be applied in full generality. We study also the applicability of the method to different problems and, in particular, in section 7.4 we match the presented method with well known results from Finance and the parametric estimation of diffusion processes. From the manifest difficulty of supply any general result, we conclude the need of studying more powerful techniques.

Chapter 8: This Chapter is an introduction to small-time expansions. In section 8.2 we introduce many possible ways to get the perturbation series, many of which well known in the literature, some new suggestions, and we compare the scope of the various derivations. An important new contribution of the chapter is a proof of the equivalence of the various derivations, while the main result is a general set of conditions for the asymptotic property of the expansion. Here we use tools from the Semigroup theory, which appear to be a new point of view, allowing our general results.

Chapter 9: More complex small-time expansions are sometimes necessary in the applications, as pointed out, e.g., in Section 9.3. We prove a general result about the asymptotic property in Section 9.2, and we use this to supply a new partial justification for the new proposal for the approximation of the option prices introduced by [MeKr06]. In section 9.5 we give a detailed and original discussion of the BS case, dealing in particular with the difficult problems of the singularities of the transition density and the convergence of the perturbation series for the small-time expansions.

Chapter 10: We thoroughly study and describe the parametrix method, a very recent proposal of the literature for the approximation of option prices, not yet published and still being studied. We just compare the method with the other techniques studied in the previous chapters, highlighting pros and cons, especially the advantages in terms of the applications from the general convergence result available for the method. We put special emphasis into the parametric estimation problem, presenting also the first partial steps of our current research in that direction. The parametrix method turns out to be the most promising technique and we collect most of our results about in the final Chapter.

Chapter 11: In this Chapter our main new contribution is the assessment the "robustness" of the parametrix method to its assumptions, which means that at least a specific model-based extension of the method is possible. Instead, we give a broader theoretical justification of this fact using a general abstract semigroup approach for the parametrix method; this abstract method could be coupled with the results presented in Chapter 7 and this is object of current research. Finally, with the advanced abstract tools considered, in section 11.3 we return to the study of the possible regularity of the asymptotic small-time expansions, while, along the way in section 11.2, we review also a set of general sufficient conditions for the well posedness hypothesis set in Chapters 2,3 and used throughout the thesis.

Appendixes

Some necessary prerequisite facts and definitions of Functional analysis, parabolic PDEs and Semigroup theory are reviewed in three separate appendixes.

Part I

Theory And Assumptions

Chapter 2

The Statistical Model

In this chapter we give the assumptions on the basic system of SDEs, which, in continuous-time Finance, is how the statistical model is typically described. In particular, for the problems and applications that we present, for both estimation and pricing ends, it is natural to consider diffusion-types models. References for SDEs are [Fri76], [KaSh92] and [Øks06]. A specific treatment of diffusion processes can be found in [Dyn65], [KaTa81] and [EtKu86]. For the Financial theory used in this chapter see, e.g., [Duf01].

2.1 Assumptions on The SDE

Let $(W(t), t \geq 0)$ be a l-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and $(\mathcal{F}_t, t \geq 0)$ be the $\mathbf{P}$-augmented filtration generated by the Brownian motion. Let $\mu_i(x,t)$ and $\sigma_{ij}(x,t)$, $i = 1, \dots, d$ and $j = 1, \dots, l$, be real measurable function on $H = \mathbb{R}^d \times [0,T]$, $T > 0$. We denote $\mu(x,t)$ the vector of components $\mu_i(x,t)$ (drift vector) and $\sigma(x,t)$ the $(d \times l)$ matrix $(\sigma_{ij}(x,t))$ (dispersion matrix). We make the following assumption:

1. There exists a $\mathcal{F}_t$-adapted stochastic process $\{S(t), t \in [0,T]\}$ unique strong solution of the system of SDEs

$$dS(t) = \mu(S(t),t)dt + \sigma(S(t),t)dW(t), \tag{2.1}$$

for $S(t) \in \mathbb{R}^d$, $t \in [0,T]$, $T > 0$, with initial condition $S(0) = \xi$ $(a.s.)$ independent of $(\mathcal{F}_t, t \geq 0)$.

(2.1) is the general form of the data generating process we use throughout the thesis. The requirement of the strong solution will be clear later. Actually, for estimation purposes, the existence of a weak solution could be enough (e.g., for a discussion see [Aït02]), but note that in Finance (in particular for derivative pricing) the choice of the probability measure plays an important role and for this reason the probability space and the risk factors $W(t)$ are usually supposed given. We review below two important classes of SDEs that admit a strong solution.

Let $P(x,t;B,s) = \mathbf{P}\left[S(s) \in B | S(t) = x\right]$ be the transition function of the Markov process $S(t)$ solution of (2.1), where $0 \leq t < s \leq T$, $x \in \mathbb{R}^d$ and B is a Borel set of $\mathbb{R}^d$. Let us assume:

2. $\{S(t), t \in [0,T]\}$ is a regular diffusion process.

3. The transition function of $S(t)$ has density $p(x,t;y,s)$, i.e.,

$$P(x,t;B,s) = \int_B p(x,t;y,s)dy.$$

4. The transition density satisfies the regularity properties

$$\int_{\mathbb{R}^d} p(x,t;y,s)dy = 1, \quad \lim_{s\downarrow t} p(x,t;y,s) = \delta(y-x),$$

where δ is the Dirac's delta distribution.

The assumption that the solution of (2.1) is a diffusion could seem restrictive, but it is a natural hypothesis for our applications. Our results do not apply to jump-diffusion processes, but the extension to Levy processes is possible and in perfect accordance with our presentation (i.e., using a semigroup approach, see [EtKu86]), despite we do not deal with in the thesis. In the recent literature there are some results in this direction (e.g., see [Sch04]). The other hypotheses on the transition density are standard (e.g., see [Bjö03] or [Jac04]). Actually, as we will see (Section 3.2), the existence of a transition density with the specified properties is not an assumption if the coefficients of the SDE (2.1) are smooth and bounded.

2.2 Infinitesimal Parameters of a Diffusion

A diffusion process is a continuous time Markov process $\{S(t), t \in [0,T]\}$ for which the sample paths are ($a.s.$) continuous functions of t. Let $|.|$ be the Euclidean norm. The continuity of the sample paths is a consequence of the following required properties: for every $\epsilon > 0$, $t \geq 0$ and $x \in \mathbb{R}^d$,

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{|x-y|>\epsilon} P(x,t;dy,t+h) = 0, \tag{2.2}$$

i.e., for a diffusion the probability of a displacement of order greater than ϵ is $o(h)$, and, assuming that, for all $i,j = 1,\ldots,d$, the following conditional expectations exist finite,

$$\lim_{h\downarrow 0} \frac{1}{h} E\left[S_i(t+h) - S_i(t) | S(t) = x\right] = b_i(x,t), \tag{2.3}$$

$$\lim_{h\downarrow 0} \frac{1}{h} E\left[\left(S_i(t+h) - S_i(t)\right)\left(S_j(t+h) - S_j(t)\right) | S(t) = x\right] = a_{ij}(x,t). \tag{2.4}$$

We give in the following sections sufficient conditions on the coefficients of the strong solution $S(t)$ of (2.1) such that (2.3) and (2.4) hold true. The vector $b = (b_1,\ldots,b_d)^\top$ is called infinitesimal mean (or drift coefficient) of the diffusion, while the matrix $a = (a_{ij})$, called *diffusion matrix*, is interpreted as the infinitesimal variance of the diffusion. $a(x,t)$ and $b(x,t)$ are continuous functions in H. The diffusion matrix is also required to be symmetric and positive semi-definite in H. If $a(x,t)$ is positive definite in H, then the diffusion is said non-degenerate; if otherwise $a(x,t)$ is singular for some $(x,t) \in H$, the diffusion is called degenerate.

Diffusions can be characterized in terms of their infinitesimal parameters and boundary conditions. In general, the infinitesimal parameters govern the realizations of the diffusion only off the boundaries of the state space. In other words, two diffusions with the same infinitesimal parameters but different boundary behavior are different processes. In the multidimensional case the analysis of the boundary behavior of diffusions is more involved and no complete characterization is available. To avoid these complications we require the following.

5. The specification of the infinitesimal parameters of the diffusion $S(t)$ is such that the boundaries of the state space are unattainable for $S(t)$.

Unattainable boundaries, such as natural and entrance boundaries, cannot be reached in a finite time $T > 0$ (our time horizon). Therefore, under this assumption, the trajectories of $S(t)$ are completely governed by the drift vector and diffusion matrix. Note that the assumption is set only for convenience and can be removed in any application for which a different boundary specification is needed, provided one supplies the study of the boundary behavior (in this connection see for the scalar case, e.g., [KaTa81], chapters 15.7-8 and 15.12.A, and [Man68]). In fact, our results will continue to hold off the boundaries.

2.3 Lipschitz Continuous Coefficients SDEs

We are ready to recall some sufficient conditions in order to ensure that

- there exists a unique strong solution $S(t)$ of (2.1);
- $S(t)$ is a diffusion process;
- the infinitesimal parameters of $S(t)$ are recovered from the drift vector and dispersion matrix of the original SDE.

The first set of conditions is given, e.g., in [Fri76], Chapter 5. The system of SDEs (2.1), with initial condition $S(0)$ such that $E\left|S(0)\right|^2 < \infty$, admits a unique strong solution if the coefficients $\mu(x,t)$ and $\sigma(x,t)$ are measurable and satisfy both the Lipschitz condition

$$|\mu(x,t) - \mu(x_0,t)| + |\sigma(x,t) - \sigma(x_0,t)| \leq K_0 \left|x - x_0\right|, \tag{2.5}$$

and the linear growth condition

$$|\mu(x,t)| + |\sigma(x,t)| \leq K_1(1 + |x|), \tag{2.6}$$

where K_0, K_1 are positive constants. These results can be weakened by a localization argument (see [Fri76], page 113). Note that if (2.1) is an *autonomous* (or time-homogeneous) SDE, i.e., the coefficients are functions only of the space-variables, say (with abuse of notation) $\mu(x,t) = \mu(x)$ and $\sigma(x,t) = \sigma(x)$, then the linear growth condition (2.6) is implied by (2.5). In turn the (local) Lipschitz continuity is ensured (by the mean value theorem) if $\mu(x), \sigma(x) \in C^1(\mathbb{R}^d)$.

If, furthermore, $\mu(x,t)$ and $\sigma(x,t)$ are continuous functions in H (which is always the case if $S(t)$ is autonomous and (2.5) holds), then the solution $S(t)$ of (2.1) is a diffusion process with drift vector $b(x,t) = \mu(x,t)$ and diffusion matrix $a(x,t) = \sigma(x,t)\sigma^\top(x,t)$.

Note that different dispersion matrices $\sigma(x,t)$, with different entries and also possibly different l-dimensions, can produce the same diffusion matrix $a(x,t)$ (for a discussion of this issue see, e.g., [Kry80], page 121.) Since we are mainly interested in the diffusion matrix, sometimes it is more natural to parametrize directly $a(x,t)$ and then go back to a dispersion matrix which satisfies (2.5) and (2.6). It is well known (see [Fri76] section 6.1) that a $(d+1 \times d+1)$ matrix $a(z)$[1], positive semi-definite for each value of the parameter $z \in \mathbb{R}^{d+1}$, has a unique square-root matrix $\widetilde{\sigma}(z)$, i.e., a unique positive semi-definite matrix $\widetilde{\sigma}(z)$ such that $\widetilde{\sigma}(z)\widetilde{\sigma}^\top(z) = a(z)$, for each $z \in \mathbb{R}^{d+1}$. If also each entry of a is Holder continuous of exponent α (we denote this by $a_{ij}(z) \in C^\alpha(\mathbb{R}^{d+1})$, $\alpha \in (0,1]$) for all $z \in \mathbb{R}^{d+1}$, then also $\widetilde{\sigma}_{ij}(z) \in C^\alpha(\mathbb{R}^{d+1})$ for all $z \in \mathbb{R}^{d+1}$. So, when we parametrize $a(x,t)$, we will assume $a_{ij}(x,t) \in C^\alpha(H)$, $\alpha = 1$, so that (note that $l = d$) $\widetilde{\sigma}(x,t)$ is a suitable dispersion matrix.

Linear SDEs

The conditions (2.6) and (2.5) are satisfied by any *linear* SDE, i.e., a SDE for which $\mu(x,t)$ and $\sigma(x,t)$ are linear functions[2] in x (see [KaSh92], Section 5.6). Linear equations have an *explicit*

[1] Here we do not distinguish between space and time-variables.

[2] For $d = 1$ we can take $\sigma(x,t) = \sum_{j=1}^{l} \alpha_j(t) + \beta_j(t)x$, but in the case $d > 1$ we are restricted to a non-random matrix $\sigma(x,t) = \sigma(t)$.

strong solution. If also $\mu(x,t), \sigma(x,t)$ are continuous in H, the solution of a linear SDE is a diffusion process for which the transition density is known in closed-form. We are particularly interested in two linear (and autonomous) equations, the Langevin equation and the geometric Brownian motion.

Let, for $d = 1$,

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \tag{2.7}$$

where $S(t) > 0$ ($a.s.$). The equation (2.7), with $\mu \in \mathbb{R}$, $\sigma > 0$ and $S(0) > 0$, is called geometric Brownian motion and is often used to model prices $S(t)$ of assets such as stocks. By Ito's formula, the logarithm of the price (the so-called instantaneous returns) satisfies

$$d \ln S(t) = \left(\mu - \sigma^2/2\right) dt + \sigma dW(t),$$

i.e., an (arithmetic) Brownian motion. Therefore, the solution of (2.7) is given by

$$S(t) = S(0) \exp\left[\left(\mu - \sigma^2/2\right) t + \sigma W(t)\right], \tag{2.8}$$

and the distribution of $S(t)$ is lognormal. The widespread use of geometric Brownian motion in Finance as a model for asset prices has origin in the fundamental work of Black and Scholes [BlSc73]. Note that the process (2.8) is positive and non-stationary (to be precise $S(t)$ is null-recurrent) for any initial value $S(0)$ and parameters μ, σ, which are characteristic features of stock prices.

The Langevin equation, in the scalar case $d = 1$, has the form

$$dS(t) = -\mu S(t)dt + \sigma dW(t), \tag{2.9}$$

where $\mu, \sigma > 0$, and its solution is given by the Ornstein-Uhlenbeck process

$$S(t) = S(0) \exp(-\mu t) + \sigma \int_0^t \exp(-\mu(t-s))\, dW(s). \tag{2.10}$$

It is well known that the process (2.10) is a Gaussian process with mean $ES(t) = S(0)e^{-\mu t}$, variance $VarS(t) = \sigma^2/(2\mu) + \left(VarS(0) - \sigma^2/(2\mu)\right) e^{-\mu t}$ and covariance function

$$Cov(S(t), S(s)) = \left[Var(S(0)) + \sigma^2/(2\mu)\left[e^{2\mu \min(t,s)} - 1\right]\right] e^{-\mu(t+s)}.$$

If $S(0)$ has a normal distribution with mean 0 and variance $\exp(-\mu t)$, then $S(t)$ is a Gaussian stationary process with covariance function $Cov(S(t), S(s)) = \sigma^2/(2\mu)e^{-\mu|t-s|}$. The existence of a stationary density is a consequence of the mean reversion in (2.9)[3]. In Finance this fact is used to model instantaneous short-term interest rates, for which the empirical evidence suggests that there exists a long-run average level (called reversion level), say $\overline{\theta} > 0$. The slight generalization of (2.9) given by

$$dS(t) = \kappa(\overline{\theta} - S(t))dt + \sigma dW(t), \tag{2.11}$$

where $\kappa > 0$ is the mean reversion parameter, is called Vasicek process, after [Vas77] who first used it to model instantaneous short interest rates. The solution of (2.11) is again an Ornstein-Uhlenbeck process. In practice it is more often used the (still linear) generalization of the Vasicek model with $\overline{\theta} = \overline{\theta}(t)$ (and possibly $\sigma = \sigma(t)$), called Hull-White model (see [HuWh90] and [Hul06], chapter 28), which allow to exactly fit the observed term-structure.

We remark that many non-linear SDE can be reduced to linear SDE by means of a suitable transformation. We refer to [KlPl99], chapter 4, for an extensive treatment of this fact.

[3] The mean reversion takes place because $\mu > 0$ in (2.9).

2.4 Affine SDEs

Affine models are the benchmark in Asset pricing as Markov term structure models (i.e., models of the spot rate), but have also wider applications in option pricing (see [Duf01], [DFS03] and [Pia04]). We give the motivation to consider this second class of SDE noticing that the Vasicek model, as a model of the short spot rate, is counterfactual in at least two dimensions. First, the nominal spot rates must be positive, while the Gaussian process solution of (2.11) can takes negative values with positive probability. Second, empirically the volatility of a spot rate tends to increase with the level of the interest rate. A model that overcomes these drawbacks is the so-called CIR process (after [CIR85], who proposed it to model the spot rate) or Feller square-root process (see [Fel51]) given by ($d = 1$)

$$dS(t) = \kappa(\overline{\theta} - S(t))dt + \sigma\sqrt{S(t)}dW(t). \tag{2.12}$$

Not surprisingly, the wide success of the SDE (2.12) in Finance does not depend on its empirical performances, but on the facts that (2.12) is solvable in closed-form (see [CIR85]) and its solution is a diffusion process (see [Fel51]).

The crucial remark about (2.12) is that the dispersion coefficient $\sigma(x) = \sigma x^{1/2}$, $\sigma > 0$, for $x \geq 0$, does not satisfy the Lipschitz condition (2.5). In fact, the derivative of $\sigma(x) = \sigma x^{1/2}$ is not defined at $x = 0$. The existence of a strong explicit solution for (2.12) in ensured by the extra condition $\kappa\overline{\theta} > \sigma^2/2$, called Feller condition (see again [Fel51]). Under this condition the Feller square root-process $S(t)$ is strictly positive ($a.s.$) for all $t \in (0, T]$ and $S(0) \geq 0$ ($a.s.$), and 0 is an entrance boundary. Uniqueness of the solution is still a different issue and here follows from the results of [YaWa71] (see also [KaSh92], page 291).

An *affine* (autonomous) diffusion is defined as the unique strong solution of the SDE (2.1) with

$$\mu(x) = \mu_0 + \mu_1 x, \tag{2.13}$$

$$a_{ij}(x) := \left(\sigma(x)\sigma^\top(x)\right)_{ij} = a_{0ij} + a_{1ij}x, \tag{2.14}$$

where $\mu_0 \in \mathbb{R}^d$, $\mu_1 \in \mathbb{R}^{d\times d}$, $a_{0ij} \in \mathbb{R}$, $a_{1ij} \in \mathbb{R}^d$, for $i, j = 1, \ldots, d$. To ensure that the solution $S(t)$ is a diffusion, we require that (2.2)-(2.4) hold, which could be checked by inspection if an explicit solution of the SDE exists. Note that μ, σ are continuous. If $a_{1ij} = 0$ for all i, j, then the affine SDE reduces to a linear SDE. So the interesting case is when $a_{1ij} \neq 0$ for some i, j, for which (2.1) is no longer linear and the state-space will be restricted by some conditions on x from the fact that the diagonal elements of $a(x)$ must be non-negative. Taking into account (2.14), the state-space must be the closure of the set $\{x \in \mathbb{R}^d : a_{0ii} + a_{1ii}x \geq 0,\ i = 1, \ldots, d\}$.

Sufficient Conditions

We supply a sets of conditions to ensure strong uniqueness and existence for a large class of SDE extending (2.12), as proposed by [DuKa96]. First note that again nothing changes if we replace σ by any square-root $\widetilde{\sigma}$ of the symmetric positive semi-definite matrix a. From this we can directly suppose $\sigma(x)$ to be symmetric for all x. It is proved in [DuKa96] that if we require that the field of matrices $a(x) = (a_{ij}(x))$ satisfying (2.14) is positive definite for each x, then (after a possible permutation of the indexes) there exists a non-singular matrix Σ and d functions

$$v_i(x) = \alpha_i + \beta_i x,$$

where $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}^d$ are such that $\alpha_i + \beta_i x > 0$ for each x in the state space and $i = 1, \ldots, d$, so that we have the representation

$$\sigma(x) = \Sigma diag\left(\sqrt{v_1(x)}, \ldots, \sqrt{v_d(x)}\right), \tag{2.15}$$

which shows again that the dispersion matrix of a non-linear affine SDE cannot satisfy (2.5).

Denote D the interior of the state-space, i.e., $D = \{x \in \mathbb{R}^d : v_i(x) > 0,\, i = 1, \ldots, d\}$, and ∂D its frontier. The central idea of [DuKa96] is to extend the Feller condition requiring that, for each component $i = 1, \ldots, d$, there exists enough positive drift $\mu(x)$ near $\partial D_i = \{x \in \overline{D} : v_i(x) = 0\}$ such that the boundary ∂D_i is unattainable for $S(t)$. The extension is:

(A) For all i and all $x \in \partial D_i$, $\beta_i^\top \mu(x) > \beta_i^\top \Sigma\Sigma^\top \beta_i / 2$.

However, we must rule out also that, when the i-th component $S_i(t)$ is near ∂D_i, the correlation structure with other components could drive $S_i(t)$ to hit ∂D_i. This is achieved requiring:

(B) For all i, j, if $\left(\beta_i^\top \Sigma\right)_j \neq 0$, then $v_i = v_j$.

Then we have:

Theorem 1 *Under conditions* (**A**) *and* (**B**) *and for all $S(0) \in D$ (a.s.), there is a unique strong solution $S(t)$ in D of the SDE (2.1) with drift vector (2.13) and dispersion matrix (2.15). Moreover, the boundaries ∂D are unattainable for $S(t)$ and, for all i, we have $v_i(S(t)) > 0$ (a.s.) for all $t \in [0, T]$.*

Extension to time dependent coefficients is possible, but we refer to [Duf01] and references therein.

Example: The Heston Model

An important affine SDE is the Heston model (see [Hes93]), a stochastic volatility model with an explicit solution. Assume that the spot asset price is governed by

$$dZ(t) = \mu Z(t)dt + \sigma_Z \sqrt{V(t)} Z(t) dW_Z(t), \tag{2.16}$$

where the stochastic volatility process $V(t)$ satisfies a square-root process

$$dV(t) = \kappa(\overline{\theta} - V(t))dt + \sigma_V \sqrt{V(t)} dW_V(t), \tag{2.17}$$

and $W_Z(t)$ and $W_V(t)$ can be correlated. The system of SDE given by (2.16) and (2.17) is not affine, but we can easily transform it. Applying Ito's lemma to $Y(t) = \ln Z(t)$, we get

$$dY(t) = \left(\mu - \frac{1}{2}\sigma_Z^2 V(t)\right) dt + \sigma_Z \sqrt{V(t)} dW_S(t), \tag{2.18}$$

and now $(Y(t), V(t))^\top$ is an affine SDE. Here we have that the state space is $\mathbb{R} \times [0, \infty)$, $v_1(Y, V) = v_2(Y, V) = V$ and $\Sigma = diag\,(\sigma_Z, \sigma_V)$. Restricting the parameters according to condition (**A**) implies the existence of a unique strong solution.

[Hes93] solved this model by guessing an exponential affine form for the Fourier transform of $(Y(t), V(t))^\top$. This is a key feature of any affine SDE and sometimes is taken as definition. We will see in the next chapter that the Fourier transform of the solution of a system of SDE, under some regularity conditions, satisfies a PDE. For affine diffusions the exponential affine form of the Fourier transform reduces the PDE to a system of ODEs, which could be analytically solvable as in [Hes93] and anyway more easy to cope with numerical methods. [Duf01] supplies also an excellent introduction to transform methods in option pricing.

Chapter 3

Cauchy Problems

There are two fundamental reasons for studying Cauchy problems in mathematical Finance. Under suitable conditions, i) the transition density of a diffusion satisfies the Kolmogorov Backward equation; ii) the price of many derivative securities are given by the solution of a Cauchy problem. These facts are related by the celebrated Feynman-Kac theorem and the concept of martingale. In this spirit, we recall also what derivatives can be priced by means of a PDE. References for the connections between SDEs and PDEs are [Fri76], [Fre85], [KaSh92] and [Kry80]. For the Financial theory, see [Duf01], [Shr04] and [Bjö03].

3.1 The Kolmogorov equations

Let $S(t)$ be the diffusion solution of (2.1), under the hypotheses of the previous chapter. Consider the operator that arises from the Ito's formula applied to any function of $C^{2,1}(H)$, $H = \mathbb{R}^d \times [0,T]$, $T > 0$, given by $L(t) = \frac{\partial}{\partial t} + A(t)$, where (recall that we denote $a(x,t) = \sigma(x,t)\sigma^\top(x,t)$)

$$A(t) = \sum_{i=1}^{d} \mu_i(x,t)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j} \tag{3.1}$$

is an elliptic, possibly degenerate, operator in H acting on the space variables of functions in $C^{2,1}(H)$.

The Backward Equation

For any measurable and bounded function $\varphi : \mathbb{R}^d \to \mathbb{R}$, under mild conditions (see [KaTa81]), the function

$$u(x,t) = E\left[\varphi\left(S(s)\right)|S(t) = x\right], \tag{3.2}$$

for $0 \le t < s \le T$, $x \in \mathbb{R}^d$, is in $C^{2,1}(H_0)$, $H_0 = \mathbb{R}^d \times [0,T)$, and satisfies the PDE

$$\left(L(t)u\right)(x,t) = 0. \tag{3.3}$$

(3.3) is called *Kolmogorov Backward equation* and defines a Cauchy problem with the terminal condition $\lim_{t\uparrow s} u(x,t) = \varphi(x)$[1]. Note that (3.3) is homogeneous, but the presence of an inhomogeneous term does not substantially complicate the problem (see, e.g., Appendixes A.2-3).

The hard part to prove is that $u \in C^{2,1}(H_0)$. If we assume that $u \in C^{2,1}(H_0)$, (3.3) is a consequence of the Markov property and Ito's lemma. By the Markov property we see that $u(S(t),t)$ is an $\mathcal{F}_t$-martingale, in fact, for $t \le r < s$,

[1] We write $u(x,t)$ instead of $u(x,t;s)$ for ease of notation. This is consistent with the fact that we keep s fixed.

$$
\begin{aligned}
E\left[u(x,r)|\mathcal{F}_t\right] &= E\left[E\left[\varphi\left(S(s)\right)|\mathcal{F}_r\right]|\mathcal{F}_t\right] \\
&= E\left[\varphi\left(S(s)\right)|\mathcal{F}_t\right] = u(x,t).
\end{aligned}
$$

Thereby, by Ito's lemma we have

$$
\begin{aligned}
u(S(r),r) - u(S(t),t) = &\int_t^r \left(L(\eta)u\right)(S(\eta),\eta)d\eta + \\
&+ \int_t^r u_x(S(\eta),\eta)\sigma(S(\eta),\eta)dW(\eta),
\end{aligned}
$$

where u_x is the $(1\times d)$ vector of the first partial x-derivatives of u, so that we can take conditional expectation with respect to $\mathcal{F}_t$ and obtain

$$
\int_t^r \left(L(\eta)u\right)(S(\eta),s)d\eta = 0,
$$

which implies (3.3) since $u \in C^{2,1}(H_0)$.

The specification $\varphi(x) = \chi_B(x)$ in (3.2) gives $u(x,t) = P(x,t;B,s)$, the transition function of the diffusion $S(t)$. On the other hand, if the transition density $(x,t) \mapsto p(x,t;y,s) \in C^{2,1}(H)$ and the operator $L(t)$ commutes with the expectation in (3.2), then

$$
\begin{aligned}
\left(L(t)u\right)(x,t) &= L(t)\int_{\mathbb{R}^d} \varphi(y)p(x,t;y,\eta)d\eta \\
&= \int_{\mathbb{R}^d} \varphi(y)L(t)p(x,t;y,\eta)d\eta,
\end{aligned} \tag{3.4}
$$

which implies that also the transition density satisfies the Kolmogorov backward equation (3.3), since in (3.4) φ is an arbitrary bounded and measurable function.

We remark that the commutativity follows provided $p \in C^{2,1}(H)$ and the tails of $p(x,t;y,\eta)$ decrease fast enough, for instance if p is Gaussian. The conditions of the Parametrix method (see [Fri64] and Chapter 10) and a version below of the Feynman-Kac theorem suffice to $p \in C^{2,1}(H)$ and the Gaussian tails behavior. The conditions of the Parametrix method are sometimes too restrictive, but only sufficient. We explicitly assume that

6. The transition density p of $S(t)$ is such that $(x,t) \mapsto p(x,t;y,s) \in C^{2,1}(H)$ and satisfies the Kolmogorov Backward equation (3.3).

Recalling 4. in the previous chapter, i.e., $p(x,t;y,s) \to \delta(x-y)$ as $t \uparrow s$, we see that under 6. the transition density of $S(t)$ is given by the fundamental solution of (3.3) (see [Fri64] and Appendix A.2), i.e., the solution of the Cauchy problem

$$
\begin{array}{ll}
L(t)p(x,t;y,s) = 0, & (x,t) \in \mathbb{R}^d \times [0,s), \\
\lim_{t\uparrow s} p(x,t;y,s) = \delta(x-y), & x \in \mathbb{R}^d.
\end{array} \tag{3.5}
$$

We stress that the all the claims of theorem 25 in Appendix A.2, but the estimates (A.16) and (A.17) (which are consequences of the parametrix method, see Chapter 10), still hold. Note also that the questions about the well posedness of (3.5) are related to the boundary conditions of $S(t)$. The uniqueness results reviewed in Appendix A.2 tacitly assume that the boundaries are unattainable, as required in 5.. In this case there is no need to specify additional boundary conditions. But any different boundary behavior must be taken into account, defining a completely different PDE problem (in fact, a boundary-initial value problem). For instance, the transition densities of the reflected and the absorbed Brownian motions on $\mathbb{R}_+$ satisfy the same Cauchy problem (3.5), but different well-posed boundary-initial value problems. For a discussion, see [KaTa81], page 214.

We notice that other versions of the Kolmogorov Backward equation hold. For instance, if we take $\varphi \in C^2(\mathbb{R}^d)$, then (3.3) is satisfied (see, e.g., [GiSk69]). We can also extend the problem to $\mathbb{C}$. E.g., taking $\varphi(x) = (2\pi)^{-\frac{d}{2}} e^{ix\cdot y}$, $y \in \mathbb{R}^d$, we see that the Fourier transform of the transition density (i.e., the Characteristic function of $S(t)$) satisfies the PDE (3.3).

The Forward Equation

Equation (3.3) is called Backward since we differentiate w.r.t. the past variables (x, t). However, the change of variable $t \mapsto s - t$ does not affect the well posedness of the problem: the effects are to invert the direction of time (now running forward), to change the sign of the time derivative, so that (with abuse of notation) $L(t) = -\frac{\partial}{\partial t} + A(t)$, and to transform the terminal condition into an initial condition. In mathematics, physics and engineering, differential problems are often solved forward in time and (3.3) is sometimes called *first* (or converse) *Kolmogorov equation* (see [GiSk69]).

In fact, there is a *second* (or direct) *Kolmogorov equation* satisfied by the transition density in the future variable (y, s) (notice, in the original times as in (3.5)), called also *Forward Kolmogorov equation* or, more often, *Fokker-Planck equation*. This equation involve the formal adjoint parabolic operator to $L(t)$ (see Appendix A.2), given by $L^*(s) = -\frac{\partial}{\partial s} + A^*(s)$, where, for any $u(y, s) \in C^{2,1}(H)$,

$$A^*(s)u = -\sum_i \frac{\partial(\mu_i u)}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2(a_{ij}u)}{\partial x_i \partial x_j}, \tag{3.6}$$

under the hypothesis that $\frac{\partial}{\partial x_i} a_{ij}$, $\frac{\partial^2}{\partial x_i \partial x_j} a_{ij}$ and $\frac{\partial}{\partial x_i}\mu_i$ exist and are continuous. The Fokker-Planck equation is

$$L^*(s)p(x, t; y, s) = 0, \tag{3.7}$$

and it is coupled with the initial condition $p(x, t; y, s) \to \delta(y - x)$ as $s \downarrow t$.

The derivation of (3.7) is quite involved and it requires more conditions than the Backward equation; for instance, it is satisfied under the Parametrix method (see the next section). An heuristic derivation of (3.7) (under our assumptions 2., 3., 4., 6. plus that $S(t)$ is autonomous) is given in [KaTa81], pages 219-220, from a formal derivation in time of the Chapman-Kolmogorov equation for the transition density. We will study again the Kolmogorov equations in Chapters 8 and 10. A detailed analysis of (3.7) can be found in [Ris89].

3.2 The Feynman-Kac Theorem

Let $S(t)$ be the strong solution of (2.1)[2]. We assume that the coefficients of $S(t)$ are continuous in H and satisfy the linear growth condition (2.6)[3]. Consider the functions $g : \mathbb{R}^d \to \mathbb{R}$, $f : H \to \mathbb{R}$ and $c : H \to \mathbb{R}$ and suppose: i) g, f, c are continuous; ii) $c(x, t) \geq 0$ for all $(x, t) \in H$; iii) f, g either satisfy a polynomial growth condition in x, i.e.,

$$|g(x)| + |f(x, t)| \leq const.(1 + |x|^\beta), \quad \text{in } H, \tag{3.8}$$

for some constant $\beta > 0$, or $g(x), f(x, t) \geq 0$ for all $x \in \mathbb{R}^d$, $t \in [0, T]$. Consider the operator $L_c(t) := L(t) - c(x, t)$. We can state the *Feynman-Kac stochastic representation formula*

[2] Actually, the existence of a weak solution is enough.

[3] Note that we do not require the Lipschitz continuity of the coefficients.

Theorem 2 *Under the above assumptions, let $v : H \to \mathbb{R}$ be continuous and such that $v(x,t) \in C^{2,1}(\mathbb{R}^d \times [0,T))$ and satisfies the Cauchy problem*

$$\begin{aligned} \left(L_c(t)v\right)(x,t) &= f(x,t), \quad (x,t) \in \mathbb{R}^d \times [0,T), \\ v(x,T) &= g(x), \qquad\qquad x \in \mathbb{R}^d, \end{aligned} \tag{3.9}$$

as well as a polynomial growth condition on H. Then $v(x,t)$ has representation

$$\begin{aligned} v(x,t) = {} & E\left[g(S(T)) \exp\left\{-\int_t^T c(S(s),s)ds\right\} |S(t) = x\right] \\ & - E\left[\int_t^T f(S(s),s) \exp\left\{-\int_t^s c(S(u),u)du\right\} ds|S(t) = x\right] \end{aligned} \tag{3.10}$$

on H and the solution $v(x,t)$ is unique.

The proof of this theorem is given in [KaSh92], page 366. We make some remarks. i) If the dispersion matrix is bounded on H, then we can weaken (3.8) requiring only that g, f are in the Tychonov class for $L_c(t)$ (see Appendix A.2), i.e., f, g satisfy the exponential growth condition

$$|g(x)| + |f(x,t)| \leq const. \exp(\gamma |x|^2), \quad \text{in } H,$$

for some suitable $\gamma > 0$, and we have that (3.10) is the unique solution of at most exponential growth of (3.9). ii) The linear growth condition is not satisfied for all non-linear affine models, but the conditions of the theorem are only sufficient and can be weakened in specific cases. For instance, a case-by-case analysis is suitable to say that the representation formula (3.10) holds in the CIR ([CIR85]) and Heston ([Hes93]) models. iii) The role of the multiplicative operator $c(x,t)$ will be explained in detail in Chapter 4. Here we notice that the assumption $c(x,t) \geq 0$ is not restrictive, because if v satisfies (3.9) and $c \geq \lambda \in \mathbb{R}$, then $w(x,t) = v(x,t)e^{-\lambda t}$ satisfies $L_c(t)w - \lambda w = fe^{-\lambda t}$. iv) If $c, f \equiv 0$, then (3.10) is in agreement with the Kolmogorov Backward equation, which could be seen as a special case of the Feynman-Kac theorem, since, under suitable conditions, formula (3.10) holds for $f \equiv 0$ and g bounded and measurable.

The hypotheses of the Feynman-Kac theorem, as stated here, are weak because the theorem does not say anything about the existence of the unique solution. In view of this fact we make the following assumption:

7. The Cauchy problem (3.9) is well posed and its unique solution is given by (3.10).

The requirement of well posedness is meant in a specified class of functions subset of the Tychonov class for $L(t)$, such as the class of functions satisfying a polynomial growth condition in the space variables. Note that the data f, g and the solution are always in the same class (but the constants of growth for the data and for the solution may be different). To support our assumption, we review three sets of sufficient conditions such that there exists a unique solution of the Cauchy problem (3.9) given by (3.10).

Lipschitz Continuous Coefficients

The assumptions of the representation theorem above can be strengthened in order to get also the existence of the solution. Following [Bal00], theorem 9.9, we have:

Theorem 3 *Assume that i) $a(x,t)$ is uniformly elliptic in H; ii) the coefficients μ, σ of $S(t)$ are continuous functions in H and satisfy (2.5) and (2.6); iii) c and f are uniformly Holder continuous in H and g is continuous in $\mathbb{R}^d$; iv) f, g satisfy the polynomial growth condition (3.8). Then there exists a unique solution $v \in C(H) \cap C^{2,1}(\mathbb{R}^d \times [0,T))$ of (3.9) of at most polynomial growth in H and is given by (3.10). Furthermore, if f, g satisfy an exponential growth condition and $\sigma_{ij}(x,t)$ are bounded functions, then v is at most of exponential growth.*

Bounded Coefficients

If we require more on the coefficients of (3.1) we can find a strong connection between the Cauchy problem for v and the Cauchy problem for the transition density of $S(t)$. From [Fri76], page 147, we have:

Theorem 4 *Assume that i) $a(x,t)$ is uniformly elliptic in H (see (A.9)); ii) the functions $a_{ij}(x,t)$, $\mu_i(x,t)$ and $c(x,t)$ are bounded on H; iii) the functions a_{ij} and μ_i are uniformly Lipschitz continuous in H, the functions a_{ij}, c and f are uniformly Holder continuous in H and g is continuous in $\mathbb{R}^d$; iv) f, g satisfy the polynomial growth condition (3.8). Then there exists a unique solution $v \in C(H) \cap C^{2,1}(\mathbb{R}^d \times [0,T))$ of (3.9) of at most polynomial growth in H and is given by (3.10).*

At first glance this could seem only a special case of the results of the previous section, but this is not the case. In fact, under the conditions of this theorem, there exists a unique fundamental solution (recall that here $t < s$) $\Gamma(x,t;y,s)$ of the operator $L(t) = \frac{\partial}{\partial t} + A(t)$ with $(x,t) \mapsto \Gamma \in C^{2,1}(H)$ (see [Fri64] and Appendix A.2). Moreover, the Cauchy problem (3.9) with $c, f \equiv 0$ has a unique solution given by

$$v(x,t) = \int_{\mathbb{R}^d} \Gamma(x,t;y,T) g(y) dy,$$

so that, for any $|g(x)| \leq const.(1 + |x|^\beta)$ in H, it must be that (the left hand side (LHS) holds by (3.10), while for the RHS see (A.20) with $t \mapsto T - t$)

$$E\left[g(S(T))|S(t) = x\right] = \int_{\mathbb{R}^d} \Gamma(x,t;y,T) g(y) dy.$$

From the arbitrary of g we have that $\Gamma(x,t;y,s)$ is the transition density p of the diffusion $S(t)$. Since $\Gamma \in C^{2,1}(H)$ and Γ has Gaussian type tails (see (A.16)), we conclude that $\Gamma(x,t;y,T)$ satisfies (3.5) in (x,t). On the other hand, under the additional assumptions that $\frac{\partial}{\partial x_i} a_{ij}$, $\frac{\partial^2}{\partial x_i \partial x_j} a_{ij}$ and $\frac{\partial}{\partial x_i} \mu_i$ exist and are bounded and continuous, the transition density Γ satisfies also the the adjoint equation (3.6) in the forward variables (y,s), showing that in this case the two Kolmogorov equations are equivalent (in the sense that they have the same solution).

Smooth Coefficients

The key assumption in the existence results above is the uniform ellipticity. Sometimes, in the applications, the operator (3.1) can be degenerate. The conditions in [Kry80], theorem 2.9.10, require more on the smoothness on all the functions involved:

Theorem 5 *Assume that i) the functions $a_{ij}(x,t)$, $\mu_i(x,t)$ are Lipschitz continuous; ii) $a_{ij}(x,t)$, $\mu_i(x,t)$, $c(x,t)$, $f(x,t) \in C^{2,0}(H)$, $g \in C^2(\mathbb{R}^d)$ and $c(x,t) \geq 0$; iii) the product of each function a_{ij},μ_i,c,f,g and their first and second derivatives with the function $(1 + |x|)^{-m}$ is bounded in H. Under these assumptions there exists a unique solution $v \in C(H) \cap C^{2,1}(\mathbb{R}^d \times [0,T))$ of (3.9) such that $|v(x,t)| \leq const.\,(1 + |x|)^m$ in H and is given by (3.10).*

3.3 No Arbitrage Option Pricing

The price of any vanilla option is given by a conditional expectation, so that, under the conditions of the Feynman-Kac theorem, we can compute the price also as the solution of a Cauchy problem. We distinguish between two fundamental cases: the generalized BS model and term-structure derivatives.

Definitions And Assumptions

A derivative security is a $\mathcal{F}_T$-measurable random variable defined by a real-valued payoff function g of some underlying financial quantities $S(t)$ in the interval $[t, T]$, where t is the current time and $S(t)$ is supposed to evolve accordingly to the SDE (2.1). We restrict ourselves to the so-called T-contingent claims, contracts to claim $g(S(T))$ at T. For instance, the payoff of an European call option is given by $g : x \mapsto (x - K)^+$, where $(.)^+$ is the positive-part function, $K > 0$ is called exercise price and x is the price of some asset (or the value of some other, not necessarily traded, economic quantity) as of time T.

The problem is to find the fair price $V(S(t), t)$ of the claim $g(S(T))$ at time t. In Arbitrage theory (e.g., see [Bjö03] or [Duf01]) a financial model to price an option is a SDE for the underlying *and* a reasonable set of assumptions. For instance, in the BS model (see [BlSc73]), the underlying of the European option is an asset which price follows a geometric Brownian motion (2.7). We assume:

- Short rate process: there exists a (positive and integrable) continually compounding spot rate process $r(S(t), t)$ and a risk-free asset $\beta(t)$ (called money market account), which price follows the ODE (w.l.o.g., set $\beta(0) = 1$)

$$d\beta(t) = r(S(t), t)\beta(t)dt.$$

- Absence of arbitrage: there not exists any possibility of certain instantaneous gain without risk.
- Completeness of markets[4]: a technical assumption that holds if, and only if, the dispersion matrix σ in (2.1) is full rank.
- Frictionless markets: the transaction costs are nought and there are not short selling constraints.

Under these assumptions, each T-claim can be perfectly replicated (or hedged), meaning that the future payoff $g(S(T))$ can be obtained investing and trading in other assets the price of the derivative at time t. In other words, the fair price of the derivative is given by the cost of the replicating strategy. This is a consequence of the fact that discounted (or deflated) asset prices, in particular the option price, are martingales under a probability measure (equivalent to $\mathbf{P}$) associated with the choice of the discount factor, called risk-neutral measure. For any choice of the numeraire[5], and we always use $\beta(t)$, there is a unique stochastic discount process. The associated risk neutral measure is also unique.

The Generalized Black-Scholes Model

This large class of models is represented by arbitrage models in which the underlying $S(t)$ is thought as a set of d traded, risky assets. Typically in this context, the coefficients of the SDE (2.1) are in the form $\mu_i(S(t), t) = b_i(S(t), t)S_i(t)$ and $\sigma_{ij}(S(t), t) = \Sigma_{ij}(S(t), t)S_i(t)$ for all i, j, i.e., one models the instantaneous returns $dS_i(t)/S_i(t)$ instead of the prices $S_i(t)$. The market is given by the underlying, the option and the riskless asset. It turns out that, if this market is complete, the derivative is a redundant asset (i.e., it does not generate new investment opportunities).

The central concept is that of risk neutral (or equivalent martingale) measure: a measure $\mathbf{Q}$, equivalent to the "statistical" measure $\mathbf{P}$, such that all the asset in the market are $\mathbf{Q}$-martingales. The so-called Fundamental theorem of Asset pricing says, roughly speaking, that

[4] The market is the set of traded assets considered.

[5] A numeraire is an asset in the model relative to which all other assets are evaluated. E.g., here we value the assets in units of the money market account. Any asset with strictly positive price can be taken as numeraire.

under the conditions above the risk neutral measure is unique and the derivative price is given by

$$V(S(t),t) = E^{\mathbf{Q}}\left[g(S(T))\exp\left(-\int_t^T r(S(s),s)ds\right)|\mathcal{F}_t\right], \tag{3.11}$$

where $E^{\mathbf{Q}}$ is the expectation under $\mathbf{Q}$. To be precise, it is proven in [HaPl81] ([HaKr79] in discrete time) that the existence of a risk neutral measure implies the absence of arbitrage, and in [HaPl83] that, under the absence of arbitrage, completeness is equivalent to the uniqueness of the risk neutral measure. The further, more interesting implication that the absence of arbitrage ensures the existence of a discount factor (i.e., a risk neutral measure) requires some technicalities. The definitive result is given by [DeSc94]. A simpler result holds under the notion of approximate arbitrage (for the definition see [Cla93], or [Duf01], page 120):

Theorem 6 *Suppose there is a bounded short rate process $r(S(t),t)$. Then there is no approximate arbitrage if, and only if, there is an equivalent martingale measure. Furthermore, the stochastic discount factor is given by $1/\beta(t)$.*

From this result we see that the discounted option price is actually a $\mathbf{Q}$-martingale. In fact, from $V(S(T),T) = g(S(T))$, we have

$$V(S(t),t)/\beta(t) = E^{\mathbf{Q}}\left[V(S(T),T)/\beta(T)|\mathcal{F}_t\right].$$

On the other hand, all the discounted assets must be $\mathbf{Q}$-martingales. By Ito's lemma applied to $Z(t) := S(t)/\beta(t)$,

$$dZ(t) = \left(-r(S(t),t)Z(t) + \frac{\mu(S(t),t)}{\beta(t)}\right)dt + \frac{\sigma(S(t),t)}{\beta(t)}dW(t),$$

so that, under technical conditions, by the Girsanov and the Martingale representation theorems, there exists a measure $\mathbf{Q}$ and a Brownian motion $W^{\mathbf{Q}}$ on $\mathbb{R}^d$ such that $dZ(t) = \frac{\sigma(S(t),t)}{\beta(t)}dW^{\mathbf{Q}}(t)$. Since $S(t) = Z(t)\beta(t)$, by Ito's formula

$$dS(t) = r(S(t),t)S(t)dt + \sigma(S(t),t)dW^{\mathbf{Q}}(t). \tag{3.12}$$

This also explains the name "risk neutral" for $\mathbf{Q}$: under $\mathbf{Q}$, all the risky assets have the same expected return r. The original drifts, that describe the investors attitudes toward risk, enter in the problem only through the assets prices $S(t)$. Note that if some assets S_i pay a dividend yield δ_i, in the equations for S_i in (3.12) there is $(r-\delta_i)$ in place of r (for other extensions see, e.g., [Hul06]).

Under our assumption 7., by (3.12), (3.11), the Markov property and the Feynman-Kac theorem we have that $V(S(t),t)$ is in $C(H)\cap C^{2,1}(\mathbb{R}^d\times[0,T))$ and solves the Cauchy problem

$$\begin{array}{ll} \left(L_{BS}(t)V\right)(x,t) = 0, & (x,t)\in\mathbb{R}^d\times[0,T), \\ V(x,T) = g(x), & x\in\mathbb{R}^d, \end{array} \tag{3.13}$$

where $L_{BS}(t) := \frac{\partial}{\partial t} + r(x,t)\sum_i x_i\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j} a_{ij}(x,t)\frac{\partial^2}{\partial x_i\partial x_j} - r(x,t)$. When we are in this pricing context we could directly assume that $\mathbf{P}=\mathbf{Q}$ and that the SDE (2.1) is specified as (3.12).

Note that, given the discount factor (which changes with the choice of the numeraire), the risk neutral probability can be recovered. This is crucial if one wants to price T-claims by Monte Carlo methods when d is high and the transition density is not known in closed form (e.g., see [Gla04]). We will follow in the second part of the thesis a complementary approach, approximating both the transition density and the derivative price by means of perturbation techniques.

The Original Black-Scholes Model

Consider the additional hypothesis that the spot process is constant, $r \in \mathbb{R}$, each assets $S_i(t)$ is governed by a geometric Brownian motion (2.7) and the dispersion matrix is given by $\sigma_{ij}(x,t) = x_i \Sigma_{ij}$, where $\Sigma \in \mathbb{R}^{d \times l}$. These are the assumptions in [BlSc73] extended to d dimensions but, for the sake of simplicity, we consider the scalar case $d = l = 1$ as in [BlSc73]. Recall that we suppose that the underlying price is $S(t) > 0$ ($a.s.$). This setting allows us to simplify a lot the pricing PDE (3.13), reducing it to the heat equation with some suitable transformations. Similar reduction techniques can be found, e.g., in [WHD95] and [Sal04].

First note that, by the Fundamental theorem of asset pricing we have that $V(S(t),t) = e^{-r(T-t)} E\left[g(S(T))|\mathcal{F}_t\right]$. Setting $r = 0$, we see from the Feynman-Kac theorem that $v(S(t),t) = E\left[g(S(T))|\mathcal{F}_t\right]$ satisfies the PDE

$$\partial_t v(x,t) + \frac{1}{2} x^2 \Sigma^2 \partial_{xx} v(x,t) = 0, \tag{3.14}$$

with the terminal condition $v(x,T) = g(x)$. We can recover the option price from $V(S(t),t) = e^{-r(T-t)} v(S(t),t)$. The same arguments are true if the spot rate is a deterministic function of t, say $r = r(t)$. The second, crucial simplification is obtained by the change of variable $y := \ln x$. From (3.14) we have that $w(y,t) = v(e^y,t)$ satisfies (we have, e.g., $\partial_x v(x,t) = \partial_y w(y,t) \frac{dy}{dx} = \partial_y w(y,t) \frac{1}{x}$)

$$\partial_t w(y,t) + \frac{1}{2} \Sigma^2 \left[\partial_{yy} w(y,t) - \partial_y w(y,t)\right] = 0 \tag{3.15}$$

and the condition $w(y,T) = g(e^y)$. (3.15) is a constant coefficients PDE. Finally, we have that $u(y,t) = w(y,t) e^{-\frac{1}{2}y - \frac{\Sigma^2}{8} t}$ solves the heat equation (backward in time)

$$\partial_t u(y,t) + \frac{1}{2} \Sigma^2 \partial_{yy} u(y,t) = 0. \tag{3.16}$$

Note that this extends to the d-dimensional case (with $\Sigma\Sigma^\top$ a $(d \times d)$ matrix and the Laplacian $\sum_{i,j=1}^d \frac{\partial^2}{\partial y_i \partial y_j}$ in place of ∂_{yy}).

If we model the returns and all the entries of the dispersion matrix $\Sigma_{ij}(S(t),t) S_i(t)$ are such that $\Sigma_{ij}(S(t),t)$ are bounded, then the log-transform $y = \ln x$ (componentwise) applied to the PDE $L_{BS}(t)V = 0$, with $r = 0$, reduces the problem to a bounded coefficients PDE, for which the parametrix method applies (see Chapter 10). But we remark that this boundedness condition is not necessary in the generalized BS model to have a closed-form solution of the option price. E.g., the so-called CEV (constant elasticity of variance) model of [Cox75] and [CoRo76], defined by ($d = 1$) $\Sigma(x,t) = \sigma_{cev} x^{-\gamma}$, where $\sigma_{cev} > 0$ and $0 \le \gamma < 1$, also allows closed-form formulas for vanilla derivative prices (see, e.g., [Epp00]), but the dispersion coefficient $\Sigma(x,t)x = \sigma_{cev} x^{1-\gamma}$ is clearly unbounded. For $\gamma = 0$ we find the classical BS model. The condition $0 \le \gamma < 1$ ensures that 0 is an unattainable barrier for any $S(0) > 0$ ($a.s.$); the same is true for any $S(0) \ge 0$ ($a.s.$) if we take more restrictively $0 < \gamma < 1$ (we are ruling out the geometric Brownian motion that cannot start from $S(0) = 0$). As the name of this model suggests, the parameter γ defines a (constant) inverse relationship between the underlying price and the volatility level $\Sigma(S(t),t)S(t) = \sigma_{cev}(S(t))^{1-\gamma}$.

Term Structure Derivatives

A term structure model is a model for the riskless spot rate process $r(S(t),t)$. The underlying SDE (2.1) is usually interpreted as a set of economic factors that drive the spot rate evolution. A term structure derivative is defined by a payoff such as $g(r(S(T),T))$. The most important derivative in this context is the so-called zero-coupon bond (or zero) maturing at a future date $T > t$: its payoff is a lump-sum payment, say 1 unit of account at T, and no dividends. Zeros are assumed riskless and are the basic assets. Denote the price at t of a zero maturing at T by $B(T,t)$[6]. Under the financial assumption above, the following version of the Fundamental theorem of Asset pricing holds

Theorem 7 *A term structure model is free of arbitrage if, and only if, there exists a risk-neutral measure* $\mathbf{Q}$ *such that, for each* $s \in (t,T]$*, the process* $B(s,t)/\beta(t)$ *is a* $\mathbf{Q}$*-martingale.*

Note that again $1/\beta(t)$ (in this context $\beta(t)$ is called accumulation factor) is the discount factor. The first consequence is that the price of any term structure derivative maturing at T (in particular $B(T,t)$) is given by

$$V(S(t),t) = E^{\mathbf{Q}}\left[g(r(S(T),T))\exp\left(-\int_t^T r(S(\tau),\tau)d\tau\right)|\mathcal{F}_t\right]. \tag{3.17}$$

The second consequence of the theorem, because it completely characterize the absence of arbitrage, is that model (2.1) is sometimes (for pricing issues) directly given under the risk neutral measure, i.e., $\mathbf{P} = \mathbf{Q}$. From assumption 7., we have that $V(S(t),t)$ is in $C(H) \cap C^{2,1}(\mathbb{R}^d \times [0,T))$ and solves the Cauchy problem

$$\begin{array}{ll} (L_r(t)V)(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [0,T), \\ V(x,T) = \varphi(x), & x \in \mathbb{R}^d, \end{array} \tag{3.18}$$

where $L_r(t) := \frac{\partial}{\partial t} + A(t) - r(x,t)$, $A(t)$ is given in (3.1) and $\varphi := g \circ r$.

We stress once again that the first step to cope with the solution of (3.18) is to apply any transformation that simplifies the problem, as we did with the BS example. An important simplification of (3.17) and (3.18) can be achieved with a change of numeraire, a technique that gives a great flexibility in option pricing. With a stochastic interest rates, a good idea is to take as numeraire a zero with maturity T (the maturity of our option), which price is $B(T,t) = E^{\mathbf{Q}}\left[\exp\left(-\int_t^T r(\tau)d\tau\right)|\mathcal{F}_t\right]$ (for ease of notation we write here $r(t)$, $V(t)$, etc.). Define a new measure $\mathbf{Q}_T$ by means of the Radon-Nikodym derivative (recall that $B(T,T) = 1$)

$$\frac{d\mathbf{Q}_T}{d\mathbf{Q}} := \frac{\exp\left(-\int_t^T r(\tau)d\tau\right)}{E^{\mathbf{Q}}\left[\exp\left(-\int_t^T r(\tau)d\tau\right)|\mathcal{F}_t\right]} = \frac{\beta(t)B(T,T)}{\beta(T)B(T,t)}. \tag{3.19}$$

Notice that $\mathbf{Q}_T$ (called T-forward martingale measure) and $\mathbf{Q}$ are equivalent since $\frac{B(T,T)}{\beta(T)}$ is a $\mathbf{Q}$-martingale and clearly $E^{\mathbf{Q}}\left[\frac{d\mathbf{Q}_T}{d\mathbf{Q}}|\mathcal{F}_t\right] = 1$ (in particular, if r is deterministic, $\mathbf{Q}_T = \mathbf{Q}$), such that (3.19) is a well defined change of measure. As a consequence, asset prices divided by (i.e., discounted by) $B(T,T) = 1$ are $\mathbf{Q}_T$-martingales. In fact

$$V(t) = E^{\mathbf{Q}}\left[g(r(T))\frac{\beta(t)}{\beta(T)}|\mathcal{F}_t\right] = B(T,t)E^{\mathbf{Q}_T}\left[g(r(T))|\mathcal{F}_t\right], \tag{3.20}$$

or $V(t)/B(T,t) = E^{\mathbf{Q}_T}[V(T)|\mathcal{F}_t]$. So we can solve the Cauchy problem (3.18) setting the reaction $r(x,t) = 0$ and then use (3.20), where $B(T,t)$ is the observed market price of the the zero[7]. This technique, if the zeros with the required maturity are traded, can be used also in the Generalized BS model with a stochastic short rate. However, it is crucial that the term-structure model produces zero prices with a (theoretical) perfect fitting of the observed zero prices. This

[6]Sometimes it could be useful to write explicitly the dependence on r as $B = B(r;T,t)$. Here we use the usual notation.

[7]If the price $B(T,t)$ is not available we can solve (3.18) with $\varphi = 1$. Anyway, this solution and the market price should always agree.

explains why the most used term-structure models are those designed to match the observed term-structure (e.g., the Hull-White model, see Section 2.3).

We make some additional remarks. i) Generally, the SDE (2.1) (under the risk-neutral measure) is taken as an affine SDE. In this case one makes the additional assumption that r is an affine function of $S(t)$. This fact implies that $B(s,t)$ have a nice exponential-affine functional form and, as a consequence, sometimes a closed-form solution. For a review of term-structure affine models, see [Duf01]. ii) If we consider derivatives paying also an instantaneous payoff process $\psi(x,t) = f(r(x,t))$, we have to add to the PDE in (3.18) an inhomogeneous term, i.e., $(L_r(t)V)(x,t) = \psi(x,t)$. In fact, in absence of arbitrage, the average instantaneous increase rate of the derivative (i.e., $\left(\frac{\partial}{\partial t} + A(t)\right) V$) plus its payoff ψ must equal rV. But note that the assumption that $f = f(r)$ is just an approximation, since typically payoffs of interest-rate derivatives are defined in terms of some discretely-compounded rate (e.g., the LIBOR rate). iii) In the scalar case $d = l = 1$, the basic SDE (2.1) models directly the short rate, i.e., it takes the form (under $\mathbf{Q}$)

$$dr(t) = \mu(r(t),t)dt + \sigma(r(t),t)dW(t).$$

Hence the operator $L_r(t)$ in (3.18) is given by $L_x(t) := \frac{\partial}{\partial t} + A(t) - x$.

Chapter 4

Transition Semigroups

We explore the properties of the conditional expectation operator and the connections between diffusion processes, semigroup theory and differential equations. This study will be carry out in connection with the applications in the second part of the thesis. Main references of the chapter are [Dyn65], [EtKu86], [KaTa81], [Fre85] and [Lam77].

4.1 Feller Semigroups

Assumption 7. means, from an abstract (i.e., functional analytic) point of view, that the solution operator[1] of (3.9) is given by the conditional expectation operator of the diffusion. In this sense (see also Appendix A.3 and Chapter 11), in this section we study the properties of the solution operator of the Kolmogorov Backward equation (i.e., we set $c \equiv 0$ in (3.9)) in the non-autonomous case. We will see that: i) The conditional expectation operator defines a two-parameters semigroup. ii) Under natural conditions, the transition semigroup is a strongly continuous semigroup of contractions in the space of continuous functions vanishing at ∞, with generator $A(t)$ given in (3.1). iii) We can reverse the problem and generate the semigroup from $A(t)$, even if $A(t)$ is degenerate. iv) It is possible to extend the transition semigroup to some L^p spaces.

The other two sections of the chapter deal with the concept of adjoint semigroup to the transition semigroup and with the extension of the results to multiplicative semigroups, which arise when $c \neq 0$ in (3.9).

The Semigroup Property

Let $P(x,t;B,s)$, where $0 \le t < s \le T$, $x \in \mathbb{R}^d$ and B is a Borel set of $\mathbb{R}^d$, be the the family of transition functions of the diffusion $S(t)$ solution of (2.1). By definition (e.g., see [Fri76], page 18, but this fact can be proved from (2.2)-(2.4)), any Markov transition function satisfies the Chapman-Kolmogorov equation

$$P(x,t;B,s) = \int_{\mathbb{R}^d} P(x,t;dy,r)P(y,r;B,s), \tag{4.1}$$

for any $r \in (t,s)$. Suppose also that $P(x,t;B,s) \to \chi_B(x) =: \delta_x B$ (the Dirac's measure) as $s \downarrow t$, in accordance to our assumption 4..

Consider the operator (sometimes called transition operator)

$$U(t,s)g(x) := E\left[g(S(s))|S(t) = x\right] = \int_{\mathbb{R}^d} g(y)P(x,t;dy,s), \tag{4.2}$$

[1] In this chapter we use many definitions and results from appendices A.1-A.3.

well defined for any $g \in B(\mathbb{R}^d)$, the space of all measurable and bounded functions $\mathbb{R}^d \to \mathbb{R}$. Then the Chapman-Kolmogorov equation is equivalent to the fact that the transition operator has the semigroup property

$$U(t,s) = U(t,r)U(r,s), \quad U(t,t) = I \ , \tag{4.3}$$

for any $r \in (t,s)$. In other words, the conditional expectation operator defines a two-parameters semigroup of operators in $B(\mathbb{R}^d)$. To see this, let B be a Borel set of $\mathbb{R}^d$. From (4.2) $U(t,s)\chi_B(x) = P(x,t;B,s)$; so by (4.1) and (4.2) again

$$\begin{aligned} U(t,s)\chi_B(x) &= \int_{\mathbb{R}^d} P(x,t;dy,r)U(r,s)\chi_B(y) \\ &= U(t,r)U(r,s)\chi_B(x). \end{aligned}$$

By linearity and monotone convergence, the first equality in (4.3) holds for any $g \in B(\mathbb{R}^d)$, while the second equality is a consequence of the additional assumption about the transition function.

It is well known that (4.1) is equivalent also to the Markov property. In fact, we can read (4.3) as $E\left[g(S(s))|S(t)=x\right] = E\left[E\left[g(S(s))|S(r)\right]|S(t)=x\right]$. This is useful also to understand the order of the composition product and of the times in (4.3) (i.e., smaller times on the left). This is related to the fact that the Cauchy problems (3.5) and (3.9) in the previous chapter are given backward in time, with a terminal condition[2].

Usually a two-parameters semigroup is called *evolution family* or *propagator*. We refer to $U(t,s)$ as the *transition propagator* associated to $S(t)$ in (2.1). The name semigroup is reserved in the case that the operator $U(t,s)$ depends only on the single parameter $\tau := (s-t) \geq 0$, writing $T(\tau) := U(t,s)$, called transition semigroup. This is the case if $S(t)$ is autonomous, when actually $P(x,t;B,s) = P(x,B;\tau)$ and $E\left[g(S(t+\tau))|S(t)=x\right] = E\left[g(S(\tau))|S(0)=x\right]$. Then the Chapman-Kolmogorov equation takes the form

$$P(x,B;\eta+\zeta) = \int_{\mathbb{R}^d} P(x,dy;\eta)P(y,B;\zeta),$$

and, in turn, the semigroup property is given by

$$T(\eta+\zeta) = T(\eta)T(\zeta), \quad T(0) = I \ , \tag{4.4}$$

for all $\eta, \zeta \geq 0$. Note that in this case, no time ordering problem arises and the change of variable $\tau = (s-t)$ is such that the new time τ runs forward.

The Feller Property

We study the family of operators $U(t,s)$, $0 \leq t < s \leq T$, in the space of continuous functions $\mathbb{R}^d \to \mathbb{R}$ vanishing at ∞, that we denote $\widehat{C}(\mathbb{R}^d)$. Performing an Alexandroff one-point compactification (see, e.g., [AlBo99]) by $\mathbb{R}^{d*} = \mathbb{R}^d \cup \{\infty\}$, we identify $\widehat{C}(\mathbb{R}^d)$ with $\widehat{C}(\mathbb{R}^{d*})$. If the state space is just a proper subset of $\mathbb{R}^d$ (see section 2.4 and recall assumption 5.), we use the convention to restrict $\widehat{C}(\mathbb{R}^d)$ to the space of continuous functions vanishing at the boundaries of the relevant state space. When necessary, nothing changes if we consider the slightly wider space of functions for which a value at the boundaries of the state space exists finitely (see [KaTa81], pages 289-292). An important point is that $\widehat{C}(\mathbb{R}^d)$ can be made into a Banach space introducing the norm $\|g\|_\infty = \sup_{x\in\mathbb{R}^d}|g(x)|$. Note that any function in $\widehat{C}(\mathbb{R}^d) = \widehat{C}(\mathbb{R}^{d*})$ is determined by its restriction as a function $\mathbb{R}^d \to \mathbb{R}$, so that (in terms of the restrictions)

[2]Sometimes (e.g., Chapters 5, 10 and 11) we will set the very same Cauchy problems forward in time, by means of the change of variable $t \mapsto s-t$. Then we will need to reverse also the time ordering of the operators to preserve the right composition product (e.g., see formula (A.44)).

$C_0(\mathbb{R}^d) \subset \widehat{C}(\mathbb{R}^d) \subset C_b(\mathbb{R}^d) \subset B(\mathbb{R}^d)$, where $C_0(\mathbb{R}^d)$ is the subspaces of $C(\mathbb{R}^d)$ of functions with compact support. Note also that we use $\mathbb{R}^d$ as space state, but we could consider any locally compact Polish space.

The evolution family $U(t,s)$ is called a *Feller propagator* if it has the following two additional conditions: i) $U(t,s) : \widehat{C}(\mathbb{R}^d) \to \widehat{C}(\mathbb{R}^d)$; ii) $U(t,t+h)g(x) \to g(x)$ (pointwise) as $h \downarrow 0$, for all $g \in \widehat{C}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. These are equivalent to ask that $(x,t) \mapsto U(t,t+h)g(x)$ is jointly continuous (see [Bal00], Chapter 5). The strongest consequence is that $U(t,s)$ is strongly continuous, i.e., $\|U(t,t+h)g - g\|_\infty \to 0$ as $h \downarrow 0$, for all $g \in \widehat{C}(\mathbb{R}^d)$ (see [Kal97], page 369, for a probabilistic interpretation of the Feller conditions and of the strong continuity). Another consequence of the Feller property is that $U(t,s)$ is a positive propagator of contractions[3]. A positive operator maps positive functions into positive functions, which is clearly true for $U(t,s)$ in $\widehat{C}(\mathbb{R}^d)$. To see that $U(t,s)$ is also a contraction, recall that the operator norm (the norm of $\mathcal{L}(\widehat{C}(\mathbb{R}^d))$) is given by $\|U(t,s)\| = \sup_{g\in\widehat{C}(\mathbb{R}^d),\|g\|_\infty\le 1} \|U(t,s)g\|_\infty$ and notice that

$$|U(t,s)g(x)| \le \int_{\mathbb{R}^d} |g(y)|\, p(x,t;y,s)dy \le \|g\|_\infty \int_{\mathbb{R}^d} p(x,t;y,s)dy.$$

In particular, since we asked in 4. that p is conservative, $U(t,s)1 = 1$[4].

We remark that the Feller property can be strengthened replacing i) with i') $U(t,s) : B(\mathbb{R}^d) \to \widehat{C}(\mathbb{R}^d)$. This is necessary for the version of the first Kolmogorov equation with a terminal datum in $B(\mathbb{R}^d)$ (see [KaTa81], 15.11).

We can introduce now the fundamental concept of *infinitesimal generator* of the Feller propagator $U(t,s)$ (also called generator of the diffusion $S(t)$). If $g \in \widehat{C}(\mathbb{R}^d)$, provided the limit exists, the family of linear operators $\mathcal{A}(t)$ defined by

$$\mathcal{A}(t)g := \lim_{h\downarrow 0} \left\| \frac{U(t,t+h)g - g}{h} \right\|_\infty, \tag{4.5}$$

is called the infinitesimal generator of $U(t,s)$. For each fixed t, the set of functions $D(\mathcal{A}(t)) \subset \widehat{C}(\mathbb{R}^d)$ such that the limit in (4.5) exists is called the domain of $\mathcal{A}(t)$. We will always ask that $D(\mathcal{A}(t)) = D$, for all $t \in [0,T]$. The full characterization of D and of $\mathcal{A}(t)$ are arduous without additional hypotheses[5] and rarely feasible for $d > 1$. But we can give the explicit form of $\mathcal{A}(t)$ for a large class of functions in D. If $g \in C_0^2(\mathbb{R}^d)$, then $g \in D$ and $\mathcal{A}(t) = A(t)$ as in (3.1), i.e.,

$$A(t) = \sum_{i=1}^d \mu_i(x,t)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{4.6}$$

The proof of this fact hinges on the definition of diffusion. From (2.2) we have, for all $x \in \mathbb{R}^d$,

$$\lim_{h\downarrow 0} \frac{U(t,t+h)g(x) - g(x)}{h} = \lim_{h\downarrow 0} \frac{1}{h} \int_{|x-y|\le\epsilon} [g(y) - g(x)]\, p(x,t;y,t+h)dy.$$

By Taylor theorem applied to $g(y)$ around x we have

$$g(y) - g(x) = \sum_{i=1}^d (y_i - x_i)\frac{\partial g(x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d (y_i - x_i)(y_j - x_j)\frac{\partial^2 g(x)}{\partial x_i \partial x_j} + o(|x-y|^2).$$

[3] This implies also that $U(t,s)$ is a family of continuous operators.

[4] If the diffusion is not conservative, one should take into account a cemetery state (see [KaTa81], Sections 15.1 and 15.11, and Section 4.3 below). Then $U(t,s)1 \le 1$.

[5] See [Man68], [KaTa81], Section 15.12, and [HST98].

Therefore, by (2.3) and (2.4) we have

$$\lim_{h\downarrow 0}\frac{U(t,t+h)g(x)-g(x)}{h}=A(t)g(x)+\lim_{h\downarrow 0}\frac{1}{h}\int_{|x-y|\le\epsilon}o(|x-y|^2)p(x,t;y,t+h)dy,$$

and we get the claim since we assumed 4. and the Feller property. In other words, for $g\in C_0^2(\mathbb{R}^d)$, the generator of the diffusion is given by the operator that arises from the Ito's formula. In fact, we can prove (4.5) also noticing that if $g\in C_0^2(\mathbb{R}^d)$, we can apply Ito's formula and obtain

$$g(S(t+h))=g(S(t))+\int_t^{t+h}(A(\tau)g)\,(S(\tau))d\tau+\sum_{i,k}\int_t^{t+h}\sigma_{i,k}\frac{\partial g}{\partial x_i}dW_k(\tau),\tag{4.7}$$

hence, taking expectation we have $E\left[\sum_{i,k}\int_t^{t+h}\sigma_{i,k}\frac{\partial g}{\partial x_i}dW_k(\tau)|S(t)=x\right]=0$, so that, again by (2.3), (2.4) and the Feller property, we get (4.5). These steps should be made rigorous taking into account a suitable sequence of stopping times (see [Øks06], pages 123 and 140). This proves also the following extension of the Dynkin's formula (for the original formula, e.g., see [KaTa81], page 298) for $g\in C_0^2(\mathbb{R}^d)$,

$$U(t,s)g(x)=g(x)+\int_t^s U(t,\tau)A(\tau)g(x)d\tau,\tag{4.8}$$

where the integral and the conditional expectation operators commute by the Fubini's theorem (see [Ash72], page 101) since, for $g\in D$, $U(t,\tau)A(\tau)g$ exists finite (actually, (4.8) holds with $\mathcal{A}(t)$ in place of $A(t)$ for all $g\in D(\mathcal{A}(t))$). Notice also that $C_0^2(\mathbb{R}^d)$ is dense in $\widehat{C}(\mathbb{R}^d)$, so that the generator of $U(t,s)$ is of course densely defined.

We conclude this section with a useful property of the generator $\mathcal{A}(t)$, called positive maximum principle. If $g\in D$ and $x_0\in\mathbb{R}^d$ is such that $g(x_0)=\sup_x g(x)\ge 0$, then we have $\mathcal{A}(t)g(x_0)\le 0$. This follows from $U(t,s)g(x_0)=\int_{\mathbb{R}^d}g(y)p(x_0,t;y,s)dy\le g(x_0)\int_{\mathbb{R}^d}p(x_0,t;y,s)dy$, so that $\frac{1}{h}\left[U(t,t+h)g(x_0)-g(x_0)\right]\le 0$. The positive maximum principle can be used to prove that the diffusion matrix $a(x,t)$ associated to $S(t)$ is necessarily positive semi-definite (see [Bal00], page 99).

Generation of Feller Semigroups

In the applications, as in this thesis, one starts from the a diffusion $S(t)$ in (2.1), in other words from its infinitesimal parameters, and then asks when the operator $A(t)$ in (4.6) actually generates a Feller semigroup. To answer it is better if we start from the autonomous case, i.e., $A(t)=A$. Clearly, all the results in the previous section hold. We can recover the general case appending the time variable to the space variables, a trick suggested, e.g., by [Fre85], page 55. However, this procedure does not transform the propagator $U(t,s)$ into a one-parameter semigroup and, furthermore, the diffusion matrix in the new augmented state space will be degenerate (see below and Chapter 9).

Recall that a Feller transition semigroup $T(\tau)$, $\tau\ge 0$, in $\widehat{C}(\mathbb{R}^d)$ is a positive, C_0 semigroup of contractions and its generator is densely defined and satisfies the positive maximum principle. On the other hand any linear operator A in $\widehat{C}(\mathbb{R}^d)$ which satisfies the positive maximum principle is dissipative. To prove this claim, let $g\in D(A)$ and $\lambda>0$. There exists a $x_0\in\mathbb{R}^d$ such that $|g(x_0)|=\|g\|_\infty$. Let $g(x_0)\ge 0$, otherwise replace g with $-g$. Since $g(x_0)=\sup_x g(x)\ge 0$, $A(t)g(x_0)\le 0$ and hence

$$\|\lambda g-Ag\|_\infty\ge\lambda g(x_0)-Ag(x_0)\ge\lambda g(x_0)=\|\lambda g\|_\infty\,.$$

It holds the following special case of the Lumer-Phillips theorem (see Appendix A.3)

Theorem 8 *The closure of a linear operator A in $\widehat{C}(\mathbb{R}^d)$ generates a positive, C_0-semigroup of contractions $T(\tau)$, $\tau \geq 0$, in $\widehat{C}(\mathbb{R}^d)$ if, and only if,*

a *$D(A)$ is dense in $\widehat{C}(\mathbb{R}^d)$.*

b *$R(\lambda I - A)$ is dense in $\widehat{C}(\mathbb{R}^d)$ for some $\lambda > 0$.*

c *A satisfies the positive maximum principle.*

For the completion of the proof, i.e., that the conditions of the theorem imply that $T(t)$ is also positive, see [EtKu86], page 165.

This theorem is the definitive answer to our question about the generation of Feller semigroups. The new question is if it is possible to give conditions on the infinitesimal parameters of the diffusion such that the closure of the operator that arises from the Ito's formula, i.e.,

$$A = \sum_{i=1}^{d} \mu_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{4.9}$$

generates a Feller transition semigroup in $\widehat{C}(\mathbb{R}^d)$. To answer, we present some results from Chapter 8 of [EtKu86]. The following theorem is a slight extension of a classical result of Dynkin (see [Dyn65], theorem 5.11)

Theorem 9 *Let $a_{ij}(x)$ and $\mu_i(x)$ be bounded for all $i, j = 1, \ldots, d$ and $x \in \mathbb{R}^d$. Suppose also that a, μ are Holder continuous, exponent $\alpha \in (0, 1]$, and that $a(x)$ is uniformly elliptic in $\mathbb{R}^d$. Then the closure of A in (4.9) on the core $C_0^\infty(\mathbb{R}^d)$ generates a Feller transition semigroup in $\widehat{C}(\mathbb{R}^d)$.*

An useful criterion for subspaces to be cores of the generator is given by the following result. Let $R(t)$ be a C_0 semigroup in a Banach space X. Any subspace $G \subset X$ invariant under $R(t)$, i.e., $R(t)G \subset G$, and dense in X is a core for the generator of $R(t)$. Then possible cores for A are $C_0^\infty(\mathbb{R}^d)$ and $\widehat{C}(\mathbb{R}^d) \cap C^2(\mathbb{R}^d)$, both dense in $\widehat{C}(\mathbb{R}^d)$ (see [EtKu86], page 370) and invariant under $T(t)$ (using the fact that $T(t)$ has the Feller property). The best choice is clearly to restrict ourselves to the smaller core $C_0^\infty(\mathbb{R}^d)$, since the restriction on a core completely determines a closed operator on its domain.

The situation is different in the scalar case. In fact, under the non-degeneracy of the diffusion $a(x) > 0$ for all x, it is possible (see [EtKu86], page 367, and also [Man68] for the details) to allow for unbounded infinitesimal parameters and to study all the possible boundary behavior of the diffusion and the associated characterizations of $D(A)$.

We turn now to the central problem of extending the generation results when the diffusion $S(t)$ is non-autonomous. We will do this in detail when we present in Chapter 9 the results about general small-time expansions. The idea is to define a new diffusion $Z(t) := (S(t), t)$ (a $d + 1$ column vector) treating the time variable as a new space variable. Notice that to ensure that $Z(t)$ has a strong solution we need that the coefficients of $S(t)$ enjoy a lot more of regularity. In particular, a sufficient condition is that $a(x, t)$ and $\mu(x, t)$ are Lipschitz continuous in (x, t). However, to apply the result above we need also that $a(x, t)$ is uniformly elliptic for all $(x, t) \in \mathbb{R}^{d+1}$. A condition that cannot be satisfied by $Z(t)$ defined above.

So consider a possibly degenerate diffusions $S(t)$. It holds the following very strong result (see [EtKu86], Section 8.1)

Theorem 10 *Let $a_{ij}(x) \in C^2(\mathbb{R}^d)$ with $\partial_{x_k} \partial_{x_l} a_{ij}$ be bounded for all $i, j, k, l = 1, \ldots, d$, and let $\mu(x)$ be Lipschitz continuous. Then the closure of A in (4.9) on the core $C_0^\infty(\mathbb{R}^d)$ generates a Feller transition semigroup in $\widehat{C}(\mathbb{R}^d)$.*

Extension to L^p

In some applications the space $\widehat{C}(\mathbb{R}^d)$ is problematic, because we want to consider conditional expectations of possibly unbounded functions. However, we can extend the Feller propagator $U(t,s)$ to a contraction semigroup to some $L^p(\mathbb{R}^d, dQ)$, $p \geq 1$, space, where Q is a given σ-finite measure on the Borel sets of $\mathbb{R}^d$. The crucial fact is that is well known (see [AlBo99], page 423) that $C_0(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d, dQ)$, $p \geq 1$, so that also $\widehat{C}(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d, dQ)$.

We say that the Feller evolution family $U(t,s)$ is symmetric w.r.t. Q if, for any $0 \leq t < s \leq T$, $\langle U(t,s)g, f\rangle_{L^2(\mathbb{R}^d,dQ)} = \langle g, U(t,s)f\rangle_{L^2(\mathbb{R}^d,dQ)}$ (where $\langle g, f\rangle_{L^2(\mathbb{R}^d,dQ)} = \int_{\mathbb{R}^d} g(x)f(x)Q(dx)$), for all $g, f \in B(\mathbb{R}^d)$. The symmetry property implies

$$\langle U(t,s)x, 1\rangle_{L^2(\mathbb{R}^d,dQ)} = \langle x, U(t,s)1\rangle_{L^2(\mathbb{R}^d,dQ)} = \langle x, 1\rangle_{L^2(\mathbb{R}^d,dQ)}.$$

Recall the representation $U(t,s)g(x) = \int_{\mathbb{R}^d} g(y)P(x,t;dy,s)$. It holds the following result

Proposition 11 *If $U(t,s)$ is a Feller transition propagator in $\widehat{C}(\mathbb{R}^d)$, symmetric with respect to the σ-finite measure Q on the Borel sets of $\mathbb{R}^d$, then $U(t,s)$ can be extended to a strongly continuous contraction propagator in $L^p(\mathbb{R}^d, dQ)$, $p \geq 1$.*

To see that the proposition is true, we have, for the strong continuity of $U(t,s)$, that $\|U(t,t+h)g - g\|_\infty \to 0$ as $h \downarrow 0$ for all $g \in \widehat{C}(\mathbb{R}^d)$. Then the Lebesgue's dominated convergence theorem implies that also for the extension $U(t,s)$ (with abuse of notation we do not distinguish between the Feller transition semigroup and its extension) $\|U(t,t+h)g - g\|_{L^p(\mathbb{R}^d,dq)} \to 0$ as $h \downarrow 0$ for all $g \in \widehat{C}(\mathbb{R}^d)$. By the density of $\widehat{C}(\mathbb{R}^d)$ in $L^p(\mathbb{R}^d, dQ)$, the strong continuity holds in $L^p(\mathbb{R}^d, dQ)$. The same density argument implies the semigroup property (4.3) for the extension. It remains to prove that the extension $U(t,s)$ is also a contraction. By the Holder's inequality applied to $U(t,s)g(x)$ we have

$$\left|\int_{\mathbb{R}^d} g(y)P(x,t;dy,s)\right|^p \leq \int_{\mathbb{R}^d} |g(y)|^p \, P(x,t;dy,s).$$

Then, by the symmetry property,

$$\int_{\mathbb{R}^d} |U(t,s)g(x)|^p \, Q(dx) \leq \int_{\mathbb{R}^d} U(t,s) \, |g(x)|^p \, Q(dx) = \int_{\mathbb{R}^d} |g(x)|^p \, U(t,s)1Q(dx),$$

i.e., $\|U(t,s)g\|_{L^p(\mathbb{R}^d,dq)} \leq \|g\|_{L^p(\mathbb{R}^d,dq)}$, proving the claim.

The importance of this result is that we can extend the class of functions in the domain D of the generator $\mathcal{A}(t)$ such that the representation $\mathcal{A}(t) = A(t)$, $A(t)$ as in (4.6), holds; in particular, we can now consider function $g \in D$ not necessarily bounded (notice that now $D \subset L^p(\mathbb{R}^d, dQ)$). Let $D_1 = L^p(\mathbb{R}^d, dQ) \cap C^2(\mathbb{R}^d)$. For each $g \in D_1$ the Ito's lemma applies, so formula (4.7) still makes sense. What we cannot infer here is that $E\left[\int_t^{t+h} \sigma_{i,k} \frac{\partial g}{\partial x_i} dB_k(\tau) | S(t) = x\right] = 0$ for all i, k. A sufficient condition for this (see [Fri76], Chapter 4) is

$$E\left[\int_t^{t+h} \left(\sigma_{i,k}(S(\tau),\tau)\frac{\partial g(S(\tau))}{\partial x_i}\right)^2 d\tau | S(t) = x\right] < \infty, \tag{4.10}$$

for all i, k and all $t, h > 0$. Then the stochastic integral in (4.7) is actually a 0-mean martingale. Let D_2 be the set of functions $g \in D_1$ such that (4.10) holds. We have $D_2 \subset D \subset L^p(\mathbb{R}^d, dQ)$ and $\mathcal{A}(t)|_{D_2} = A(t)$ as follows from the proof of (4.8). Condition (4.10) is satisfied if: i) $\sigma_{i,k}$ and $\partial_{x_i} g$ are bounded for all i, k; ii) $g \in C_0^2(\mathbb{R}^d)$ and $\sigma_{i,k}$ continuous but unbounded (as done above); iii) g is polynomial growth, in the case that diffusion is autonomous and μ, σ are of linear growth, or that the tails of Q decrease at least exponentially fast.

We remark that sometimes a *formal* extension of domain of the operator $A(t)$ as in (4.6) is defined without any reference to the space of definition of $A(t)$; in that case, we call the set D_3 of functions $g \in C^2(\mathbb{R}^d)$ satisfying (4.10) the *extended domain* of $A(t)$. This is reasonable, since actually $D_3 \supset C_0^2(\mathbb{R}^d)$, which is always a core for $A(t)$ whatever is the Banach space of definition ($\widehat{C}(\mathbb{R}^d)$ or $L^p(\mathbb{R}^d, dQ)$) of the Markov generator. Moreover, if we can define $\mathcal{A}(t)$ in $L^p(\mathbb{R}^d, dQ)$, we have $D_3 = D_2$.

The main question is how to choose the measure Q in practice. In the next section we see what is the most natural choice, when available, and that, to some extent, it is also possible to take a measure Q dependent on t (namely, we could use $Q(dx,t) = \Pi(dx,t)$ as in formulas (4.11)-(4.13) below).

4.2 The Adjoint Semigroup

There is a second evolution family associated to the transition function of the diffusion $S(t)$ solution of (2.1). Denote $\Pi(.,t)$, $t \geq 0$, the distribution of $S(t)$, i.e., $\Pi(B,t) = \mathbf{P}\left[S(t) \in B\right]$, where B a Borel set of $\mathbb{R}^d$[6]. In particular, $\Pi(.,0)$ is the initial law of the Markov process. It is well known that the knowledge of $\Pi(.,0)$ and of the family of transition functions $P(x,0;B,t)$, $t \geq 0$, is equivalent to the knowledge of $\Pi(.,t)$, $t \geq 0$. This fact, equivalent to the Markov property of the process $(S(\tau), \tau \geq 0)$, can be expressed by the formula

$$\Pi(B,t) = \int_{\mathbb{R}^d} \Pi(dx,0) P(x,0;B,t). \tag{4.11}$$

From (4.11) we see that actually $\Pi(.,t)$ is a measure on the Borel set of $\mathbb{R}^d$ (a probability measure under our assumption 4.): it is non-negative, $\Pi(\mathbb{R}^d,t) = 1$ and inherits the countable additivity of the integral. Thereby we can define an operator $W(t,s)$, $0 \leq t < s$, that maps the probability measure $\Pi(.,t)$ into $\Pi(.,s)$ by the formula

$$\Pi(B,s) = \int_{\mathbb{R}^d} \Pi(dx,t) P(x,t;B,s). \tag{4.12}$$

Then by (e.g.) [GiSk69], Lemma 1, page 300, (4.12) and the Chapman-Kolmogorov equation (4.1), we have that $W(t,s)$ enjoys the semigroup property (4.3). Moreover, $W(t,s)$ is a contraction and (from 4.) $W(t,s)1 = 1$ (with obvious meaning). As a function of t, s, $W(t,s)$ defines a propagator in the space of probability measures on the Borel σ-algebra of $\mathbb{R}^d$, say $\mathcal{X}$. We can make $\mathcal{X}$ a Banach space with the norm of total variation of the set-function $\Pi(.,t)$. For details (in which we do not enter), see [Fre85], page 71, and [AlBo99], Chapter 9.

The interesting fact is that the propagators $U(t,s)$ and $W(t,s)$ are mutually adjoint in the space $L^2(\mathbb{R}^d, \Pi(dx,t))$: for all $0 \leq t < s \leq T$,

$$\int_{\mathbb{R}^d} \Pi(dx,t) \left(U(t,s)g\right)(x) = \int_{\mathbb{R}^d} g(x) \left(W(t,s)\Pi\right)(dx,t), \tag{4.13}$$

for all $g \in B(\mathbb{R}^d)$ and $\Pi(.,t) \in \mathcal{X}$. (4.13) is again a consequence of Lemma 1, page 300, [GiSk69]. Let $\mathcal{A}^*(t)$ be the generator of $W(t,s)$. Then it turns out that $\mathcal{A}^*(t)$ is an extension of the operator $A^*(t)$, the formal adjoint (3.6) of the operator given in (4.6), in the following sense: suppose exists $\pi(x,t)$ such that $\Pi(B,t) = \int_B \pi(x,t)dx$ and that $\pi(x,t)$ is smooth enough, then $\left(\mathcal{A}^*(t)\Pi\right)(B,t) = \int_B A^*(t)\pi(x,t)dx$. In particular, if $A^*(t)\pi(x,t) = 0$ and $\pi \geq 0$, we have that $\pi(x,t)$ is the density function corresponding to $S(t)$. In other words, $\pi(x,t)$ satisfies the Fokker-Planck equation (3.7) with initial condition $\pi(x,0)$ (supposed known).

The study of the conditions such that the density function $\pi(x,t)$ of the random variable $S(t)$ satisfies the Fokker-Planck equation (3.7), when $(S(\tau), \tau \geq 0)$ is a diffusion process, can be found

[6] In other words, $\Pi(t,.)$ the one-dimensional distribution of $\{S(\tau), \tau \geq 0\}$ at t fixed.

in [Ris89][7], Chapter 4. We will deepen later on, when we will discuss small-time techniques in Chapter 8, the connection between the two Kolmogorov equations and their equivalence for the transition density.

Stationary Density

An invariant measure of the autonomous Markov process $S(t)$ with transition function $p(x, dy; \tau)$, $\tau \geq 0$, is a measure Q on the Borel sets of $\mathbb{R}^d$ with the property that, for all $\tau > 0$, $g \in B(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} T(\tau)g(x)Q(dx) = \int_{\mathbb{R}^d} g(x)Q(dx). \tag{4.14}$$

Now, the adjoint semigroup of $T(\tau)$ takes the form $V(s-t) = W(t,s)$. The generator of $V(\tau)$ is given by $\mathcal{A}^*$ and its restriction

$$A^*u = -\sum_i \frac{\partial}{\partial x_i}(\mu_i(x)u) + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}(x)u),$$

is the formal adjoint to A, the generator of $T(\tau)$ on $C_0^2(\mathbb{R}^d)$. A stationary distribution is an invariant measure Q for the transition density with the additional property that $Q(\mathbb{R}^d) = 1$. If the initial distribution of $S(0)$ is Q, then the distribution of $S(t)$ does not change with time. Notice also that a linear combination of invariant measures with non-negative coefficients is again an invariant measure, so that the collection of all the invariant measure of a Markov process form a cone. Everything we said above applies to any invariant measure. In particular, if exists $q(x)$ such that $A^*q(x) = 0$ and $q \geq 0$, then $q(x)$ is the density function of an invariant measure of $S(t)$. If also $q(x)$ is a probability density, $Q(dx) = q(x)dx$ is a stationary distribution. Finally, if $A^*q(x) = 0$ is well posed, the stationary density is unique.

The L^p extension again

Then a natural choice for the L^p extension is given by a stationary distribution $Q(dx)$ with density $q(x)$, provided it exists, in the case the diffusion $S(t)$ is autonomous. This is a standard selection, see, e.g., [HaSc95] and [HST98] for some statistical interesting applications tuned on financial data and problems. [HaSc95] and [HST98] define the transition semigroup directly in $L^2(\mathbb{R}^d, q(x)dx)$ requiring the existence of a stationary distribution and that the diffusion has been correctly initialized too. To get their important results, they ask also that the diffusion is time-reversible (but note that any diffusion with self-adjoint generator in $L^2(\mathbb{R}^d, dQ)$ is time reversible and vice versa, see [HaSc95], [HST98]). Similar assumptions are made, e.g., in [KeSo99] and [DaGu01], with regard to the estimation of parameters of a discretely observed diffusion. These celebrated examples stress the importance of the extension result for estimation ends. In this setting we have also a strong implication about condition (4.10). Using the relation (see formula (4.14))

$$\int_{\mathbb{R}^d} \varphi(x)q(x)dx = \int_{\mathbb{R}^d} T(\tau)\varphi(x)q(x)dx, \tag{4.15}$$

which holds for all $\varphi \in L^2(\mathbb{R}^d, q(x)dx)$ (and, actually, for any invariant measure $q(x)$ for $S(t)$), provided the process is correctly initialized, i.e., $S(0) \sim q(x)$, we see that (4.10) is implied by the requirement that the vector $(\sigma\partial_x g)_k \in L^2(\mathbb{R}^d, q(x)dx)$ (where $\partial_x g$ is the vector of the first partial derivatives of g). This is a standard remark used, e.g., in [DaFl86] and [HaSc95].

[7] We refer to [Ris89] for find many techniques of solutions of (3.7) for $\pi(x,t)$ and how to use the boundary conditions when available (i.e., when assumption 5. is violated).

4.3 Multiplicative Semigroups

In section (3.3), by means of some analytic and stochastic techniques, we showed that we can often solve the Cauchy problem (3.9) setting the reaction term $c \equiv 0$ and then recover the general solution. Now we explain the role of the potential term $c(S(t), t)$ in terms of Markov processes and we will see the effect of $c \neq 0$, and why we can safely suppose $c = 0$. We sketch only what we need, for a detailed treatment see [BlGe68] (especially Chapter 3), [GiSk74], Vol.2, and [KaTa81], Section 15.12. An introduction for the connections between Classical and Stochastic Potential theories[8] is [Kal97], Chapter 24.

Multiplicative Functionals A family of random variables $\{M(t,s), 0 \leq t \leq s \leq T\}$ is called a multiplicative functional of the Markov process $S(t)$ solution of (2.1) if, for all t, s: i) $M(t,s)$ is $\mathcal{F}(s)$ measurable; ii) for all $r \in [t,s]$, $M(t,s) = M(t,r)M(r,s)$ $(a.s.)$; iii) $M(t,s) \in [0,1]$ $(a.s.)$. It follows that $M(t,s)$ is monotonically non-decreasing in t. If $S(t)$ is autonomous, then property ii) (with abuse of notation write $M(\tau) = M(t,s)$, $\tau = (s-t)$) is replaced by $M(t+s) = M(t)\left(M(s) \circ \theta^+(t)\right)$, where $\theta^+(t)$ is the shift operator in the space of events Ω[9] such that $\omega(s) \overset{\theta^+(t)}{\longmapsto} \omega(s+t)$. We are interested in multiplicative functionals of integral type, given, for any measurable and non-negative function $c(x,t)$, by

$$M(t,s) = \exp\left[-\int_t^s c(S(\tau),\tau)d\tau\right], \tag{4.16}$$

provided the integral is finite for all t, s. Condition iii) can be relaxed asking that c is only bounded below (or strengthened with $c \in B(\mathbb{R}^d)$).

A family of operators in $B(\mathbb{R}^d)$ is associated with (4.16) by means of the formula

$$R(t,s)g(x) := E\left[g(S(s))M(t,s)|S(t) = x\right].$$

$R(t,s)$ has the semigroup property (4.3). In fact, for all $r \in [t,s]$ and $g \in B(\mathbb{R}^d)$,

$$R(t,s)g(x) = E\left[E\left[g(S(s))M(r,s)|S(r)\right]M(t,r)|S(t) = x\right] = R(t,r)R(r,s)g(x). \tag{4.17}$$

We call $R(t,s)$ the propagator generated by $M(t,s)$. Then, what is the diffusion associated to the evolution family $R(t,s)$? To answer the question and see the relation between $U(t,s)$ and $R(t,s)$, we can define a new transition function $R(x,t;B,s) := E\left[\chi_B(S(s))M(t,s)|S(t) = x\right]$, where B is any Borel set of $\mathbb{R}^d$. The proof that $R(x,t;B,s)$ is a genuine transition function is equivalent to prove that $R(x,t;B,s)$ satisfies (4.1), but this follows from (4.17). Hence we have the representation $R(t,s)g(x) = \int_{\mathbb{R}^d} g(y)R(x,t;dy,s)$. The important point is that $R(x,t;B,s) \leq P(x,t;B,s)$ for any Borel set B of $\mathbb{R}^d$, so that $R(t,s)g \leq U(t,s)g$ for all $g \in B(\mathbb{R}^d)$. Any semigroup with this property is called a subordinated semigroup to $U(t,s)$. The associated diffusion process, say $S^c(t)$, is called subordinated process of $S(t)$.

We turn to a probabilistic interpretation of $S^c(t)$. The transition function $R(x,t;dy,s)$ of $S^c(t)$ is not stochastic, i.e., there is a leakage of probability, which depends on c, t and s (from the definition of $M(t,s)$). From (4.16), after some manipulations (see [GiSk74], Vol.2, page 69), we find that $R(x,t;B,s)$ must satisfy the Volterra integral equation

$$R(x,t;B,s) = P(x,t;B,s) - \int_t^s c(y,r)R(y,r;B,s)P(x,t;dy,r). \tag{4.18}$$

[8] We just mention that the Newtonian potential in $\mathbb{R}^d$ is given by the resolvent operator (computed at $\lambda = 0$) of the Laplacian in $\widehat{C}(\mathbb{R}^d)$, which in turn is the generator of the semigroup associated with the Brownian motion.

[9] We always identify Ω with $\left(\mathbb{R}^d\right)^{[0,T]}$.

The second term amounts the loss of probability. If P and c are known, (4.18) could be solved by the method of successive approximations. A similar problem arises in the parametrix method (see Chapters 10 and 11).

An alternative, equivalent approach sees the subordinated process $S^c(t)$ as a killed version of $S(t)$. Introduce a cemetery state Δ that we append to the state space, so that $S^c(t)$ is conservative on $\mathbb{R}^d \cup \{\Delta\}$. Each function of $g \in B(\mathbb{R}^d)$ is extended to $\mathbb{R}^d \cup \{\Delta\}$ by $g(\Delta) = 0$ and we identify $B(\mathbb{R}^d)$ with $B(\mathbb{R}^d \cup \{\Delta\})$. Δ is a trap point and if $S^c(t)$ enters there we say that the diffusion is killed. $S^c(t)$ has the same infinitesimal parameters of $S(t)$, plus a the infinitesimal killing rate

$$\lim_{h \downarrow 0} \frac{1}{h} P\left[S^c(t+h) = \Delta | S^c(t) = x\right] = c(S(t), t) \geq 0.$$

Therefore, the subordinated process $S^c(t)$ behaves like $S(t)$ up to the time at which it is killed, called killing time. The (possibly stochastic) killing time is completely characterized by the infinitesimal parameters of $S^c(t)$ (see [KaTa81], pages 313-314). On the other hand, given the diffusion $S(t)$, if $S^c(t)$ is a subprocess of $S(t)$, then there exists a multiplicative functional of $S(t)$ which generates the semigroup associated to $S^c(t)$ (see [BlGe68], Chapter 3, and [GiSk74], Vol.2).

In terms of our financial applications (see Chapter 3), the deflated stochastic process under which the assets in the market are martingales is just the subordinated process corresponding to the given (possibly) stochastic spot rate, interpreted as the killing rate of our initial financial model as in (2.1).

Generator of The Subordinated Semigroup

Formula (4.18) shows that (in principle) the knowledge of P and c is equivalent to the knowledge of R. Hence it is reasonable to suppose that also the generator of the semigroup $R(t,s)$ (i.e., the generator of the subordinated diffusion) is completely determined by the generator of $U(t,s)$ and the function $c(x,t)$. This is true, at least in the autonomous case. Write $T(\tau)$, $R(\tau)$, $\tau \geq 0$, for the semigroup in $\widehat{C}(\mathbb{R}^d)$ associated to the time-homogeneous diffusion $S(\tau)$ and to the subordinated process $S^c(\tau)$, where $c(x) \geq 0$ is the infinitesimal killing rate of $S^c(\tau)$. The proof of the following theorem, based on the resolvent equation (A.3), is given in [KaTa81], page 314.

Theorem 12 *Let $\mathcal{A}$ be the generator of $T(\tau)$ and $\mathcal{A}_R$ be the generator of $R(\tau)$. Then $\mathcal{A}_R = \mathcal{A} - c(x)$ on the domain $D(\mathcal{A}_R) = D(\mathcal{A})$.*

A partial, easier proof of this fact can be found in [Øks06] (Chapter 8), by means of the Dynkin's formula (as in (4.8)). The central idea of this proof is to fix $(s,x,z) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$ and define the new diffusion $Y(\tau) = (s - \tau, S(\tau), Z(\tau))$ (column vector), where we suppose that the autonomous diffusion $S(t)$ starts at x ($a.s.$) and we set $-Z(\tau) := z + \int_0^\tau c(S(u))du$. Then, as in our sketched proof of (4.8), it turns out that for any $h \in C_0^2(\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d)$, h is in the domain of the generator of $Y(\tau)$ while the generator is $A_Y = -\frac{\partial}{\partial s} + A - c(x)\frac{\partial}{\partial z}$, where A is given in (4.9).

Chapter 5

Model Error

In this chapter we introduce the central concept of model error. The key result is that the model error can be estimated by means of an a-priori inequality. This is a first step towards the applications considered in the second part of the thesis. References are [Cor06a], [Fri76] and [Eva98]. For a clear introduction of the financial issues, see [Cor06b] and references therein.

5.1 Model Error Representation

The central assumption of Chapter 3, namely 7. (we can see 6. as a special case), is that the Cauchy problem (3.9) is well posed and that its solution $v(x,t)$ is given by the Feynman-Kac representation formula (3.10). For ease of exposition we recall that (3.9) is given by

$$\begin{aligned} (L_c(t)v)(x,t) &= f(x,t), \quad & (x,t) &\in \mathbb{R}^d \times [0,T), \\ v(x,T) &= g(x), & x &\in \mathbb{R}^d, \end{aligned} \tag{5.1}$$

where $L_c(t) = \frac{\partial}{\partial t} + A(t) - c(x,t)$ and $A(t)$ is the generator

$$A(t) = \sum_{i=1}^{d} \mu_i(x,t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j}. \tag{5.2}$$

of the diffusion $(S(t), t \geq 0)$ solution of (2.1). We suppose that a set of sufficient conditions on $L_c(t)$, f and g for the fulfillment of assumption 7. holds (possibly after some suitable transformations, e.g., the log-transform in the BS model).

Of course, that (5.1) is well posed does not mean that we can solve it in closed-form. The problems stem from the functional form of the coefficients of $A(t)$. Only under particular choices of the infinitesimal parameters μ and a of $(S(t), t \geq 0)$ there exists an analytic solution of (5.4). Usually, these are the cases when the transition density $p(x,t;y,s)$ of $(S(t), t \geq 0)$, i.e., the fundamental solution associated to $L_c(t)$, can be obtained also by some stochastic arguments. In Finance, the most remarkable examples are the solutions of the SDE models we have reviewed in Chapters 2 and 3: the BS model on $\mathbb{R}^d$, the Vasicek and CIR models (for term structure derivatives), the Heston model (for stochastic volatility).

In general, under mild assumptions, the solution of (5.1) can be computed numerically. One can either solve the PDE via numerical methods (e.g., the Finite differences methods, see [Duf06], [RiMo67] and [Smi85]), or exploit the representation formula (3.10) and the arguments in Chapter 3 (see formula (3.20)) to use Monte Carlo and simulation methods (see [Gla04] and [KlPl99]).

In practice, as explained in the introduction, due to the costs of finding an efficient numerical solution and to estimate a complex model, often the estimation and pricing are performed using an auxiliary, simple and solvable model, in the belief that a careful calibrated model can perform well[1]. For instance, in the generalized BS model (see Section 3.3) one could make the additional assumption that $\sigma_{ij}(x,t) = x_i \Sigma_{ij}$, where $\Sigma \in \mathbb{R}^{d\times l}$, obtaining the classic, solvable BS model.

The idea is to approximate the solution of (5.1) with the solution of a similar Cauchy problem (i.e., with the same c, f, g) where we have replaced the generator $A(t)$ with another diffusion generator $A_0(t)$ such that the auxiliary Cauchy problems admits a closed-form solution. This is a standard practice in Markov valuation theory, supported by a great empirical success[2]. In particular, calibration does produce some inconsistencies (e.g., when one performs a recalibration with new data, the parameters could change substantially, even if the recalibration is done with few new data), but allows to cope with the issues discussed in Section 3.3 (e.g., the necessity of a perfect fit of the data when the model is used to price a derivative). The analysis here shed some light, in terms of the Cauchy problems, on the reasons why this practice works. But most of all, we can introduce with a concrete application some of the key concept we need in the next Chapter and in the second part of the thesis.

Error Representation

In order to emphasize the essential features of the analysis, we proceed to simplify as much as possible problem (5.1). From the results of the previous two chapters, we can assume that $c \equiv 0$. So we recover that (5.1) is an inhomogeneous Kolmogorov Backward equation for $v(x,t)$, with terminal condition $v(x,T) = g(x)$. By the linearity of (5.1) and the Duhamel's principle (see Appendix A.2), we can also set the source $f \equiv 0$ and obtain the inhomogeneous solution from the solution of a second homogeneous Cauchy problem. Finally, we can apply the change of variable[3] $t \mapsto T - t$ and reduce the Cauchy problem (5.1) to

$$\begin{aligned} -\tfrac{\partial}{\partial t} v(x,t) + A(t)v(x,t) = 0, &\quad (x,t) \in \mathbb{R}^d \times (0,T], \\ v(x,0) = g(x), &\quad x \in \mathbb{R}^d. \end{aligned} \tag{5.3}$$

The advantage of working with (5.3) is twofold. First, we have that both the option price and the transition function $p(x,t;y,s)$ of $(S(\tau), \tau \geq 0)$ satisfy (5.3) (p in the (x,t) variables, where $t > s$ after the change of variables). The initial condition is $g(x) = \delta(x-y)$ for the fundamental solution $p(x,t;y,s)$ and $g(x)$ equal to the terminal payoff for the T-contingent claim. We recall that the option price is obtained by averaging with respect to p (see (3.10), but also (A.26)). Second, we see clearly from (5.3) that the solvability in closed-form of the PDE accounts to the specific functional form of the linear operator $A(t)$: we denote the solution of (5.3) by $v^{A(t)}(x,t)$.

Now, let $A_0(t)$ be a second diffusion generator (in the same space of $A(t)$) to which we associate a second, well-posed but also analytically solvable Cauchy problem (with the same initial condition of (5.3))

$$\begin{aligned} -\tfrac{\partial}{\partial t} v(x,t) + A_0(t)v(x,t) = 0, &\quad (x,t) \in \mathbb{R}^d \times (0,T], \\ v(x,0) = g(x), &\quad x \in \mathbb{R}^d. \end{aligned} \tag{5.4}$$

Let $v^{A_0(t)}(x,t)$ be the explicit solution of (5.4). Note that, by linearity of $A(t)$ and $A_0(t)$, we can define the operator $A_1(t) := A(t) - A_0(t)$, i.e., it holds the decomposition

$$A(t) = A_0(t) + A_1(t). \tag{5.5}$$

Clearly, $A_1(t)$ is a linear partial differential operator of at most second order. Notice that $A_1(t)$ does not need to be a Markov generator. It is natural to interpret $v^{A_0(t)}(x,t)$ as an approximation to $v^{A(t)}(x,t)$ and to call the difference

[1] Generally the auxiliary model is far from a more complex statistical model, which could be supposed to hold by some a-priori information or some advanced but cumbersome statistical technique used only occasionally.

[2] The widespread use of the BS and Black [Bla76] models stems form their empirical success (see [Hul06], Chapters 26, for the use of Black model, and [Reb04] for a discussion of the performances of the BS model).

[3] It would be more proper to set $\tau := T - t$ and use this new variable, but since confusion can not arise here, we use the standard change of variable used in PDE theory.

$$e(x,t) := v^{A(t)}(x,t) - v^{A_0(t)}(x,t)$$

model error. In fact, the generators $A(t), A_0(t)$ characterize the diffusions involved. By simple algebra, we see that the model error (5.12) must satisfy the initial value problem

$$\begin{aligned} -\tfrac{\partial}{\partial t}e(x,t) + A(t)e(x,t) &= -A_1(t)v^{A_0(t)}(x,t), \quad && (x,t) \in \mathbb{R}^d \times (0,T], \\ e(x,0) &= 0, && x \in \mathbb{R}^d. \end{aligned} \tag{5.6}$$

In other words, $e(x,t)$ solves a Kolmogorov equation with homogeneous initial condition and a well defined inhomogeneous source term given by $-A_1(t)v^{A_0(t)}(x,t)$. Of course, if we cannot solve analytically (5.3), the same is true for (5.6). Our goal here is to show that, under mild hypotheses, the error $e(x,t)$ committed in replacing $A(t)$ with $A_0(t)$ can be explicitly evaluated and is bounded. Furthermore, the estimate of the error depends on the difference $A_1(t)$ between the two models, so that if the error is assessed as too big, one can use a different auxiliary model, at least following an ad-hoc procedure.

The main use of the model error and of its estimate (as discussed below) in PDE theory (also from an abstract point of view, see [ItKa02] Chapter 4, [Paz83], Chapter 3) is in the approximation of a complex problem (say involving $A(t)$ in (5.2)) by means of a sequence of reduced problems, typically projections of the PDE problem on finite dimensional subspaces, with the aim of applying some convergence results to show that: i) the sequence of model errors e_n tends to 0 as $n \to \infty$; ii) some key properties of the sequence of reduced problems are enjoyed by the limit. Notice that all the techniques for the numerical solution of PDEs could be seen as an application of this idea (e.g., see [Eva98], Chapter 7, and [ItKa96]).

5.2 A-priori Inequalities of Integral Type

In PDE theory the so-called a-priori estimates play a central role. An a-priori estimate is an inequality for the solution that does not depend on the fact that the corresponding PDE problem is well-posed and even that any solution exists. In other words, the estimate holds without any preliminary knowledge about the existence of the solution and is based on the coefficients of the PDE, the data considered and the class of functions in which we are looking for the solution. In fact, an important use of such inequalities is to prove the actual existence of a solution.

There are many types of a-priori inequalities for parabolic PDEs, each suited for a different type of solution and space of function in which looking for, so that the main difference is the norm used (see [IKO62], [Fri64] and [Kry96]). However, the estimates are typically stated for initial-boundary problems, i.e., Cauchy problems on a bounded domain coupled with some boundary conditions. Let D be a bounded domain[4] in $\mathbb{R}^d$. Denote the boundary of D by ∂D. In accordance to the discussion in the previous section, we consider initial-boundary problems of the following type (boundary condition of Dirichlet type)

$$\begin{aligned} -\tfrac{\partial}{\partial t}u(x,t) + A(t)u(x,t) &= f(x,t), \quad && (x,t) \in D \times (0,T], \\ u(x,0) &= g(x), && x \in D, \\ u(x,t) &= h(x,t), && (x,t) \in \partial D \times (0,T), \end{aligned} \tag{5.7}$$

where, w.l.o.g., we can suppose that $A(t)$ in (5.2) is the infinitesimal generator of a diffusion (see Appendix A.2 and Chapter 4).

When functional methods are used to solve (5.7), i.e., when one is looking for some kind of weak solution of the initial-boundary value problem, the most important estimates are the ones in the norms of L^p spaces. In particular, let $L^p(D)$ be the space of functions $\varphi : D \to \mathbb{R}$ such that $\int_D |\varphi(x)|^p\,dx < \infty$, where dx is the Lebesgue measure, and, for any Banach space X, let $L^p(0,T;X)$ be the space of functions $\Psi : (0,T) \to X$ with finite norm

[4] In PDE theory a domain is an open, connected subset of an Euclidean space.

$$\|\Psi(t)\|_{L^p(0,T;X)} = \left[\int_0^T (\|\Psi(t)\|_X)^p \, dt \right]^{1/p}.$$

See Appendix A.3 for similar definition of $C^n(0,T;X)$ and for the remark that any classical solution $u(x,t)$ of (5.7) can be seen as a function $U(t) := u(.,t)$ in $C^1(0,T;X)$, X a suitable Banach space. Weak solutions allow $U(t)$ to be in some more general subspace of $L^p(0,T;L^p(D))$ (e.g., see [Sal04], page 338, for a classification of the weak solutions). We are interested in the following result.

Theorem 13 *Let ∂D be of class C^2 and $u(x,t)$ a classical solution of (5.7) with $g = h \equiv 0$, where $A(t)$ is given by (5.2). Suppose that $a(x,t)$ is uniformly elliptic in $D \times (0,T)$ (constants $m \le M$, see (A.9)), $a_{ij}(x,t)$ is Lipschitz continuous in (x,t) on $\overline{D \times (0,T)}$ for all i,j (constant $\alpha(r) \downarrow 0$ as $r \downarrow 0$) and $\mu_i(x,t)$ is bounded for all (x,t) (say $\sum_i |\mu_i| \le K$). Then, for any $p \in (1,3/2) \cup (3/2,\infty)$ and $f \in L^p(0,T;L^p(D))$, there exists a constant C (depending only on m, $\alpha(r)$, K, T and D) such that*

$$\int_0^T \int_D \left(|u|^p + \sum_i |\partial_{x_i} u|^p + \sum_{ij} \left|\partial_{x_i}\partial_{x_j} u\right|^p + |\partial_t u|^p \right) dxdt \le C \int_0^T \int_D |f|^p \, dxdt. \tag{5.8}$$

The theorem is stated in [Fri76] (page 240, Vol.2) for the so-called strong solutions of (5.7) (see [Fri76], page 236). In that case the conditions are sufficient also to say that there exists a unique strong solution. The uniqueness for the strong solution and the sufficient regularity of the data f imply that there could be at most one classical solution. If a classical solution exists, it must satisfy (5.8). As said, the a-priori inequality (5.8) holds independently of the knowledge that a classical solution actually exists.

We make some additional remarks. i) The general case with $c \ne 0$ is identical, requiring that $|c| \le K_1$; then C depends also on K_1. Other similar estimates allow for $g \ne 0$ (see [IKO62]) and the estimates depend also on the L^p-norm of g. ii) Integral estimates in $L^2(0,T;L^2(D))$ are called energy estimates. A fundamental use of energy estimates is to prove the well posedness of weak problems by showing the convergence of a Finite element approximation of the problem (see [Eva98], Chapter 7, and [QuVa97]). iii) The only strong condition in the theorem is the uniform ellipticity of the diffusion matrix, but this is a typical requirement for a-priori inequalities (see [IKO62], Section 2, [Fri64], Chapter 3, and again [Eva98] Chapter 7). When the problem is degenerate, it is a good idea to use some abstract method instead (e.g., see [ItKa02]). iv) An extension of (5.8) to unbounded domain, i.e., to pure initial-value problems, is given, e.g., in [Aro68]. The idea is to set a sequence of problems which limit is the infinite domain case (note that in theorem 13 the Dirichlet condition is homogeneous, i.e., $u(x,t) = 0$ on $\partial D \times (0,T)$). We give below a different extension directly applicable to pure initial value problems.

A Weighted Energy Inequality

The following result is an energy estimate for a Cauchy problem, from [Cor06a].

Proposition 14 *Let $u(x,t)$ solve*

$$\begin{array}{rr} -\frac{\partial}{\partial t} u(x,t) + A(t)u(x,t) = f(x,t), & (x,t) \in D \times (0,T], \\ u(x,0) = g(x), & x \in D, \end{array} \tag{5.9}$$

where D is a domain in $\mathbb{R}^d$ of the form $D = D_1 \times \cdots D_d$, each D_i a possible unbounded interval of $\mathbb{R}$ and $A(t)$ is is given in (5.2). Assume that $a(x,t)$ is uniformly elliptic in $D \times (0,T)$ (constants $m \le M$) and there is a constant K such that $\sum_j \sum_i \left(\partial_{x_j} a_{ij}(x,t) + \mu_j(x,t)\right)^2 \le K^2$. Let Q be *a probability measure on D endowed with density $q(x) \in C^1(D)$, such that $\partial_{x_i} \ln q(x) \leq \kappa \in \mathbb{R}_+$ for all i. If*

$$\lim_{\substack{x_i \uparrow \sup D_i \\ x_i \downarrow \inf D_i}} \sum_j a_{ij}(x,t) \left(\partial_{x_j} u(x,t)\right) u(x,t) q(x) = 0 \tag{5.10}$$

for all i,and if $g, f \in L^2(\mathbb{R}^d, dQ)$ and $\int_0^t f^2 d\tau < \infty$, then

$$\begin{aligned} &\int_D |u(x,t)|^2 q(x)dx + m \sum_i \int_0^t \int_D |\partial_{x_i} u(x,\tau)|^2 q(x)dxd\tau \\ &\leq e^{Ct} \left[\int_D |g(x)|^2 q(x)dx + \int_0^t \int_D |f(x,\tau)|^2 q(x)dxd\tau \right], \end{aligned} \tag{5.11}$$

where $C = \left[(\kappa M + K)^2 / m + 1 \right]$.

Notice that the a-priori estimate (5.11) of the $L^2(\mathbb{R}^d, dQ)$-norm[5] (of any solution of (5.9) and of its first partial space-derivatives) is explicit and depends on three factors: a) The constants associated to the operator $A(t)$, in particular the constants of uniform ellipticity m, M. b) The time span $(0,t)$; back to the original time variables we would have e^{T-t} and the time-integrals running from t to T. c) The data f, g.

The role of (5.10) is the same as $u(x,t) = 0$ on $\partial D \times (0,T)$ in theorem 13, but now we can consider solutions $u(x,t)$ not necessarily vanishing on ∂D_i. The proposition extends to the case in which Q depends on t, say $Q(dx,t) = q(x,t)dx$, if m, M, K, κ are independent of t and the straightforward modification of (5.10) holds uniformly in t.

Estimation of The Model Error

Under the required conditions, we can apply the estimate (5.11) to the solution of the Cauchy problem (5.6). We obtain

$$\begin{aligned} &\int_{\mathbb{R}^d} |e(x,t)|^2 q(x)dx + m \sum_i \int_0^t \int_{\mathbb{R}^d} |\partial_{x_i} e(x,\tau)|^2 q(x)dxd\tau \\ &\leq \exp\left(\left[(\kappa M + K)^2 / m + 1 \right] t \right) \int_0^t \int_{\mathbb{R}^d} \left| A_1(\tau) v^{A_0(t)}(x,\tau) \right|^2 q(x)dxd\tau. \end{aligned} \tag{5.12}$$

This bound explains why a model that is known to be wrong can actually perform well. If the auxiliary model $A_0(t)$ is near enough to $A(t)$ in the sense that the conditions of the proposition apply to the model error, then the error is bounded. If the error is assessed as too large, one can change the auxiliary model in accordance to (5.12). As suggested in [Cor06a], one could even implement some minimization procedures in order to choose the optimal auxiliary model $A_0(t)$ with respect to $A(t)$.

A second advantage is that we have at once an error estimate for both the transition functions and the option prices. Furthermore, if we require that the option price is smooth as necessary, we can find also error estimates for the sensitivities (or Greeks, see [Hul06], Chapter 15) of the derivative price. The sensitivities are fundamental in the hedging of derivatives products. Say that $v^{A(t)}(x,t)$ is the option price and $x \in \mathbb{R}$. Two important Greeks are the so-called delta $\Delta^{A(t)}$ and gamma $\Gamma^{A(t)}$ of $v^{A(t)}(x,t)$, defined as $\Delta^{A(t)}(x,t) := \partial_x v^{A(t)}(x,t)$ and $\Gamma^{A(t)}(x,t) := \partial^2_{xx} v^{A(t)}(x,t)$. We refer to [Cor06a] for a discussion about the error in the deltas $e^{\Delta} = \Delta^{A(t)} - \Delta^{A_0(t)}$ and other Greeks.

[5] For the issues about the choice of Q see chapter 4.

Example: The Generalized Black-Scholes Model

To complete the discussion of (5.12) we give an example where the conditions are met: the generalized BS model (section 3.3) that can be proficiently approximated by the classic BS model, giving a possible explanation of the empirical success of the latter model. As a by-product we give a nice interpretation of the function $A_1(t)v^{A_0(t)}(x,t)$ for this special, but important case.

For the ease of exposition suppose that we are in the scalar case $x \in \mathbb{R}_+$ and let $H_0 = \mathbb{R} \times (0,T]$. In the generalized BS model (where recall that we can set $r = 0$), the dispersion coefficient of $S(t)$ is given by $\Sigma(S(t),t)S(t)$. To stress the implications of the log-transform, we denote the generator of $S(t)$ by $A(x,t) = \frac{1}{2}a(x,t)x^2\partial^2_{xx}$, where $a(x,t) := (\Sigma(x,t))^2$. Let $v^{A(x,t)}(x,t)$ the option price, solution of (5.3) with $A(t) = A(x,t)$.

Let us apply the log-transform $y = \ln x$. We have that $A(x,t) = A(y,t) = \frac{1}{2}a(e^y,t)\left[\partial_{yy} - \partial_y\right]$ and so $v^{A(x,t)}(x,t) = w^{A(y,t)}(y,t)$ satisfies (see equation (3.15))

$$\begin{aligned} -\partial_t w^{A(y,t)} + A(y,t)w^{A(y,t)} &= 0, \quad (y,t) \in H_0, \\ w^{A(t)}(y,0) &= g(e^y), \qquad y \in \mathbb{R}. \end{aligned} \tag{5.13}$$

If we set $a(x,t) = \alpha > 0$, then (5.13) reduces to a Cauchy problem for the heat equation and admits an explicit solution (see Appendix A.2 and Chapter 7). So it is natural to take $A_0(x) = \alpha x^2\partial^2_{xx} = A_0(y) = \alpha\left[\partial^2_{yy} - \partial_y\right]$, where $w^{A_0(y)}$ solves

$$\begin{aligned} -\partial_t w^{A_0(y)} + A_0(y)w^{A_0(y)} &= 0, \quad (y,t) \in H_0, \\ w^{A_0}(y,0) &= g(e^y), \qquad y \in \mathbb{R}. \end{aligned}$$

In turn, the model error $e(y,t) = w^{A(y,t)}(y,t) - w^{A_0(y)}(y,t)$ satisfies

$$\begin{aligned} -\partial_t e(y,t) + A(y,t)e(y,t) &= -A_1(y,t)w^{A_0}(y,t), \quad (y,t) \in H_0, \\ e(y,0) &= 0, \qquad y \in \mathbb{R}, \end{aligned}$$

where, by the decomposition (5.5), $A_1(y,t) = \frac{1}{2}\left(a(e^y,t) - \alpha\right)\left[\partial^2_{yy} - \partial_y\right]$.

If we assume that $a(e^y,t)$ is bounded (i.e., $\Sigma(x,t)$ is bounded) and uniformly elliptic in $\mathbb{R} \times [0,T]$, then we can apply the proposition and get the estimate

$$\begin{aligned} &\int_{\mathbb{R}} |e(y,t)|^2 q(y)dy + m\int_0^t \int_{\mathbb{R}} |\partial_y e(y,\tau)|^2 q(y)dyd\tau \\ &\leq \exp\left(\left[(\kappa M + K)^2/m + 1\right]t\right)\int_0^t \int_{\mathbb{R}} \left|A_1(y,\tau)w^{A_0(y)}(y,\tau)\right|^2 q(y)dyd\tau. \end{aligned}$$

Now, if we go back to the original variable x, $A_1(y,t) = \frac{1}{2}\left(a(x,t) - \alpha\right)x^2\partial^2_{xx}$ and recall $w^{A_0(y)}(y,t) = v^{A_0(x)}(x,t)$, we have

$$A_1(y,\tau)w^{A_0(y)}(y,\tau) = \frac{1}{2}\left(a(x,t) - \alpha\right)x^2\Gamma^{A_0(x)}(x,t).$$

From this we see that the contribution to the ($L^2(\mathbb{R}^d, dQ)$-norm of the) error is substantial in the regions where the gamma of the auxiliary model (in the original variables) is relevant. This result gives a theoretical justification of the widespread practice of using the gamma of the classic BS model to assess its pricing and, most of all, hedging performances. We will study in detail the gamma of the BS model in Chapter 9.

Chapter 6

Perturbations

Perturbation theory is a vast and complex field of modern Mathematics. It is more precise to say that there exist many perturbation theories, since a lot of perturbation methods have been successfully devised in many branches of not only applied, but also pure Mathematics. This chapter is just an introduction to some very basic facts that we need in the applications. Reference for the perturbations of differential equations are [Hin91], [Nay73] and [KeCo96]. For the abstract results see [Kat80], [EnNa99] and [ReSi80], Vol.4.

6.1 The Problem And The Aims

Let $A(t)$, given in (4.6) or (5.2), be the generator associated to the statistical diffusion-type model (2.1). The problem is to give an analytic approximation of the solution of the Cauchy problem

$$\begin{array}{ll} \frac{\partial}{\partial t}u(x,t) + A(t)u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [0,T), \\ u(x,T) = g(x), & x \in \mathbb{R}^d, \end{array} \tag{6.1}$$

which we suppose well posed but not solvable in closed-form. In the previous chapter we saw that, under general conditions, a bounded error in the specification of the diffusion model implies a bounded model error, and that the estimate can be explicitly evaluated. The key was the decomposition (5.5), i.e.,

$$A(t) = A_0(t) + A_1(t), \tag{6.2}$$

where $A_0(t)$ is the generator of a second, auxiliary diffusion model such that the Cauchy problem

$$\begin{array}{ll} \frac{\partial}{\partial t}u(x,t) + A_0(t)u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [0,T), \\ u(x,T) = g(x), & x \in \mathbb{R}^d, \end{array}$$

has an explicit solution $u^{A_0(t)}(x,t)$. The approximation of $u^{A(t)}(x,t)$ by means of (6.2) and $u^{A_0(t)}(x,t)$ can be carried out with some perturbation techniques. This is what we do in our applications.

Perturbation theory is not a sharply defined theory (for a discussion see, e.g., the introduction of [Kat80]), but an extremely large class of methods, used to solve very different (sometimes extremely complicated) problems, not only about differential equations. Each method is well suited for very specific problems (this is true especially for the most powerful asymptotic techniques), but the fundamental features are the same. Before we explain these features, we stress that even the goals of different branches of perturbation theory can appear widely different. We make the following partial remarks:

- In applied Mathematics the aim is to solve a specific complex problem from a simpler well-known result (the unperturbed problem) obtained as limiting case of the complex problem. To solve means to give an approximation and to study the qualitative behavior of the solution of the full problem. Often the idea is to use a perturbation technique in combination with a numerical technique to check and, if possible, improve the numerical solution. This is necessary in particular if there are singularities (a numerical procedure can underperform substantially in the part of the state space near a singularity of the solution; e.g., see the discussion about the so-called boundary layer techniques, [Hin91] or [Nay73]). A second example is if the numerical problem is posed in an unbounded domain. Since one cannot take into account infinity in numerical calculations, one can resort on some perturbation to yield the asymptotic behavior at infinity and pose appropriate boundary condition for an appropriate bounded subdomain.

- In pure Mathematics (e.g., see [EnNa99] for semigroup theory) the goal is to see if it is possible to infer the well posedness (or other properties) of the complex problem from the well posedness of a simpler problem. In this case one starts from the simpler problem and then perturb it, as in (6.2). An important connection with applied Mathematics, in particular Physics, is the study of the spectral properties of the operators involved, usually linear operators. The relations between eigenvalues of the perturbed and unperturbed operators are of special interest (e.g., see [ReSi80], Vol.4). The advantage of an abstract approach is given by the broad applicability of the results.

Referring to our problem of solving (6.1), we are interested in the following perturbations: i) the additive perturbation of linear differential equations, in particular PDEs (e.g., see [KeCo96] and [Geo95]); ii) the perturbation of linear operators in Banach spaces, in particular the perturbation of semigroups of operators (see [Kat80] and [EnNa99]). There are strong connections, so that the basic ideas are essentially the same. The fundamental two ingredients of any perturbation method are:

(a) Find an expansion which defines a formal series solution of the problem.

(b) Truncate suitably the formal series solution to get an approximation and study the error committed in the truncation.

We define the formal solution of (6.1) below. Here we make the following remarks. i) Often the expansion is obtained when some parameter of the problem is small or large (typically for the parameter that tends to 0 or ∞, otherwise we could rescale it), calling the expansion parameter perturbation. Alternatively, the expansion can be carried out in terms of some coordinate (i.e., an independent variable) of the problem, say y, for $y \to y_0$ (again, it is often possible to rescale y_0 to 0 or ∞), referred as coordinate perturbations. Note that generally the expansion will hold only in a limited range of values of that parameter or coordinate; but for some results it is possible to match different available expansions that hold for different ranges (e.g., see [Hin91], Chapter 5 or [Nay73], Chapter 4). ii) In order to get the expansion, if there are not suitable coordinate or parameters in the problem, sometimes it is convenient to artificially introduce a small dimensionless parameter ϵ. In this case the decomposition (6.2) will take the form (e.g., see [KeCo96] and [Fre85])

$$A(t;\epsilon) = A_0(t) + \epsilon A_1(t), \tag{6.3}$$

with $A(t;\epsilon) \to A_0(t)$ as $\epsilon \downarrow 0$. iii) In connection with (**b**), a nice feature of a good approximation is that only the first few terms in the expansion are needed to get the desired precision.

Formal Solutions For PDEs

A *formal series solution* of (6.1) is defined as follows (e.g., see [Pol02]). Let $\{u_n(x,t)\}$, $n = 0, 1, \ldots$, be a sequence of particular, classical solutions of the linear, homogeneous PDE

$$\frac{\partial}{\partial t}u(x,t) + A(t)u(x,t) = 0. \tag{6.4}$$

Then, because for (6.4) holds the principle of linear superposition (see Appendix A.2), the series

$$\sum_{n=0}^{\infty} u_n(x,t), \tag{6.5}$$

irrespective of its convergence, is called a formal solution of (6.4). To match the terminal condition in (6.1) we ask also that $u_0(x,T) = g(x)$ and $u_n(x,T) = 0$ for all $n \geq 1$.

If (6.5) is uniformly convergent and the sum of (6.5) is in $C^{2,1}(\mathbb{R}^d \times [0,T))$, then (6.5) is a genuine classical solution of (6.1). A sufficient condition for this is that $\sum_n u_n(x,t)$ and the series of all the derivatives, $\sum_n \partial_t u_n(x,t)$, $\sum_n \partial_{x_i} u_n(x,t)$, etc., converge uniformly[1] (e.g., see [Fri71] or [PaSa90], Vol.2). Clearly, if the series (6.5) diverges, then it cannot be a classical solution of (6.4). Anyway, to find a good perturbation it is not necessary to prove that a formal solution converges uniformly and even that converges at all.

6.2 Regular And Singular Perturbations

There are essentially only two kinds of perturbations: regular or singular. And there are only two general methods to get a formal series solution of the problem: expansion or iterative methods. In the next two sections we explain, by means of simple algebraic or ODE-type examples, these overlapping classifications.

Regular Perturbations

A perturbation series is called *regular* (or *analytic*) if the perturbation series converges uniformly. The regularity of the perturbation not only implies that, under minimal conditions, the formal series solution is a classical solution of the differential problem and that adding more terms in the approximation we can gain any precision, but also that the problem is stable (in the sense that there is some continuity, suitable for the context, e.g., in terms of semigroups see [HiPh57], Chapter 13) in the considered perturbation.

Let y be the independent variable(s) of the problem and ϵ a parameter. Call $v(y;\epsilon)$ the solution of the problem, such that $v(y;0) = v_0(y)$ solution of the problem with $\epsilon = 0$. For instance, consider the problem to find the roots of

$$y^2 + \epsilon y - 1 = 0, \tag{6.6}$$

for small values of the parameter ϵ. Clearly the exact solution is the vector $v(y;\epsilon) = (y_+, y_-)$ where $y_\pm = -\epsilon/2 \pm \sqrt{1 + (\epsilon/2)^2}$. We can expand $\sqrt{1+(\epsilon/2)^2}$ in series obtaining, e.g., for the first root,

$$y_+ = 1 - \epsilon/2 + \epsilon^2/8 + O(\epsilon^4). \tag{6.7}$$

convergent for $|\epsilon| < 2$.

In the *expansion method* we look for a solution $v(y;\epsilon) = \sum_{n=0}^{\infty} \epsilon^n v_n(y)$, starting from the problem with $\epsilon = 0$. In the example, $v_0(y) = (1, -1)$. Note that if we prove that the series converges absolutely for some $\epsilon \neq 0$, then the solution $v(y;\epsilon)$ of the full problem is analytic in the parameter ϵ for a certain ray of convergence, explaining the name of this method. For simplicity, choose one root, e.g., 1. Then $v^1(y;\epsilon) = y_+(\epsilon) = \sum_{n=0}^{\infty} \epsilon^n y_n$, where $y_0 = 1$ and y_n, $n \geq 1$, have to be found. Substituting formally the series solution $\sum_{n=0}^{\infty} \epsilon^n y_n$ into (6.6), we have (ordering by column the same powers of ϵ)

[1] Then $\partial_t \sum_n u_n(x,t)$, etc., are well defined and equal to $\sum_n \partial_t u_n(x,t)$, etc.

$$\begin{array}{ccccccccc} 0 & = & 1 & + & \epsilon 2y_1 & + & \epsilon^2(y_1^2+2y_2) & + & \cdots \\ & & & + & \epsilon & + & \epsilon^2 y_1 & + & \cdots \; . \\ & & -1 & & & & & & \end{array}$$

Hence, equating the same powers of ϵ, we have: at ϵ^0, $1-1=0$, i.e., $y_0 = 1$ (as should be); at ϵ^1, $2y_1+1=0$, i.e., $y_1 = -1/2$; at ϵ^2, $2y_1^2+2y_2+y_1 = 0$, i.e., $y_2 = 1/8$, using y_0 and y_1. Without bothering to check the higher order terms, we have found the right expansion (6.7). Of course we know that the result is correct because we know the right answer. The interesting case, in which the right answer is not available, follows true if we prove that the power series solution in ϵ converge for a positive ray (in the example, for all $|\epsilon| < 2$). This could not be easy in a complex problem (see Chapter 7).

The first step of the alternative *iterative method* is to find some possible recursive formula. The standard idea is to try a *rearrangement* of the original problem. We propose to rearrange (6.6) as $y = \pm\sqrt{1-\epsilon y}$. So, working with the positive root, a possible iterative process is

$$y_{n+1} = \sqrt{1-\epsilon y_n},$$

for all $n \geq 1$, and the initial condition $y_0 = 1$, again obtained form $\epsilon = 0$ in (6.6). The first iteration gives $y_1 = \sqrt{1-\epsilon}$, which can be expanded in series as

$$y_1 = 1 - \epsilon/2 - \epsilon^2/8 + \cdots .$$

The heuristic rule is to stop each iteration at the first new term, here $y_1 = 1-\epsilon/2$. In fact, note that the first term excluded (i.e., $-\epsilon^2/8$) is wrong. The idea is to check the correctness of the last term included making one more iteration. The second iteration gives $y_2 = \sqrt{1-\epsilon y_1} = \sqrt{1-\epsilon(1-\epsilon/2)}$, so

$$y_2 = 1 - \epsilon/2 + \epsilon^2/8 + \cdots ,$$

which confirms the first term $\epsilon/2$ and adds another right term, that in turn can be checked with another iteration. Clearly, the iterative procedure starts with less rigor and requires more work to check its correctness. However, often, in a complex problem, an iterative procedure is the only way to get formally a power series expansion for $v(y;\epsilon)$. We clarify this assertion and its importance with an example, that we will use and extend in section (6.3) and Chapters 7,10.

Consider the problem of solving the ODE

$$\frac{d}{dy}v(y) = (\alpha+\epsilon\beta)v(y), \tag{6.8}$$

where $y > 0$, $\alpha, \beta \in \mathbb{R}$ and ϵ is a small parameter. We couple (6.8) with the initial condition $v(0) = 1$. The exact solution is $v(y;\epsilon) = e^{(\alpha+\epsilon\beta)y}$, which admits the expansion $v(y;\epsilon) = e^{\alpha y} + \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!}(\beta y)^n$, convergent for any $\epsilon \in \mathbb{R}$. We would like to use a similar, very appealing formula for $(\alpha+\epsilon\beta)$ replaced by $A = A_0 + \epsilon A_1$, where A, A_0 and A_1 are unbounded linear operators. Since A_1 is unbounded it is clear (see Appendix A.3) that $\sum_n \frac{\epsilon^n}{n!}A_1^n$ cannot converge in uniform operator topology, if actually A_1^n, $n > 2$, makes any sense. Moreover, even when A_1^n, $n > 2$, is well defined, since A_1 needs not to be a Markov generator (while we require A, A_0 to be generators), it could be that the domain of A_1^∞ is too small. We will use a similar (in the form), but substantially different formula in Chapters 8 and 9, in terms of a singular perturbation. Here, since for a regular perturbation we want the convergence of the series, we need to follow another route. Multiply both sides of (6.8) by the integrating factor $e^{-\alpha y}$. Then we obtain the equivalent integral equation (loosely speaking, a rearrangement of (6.8))

$$v(y;\epsilon) = e^{\alpha y} + \epsilon\int_0^y e^{\alpha(y-x)}\beta v(x)dx. \tag{6.9}$$

Iterating, we find the perturbation expansion

$$v(y;\epsilon) = e^{\alpha+\epsilon\beta} = e^{\alpha y} + \epsilon \int_0^y e^{\alpha(y-x)}\beta e^{\alpha x}dx + O(\epsilon^2). \tag{6.10}$$

Going on and calling $v_0(y) = e^{\alpha y}$, $v_1(y) = \int_0^y e^{\alpha(y-x)}\beta e^{\alpha x}dx$, etc., we have formally defined the expansion $v(y;\epsilon) = \sum_{n=0}^{\infty} \epsilon^n v_n(y)$. Formula (6.10) extends not only to bounded operators (e.g., α, β replaced by suitable matrices; this is the perturbation of the exponential matrix or, equivalently, of uniformly continuous semigroups), but also to the abstract operator setting, with unbounded, time-dependent operators. To stress the superiority of the iterative method and the use of integral equations, we make the following citation from [Bel64] (page 18):

> `The conversion of a differential equation into an integral equation is one of the most powerful devices available in connection with the study of approximate solutions, and one of the few standard techniques.`

We need to make some remarks. i) It is better to use integral operators rather than differential operators in carrying out an iteration method since the former are smoothing operators (in the sup-norm) and the latter not. E.g., consider the functions $v(y) = 1/(1+y)$ and $u(y) = 1/(1+y) + \epsilon \sin e^y$: then $\max_{y\in[0,Y]} |v(y) - u(y)|$ is finite and also the difference of the integrals over $[0,Y]$ is finite, but $v'(y)$ and $u'(y)$ are wildly different. ii) Formula (6.9) is strongly related with the so-called variation of constants technique to solve inhomogeneous ODEs. We refer to it as the *variation of parameters formula for the perturbed problem*. iii) Note that if the perturbation series converges uniformly for $\epsilon = 1$, then we have found a classical solution of the problem. See Section 6.3 for further comments. iv) A regular perturbation is often possible in the study of the point spectra of a scalar Sturm-Liouville problem (e.g., see [Kev00], Chapter 8, and [Zet97]). This will be crucial in Chapter 7. The extension to the study of the perturbation of eigenvalues (and of the relative eigenfunctions) of an unbounded linear operator is called Analytic perturbation theory of Kato-Rellich (references are [Kat80], Chapter 7, and [ReSi80], Chapter 12).

Asymptotic (or Singular) Perturbations

In differential problems the uniform convergence of the perturbation series is extremely rare and so, often, regular perturbations are unfeasible. Typically the series diverges when there is a singularity in the problem (so the name *singular perturbations*), i.e., when the unperturbed problem differs in an important way from the problem with the parameter (or coordinate) different from its limiting value[2], but this is not necessary. The salient point is that a divergent perturbation series (when a regular perturbation is not available) to be of any use must have some very specific properties and the crucial is the asymptotic property, from which the name *asymptotic perturbations*.

To highlight the main points we develop now an apparently regular differential example involving a coordinate perturbation. We look for a solution of

$$\frac{d}{dy}v(y) + v(y) = \frac{1}{y}, \tag{6.11}$$

for large y. We seek a regular series solution in the form $v(y) = \sum_{n=1}^{\infty} y^{-n}\alpha_n$ (intuitively, the large y takes the place of the small ϵ in the examples above; we could write $\epsilon = y^{-1}$). We have to find the coefficients α_n. Substituting this in (6.11) gives

$$\sum_{n=1}^{\infty} -n\alpha_n y^{-n-1} + \sum_{n=1}^{\infty} \alpha_n y^{-n} = \frac{1}{y},$$

[2] E.g., consider the problem to find the roots of $\epsilon y^2 + y - 1 = 0$, for small values of the parameter ϵ. This is a typical singular problem. The degeneracy is that for $\epsilon = 0$ there is just one root.

which is equivalent to $\sum_{n=1}^{\infty}(\alpha_{n+1}-n\alpha_n)\,y^{-n-1}=(1-\alpha_1)y^{-1}$. Since we have an identity in y, the same powers of y must have the same coefficients, i.e., $\alpha_1=1$, $\alpha_{n+1}=n\alpha_n$ for $n\geq 1$, so that $\alpha_n=(n-1)!$. It is clear that the series expansion $v(y)=\sum_{n=1}^{\infty}y^{-n}(n-1)!$ now diverges for all value of y. Despite its divergence the series is useful in numerical calculations. To see this we need to stop the series at a fixed N and prove that the remainder can be made arbitrarily small. Notice that $v(y)=e^{-y}\int_{-\infty}^{y}x^{-1}e^{x}dx$, is a particular solution of (6.11) provided $y<0$ (we read the requirement y large as $|y|$ large). Integrating by parts we have

$$\begin{aligned} v(y) &= y^{-1}+e^{-y}\int_{-\infty}^{y}x^{-2}e^{x}dx = y^{-1}+y^{-2}+2e^{-y}\int_{-\infty}^{y}x^{-3}e^{x}dx \\ &= \sum_{n=1}^{N}y^{-n}(n-1)!+N!e^{-y}\int_{-\infty}^{y}x^{-N-1}e^{x}dx. \end{aligned}$$

Call the second term in the right hand side (RHS), i.e., $R_{N+1}(y):=v(y)-\sum_{n=1}^{N}y^{-n}(n-1)!$, the remainder[3]. By the factorial in the numerator, the remainder cannot tend to 0 as $N\to\infty$, but, for N fixed and $y<0$,

$$|R_{N+1}(y)|\leq N!\left|y^{-N-1}\right|e^{-y}\int_{-\infty}^{y}e^{x}dx = N!\left|y^{-N-1}\right|\downarrow 0$$

as $|y|\to\infty$. In this sense, for sufficiently large $|y|$, the series expansion makes sense. A divergent power series with these properties is called an asymptotic series, writing, as $|y|\to\infty$, $v(y)\sim\sum_{n=1}^{\infty}y^{-n}(n-1)!$.

Asymptotic Approximations

The definition of asymptotic series is the mathematical formalization of the concept of a good approximation with an error understood and controllable, so that it could be made small as desired (see [Hin91], [Nay73] or [Geo95]). It is convenient to introduce a small real parameter ϵ (it may be a parameter or coordinate of the problem), such that $\epsilon\downarrow 0$ (possibly after rescaling if originally $\epsilon\to\epsilon_0$). Let $f(y;\epsilon)$ and $\{f_n(y;\epsilon)\}_{n=0}^{N}$, $N\leq\infty$, be real functions of ϵ and other independent variables y[4]. We say that $\sum_{n=0}^{N}f_n(y;\epsilon)$ is an asymptotic approximation (or expansion, if $N=\infty$) of $f(y;\epsilon)$ as $\epsilon\downarrow 0$ if, for each $M\leq N$,

$$\frac{f(y;\epsilon)-\sum_{n=0}^{M}f_n(y;\epsilon)}{f_M(y;\epsilon)}=\frac{R_{M+1}(y;\epsilon)}{f_M(y;\epsilon)}\to 0, \tag{6.12}$$

as $\epsilon\downarrow 0$, i.e., if the remainder $R_{M+1}(y;\epsilon)$ is smaller than the last term included for ϵ *small enough*. If (6.12) holds we write $f(y;\epsilon)\sim\sum_{n=0}^{N}f_n(y;\epsilon)$. The case $N<\infty$ accounts for the situations in which is not possible in principle to obtain via an iterative method an infinite number of terms. However, we will see that to have a possibly infinite number of terms is quite irrelevant.

We are specially interested in the so-called *asymptotic series*, for $N=\infty$ and $f_n(y;\epsilon)=\alpha_n(y;\epsilon)\delta_n(\epsilon)$, where $\{\delta_n(\epsilon)\}$ is an asymptotic sequence and $\alpha_n(y;\epsilon)$ are coefficients, often $\alpha_n=\alpha_n(y)$. A sequence of functions $\{\delta_n(\epsilon)\}$ is called *asymptotic* as $\epsilon\downarrow 0$ if, for any n, $\delta_{n+1}(\epsilon)=o\left[\delta_n(\epsilon)\right]$ as $\epsilon\downarrow 0$. The most relevant example is $\delta_n(\epsilon)=\epsilon^n$ (as in the example about (6.11), where $\epsilon=y^{-1}$) in which case we speak of *asymptotic power series*. If $\alpha_n=\alpha_n(y)$, then we have $f(y;\epsilon)\sim\sum_{n=0}^{\infty}\alpha_n(y)\delta_n(\epsilon)$ if, and only if, for any n,

[3] We stress that the remainder is defined by this formula and it is not $\sum_{N+1}^{\infty}y^{-n}(n-1)!$, as it would in the study of convergent series. For a convergent series the two definitions coincide.

[4] Many problems are naturally posed for complex function of complex variables, but we do not need this and so we do not consider this generalization.

$$R_{n+1}(y;\epsilon) = o(\delta_n(\epsilon)) = O(\delta_{n+1}(\epsilon)), \tag{6.13}$$

as $\epsilon \downarrow 0$. In the general case we need (6.12), namely (for each n) $R_{n+1}(y;\epsilon) = O(\alpha_{n+1}(y;\epsilon)\delta_{n+1}(\epsilon))$ as $\epsilon \downarrow 0$, but (6.13) suffices if, for fixed n, $|\alpha_{n+1}(y;\epsilon)| \le const.$ as $\epsilon \downarrow 0$. We notice (for a detailed discussion see Chapter 8) that any analytic function in ϵ around $\epsilon = 0$ (i.e., with an absolutely convergent representation $f(y;\epsilon) = \sum_{n=0}^{\infty} \alpha_n(y)\epsilon^n$, for some $\epsilon > 0$) admits an asymptotic expansion, as follows from the Taylor's theorem with remainder in Peano's form.

Properties of Asymptotic Series

The key properties of an asymptotic approximation are better understood by examples. We discuss here with algebraic examples the following issues: 1) The quality of the approximation of an asymptotic expansion. 2) How many terms one should use. 3) Uniqueness of the representation.

1) *Convergent versus divergent series.* We want to show that a divergent series can be sometimes more useful of a convergent one. Consider the error function $Erf(z) := 2/\sqrt{\pi}\int_0^z e^{-t^2}dt$. Since $e^{-t^2} = \sum_{n=0}^{\infty} \left(-t^2\right)^n/n!$, for all $t \in \mathbb{R}$ (note that this series is absolutely and uniformly convergent for all $t \in \mathbb{R}$), we can integrate termwise and get

$$Erf(z) = 2/\sqrt{\pi}\sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)n!}, \tag{6.14}$$

convergent for all $z \in \mathbb{R}$. For a small $|z|$ the convergence is very good, but as $|z|$ increases one needs a very large number of terms to maintain a given accuracy. E.g., for $z = 5$, 75 terms are needed to get an accuracy of 10^{-5}. The reason is that (6.14) is alternate and the intermediate terms are very large. A converging series with a so slow rate of convergence, here for z large (actually 5 is not large!; this is an extreme example), cannot be useful. An alternative expansion for Erf for large z is (set $\epsilon = 1/z$ to use the notation of the previous section) can be obtained (recall that $\int_0^{\infty} e^{-t^2}dt = \sqrt{\pi}/2$) from $Erf(z) = 1 - 2/\sqrt{\pi}\int_z^{\infty} e^{-t^2}dt$. Integrating by parts $\int_z^{\infty} e^{-t^2}dt = e^{-z^2}/(2z) - \int_z^{\infty}\left(e^{-t^2}/2t^2\right)dt$. By iteration

$$Erf(z) = 1 - e^{-z^2}/\left(z\sqrt{\pi}\right)\left[1 - 1/\left(2z^2\right) + 3/(2z^2)^2 + O(z^{-6})\right].$$

For $z \ge 3$, the first two terms give an accuracy of 10^{-5}. For z very large the first term is enough for any practical purpose.

Another famous example of asymptotic expansion is the DeMoivre-Stirling formula $n! \sim n^n e^{-n}(2\pi n)^{1/2}$ for $n \to \infty$. This is a special case of the asymptotic divergent expansion of the gamma function (see [Hoc86], Chapter 3): $\Gamma(z) = z^z e^{-z}(2\pi/z)^{1/2}\sum_{n=0}^{\infty}\alpha_n/z^n$, as $z \to \infty$, where $\alpha_0 = 1$, $\alpha_1 = 1/12$,... The first term alone of the series gives good accuracy for z as large as 10.

An asymptotic series $f(y;\epsilon) \sim \sum_{n=0}^{\infty}\alpha_n(y;\epsilon)\delta_n(\epsilon)$ is characterized by the properties that for ϵ small enough the first few terms of the expansion, in particular the leading term, are roughly correct. This is the intuition behind the asymptotic property. The nearer ϵ to 0, the fewer terms are necessary. For this reason, asymptotic series are also called series convergent at the beginning. Even if $f(y;\epsilon)$ is analytic in $0 < |\epsilon| \le const.$, then, for ϵ small enough, adding further terms does not really improve the approximation substantially (see Chapter 8), while it does if ϵ is large but inside the ray of convergence, as for (6.14) with $\epsilon = 1/z$ large (i.e., z small). On the other hand, if $f(y;\epsilon)$ is not analytic, adding more terms is always counterproductive, since the high order terms will tend to oscillate wildly, destroying any approximation (for an example and discussion, see [ReSi80], Vol.4, page 27).

2) *How many terms*? The definition of asymptotic property does not say how "small" is "small enough". In general the number of terms and the range of usable $\epsilon > 0$ depend on both $f(y;\epsilon)$ and $\delta_n(\epsilon)$. If f is known and the expansion does not converges, we can compute $R_{n+1}(y;\epsilon)$

for fixed y and find an optimal value M of terms to include for any small ϵ (or we could keep ϵ fixed and optimize M for y is some relevant region). The error will often decrease monotonically up to M and then increase. Sometimes it is possible that the error oscillates for a while, but it must increase definitely. If f is not known, as in the case of interesting differential problems, one cannot forecast a-priori the behavior of the remainder. However, we always know the order of the error.

3) *Uniqueness.* When an asymptotic expansion turns out to be unsatisfactory, one can always look for a different asymptotic representation (in fact, if we are relying on asymptotic results is just because that is the only possible source of qualitative information). We have two facts:

a) If a function has an asymptotic expansion in terms of a given asymptotic series and $\alpha_n = \alpha_n(y)$, then the approximation is unique (this property is lost in the general case $\alpha_n = \alpha_n(y;\epsilon)$, but it is not important). However, many different functions will share the same asymptotic approximation. For instance, consider $\delta_n(\epsilon) = \epsilon^n$. As $\epsilon \downarrow 0$, both e^ϵ and $e^\epsilon + e^{-1/\epsilon}$ have the same asymptotic power series $\sum_{n=0}^{\infty} \epsilon^n/n!$. We notice that the sum of an asymptotic series, even if convergent, could have nothing to do (for $\epsilon > 0$) with the function it approximates. Notice also that two functions with the same asymptotic power series can only differ by a non-analytic quantity, since two analytic functions with the same power series are identical.

b) A function with an asymptotic expansion will have many other possible representations in terms of different asymptotic sequences. E.g., let $\delta'_n(\epsilon) = (\sin\epsilon)^n$, which is asymptotic as $\epsilon \downarrow 0$. Then we have $e^\epsilon \sim \sum_{n=0}^{\infty} (\sin\epsilon)^n/n!$ as $\epsilon \downarrow 0$. Once again, the quality of the approximation for different asymptotic sequences will be identical for ϵ close to 0, but (this is the point) can be substantially different for larger ϵ (see the example on Erf).

6.3 Perturbations of Semigroups of Operators

We present here the basic results of the *regular* perturbation theory for semigroups, which we believe are useful to understand and to use the parametrix method. Recall that, under suitable assumption (see Appendix A.3), the solution operator of an autonomous Cauchy problem (i.e., (6.1) with $A(t) = A$) is given by a semigroup of operators, in accordance to our assumption 7.. In this section we do not use any parameter but we still ask that the perturbation series solution to be uniformly convergent, so that the perturbation series is a classical solution of the problem associated to the considered perturbation, proving the existence of a solution of the full problem. The truncation of the convergent perturbation series will give an approximation that can be in principle can be good as desired (but recall the issues about the speed of convergence of a regular perturbation series). Furthermore, the regularity of the perturbation ensures the stability of the solution operator of the abstract Cauchy problem w.r.t. that kind of perturbation; this is a key issue in the semigroup theory (see [HiPh57], Chapter 13, and [Kat80], Chapter 9).

Let A_0 be the generator of a strongly continuous semigroup $T(t)$, $t \geq 0$, in a given Banach space X. The idea is to build new semigroups from $T(t)$. So we ask under what conditions on the linear operator A_1 (called a perturbation to A_0) the sum $A := A_0 + A_1$ generates a strongly continuous semigroup $U(t)$, $t \geq 0$, in X. Since A is defined on $D(A) = D(A_0) \cap D(A_1)$ and $D(A_0)$ is only dense in X, if A_1 is unbounded it could be that $D(A)$ is not dense in X and therefore A could not be a generator. Here we consider the simple case in which $A_1 \in \mathcal{L}(X)$, such that $D(A_1) = X$. We have

Theorem 15 *Let A_0 be the generator of a strongly continuous contraction semigroup $T(t)$, $t \geq 0$, in X (i.e., with $\|T(t)\| \leq 1$). If $A_1 \in \mathcal{L}(X)$, then $A = A_0 + A_1$ on $D(A) = D(A_0)$ generates a strongly continuous semigroup $U(t)$, $t \geq 0$, in X such that $\|U(t)\| \leq e^{\|A_1\|t}$. Furthermore, we have*

$$U(t)g = T(t)g + \int_0^t T(t-s)A_1U(s)g\,ds, \tag{6.15}$$

for every $g \in X$ and $t \geq 0$.

Note the similarities between formulas (6.15) and (6.9) (with $\epsilon = 1$). We call the integral equation of Volterra-type (6.15) the *variation of parameter formula for the perturbed semigroup.* Essentially, we have converted again the perturbed differential problem into an integral equation. The method of conversion of a differential equation (here an abstract ODE) into an integral problem that can be solved by successive substitution obtaining a uniformly convergent series is often called Picard's method of successive approximation. It finds application in any differential problem (e.g., see 5.1.1, in [Fri76] for SDEs and Chapters 10 and 11 of the thesis for PDEs).

We can iterate (6.15) (recall that $D(A_1) = X$), finding

$$U(t)g = T(t)g + \int_0^t T(t-s)A_1T(s)gds + \int_0^t T(t-s)A_1 \int_0^s T(s-z)A_1U(z)gdsdz,$$

and so on, so that we can define a formal series solution of (6.15), for any $g \in X$, given by

$$U(t)g = \sum_{n=0}^{\infty} U_n(t)g, \tag{6.16}$$

or, since g is arbitrary, $U(t) = \sum_{n=0}^{\infty} U_n(t)$, where $U_0(t) = T(t)$, and, for $n \geq 1$,

$$U_{n+1}(t) = \int_0^t T(t-s)A_1U_n(s)ds. \tag{6.17}$$

The perturbation series[5] (6.16) is called Dyson series (or Dyson-Feynman-Phillips series). We need that (6.16) converges in the uniform operator topology. In fact in that case we can interchange integral and summation such that

$$\begin{aligned}\sum_{n=0}^{\infty} U_n(t)g &= T(t)g + \sum_{n=1}^{\infty} \int_0^t T(t-s)A_1U_n(t)gds \\ &= T(t)g + \int_0^t T(t-s)A_1 \sum_{n=1}^{\infty} U_n(t)gds,\end{aligned}$$

proving that (6.16) solves (6.15). The uniform convergence in this case is easy to see by means of the recurrence assumption that $\|U_n(t)\| \leq (\|A_1\|\,t)^n/n!$. To prove this notice that $\|U_0(t)\| \leq 1$ and, for any $n \geq 0$,

$$\begin{aligned}\|U_{n+1}(t)\| &\leq \int_0^t \|T(t-s)\|\,\|A_1\|\,\|U_n(s)\|\,ds \\ &\leq \frac{\|A_1\|^{n+1}}{n!}\int_0^t s^n ds = \frac{(\|A_1\|\,t)^{n+1}}{(n+1)!}.\end{aligned}$$

The name Dyson series is used also for the case the generators are time-dependent (then the solution operator will be a propagator). This extension will be carried out by means of the parametrix method which also allow to consider an unbounded time-dependent perturbation $A_1(t)$, provided we choose a special unperturbed problem $A_0(t)$ (see Chapters 10 and 11). We stress that the proof above of the uniform convergence of the perturbation series relies on the fact that in the variation of parameter formula for the perturbed semigroup (6.15) are involved only continuous operators. This is one of the key feature of the parametrix method.

[5]Sometimes the perturbation series (6.16) is called also Neumann series, because of its convergence (the key step of the successive approximation method). In fact, any series of operators convergent in the norm of $\mathcal{L}(X)$ is called Neumann series. E.g., let V be an operator in X with $\|V\| \leq 1$, then $\sum_n V^n$ converges and extends the concept of geometric series.

We remark also that, in the autonomous case, it holds also a second integral equation of Volterra-type, equivalent to (6.15). It is given by

$$U(t)g = T(t)g + \int_0^t U(s)A_1T(t-s)gds, \tag{6.18}$$

for every $g \in X$ and $t \geq 0$. Then we have the second Dyson series $U(t) = \sum_{n=0}^{\infty} V_n(t)$, where $V_{n+1}(t) = \int_0^t V_n(s)A_1T(t-s)ds$, equivalent to (6.16). We note also that in the time-dependent case, where generally the operators do not commute and must be time-ordered, (6.18) is related to the adjoint operators of the generators involved (see Chapter 10).

Part II

Perturbation Techniques And Applications

Chapter 7

Regular Perturbations

We present a regular perturbation expansion for a PDE problem with a small parameter ϵ. The main result is a set of sufficient conditions for the perturbation of the diffusion coefficient for a scalar BS PDE, recently given in [CaCo06]. References for regular perturbations of differential problem are [Bel64] and [Geo95]. About the perturbation of the spectra of a linear operator see [Kat80] and [ReSi80]. For the Sturm-Liouville theory, references are [Wei65], [Zet97] and [Zet05].

7.1 The Perturbation Expansion

We consider the classical perturbation obtained introducing a small parameter $\epsilon \geq 0$ in the PDE problem (see Section 6.2). In other words, we are interested in small perturbations of a well known auxiliary result, say that $u_0(x,t)$ is its solution, used as the centre of the expansion. We aim to find a regular perturbation, so we have to prove that the problem admits an absolutely convergent series solution representation of the form $\sum_{n=0}^{\infty} \epsilon^n u_n(x,t)$, for a strictly positive ray of convergence. Potentially any solvable model can be used as auxiliary model, but the proof of the convergence of the power series is in general quite tricky, if feasible at all, even for very simple perturbations of simple solvable models (see Section 9.3). Furthermore, no general result is available and one has to study the problem on a case by case basis.

The Problem

We start from the centre of the approximation, that is a solvable well known model that we want to perturb. We make the (sometimes restrictive, if we are interesting pricing derivatives, less for estimation ends) assumption that the considered diffusions are autonomous. The reason will be clear later and it is not strictly required to get the regularity of the perturbation expansion (i.e., the absolute convergence of the perturbation series in ϵ), but it is crucial in the specific result that we present here. In particular, let

$$A_0 = \sum_{i=1}^{d} \mu_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} \tag{7.1}$$

be the Markov generator of a time-homogeneous diffusion, say $(S_0(t), t \geq 0)$, strong solution of the associated SDE (see (2.1), with the coefficients as given in (7.1)). It is crucial that we know in closed-form the transition density $p_0(x,y;t)$ of $S_0(t)$. We stress that, potentially, (7.1) can be the generator of any auxiliary solvable SDE-type model, such those reviewed in Chapters 2,3.

The problem is to approximate the solution of the autonomous Cauchy problem, written here forward in time[1],

[1] Recall that the Cauchy problems connected to assumptions 6. − 7., e.g., (3.5), are set backward in time. Here

$$\begin{aligned} -\frac{\partial}{\partial t}u(x,t;\epsilon) + A(\epsilon)u(x,t;\epsilon) = 0, &\quad \mathbb{R}^d \times (0,T], \\ u(x,0;\epsilon) = g(x), &\quad \mathbb{R}^d, \end{aligned} \tag{7.2}$$

where $\epsilon \geq 0$ is the dimensionless small perturbation parameter and the operator $A(\epsilon)$ is defined by

$$A(\epsilon) := A_0 + \epsilon A_1, \tag{7.3}$$

A_0 given by (7.1). A_1 is a suitable perturbation[2] of A_0 such that the Cauchy problem (7.2) is well posed in a (right) neighborhood of $\epsilon = 0$, say $\mathcal{U}(\epsilon) = [0, \overline{\epsilon})$, for some $\overline{\epsilon} > 0$. Under our assumptions 1. − 7. (see Chapters 2 and 3), this is equivalent to ask that, for all $\epsilon \in \mathcal{U}(\epsilon)$, there exists a diffusion $(S_\epsilon(t), t \geq 0)$ strong solution of the associated SDE, so that $A(\epsilon)$, ϵ fixed, is the Markov generator of $S_\epsilon(t)$. In fact, the solution of problem (7.2), $\epsilon \in \mathcal{U}(\epsilon)$ fixed, with the special initial condition $u(x,0;\epsilon) = \delta(x)$ is given by the transition density of $S_\epsilon(t)$, that we suppose we cannot compute in closed-form, unless we turn the perturbation off, i.e., $\epsilon = 0$.

A regular perturbation method in this context accounts to prove that such a $\overline{\epsilon} > 0$ exists, given the decomposition (7.3). The idea is to show: 1) that the solution $u(x,t;\epsilon)$ of (7.2) has the power series representation

$$u(x,t;\epsilon) = \sum_{n=0}^{\infty} \epsilon^n u_n(x,t), \tag{7.4}$$

where $u_n(x,t)$, $n \geq 0$, will solve suitable PDEs related to (7.2); 2) that (7.4) is absolutely convergent for a positive ray of convergence $\overline{\epsilon}$, or, equivalently, that the solution $u(x,t;\epsilon)$ of (7.2) is analytic in ϵ, for $\epsilon \in [0, \overline{\epsilon})$.

We make some remarks: i) If 1)-2) above are satisfied, then we see that (7.4) defines a classical solution of (7.2), for $\epsilon \in [0, \overline{\epsilon})$. In fact, by a standard fact about power series (e.g., see [PaSa90], Vol.2, Section 3.2.2), if (7.4) converges absolutely in $[0, \overline{\epsilon})$, then it also converges uniformly in $[0, \overline{\epsilon} - \varepsilon]$, for all $0 < \varepsilon < \overline{\epsilon}$, and the same is true for the series of the derivatives when each term in (7.4) solves a particular PDE (see the next Section). ii) The analyticity in $\epsilon \in [0, \overline{\epsilon})$ of $u(x,t;\epsilon)$ means that the problem varies smoothly in ϵ, i.e., there is stability for A_0 w.r.t. the perturbation A_1. The main role of ϵ in (7.3) is to bound the effects of the unbounded perturbation A_1. iii) In particular, no jump in the mapping $\epsilon \mapsto u(x,t;\epsilon)$ can occur when the perturbation is turned on, i.e., for $\epsilon \neq 0$ near $\epsilon = 0$. This also means that, for ϵ very small, the leading term $u_0(x,t)$ is virtually correct. In fact, if a series solution such as (7.4) can be derived, it is clear (provided $u_n(x,t)$, $n \geq 0$, do not depend on ϵ), that the series (7.4) is always at least asymptotic as $\epsilon \downarrow 0$. This is another strong reason to introduce the dimensionless parameter ϵ in (7.3).

Derivation of The Series Solution

There are several ways to find a formal series solution of (7.2) in the form (7.4). We propose to use an iterative method. As already stressed, we have that $u_0(x,t) := u(x,t;0)$ solves the Cauchy problem (in fact, (7.2) with $A(0) =: A_0$)

$$\begin{aligned} -\frac{\partial}{\partial t}u_0(x,t) + A_0 u_0(x,t) = 0, &\quad \mathbb{R}^d \times (0,T], \\ u_0(x,0) = g(x), &\quad \mathbb{R}^d, \end{aligned} \tag{7.5}$$

with explicit solution given by the averaging formula (recall assumption 7.)

$$u_0(x,t) = \int_0^t g(y) p_0(x,y;t) dy, \tag{7.6}$$

we set (7.2) forward in time since the generators are autonomous, so that we can apply the change of variables $\tau = s - t$ in (3.5), which inverts the time direction.

[2]Notice also that we do not need to ask A_1 to be a Markov generator.

for any suitable initial condition $g(x)$. Now, we notice that we can write the homogeneous PDE in (7.2) as the inhomogeneous PDE (for $t > 0$)

$$-\frac{\partial}{\partial t}u(x,t;\epsilon) + A_0 u(x,t;\epsilon) = -\epsilon A_1 u(x,t;\epsilon). \tag{7.7}$$

Then we can solve (7.7) w.r.t. the Markov generator A_0: using again assumption 7. (see also (A.26), (6.9) and (6.15)), the solution of (7.7) subject to the initial condition $u(x,t;\epsilon) = g(x)$ has the following formal representation[3]

$$\begin{aligned} u(x,t;\epsilon) &= \int_{\mathbb{R}^d} g(y)p_0(x,y;t)dy + \epsilon \int_0^t \int_{\mathbb{R}^d} p_0(x,y;\tau)A_1 u(y,\tau;\epsilon)dyd\tau \\ &= u_0(x,t) + \epsilon \int_0^t \int_{\mathbb{R}^d} p_0(x,y;\tau)A_1 u(y,\tau;\epsilon)dyd\tau. \end{aligned} \tag{7.8}$$

Of course we do not know explicitly the second term of (7.8). But we can iterate the integral equation (7.8) as in (6.10) (or in the Dyson series (6.17), multiplying each term by the small parameter ϵ) and get the terms of expansion (7.4) by

$$u_1(x,t) = \int_0^t \int_{\mathbb{R}^d} p_0(x,y;\tau)A_1 u_0(y,\tau)dyd\tau,$$

and in general, for $n \geq 1$,

$$u_n(x,t) = \int_0^t \int_{\mathbb{R}^d} p_0(x,y;\tau)A_1 u_{n-1}(y,\tau)dyd\tau. \tag{7.9}$$

If the perturbation A_1 has been carefully chosen such that $A_1 u_0(x,t) \in L^1(\mathbb{R}^d \times [0,T], p_0(x,y;t)dy \times dt)$, then, provided we prove that the series (7.4) converges absolutely in $[0,\overline{\epsilon})$, for some $\overline{\epsilon} > 0$, $u_n(x,t)$, $n \geq 1$, in (7.9) are well defined and, from (7.9) and once again assumption 7., we see that $u_n(x,t)$, $n \geq 1$, is the unique solution of the Cauchy problem

$$\begin{array}{cc} -\frac{\partial}{\partial t}u_n(x,t) + A_0 u_n(x,t) = -A_1 u_{n-1}(x,t), & \mathbb{R}^d \times (0,T], \\ u_n(x,0) = 0, & \mathbb{R}^d, \end{array} \tag{7.10}$$

and (7.4) is a genuine solution of (7.2).

We stress that the solution $u_n(x,t)$ of (7.10), $n \geq 1$, can in principle be computed in closed form if we know explicitly $u_{n-1}(x,t)$. Anyway, it could be that the term $A_1 u_{n-1}(x,t)$ cannot be integrated in closed-form, and one has to use some numerical approximations.

7.2 An Introduction to Sturm-Liouville Problems

We supply the basic elements of the Sturm-Liouville theory that we need. Consider the ODE

$$au'' + bu' + cu = f, \tag{7.11}$$

where all the function are real valued, defined on the (possibly infinite) real interval $\alpha < x < \beta$[4]. If (7.11) is normal, i.e., $a > 0$ for all $x \in (\alpha,\beta)$, introducing the function (the integral is indefinite)

$$p(x) = \exp\left[\int^x \frac{b(y)}{a(y)}dy\right], \tag{7.12}$$

[3]The representation formula (7.8) is only formal since we can not check at this stage that the source term $-\epsilon A_1 u(x,t;\epsilon)$ in (7.7) satisfies the necessary regularity conditions.

[4]It is often convenient to use complex valued functions and is also possible to define the problem on a domain in $\mathbb{R}^d$, but we do not use these extensions.

we can always reduce (7.11) to the equation

$$(pu')' + ru = h, \tag{7.13}$$

where $r = (c/a)p$ and $h = (f/a)p$. (7.13) is called self-adjoint form of (7.11).

An eigenvalue problem is given by the differential equation (7.11) with $a > 0$, a, b, c continuous (at least a.e.) on $[\alpha, \beta]$ and $f(x) = \lambda u(x)$[5], $\lambda \in \mathbb{C}$ undetermined, together with some boundary conditions. The goal is to find the λ's such that the equation admits a non-zero solution. Any of these λ's is called eigenvalue and the associated solution is called eigenfunction (or eigenvector). Consider the function

$$w(x) = p(x)/a(x). \tag{7.14}$$

We have that the eigenvalue problem can be written in self-adjoint form as

$$(pu')' - qu = \lambda wu, \tag{7.15}$$

where $q = -cw$. The problem is unaltered by the transformations in the sense that eigenvalues and eigenfunctions remain the same. To solve an eigenproblem is necessary to take into account some boundary conditions. The more general conditions consider the behavior of u and u' at α, β. Let $U := (u, pu')^\top$ and M, N be (2×2) matrices with real entries (again, it is possible and often convenient to consider complex entries). We suppose that the boundary conditions of problem (7.15) are given by $MU(\alpha) + NU(\beta) = 0$, where at least one of the entries of each matrix M and N is different from 0. An eigenproblem in self-adjoint form with the described boundary conditions is called Sturm-Liouville (SL) problem. A SL problem, let's call it ω, is characterized by the quantities above, so that it is sometimes used the notation $\omega = \omega(\alpha, \beta, M, N, p, q, w)$. Later on we simply write ω.

Regular SL Problems

SL problems are classified as either regular or singular. A SL problem is said regular[6] if i) the interval (α, β) is finite, ii) the functions $1/p$, q and w are absolutely integrable in (α, β), i.e., $1/p, q, w \in L^1(\alpha, \beta)$, (this is true, e.g., if p, p', q and w are continuous a.e. on $[\alpha, \beta]$) and iii) p, w are positive (at least a.e.). Any non-regular SL problem is called singular.

Denote $\langle f, g \rangle$ the natural inner product of the Hilbert space $L^2(\alpha, \beta; w(x)dx)$. We have the following result

Theorem 16 *For a regular SL problem ω, the following are true:*

1. *The spectrum of ω, i.e., the spectrum of each of the equivalent differential operators in the LHSs of (7.11) and (7.15), is purely discrete: $\sigma = \sigma_{disc}$. In particular, there exists a countable number of eigenvalues $\{\lambda_n\}_{n \geq 1}$ that are real, simple, bounded above and unbounded below. They could be ordered such that $\infty > \lambda_1 > \lambda_2 > \cdots$, $\lambda_n \to -\infty$.*

2. *The eigenfunctions $\{u_n\}$ associated to $\{\lambda_n\}$ form an orthonormal basis for $L^2(\alpha, \beta; w(x)dx)$. So any function $f \in L^2(\alpha, \beta; w(x)dx)$ have representation $f = \sum_{n=0}^{\infty} c_n u_n$, $c_n = \langle f, u_n \rangle$, convergent in the norm of $L^2(\alpha, \beta; w(x)dx)$. However, if we take $c_n = \langle f, u_n \rangle / \langle u_n, u_n \rangle$, then the convergence is pointwise in (α, β) (but other conditions are needed such that the series converges pointwise also at α, β).*

[5] Typically, eigenproblems are defined by $f(x) = -\lambda u(x)$. The only effect of the $-$ is to mirror about the imaginary axis the spectrum of the differential operators studied.

[6] There are other possible definitions of regular SL problem allowing the interval (α, β) to be infinite (see [Zet97] and [Zet05] for the maximum generality).

3. *We can associate to* ω *the norm*

$$\|\omega\|_\Omega = |\alpha + \beta| + \|M + N\| + \|1/p + q + w\|_{L^1(\alpha,\beta)}\,,$$

making the problem space $\Omega = \{\omega\}$ *a Banach space.*

4. *Eigenvalues and eigenvectors depend continuously on the problem* ω*. In other words, for any* $\varepsilon > 0$*, given a problem* $\omega_0 \in \Omega$ *near* ω *in the sense that exists* $\delta > 0$ *such that if* $\|\omega - \omega_0\|_\Omega < \delta$*, then, for each eigenvalue* $\lambda(\omega_0)$ *of* ω_0*, there is an eigenvalue* $\lambda(\omega)$ *such that* $|\lambda(\omega) - \lambda(\omega_0)| < \varepsilon$*. The same continuity holds for the eigenfunctions.*

We stress that the regularity of the problem is only sufficient for the theorem. In particular it is possible that a singular problem have a purely discrete spectrum. We give now the simplest example of regular SL problem, that explain some typical features of any regular problem, even if we do not enter in these details. Consider the Fourier equation

$$u'' = \lambda u, \tag{7.16}$$

on the bounded interval $(0, 1)$, with boundary conditions $u(0) = u(1) = 0$. In accordance to the theorem, it is well known that (7.16) has eigenvalues $\lambda_n = -n^2\pi^2$, $n = 1, 2, \ldots$, with associated eigenfunctions $u_n(x) = \alpha_n \sin(n\pi x)$ (to be normalized). Notice that the eigenfunction $u_n(x)$ oscillates (and has $n - 1$ zeros in $(0, 1)$). This explains why the expansion in point 3. of the theorem is often called Fourier series of $f \in L^2(\alpha, \beta; w(x)dx)$, despite the eigenfunctions do not need to be sine functions. We recall that a series representation such $f = \sum_{n=0}^{\infty} c_n u_n$ holds in principle for any f in any (separable) Hilbert space (see Appendix A.1), but the advantage of the Sturm-Liouville approach is that the problem of finding a basis is reduced to an eigenproblem. Moreover, at least as long as the problem is regular, it is possible to find explicitly this basis and, anyway, not extremely complicated to find a numerical solution. In fact, since the problem is set on a finite interval, one can discretize the SL problem, reducing it to a (possibly very large) eigenproblem on a finite-dimensional space (i.e., the diagonalization of a square matrix or, more generally, to a singular value decomposition of a matrix).

Singular Problems

Singular problems are extremely more involved of regular ones. Unfortunately, in the financial applications, since the state space of the statistical SDE model is typically unbounded, one has to face almost only singular problems. The spectra of a singular SL problem can be purely discrete $\sigma = \sigma_{disc}$, purely continuous $\sigma = \sigma_{ess}$, or, more often, mixed. But, in general, the basic examples have either $\sigma = \sigma_{disc}$ or $\sigma = \sigma_{ess}$. The nature of the spectrum depends on the boundary conditions and, most of all, the behavior of the problem near each infinite end-point (see, e.g., [Zwi98]).

We are interested in the simplest example of singular SL problem, given by the Fourier equation (7.16) on $(\alpha, \beta) = \mathbb{R}$. This is clearly a singular problem because (α, β) is unbounded, but notice also that $1/p = w = 1 \notin L^1(\mathbb{R})$, ensuring that the Fourier equation is singular in any possible definition of regular SL problem. It is well known that the spectrum of this SL problem is given by $\sigma = \sigma_{ess} = (-\infty, 0]$. It is manifest the huge difference between the regular and singular cases for the Fourier equation.

The point is that, under suitable conditions (e.g., see [Zet97], section 5), the spectra of a singular problem can be approximated by the eigenvalues of a regular version of the problem. This is possible for the Fourier equation. Let $\{\alpha_k\}_{k\geq 0}$ and $\{\beta_k\}_{k\geq 0}$ be real sequences such that $\alpha_k \to -\infty$ and $\beta_k \to \infty$, and let $\{\lambda_n(\alpha_k, \beta_k)\}_n$ be the sequence of eigenvalues of the of the Fourier equation on (α_k, β_k). We saw in the previous section that $\lambda_n(0, 1) = -n^2\pi^2$, $n = 1, 2, \ldots$.

In general, possibly with different boundary conditions, the eigenvalues of the regular Fourier equation are $\lambda_n(\alpha_k, \beta_k) = -n^2\pi^2/(\beta_k - \alpha_k)^2$ for $n = 0, 1, \ldots$[7]. Since the set

$$E = \{\lambda_n(\alpha_k, \beta_k) : n, k = 0, 1, \ldots\},$$

is dense in $\mathbb{R}$, we see that each element of the spectrum of the singular Fourier equation can be approximated by the limit of a sequence of eigenvalues of regular problems on the interval (α_k, β_k), $k = 0, 1, \ldots$.

Application to Diffusion Processes

Now we explain why we assumed in Section 9.1 that the differential operators considered, i.e., the generator A_0 of the auxiliary model and the perturbation A_1 in (7.3), are autonomous. We suppose for simplicity, in agreement with the previous section, that the diffusion $S_0(t)$ associated to A_0 is scalar. We recall that this is not strictly necessary, but the general case is more involved. Since A_0 is time-homogeneous, the Cauchy problems (7.5) (the same is true for (7.2)) can be tackled by the separation of variables method[8]. Let $u_0(x, t)$ be the solution of (7.5), we can set

$$u_0(x, t) = c(t)\varphi(x), \tag{7.17}$$

substitute this into the PDE in (7.5) and get the equation

$$\frac{c'(t)}{c(t)} = \frac{A_0\varphi(x)}{\varphi(x)}. \tag{7.18}$$

Since the LHS of (7.18) depends only on the variable t, while the RHS only on x, it must be that the unique solution (supposing (7.5) is well posed) of (7.18) is constant, say λ, so that (7.18) is equivalent to each of the following two ODEs

$$c'(t) = \lambda c(t), \tag{7.19}$$

$$A_0\varphi(x) = \lambda\varphi(x). \tag{7.20}$$

From (7.19) we have clearly $c(t) = const.e^{\lambda t}$, so that all the informations about λ must be derived from (7.20). Since the operator A_0 is a second order differential operator, (7.20) is an ODE in the form (7.11). If the diffusion coefficient in A_0 is strictly positive on the support of the diffusion, i.e., the diffusion is non-degenerate, than (7.20) is also normal and defines a SL problem that can be written in self-adjoint form (7.15). We remark that, in this context, the function $1/p(x)$, p given in (7.12), is called the scale density of $S_0(t)$, while the weight function $w(x)$ in (7.14) is called the speed density of $S_0(t)$.

In the financial applications, the support (i.e., the state space) of the diffusion $S_0(t)$ is typically unbounded, so that the SL problem (7.20) is singular. However, some of the most used SDE-type models, such as the Ornstein-Uhlenbeck process (solution of the SDE (2.9)) or the Feller square-root process (solution of (2.12)) under the Feller condition, admit a purely discrete spectrum, composed of only a countable number of real, non-positive (since the associated operator are Markov generators, which spectra is included in the negative complex plane, 0 necessarily included (see Appendix A.3)) eigenvalues, say $\{\lambda_n\}_{n\geq 0}$. Other diffusions with this feature (e.g., the Jacobi process), also non-scalar, are reviewed in [KaTa81], Section 15.12, where it is also explained how to use the initial condition of the Cauchy problem (7.5) in order to obtain a spectral representation (in series form) of the solution. We sketch this. For each eigenvalue λ_n,

[7] E.g., 0 is an eigenvalue of (7.16) on (α_k, β_k) if we take $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $M = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, see [Zet97]. When 0 is an eigenvalue, the n-th eigenfunction has always n zeros.

[8] A general reference for the separation of variables methods in parabolic problems is [Pol02].

the solution of (7.19) takes the form $c_n(t) = c_n e^{\lambda_n t}$, so that, according to (7.17), $U_n(x,t) = c_n e^{\lambda_n t}\varphi_n(x)$, where $\{\varphi_n(x)\}_{n\geq 0}$ is the set of orthonormal eigenfunction associated to $\{\lambda_n\}_{n\geq 0}$, is a particular solution of the PDE in (7.5). By the linearity of the PDE, we can invoke the superposition principle (see Appendix A.2) and obtain

$$u(x,t) = \sum_{n=0}^{\infty} c_n e^{\lambda_n t}\varphi_n(x), \tag{7.21}$$

a formal solution of (7.5). We have to choose c_n according to the initial condition in (7.5) $u_0(x,0) = g(x)$. A clever choice is $c_n = \langle g, \varphi_n\rangle / \langle \varphi_n, \varphi_n\rangle$ as in point 2. of the theorem 22. Then (7.21) converges and so it is the solution of (7.5). Taking the initial condition $g(x) = \delta(x)$, under some technical conditions it holds a series spectral representation of the transition density of $S_0(t)$, given by

$$p(x,y;t) = w(y)\sum_{n=0}^{\infty} e^{\lambda_n t}\frac{\varphi_n(x)\varphi_n(y)}{\langle u_n, u_n\rangle}. \tag{7.22}$$

Apart from those cited examples, generally, the spectra of a diffusion generator on an unbounded support will have a continuous part, i.e., $\sigma_{ess} \neq \emptyset$. In this case, sometimes it is still possible to find a spectral representation of the solution of a Cauchy problem (using some functional calculus, see Appendix A.1), but now in integral form. We are interested in the following remark. Consider the generator of the Brownian motion, given by $A = \frac{1}{2}\partial^2_{xx}$. The associated SL problem is just the singular Fourier equation: $u'' = 2\lambda u = \widetilde{\lambda} u$. Therefore, the spectra of A is $\sigma = \sigma_{ess} = (-\infty, 0]$. We could say more. Since the transition density $p(x,y;t)$ of the Brownian motion is explicit, a Gaussian random variable with expectation x and variance t, we have a spectral representation of $p(x,y;t)$ given by the inverse Fourier transform of that Gaussian kernel, i.e.,

$$p(x,y;t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{i\nu x - \nu^2 t/2} e^{-i\nu y} d\nu. \tag{7.23}$$

From (7.23) (looking also at the series representation (7.22)) we see that the eigenfunctions of the Brownian motion should be given by $\varphi(x;\nu) = e^{i\nu x}$. Actually, this function satisfies the equation $\varphi''(x;\nu) = -\nu^2\varphi(x,\nu)$, for all $\nu \in \mathbb{R}$. Then, form the result about the Fourier equation, taking $-\nu^2 = \widetilde{\lambda}$, we have found eigenvalues and eigenfunction associated to the Brownian motion.

7.3 Perturbation of The Black-Scholes Equation

We turn to study a specific example, the classical BS model (see Sections 3.3). As explained in the first part of the thesis, this is by far the most important financial model, so it makes sense to start form this special auxiliary model. The generator of the scalar BS model (in reduced form, as shown in Section 3.3 and Chapter 4, there is not loss of generality if we consider the problem (3.13) with deterministic spot rate $r = 0$) is given by

$$A_0 := \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2}, \tag{7.24}$$

where $\sigma > 0$, $x \in (0,\infty)$. Problem (7.5) with A_0 in (7.24) is solvable in closed-form for weak conditions on the initial datum $g(x)$. In particular, we recall that the transition density associated to (7.24) is explicitly given by a lognormal distribution (see Section 2.3). We consider a perturbation problem (7.2) in the form

$$-\frac{\partial}{\partial t}u(x,t;\epsilon) + \frac{1}{2}\sigma^2 x^2\left[1 + \epsilon\Phi(x/K)\right]\frac{\partial^2}{\partial x^2}u(x,t;\epsilon) = 0, \tag{7.25}$$

for $x > 0$ and $t \in (0, T]$, with the initial condition $u(x, 0; \epsilon) = g(x)$. In other words, the perturbation is given by $A_1 = \frac{1}{2}\sigma^2 x^2 \Phi(x/K)\frac{\partial^2}{\partial x^2}$, where $K > 0$ has the same dimension of x (if $u(x, t; \epsilon)$ is the price of an European option, we can take as K the strike price). It is useful to introduce explicitly K so that x/K is dimensionless (as ϵ). Notice that we consider a function Φ independent of t, so that also the perturbation A_1 is time-independent. It is clear that the study of A_1 coincides with that of the function Φ, that we call the perturbation to the the BS equation $-\frac{\partial}{\partial t}u_0 + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2}u_0 = 0$, which clearly is obtained form (7.25) setting $\epsilon = 0$.

The goal is to find the most general perturbation Φ such that the solution $u(x, t; \epsilon)$ of (7.25) is analytic in ϵ in a (right) neighborhood of $\epsilon = 0$. A set of sufficient conditions from [CaCo06] is given by

Proposition 17 *The solution of (7.25), coupled with an appropriate initial condition, is analytic in the perturbative parameter ϵ, for ϵ in a neighborhood of $\epsilon = 0$, if: i) $\Phi(x) > 0$ and is bounded for $x \in (0, \infty)$, and ii) $\int_0^\infty \frac{\epsilon\Phi(x)}{1+\epsilon\Phi(x)}\frac{dx}{x} < \infty$.*

We make some remarks. The key condition is i), requiring the perturbation $\Phi(x)$ to be bounded. This condition is not be surprising, since the goal here is to prove the convergence of the perturbation series, and the boundedness of the diffusion matrix is one of the key conditions in the parametrix method presented in Chapter 10 (that, however, works also in d spacial dimensions). Actually, we do not require the continuity of $\Phi(x)$, and so we could, e.g., consider the perturbation $\Phi(x) = C\chi_B(x)$, χ_B the characteristic function of the Borel set $B \subset (0, \infty)$, C any positive constant. This correspond to a jump in the volatility of amplitude C times the entity of the perturbation ϵ, when this is turned on. The second assumption ii) is just a technical, but crucial, condition that will be clear later.

We stress that if the considered perturbation fails the boundedness condition i), e.g., if $\Phi(x)$, $x > 0$, is polynomial of order grater than 1, then, as already noticed, the expansion (7.4) still makes sense as a singular perturbation (see Section 6.2).

The Steps in The Proof

The technical details of the proof are quite involved and out of the scope of the thesis, but the ideas and techniques used are likely to apply to more general result and so we discuss them. This shows also the complexity of proving the regularity of a perturbation even in a simple setting. The main ideas are: 1) To reduce the Cauchy problem to a singular SL problem; 2) To infer the analyticity of the series solution (7.4) studying the spectral properties of a regular perturbed subproblem; 3) To get back to the singular SL problem and reduce its study to that of the unperturbed Fourier equation; 4) To extend the analyticity to the solution of the perturbed PDE. We explain now these 4 steps.

1) *The SL problem.* The first step in facing any PDE problem is to apply all the possible simplification of the problem. Consider the change of variables $x/K \mapsto y$, $\sigma^2 t/2 \mapsto \tau$ (note that y, τ are dimensionless) and $v(y, \tau; \epsilon) = u(x, t; \epsilon)$. The PDE (7.25) is equivalent to

$$-\frac{\partial}{\partial \tau}v(y, \tau; \epsilon) + y^2\left[1 + \epsilon\Phi(y)\right]\frac{\partial^2}{\partial y^2}v(y, \tau; \epsilon) = 0, \tag{7.26}$$

for $y > 0$ and $\tau \in (0, \sigma^2 T/2]$, with the initial condition $v(y, 0; \epsilon) = g(Ky)$. Supposing for the moment that this Cauchy problem is well posed, one can try to solve it by separation of variables. From (7.26), equation (7.20) takes the form (denote $\varphi' = \frac{\partial}{\partial y}\varphi$)

$$y^2\left[1 + \epsilon\Phi(y)\right]\varphi''(y; \epsilon) = \lambda\varphi(y; \epsilon). \tag{7.27}$$

The goal now is to write this eigenproblem into a SL problem in form (7.15). Perform a second change of variables $z = \ln y$ and set $\Psi(z) = \Phi(y)$ and $\psi(z; \epsilon) = \varphi(y; \epsilon)$. We find (we have, e.g., $\varphi'(y; \epsilon) = \psi'(z; \epsilon)\frac{dz}{dy} = \psi'(z; \epsilon)\frac{1}{y}$)

$$[1+\epsilon\Psi(z)]\left(\psi''(z;\epsilon)-\psi'(z;\epsilon)\right)=\lambda\psi(z;\epsilon), \tag{7.28}$$

for $z \in \mathbb{R}$. Finally consider the further transformation $\phi(z;\epsilon) = e^{-z/2}\psi(z;\epsilon)$. Since $\phi'' = e^{-z/2}[\psi''-\psi'] + \frac{1}{4}e^{-z/2}\psi$, by (7.28), we have

$$\phi''(z;\epsilon) = \frac{\lambda}{1+\epsilon\Psi(z)}\phi(z;\epsilon) + \frac{1}{4}\phi(z;\epsilon),$$

which is equivalent to

$$\phi''(z;\epsilon) - \frac{1}{4}\frac{\epsilon\Psi(z)}{1+\epsilon\Psi(z)}\phi(z;\epsilon) = \frac{\eta}{1+\epsilon\Psi(z)}\phi(z;\epsilon), \tag{7.29}$$

for $z \in \mathbb{R}$, where $\eta = \lambda + 1/4$.

Equation (7.29) is the singular SL problem, with (in the notation of equation (7.15)) $p(z) = 1$ (not perturbed), $q(z) = \frac{1}{4}\frac{\epsilon\Psi(z)}{1+\epsilon\Psi(z)}$ and $w(z) = \frac{1}{1+\epsilon\Psi(z)}$. Notice that for $\epsilon = 0$, the SL problem (7.29) simplifies to

$$\phi''(z;0) = \eta\phi(z;0), \tag{7.30}$$

for $z \in \mathbb{R}$, i.e., to the Fourier equation. So the considered perturbation of the BS PDE in (7.25) is in fact equivalent to a perturbation of the Fourier equation in the functions q and w (that, for the classical Fourier equation, are $q = 0$ and $w = 1$). This should not be surprising for two facts: i) as we saw, the Fourier equation is the singular SL problem associated to the arithmetic Brownian motion; ii) geometric and arithmetic Brownian motions (and their generators, i.e., respectively, (7.24) and the scalar Laplacian $\Delta = \partial_{xx}$) are one-to-one by means of the log-transform (see Section 3.3), that we have applied here.

2) *The regular subproblem.* The SL problem (7.29) has to be reduced to a finite interval (α,β), $-\infty < \alpha < 0 < \beta < \infty$. Set the boundary conditions $u(\alpha) = u(\beta) = 0$. The conditions of the proposition, after the transformations at 1) and the reduction here, take the form: i) $\Psi(z) > 0$ and bounded on $[\alpha,\beta]$, say $\sup_{[\alpha,\beta]}\Psi(z) = m$; ii) $\int_\alpha^\beta \frac{\epsilon\Psi(z)}{1+\epsilon\Psi(z)}dz < \infty$. Under these conditions, a long and complex analysis (we refer to [CaCo06]) ensures that perturbed eigenvalues and eigenfunction of the regular problem are analytic in ϵ in a neighborhood of $\epsilon = 0$. From the theorem in Section 9.1, we have also that the perturbed regular problem is continuous in the eigenvalues. In particular, the regular SL problem is continuous in (η,ϵ) jointly.

3) *Back to the singular case.* To exploit the nice properties of the Fourier equation to extend the analyticity to the singular problem, one has to show that the spectrum of (7.29) coincides with that of (7.30). First, by a theorem in [HiSh86], under $\int_\alpha^\beta \frac{\epsilon\Psi(z)}{1+\epsilon\Psi(z)}dz < \infty$, we have that the essential spectrum of (7.29) is the same of that of the singular SL problem

$$\phi''(z;\epsilon) = \frac{\eta}{1+\epsilon\Psi(z)}\phi(z;\epsilon), \tag{7.31}$$

since (7.30) and (7.31) differ only for a $L^1(\alpha,\beta)$ perturbation of the function q (here the scope of condition ii) in the proposition). Then (7.31) and (7.30) differ only for a perturbation of the weight function of the singular Fourier equation. But the norms of $L^2(\alpha,\beta)$ and $L^2(\alpha,\beta;w(z)dz)$ are equivalent thanks to $w(z) = \frac{1}{1+\epsilon\Psi(z)}$ and $\sup_{[\alpha,\beta]}\Psi(z) = m$; in fact,

$$\frac{1}{1+\epsilon m}\int_\alpha^\beta |f(z)|^2\,dz \le \int_\alpha^\beta |f(z)|^2\,w(z)dz \le \int_\alpha^\beta |f(z)|^2\,dz.$$

As a result, the essential spectra of (7.31) and (7.30) coincide. Since the spectra of (7.30) is just $\sigma = \sigma_{ess} = (-\infty, 0]$, the same is true for (7.29).

4) *Solution of the PDE.* To extend the analyticity in η (i.e., in λ in (7.27)) of the singular SL problem (7.29) to the solution of the Cauchy problem it is convenient to use a spectral

representation such as (7.23). If we apply the Fourier transform to both sides of equation (7.27), we reduce it to a Cauchy problem for an ODE in the Fourier transform[9]. Inverting the Fourier transform one gets an integral equation, which kernel is given by the solution of the singular SL problem. Since the kernel is analytic in ϵ in a neighborhood of $\epsilon = 0$ and continuous in both variables, if also the initial condition is integrable, then also the integral is analytic in ϵ.

7.4 Applications And Related Techniques

The perturbation result of the BS PDE in the previous section can be directly applied to the approximation of option prices when the underlying is described by a diffusion with generator $A = \frac{1}{2}\sigma^2 x^2 \left[1 + \epsilon\Phi(x/K)\right] \frac{\partial^2}{\partial x^2}$. The zero-order term of (7.4) is the BS price, while the other terms as in (7.9) are prices of options with null terminal payoff and continuous payoff function $-A_1 u_{n-1}(x,t)$ (also interpretable as a transaction cost function). The regularity of the perturbation is a possible explanation of why the BS model is robust to specification errors (see [Reb04]) and works in practice also when is clearly a misspecified model. As far the AMLE problem, we notice that if the perturbation in (7.26) is turned off, i.e., in the classical BS case, the estimation is not difficult, at least if the observations are evenly spaced and we aim to estimate the scalar drift and diffusion coefficients, since the returns follow a drifted Brownian motion and so are i.i.d.: the problem is that of maximizing a Gaussian likelihood. For $\epsilon \neq 0$, the problem is instead quite complicated. In this direction we prefer to study the alternative approximation of the perturbed BS model obtained from the parametrix method (Chapter 10).

However, we can suggest some useful uses of regular perturbations. There are some important applications of the Sturm-Liouville theory to the estimation of the parameters of a discretely observed diffusion and the approximation of option prices. In the well known techniques considered the eigenfunctions of the problem are supposed explicitly known. The point is that, when an explicit eigenexpansion of the transition density p of a scalar diffusion $S(t)$ on an unbounded state space (as in our problems) is available, i.e., when we know in closed-form eigenfunctions and eigenvectors, then it is typically the case that we already know in closed-form the transition density p. So, both the estimation, for which we can directly apply the maximum likelihood method, and the pricing/hedging problem are already solved as well. Therefore, the interesting cases occur a) either if the sum of the eigenexpansion (7.22) is not known, so that a truncation of the expansion gives an approximation of p, b) or, more interestingly, when one of those solvable models is suitably perturbed. We explain why the regular perturbation results of this chapter can be used to get an approximate result when the eigenfunctions are not known in closed-form, with particular regard to the BS case.

Estimating Equations Based On Eigenfunctions

Let $S(t) = S(t;\theta)$ be an autonomous scalar diffusion described by the SDE $dS(t;\theta) = \mu(x;\theta)dt + \sigma(x;\theta)dW(t)$, where θ is a real vector of parameters and μ, σ are known functions, smooth enough in their arguments to ensure the existence of a strong solution for each θ in the parameter space $\Theta \subset \mathbb{R}^m$. In this parametric setting an estimator of θ can be found by means of an estimating function, i.e., an equation $G_N(\theta) = 0$ solved for θ, for some function $G_N(\theta)$ of $N+1$ discrete observations $S(t_i;\theta)$, $i = 0, \ldots, N$, $0 = t_0 < t_1 < \cdots < t_N$, and θ. The prototype of estimating function is the score function, which is a martingale. Typically the transition function is not known explicitly, so an idea is to look for a different martingale $G_N(\theta)$, possibly with the same asymptotic optimal properties of the maximum likelihood estimator (see [BJS04] and [Sør98]).

Suppose that the SL problem associated to $S(t)$ (henceforth, for ease of notation, we suppress again the explicit dependence on θ) has a countable set of solutions $(\varphi_k(x), \lambda_k)_{k\geq 0}$ as described

[9] For an introduction to the Fourier transform method in the solution of parabolic PDEs in two variables, we refer to [Pol02].

in section 9.2. If φ_k belongs to the extended domain of the generator of $S(t)$ (denoted D_3 in section 4.1, see in particular (4.10) and the discussion there), condition satisfied if, e.g., φ_k are polynomial and μ, σ are of linear growth, then the process $Y(t) = e^{-\lambda_k t}\varphi_k(S(t))$ is a martingale (this follows by the Ito's lemma and $A(x)\varphi_k(x) = \lambda_k \varphi_k(x)$), and so (the expectation is taken under the true value of the parameter θ)

$$E\left[\varphi_n(S(t_n)) - e^{\lambda_n(t_i - t_{i-1})}\varphi_n(S(t_{i-1}))|S(t_{i-1})\right] = 0, \tag{7.32}$$

for $i = 1, \dots, N$. Hence, we can use the first K eigenfunctions (excluding the pair $(\varphi_0(x), \lambda_0)$) to define an empirical moment estimator of θ given by

$$G_{K,N}(\theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} f_k(S(t_{n-1});\theta)\left[\varphi_k(S(t_n)) - e^{\lambda_k(t_n - t_{n-1})}\varphi_k(S(t_{n-1}))\right] = 0, \tag{7.33}$$

for some known functions f_k. The simplest, natural choice is $f_k = 1$ for all k, but we refer to [KeSo99] for the optimal choice of the weight functions f_k (such that the estimating function (7.33) is *optimal* in the sense of Godambe-Heyde, see [GoHe87] and [Sør98]). In [KeSo99] it is also proven that, if the diffusion is ergodic with an invariant measure, under some technical conditions, then (7.33) is consistent and with asymptotic Gaussian distribution, as $N, K \to \infty$.

We make some remarks. i) The pros of (7.33) are that 1) the formula is invariant under twice continuously differentiable one-to-one transformations of the data[10] and 2) the asymptotic properties hold without the unrealistic assumption that $(t_n - t_{n-1}) \to 0$. The main con is that, as already remarked, if the eigenfunctions are known, it is likely that the transition density is known as well (at most we could not know the exact sum of (7.22)), and we can consider the MLE for the ergodic case (or the AMLE truncating the spectral representation (7.22)). In particular, if the MLE is available, this will have at least the same (asymptotic) efficiency of the estimator from (7.33). ii) There is a clear trade-off, for K finite, between the gain in efficiency and the complexity cost. Moreover, unless the eigenfunctions are polynomial (e.g., as in some regular SL problem, but also for the most used diffusion models, such as the Ornstein-Uhlenbeck and the CIR processes, when maximum likelihood estimation is possible), then the estimator could be find only numerically. And there is a second trade-off in the choice of the weights f_k. A non-optimal, but easy choice (like $f_k = 1$ for all k) affects the efficiency, but not the consistency. iii) If the spectra of the generator is continuous, then [KeSo99] suggest to use anyway (7.33), for some elements $\lambda_k < \cdots < \lambda_1 < 0$ in σ_{ess}, provided the eigenfunctions are known. But, in general, the eigenfunctions need not be real valued, as for the BS model (see (7.23)), which make this proposal infeasible. As a consequence we cannot apply directly the results in Section 9.3 to the setting here and it is necessary to look for other ideas and other perturbation techniques.

So, the estimator from (7.33) is interesting in particular if the diffusion is ergodic with invariant measure and the spectra of its generator is discrete, at least in a left neighborhood of 0 (i.e., the generator has a mixed spectrum; some sufficient conditions for this kind of spectra, considering also different boundary behavior, are given in [HST98], Section 4, under the additional assumptions that the diffusion is time-reversible and correctly initialized, and therefore the generator is self-adjoint on $L^2(\mathbb{R}, q(x)dx)$, where $q(x)$ is the stationary density[11]). Two relevant cases are the already mentioned Ornstein-Uhlenbeck and the CIR processes. The possibility to find a regular perturbation result for these diffusions is a very interesting open question (of course object of further research) as shown by the following discussion.

[10] Notice that this feature is crucial, since to solve a regular SL problem is often necessary to apply some transformation such the log-transform used in Section 9.3. This property follows by the Ito's formula, see [KeSo99].

[11] Notice also, e.g., from [KaTa81], page 221, that any invariant measure $q(x)$ is proportional to the speed density (i.e., the weight function $w(x)$ of the SL problem). Thereby, if the speed density $w(x)$ is strictly positive and $\int w(x)dx < \infty$ on the support of the diffusion, we can directly take $q(x) = w(x)$.

Suppose we could prove a regular perturbation result for, say, the CIR process (or any other generator for which we know in closed-form a discrete set of eigenelements). In the notation of section 9.1, let A be the perturbed generator given by (7.3) and A_0 be the generator of the CIR process. Then, by the convergence of the perturbation series (7.4), we have that any eigenpair $(\varphi_k^\epsilon(x), \lambda_k^\epsilon)$ solution of the SL problem for A can be expanded in convergent power series in the perturbation parameter ϵ in the following fashion

$$\varphi_k^\epsilon(x) = \textstyle\sum_{n=0}^\infty \epsilon^n \varphi_k^n(x), \quad \lambda_k^\epsilon = \sum_{n=0}^\infty \epsilon^n \lambda_k^n, \tag{7.34}$$

where $(\varphi_k^0(x), \lambda_k^0)$ is the k-th eigenpair of the unperturbed problem A_0, while $(\varphi_k^n(x), \lambda_k^n)$, for any $n \geq 1$, solve the SL problem associated to the PDE in (7.10), $n \geq 1$. These SL problem have a explicit solutions and so a truncation of the series in (7.34) at $M > 0$ provides an approximation of $\varphi_k^\epsilon(x)$ and λ_k^ϵ, for any fixed k. Using these approximations in (7.33) we have found an estimator of θ that, though very complicated, enjoys the same nice properties of the estimator of [KeSo99] for $N, K, M \to \infty$. Furthermore, the estimator is explicit, given the polynomial nature of the eigenfunctions of an auxiliary model such the CIR process (at least, if we take $f_k = 1$ for all k), a rare feature for a martingale estimating functions (see [Sør98]). We can also estimate the perturbative parameter ϵ appending it to θ. Finally, one could even exploit the asymptotic normality of the estimator to set an asymptotically normal test for the null $H_0 : \epsilon = 0$.

The same consideration apply to the following situation. [HaSc95] propose the estimating function

$$G_N(\theta) = \sum_{n=1}^N A_0 \phi(S(t_n)) = 0, \tag{7.35}$$

for any function $\phi \in D(A_0)$, where A_0 is given in (7.1). The obtained estimator, much easier than (7.12), is consistent pretty much under the same conditions above (in particular, it is required that the generator is self-adjoint). In [HST98] it is also suggested how to chose ϕ in practice: the best choice is the first non-constant eigenfunction and, clearly we can directly consider K pairs of eigenelements as in (7.33). When the eigenfunctions are known, the optimal estimator is (7.34), but it could be better to have an easier explicit estimator from (7.35), that remains explicit also if the statistical model is regularly perturbed.

Pricing Using Eigensecurities

[DaLi03] propose to approximate the price of derivatives using the spectral representation of the transition density. They call eigensecurities any eigenfunction of the generator[12]. If the SL problem is regular or if the diffusion is ergodic, with invariant measure and purely discrete spectra, one can clearly approximate the price of any option with payoff in $L^2(\mathbb{R}, q(x)dx)$, $q(x)$ the invariant measure, with a truncated eigenexpansion from (7.21). In particular, [DaLi03] examine the singular and regular SL problems associated with the CEV and CIR processes, which have a purely discrete spectra and known eigenfunctions under some restrictions of the parameters. Actually the eigenelements of the singular problem for the CEV process have to be found numerically, which could be against the philosophy of the method.

The key idea in [DaLi03] is to exploit that the eigenfunctions are explicitly known for most regular SL problems[13]. In fact, regular SL problems can be recovered in Valuation theory considering double barrier knockout options, i.e., options defined in a compact subset of the state

[12] That coincide with the eigenfunctions of the generated semigroup, see (11.11). This is relevant because the option price is given by the application of the semigroup to the terminal payoff of the option.

[13] The idea to reduce the SL problem truncating the state space of the diffusion has been proposed also by [DaGu01] for the inferential problem. In this case the formal justification is that, if the diffusion is ergodic and correctly initialized, then the "practical" support should be finite. In terms of our BS problem, which is null-recurrent, this arguments in [DaGu01] are not valid.

space, say $[\alpha, \beta]$, expiring if the underlying diffusion touches one of the barriers (that therefore are exit for the regular SL problem, i.e., the associate boundary conditions are $\varphi(\alpha) = \varphi(\beta) = 0$, as in Section 9.3.2). In the case the double barrier knockout options is written on a perturbed BS model, provided the payoff of the option is in $L^2(\mathbb{R})$ (since $w = 1$ is the weight function of the regular Fourier equation (7.16)), a condition satisfied, e.g., by any European put option, we can apply the result in proposition 17 (to the regular subproblem).

Chapter 8

Small-Time Expansions

The next two chapters are about the study of the so-called small-time expansions. Here we develop the results for the cases typically considered in the applications, and useful to pave the way for the more involved general results in the next chapter. General references for the semigroup approach to diffusions are [KaTa81] and [EtKu86]. Main references for asymptotic perturbation theory are [Nay73], [Hin91] and [ReSi80], Vol.4.

8.1 Introduction: Definitions And Goals

Let $(S(t), t \geq 0)$ is a d-dimensional diffusion, strong solution of the SDE (2.1). Consider the family of operators

$$A(t) = \sum_{i=1}^{d} \mu_i(x,t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{8.1}$$

for $t \in [0,T]$. In this chapter we always require that the diffusion matrix $a(x,t)$ is uniformly elliptic in $\mathbb{R}^d$ (see (A.9)). The transition density of $S(t)$ and the price of any European-style option written on the underlying $S(t)$ are given by the solution of the Cauchy problem (set here backward in time, as in Chapter 3)

$$\begin{array}{ll} \frac{\partial}{\partial t} u(x,t) + A(t)u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [0,s), \\ u(x,s) = g(x), & x \in \mathbb{R}^d, \end{array} \tag{8.2}$$

where, for the transition density $u(x,t) = p(x,t;y,s)$, $t < s := t + \tau \leq T$, the datum is $g(x) = \delta(x-y)$ as $t \uparrow s$ (see Section 3.1), while for the option price the payoff g depends on the specific contract (see Section 3.3). Furthermore, we recall that, from the results in Sections 3.3 and 4.3, we can set the reaction term to 0 in the pricing PDEs (3.13) and (3.18) (see also (3.9)). We consider explicitly the two different problems here since it turns out that the special terminal condition $g(x) = \delta(x-y)$ for the transition density helps a lot. This is true also because of the lack of smoothness for many option payoffs (see Section 8.4 below).

As explained in Chapter 4, under the Feller property, the conditional expectation operator associated to $S(t)$ generates a Feller propagator of contractions $U(t,s)$, $t < s = t + \tau$, in the Banach space $\widehat{C}(\mathbb{R}^d)$ (see (4.2)), while the (closure of the) operator $A(t)$ in $\widehat{C}(\mathbb{R}^d)$, for any t, is the generator of $S(t)$, at least when restricted to $C_0^2(\mathbb{R}^d)$. We recall also that the representation (8.1) of the generator of $S(t)$ holds also in some extended domains (see Section 4.1), in particular if we can extend $U(t,s)$ to some $L^p(\mathbb{R}^d, dQ)$ space (Q also possibly time-dependent). For the sake of generality denote X the Banach space (either $\widehat{C}(\mathbb{R}^d)$ or $L^p(\mathbb{R}^d, dQ)$) in which $U(t,s)$ and $A(t)$ are defined. Then, under assumption 7., $U(t,s)$ is the solution operator of (8.2), i.e.,

$u(x,t) = U(t,s)g(x)$, for any $g \in D(A(t)) = D$ (a subspace of X), always supposed independent of $t \in [0,T]$ (see Appendix A.3 and Chapter 11 for the conditions on the well posedness).

The goal of this chapter is to find a perturbation series expansion of the evolution family $U(t,t+\tau)$, for any $t \geq 0$, in the form

$$U(t,t+\tau)g(x) = \sum_{n=0}^{N} \frac{\tau^n}{n!} [A(t)]^n g(x) + R_{N+1}(x,\tau), \tag{8.3}$$

for suitably $g \in D$, where $R_{N+1}(x,\tau)$ is the remainder, defined as the difference between $U(t,t+\tau)g$ and its power series approximation truncated at $N > 0$. We interpret the time-span $\tau = (s-t)$ as the perturbative parameter, i.e., (8.3) is seen as a coordinate perturbation. We aim to prove that (8.3) makes sense as an asymptotic series in the parameter τ (see Section 6.2): $U(t,t+\tau)g \sim \sum_{n=0}^{N} \frac{\tau^n}{n!} [A(t)]^n g$ as $\tau \downarrow 0$. In this case, for obvious reasons, the perturbation series (8.3) is often called a small-time expansion of $U(t,t+\tau)g$. Under the assumptions stated in this section, we can easily prove the asymptotic property of (8.3) using some results from the semigroup theory.

We are interested in the study of the asymptotic series (8.3) since it clearly supplies an analytical approximation of the solution of (8.2) for small τ. In fact, (8.3) with $N = \infty$ is a formal series solution of (8.2). It would be optimal to prove the absolute convergence of the series (8.3) for a positive ray $\tau > 0$ and at least for some specific g, because in that case not only the perturbation would be regular, but also the formal solution would be a true classical solution of (8.2). However, since any diffusion generator $A(t)$ in a Banach space X is necessarily unbounded, the convergence of the perturbation series (8.3) is a delicate issue and, in general, does not hold (for a discussion see, e.g., [EnNa99], Section 2.3.a), though not excluded a priori. In the autonomous case, under very restrictive conditions it is possible to conclude the convergence of the perturbation series (see Section 11.3). In general, one has to supply some estimates of $\|[A(t)]^n g\|_X$ for A, g given.

The fact that the perturbation series is typically only asymptotic is offset by the extreme weakness of the hypothesis required: we do not strictly require any boundedness of the coefficients in (8.1)), while there is complete freedom in the choice of the model $A(t)$ (in this connection see the parametrix method, Chapter 10). Even the uniform ellipticity can be removed (see the next Chapter). Actually, the asymptotic property (and its strong consequences in terms of the transition density) of (8.3) has been often (in the autonomous case) in order to get an approximation of the transition density to compute AMLEs. We present also an application for the approximation of particular derivative prices. However, we remark that in some of those financial applications, the small-time series expansion is justified by an heuristic appealing of formula (8.3), rather than studying its asymptotic property as $\tau \downarrow 0$. We believe that the stress on this key fact and the use of the semigroup approach can improve the understanding of the applications presented.

8.2 The Autonomous Case

We start with the simpler time-homogeneous case, not only to get intuition, but also because we will need these results in for more general small-time expansions considered in Chapter 9. So, let $(S(t), t \geq 0)$ is a d-dimensional autonomous non-degenerate diffusion with drift vector $\mu(x)$ and diffusion matrix $a(x)$. In this case the generator of $S(t)$ is given by the closure of the differential operator

$$A = \sum_{i=1}^{d} \mu_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{8.4}$$

defined in the Banach space X. We recall that if the coefficients of (8.4) are regular as required in theorem 9 in Section 4.1, then the family of operators defined by

$$T(\tau)g := E\left[g(S(\tau))|S(0)\right] = E\left[g(S(\tau+t))|S(t)\right] = U(t, t+\tau)g, \tag{8.5}$$

for any $\tau, t \geq 0$, is a one-parameter Feller semigroup in $X = \widehat{C}(\mathbb{R}^d)$. More generally, let us assume that A generates a Feller semigroup (allowing possibly the coefficients in (8.4) to be unbounded). The extension of $T(\tau)$ to a semigroup of contraction in $X = L^2(\mathbb{R}^d, dQ)$ is straightforward if $S(t)$ is stationary and strongly ergodic, using an invariant distribution Q. We have also that $C_0^\infty(\mathbb{R}^d)$ is a core for A (this is implied by the fact that $C_0^\infty(\mathbb{R}^d)$ is dense in $D(A)$ and $C_0^\infty(\mathbb{R}^d)$ is invariant under the semigroup $T(\tau)$ in X generated by A). For $g \in D(A)$ the Cauchy problem (8.2) becomes

$$\begin{array}{ll} -\frac{\partial}{\partial \tau} u(x,\tau) + Au(x,\tau) = 0, & \mathbb{R}^d \times (0,T] \\ u(x,0) = g(x), & \mathbb{R}^d \end{array}, \tag{8.6}$$

due to the change of variables $\tau = s - t$, and $T(\tau)$ in (8.5) is the associated solution operator: $u(x,\tau) = T(\tau)g$, $\tau \in [0,T]$ (see Appendix A.3). Finally, the series expansion (8.3) takes the special form,

$$T(\tau)g(x) = \sum_{n=0}^{N} \frac{\tau^n}{n!} A^n g(x) + R_{N+1}(x,\tau), \tag{8.7}$$

for a suitable $g \in X$. In this special setting and the minimal, necessary (for the formula (8.7) to be derived) conditions on g, the asymptotic property as $\tau \downarrow 0$ of (8.7) easily follows. We will see that this key property is quite natural, since we will show that we can always interpret (8.7) as the Taylor series of the function $\tau \mapsto T(\tau)g$ around $\tau_0 = 0$.

8.2.1 Formal Expansions

We aim to present some ways to derive a formal series solution of (8.6) in the form (8.7). Many of these expansions could be found in the literature in different contexts (e.g., (8.7)-(8.9) is considered in [Ris89], while (8.7)-(8.14) is used in [DaFl86], in the ergodic case, for estimation ends). The goal of this section is to prove the equivalence of these expansions under the crucial hypothesis the generator A of the diffusion $S(\tau)$ is autonomous. In the general non-autonomous case, this equivalence is no longer true.

First of all we have to set some natural conditions such that the iterates of the generator A, i.e., A^n, $n \geq 2$, are well defined operators in the Banach space X. Given the explicit formula (8.4), we must ask that the coefficients $\mu_i(x), a_{ij}(x) \in C^\infty(\mathbb{R}^d)$. We noticed in Section 6.2 that to possibly have an infinite asymptotic expansion could be quite useless, at least whenever the convergence of the perturbation series is excluded. So, supposing we are interested in an expansion truncated at $N > 0$, it suffices to ask that $\mu_i(x), a_{ij}(x) \in C^{2N}(\mathbb{R}^d)$. We always suppose this later on.

The second issue are the natural conditions on the initial datum g in (8.6) such that $A^n g$, $n \geq 1$, in turn, are well defined. We define here, given the Banach space X, $D(A) \subset X$ the largest subspace in the domain of the generator of the diffusion $S(t)$ such that the generator has representation (8.4). We clearly need that $g \in D(A^n)$, $n \geq 1$. The point is that, since A is a Markov generator, provided as already required that A^n, $n \geq 2$, are well defined, then the subspace $D(A^\infty) = \cap_{n=0}^\infty D(A^n)$, and hence each $D(A^n) := \{g \in D(A^{n-1}) : A^{n-1}g \in D(A)\}$, are cores for A (e.g., see [EnNa99], Theorem 1.1.8). In particular, by the Feller property (see Section 4.1) and theorem 4, we see that $C_0^{2N}(\mathbb{R}^d)$, for each $N > 0$, is a core for A. Notice that there are two advantages when one works directly on $C_0^{2N}(\mathbb{R}^d)$ or $C_0^\infty(\mathbb{R}^d)$: i) since any closed operator is characterized by any of its restrictions to a core, it makes sense to use the smallest possible core

$C_0^\infty(\mathbb{R}^d)$ (or $C_0^{2N}(\mathbb{R}^d)$ if we are interested in (8.7) truncated at $N > 0$); ii) the characterization of $D(A^n)$, $n \geq 1$ is rarely feasible, so we do not know what actually is in $D(A^N)$ but not in $C_0^{2N}(\mathbb{R}^d)$ without additional hypotheses (see Section 4.1).

Iterative Methods

The simplest way of deriving the perturbation series (8.7) is to notice that (8.6), $g \in D(A)$, is formally equivalent to the integral equation $\int_0^\tau Au(x,t)dt = \int_0^\tau \frac{\partial}{\partial t}u(x,t)dt = u(x,\tau) - u(x,0)$, i.e.,

$$u(x,\tau) = g(x) + \int_0^\tau Au(x,t)dt. \tag{8.8}$$

Let $g \in D(A^\infty)$. We can iterate (8.8) and, if $A\int_0^\tau A^n u(x,t)dt = \int_0^\tau A^{n+1}u(x,t)dt$, for all $n \geq 1$, we get (recall that $u(x,\tau) = T(\tau)g(x)$) (8.7) with

$$R_{N+1}^{(1)}(x,\tau) = \int_0^\tau \int_0^{t_1} \cdots \int_0^{t_N} A^{N+1}T(t_{N+1})g(x)dt_{N+1}\cdots dt_1, \tag{8.9}$$

where $t_N < t_{N-1} < \cdots < t_1$. Actually, the commutativity required holds when $g \in D(A^\infty)$, but to see this is convenient to use the semigroup theory and derive (8.7)-(8.9) by means of the Dynkin's formula.

The Dynkin's formula says that, if $g \in D(A)$ (and τ is a Markov time), then

$$E\left[g(S(\tau))|S(0)=x\right] = g(x) + E\left[\int_0^\tau Ag(S(t))dt|S(0)=x\right]. \tag{8.10}$$

For $g \in D(A)$, $Ag \in X$, and so we have $E\left[Ag(S(t))|S(0)=x\right] = T(t)Ag(x) \in X$, i.e., the expectation exists finite. Then, by Fubini's theorem (e.g., see [Ash72], page 101), we have that (8.10) has the equivalent representations

$$T(\tau)g(x) = g(x) + \int_0^\tau T(t)Ag(x)dt \tag{8.11}$$

$$= g(x) + \int_0^\tau AT(t)g(x)dt, \tag{8.12}$$

where the second inequality follows since $T(t)Ag = AT(t)g$, for all $t \geq 0$ and $g \in D(A)$ (see Appendix A.3).

Now, if $g \in D(A^\infty)$, we can iterate, either (8.12) directly obtaining (8.7)-(8.9), or (8.11) with $Ag(x)$ replaced by $A^n g(x)$ at the n-th iteration obtaining

$$R_{N+1}^{(2)}(x,\tau) = \int_0^\tau \int_0^{t_1} \cdots \int_0^{t_N} T(t_{N+1})A^{N+1}g(x)dt_{N+1}\cdots dt_1, \tag{8.13}$$

which, by the same arguments above, coincides with

$$R_{N+1}^{(3)}(x,\tau) = E\left[\int_0^\tau \int_0^{t_1} \cdots \int_0^{t_N} A^{N+1}g(S(t_{N+1}))dt_{N+1}\cdots dt_1|S(0)=x\right]. \tag{8.14}$$

Notice that (8.13) is also equal to (8.9) because, e.g., for $g \in D(A^2)$ and all $t \geq 0$, $g, Ag \in D(A)$, so that $T(t)A^2g = T(t)A\,(Ag) = AT(t)Ag = A^2T(t)g$.

To see that all the expansions derived in this section are equivalent (i.e., their remainders $R_{N+1}^{(k)}(x,\tau)$, $k = 1,2,3$, are equal), it remains to show the assumption made in the derivation of (8.9). Consider $A\int_0^\tau A^nT(t)g(x)dt$, $n \geq 1$. If $g \in D(A^\infty)$, since, for any n, $t \mapsto A^nT(t)g$:

$[0, T] \to D(A)$ and $A^n T(t)g$, $A^{n+1}T(t)g$ are integrable (by the Dynkin's formula (8.12) with A replaced by A^n, A^{n+1}), we can apply formula (A.29), proving the assumed commutativity.

We remark that an expansion of $T(\tau)g(x)$ in form (8.7)-(8.13) or (8.14), basically obtained form the iterative application of the Dynkin's formula (8.10)-(8.12) or, only apparently more generally of the Ito's lemma to $A^n g(x)$, $n \geq 1$, is sometimes called in the literature stochastic Taylor (or Ito-Taylor) expansion (see [KlPl99], Chapter 5). We will see in the next section that (8.7)-(8.13) is actually a genuine Taylor expansion[1] of $T(t)g$. The key observation is that in the representation formulas for $R^{(k)}_{N+1}(x,\tau)$, $k = 1, 2, 3$, we can change the order of integration. E.g., for (8.9), since $A^{N+1}T(t_{N+1})$ depends only on t_{N+1}, we have

$$R^{(1)}_{N+1}(x,\tau) = \frac{1}{N!}\int_0^\tau (\tau - t)^N A^{N+1}T(t)g(x)dt. \tag{8.15}$$

Taylor Expansions

We start recalling (see Appendix A.3) that, since A is the generator of the strongly continuous semigroup $T(\tau)$ in (8.5), we have that $T(\tau)g$ is derivable in τ, for $g \in D(A)$ and all $\tau \geq 0$. In particular, for $g \in D(A)$, $\frac{d}{d\tau}T(\tau)g = T(\tau)Ag = AT(\tau)g$ (see (A.35)). We recall that the derivative is meant in the norm of the Banach space X, i.e., $\frac{d}{d\tau}T(\tau)g = \lim_{h\to 0} \|(T(\tau+h)g - T(\tau)g)/h\|_X$. In general, for any $n \geq 0$, if $g \in D(A^n)$, we have $\frac{d^n}{d\tau^n}T(\tau)g = T(\tau)A^n g = A^n T(\tau)g$. E.g., for $g \in D(A^2)$, $\frac{d^2}{d\tau^2}T(\tau)g = \frac{d}{d\tau}(T(\tau)Ag) = T(\tau)A^2 g$. Furthermore, the strong continuity of $T(\tau)$ implies that $\lim_{\tau\downarrow 0}\|T(\tau)A^n g - A^n g\|_X = 0$. So, under $g \in D(A^{N+1})$, for all $N > 0$, we can expand expand $\tau \mapsto T(\tau)g$, for $\tau > 0$, around $\tau = 0$ in Taylor series up to the N-th order, obtaining (8.7), with remainder (in Lagrange form)

$$R^{(4)}_{N+1}(x,\tau) = \frac{\tau^{N+1}}{(N+1)!}\left(\frac{d^{N+1}}{d\tau^{N+1}}T(\tau)g(x)\right)_{\tau=\eta} = \frac{\tau^{N+1}}{(N+1)!}A^{N+1}T(\eta)g(x), \tag{8.16}$$

where $0 < \eta < \tau$.

The equivalence of (8.16) and (8.15) follows from the mean value theorem for operator valued functions (see, e.g., [HiPh57]). In fact, for $g \in D(A^{N+1})$, $A^{N+1}T(\tau)g = \frac{d}{d\tau}A^N T(\tau)g$ and so

$$\int_0^\tau (\tau - t)^N A^{N+1}T(t)g(x)dt = A^{N+1}T(\eta)g(x)\int_0^\tau (\tau - t)^N dt,$$

with $0 < \eta < \tau$. In other words, (8.7)-(8.15) is just the same Taylor expansion (8.7)-(8.16) with remainder in integral form. Then the same is true for all the perturbation series derived in the preceding section.

We make now some remarks useful for dealing with the analogous of formula (8.7) in the non-autonomous case. The result is that we soon understand that we cannot use the direct Taylor expansion method in the time-inhomogeneous case. Consider explicitly the initial $t \geq 0$ and terminal $s = t + \tau \geq t$ times in formula (8.5), which we recall do not play any particular role here, but they do in the general case. Then we can write (8.7)-(8.16) as

$$T(s-t)g(x) = \sum_{n=0}^{N}\frac{(s-t)^n}{n!}A^n g(x) + \frac{(s-t)^{N+1}}{(N+1)!}A^{N+1}T(\nu)g(x), \tag{8.17}$$

where we have to take $t < \nu < s$. In fact, (8.17) is the Taylor expansion of both mappings: i) $t \mapsto T(s-t)g$, for $0 \leq t < s$, around $s > 0$; and ii) $s \mapsto T(s-t)g$, for $s > t$, around $t \geq 0$. To see this, let $g \in D(A)$ and notice that, for $t < s$,

[1] This is not necessary for a stochastic Taylor expansion. In particular, we can devise an Ito-Taylor expansion also in the time-inhomogeneous case, but in that case it will not be the Taylor expansion of the propagator considered (see Section 9.4).

$$\begin{aligned}\frac{\partial^+}{\partial t}T(s-t)g &= \lim_{h\downarrow 0}\left\|\frac{T(s-[t+h])g - T(s-t)g}{h}\right\|_X \\ &= \lim_{h\downarrow 0}\left\|T(s-[t+h])\left(\frac{I-T(h)}{h}\right)g\right\|_X = -T(s-t)Ag,\end{aligned}$$

using the semigroup property $T(s-t-h+h) = T(s-t-h)T(h)$ and the strong continuity of $T(s-t)$. On the other hand, for $t \le s$,

$$\begin{aligned}\frac{\partial^-}{\partial t}T(s-t)g &= \lim_{h\downarrow 0}\left\|\frac{T(s-t)g - T(s-[t-h])g}{h}\right\|_X \\ &= \lim_{h\downarrow 0}\left\|T(s-t)\left(\frac{I-T(h)}{h}\right)g\right\|_X = -T(s-t)Ag.\end{aligned}$$

In conclusion $\frac{\partial}{\partial t}T(s-t)g = -T(s-t)Ag$. The same steps prove also that, if $g \in D(A^n)$, for any $n \ge 1$, $\frac{\partial^n}{\partial t^n}T(s-t)g = (-1)^n T(s-t)A^n g$ and $\frac{\partial^n}{\partial s^n}T(s-t)g = A^n T(s-t)g$. Therefore it is clear, by the commutativity of A and $T(s-t)$, that (8.17) holds as claimed. We notice here that we can rewrite in terms of s, t also the remainders $R^{(k)}_{N+1}(x,\tau)$, $k = 1,2,3$, in (8.9), (8.13) and (8.14); the only change is in the extremes of integration (s in place of τ and t in place of 0).

In the non-autonomous case where the propagator $U(s,t)$ takes the place of $T(s-t)$, one can still prove that $\frac{\partial}{\partial t}U(s,t)g = -U(s,t)A(t)g$ and $\frac{\partial}{\partial s}U(s,t)g = A(s)U(s,t)g$ as above (e.g., see [Paz83], Chapter 5). But then we have, for instance, $\frac{\partial^2}{\partial t^2}U(s,t)g = -\frac{\partial}{\partial t}\left[U(s,t)A(t)\right]g = U(s,t)\left[A(t)\right]^2 g - U(s,t)A'(t)g$ (formal writing), making the direct Taylor approach infeasible since in general $A'(t)g \neq 0$. However, the iterative methods extend with only small complication, showing once again their superiority in the derivation of formal perturbation series.

8.2.2 The Asymptotic Property

We prove now that the coordinate perturbation series (8.7)-(8.16) (and hence also (8.7) with each of its equivalent remainder representation) is an asymptotic series as $\tau \downarrow 0$. Once noticed that (8.7)-(8.16) is a genuine Taylor series, then the asymptotic property is an easy consequence of the strong continuity of the semigroup $T(\tau)$, $\tau \ge 0$, and the following remark.

Let $\varphi \in C^\infty(\mathbb{R}_+ \cup \{0\})$, meaning that φ is infinitely differentiable from the right at 0 and the derivatives extend continuously to 0. By Taylor's theorem, for any $N > 0$, we have the local approximation for $z > 0$, around $z = 0$, $\varphi(z) = \sum_{n=0}^N \alpha_n z^n + R_{N+1}(z)$, where $\alpha_n = \varphi^{(n)}(0)/n!$ and

$$|R_{N+1}(z)| = \left|\varphi(z) - \sum_{n=0}^N \frac{\varphi^{(n)}(0)}{n!}z^n\right| \le \frac{z^{N+1}}{(N+1)!}\sup_{\eta\le z}\left|\varphi^{(N+1)}(\eta)\right|.$$

Hence $|R_{N+1}(z)| \to 0$ as $z \downarrow 0$, proving that $\varphi(z) \sim \sum_{n=0}^N \alpha_n z^n$ for $z \downarrow 0$.

Notice that φ does not need to be analytic. Anyway, even if $\varphi(z)$ is analytic in a right neighborhood of $z = 0$, then the quality of the approximation is not improved adding more terms if z is "small enough", because the Taylor approximation is local and the leading term ($n = 0$) is virtually correct for z very near to $z = 0$. The situation does not substantially change also for mildly, practically small $z > 0$, because the proved asymptotic property ensures that the remainder is smaller (in absolute value) than the last term included as $z \downarrow 0$. On the other hand, in the unlucky event that just the first few terms oscillates wildly, then clearly the asymptotic approximation oscillates too, being completely useless. However, in that case, pretty much the same is true for an absolutely convergent series, for many terms will be necessary to match any precision required. As discussed in Section 6.2, a slowly convergent series could prove to be of small use, unless, of course, it is the only perturbation series available.

Now, our function $\varphi(z)$ is given by $T(\tau)g$, with τ playing the role of z. We have $g \in D(A^\infty) \subset X$ and, we can directly take $X = L^p(\mathbb{R}^d, dQ)$, so that g does not need to be bounded. Then, since $T(\tau)g$ is infinitely derivable and strongly continuous in $\tau \geq 0$, we see that

$$\left\| R^{(4)}_{N+1}(x,\tau) \right\|_X = \left\| \frac{\tau^{N+1}}{(N+1)!} A^{N+1} T(\eta) g(x) \right\|_X = O(\tau^{N+1}), \tag{8.18}$$

as $\tau \downarrow 0$, because $\eta < \tau$ and, by strong continuity, $T(\eta)g \to g$ in the strong operator topology. This is the mathematical reason to let $\tau \downarrow 0$: allow to evaluate the (order of the) remainder avoiding to compute the function $T(\eta)g$, which we cannot solve in closed-form (since $T(\eta)$ is the solution operator of the full, unsolvable Cauchy problem (8.6)). Noticing that g must be continuous to be in $D(A^\infty)$, we have that (8.18) holds uniformly in compact sets containing $x \in \mathbb{R}^d$. If, more restrictively, $X = \widehat{C}(\mathbb{R}^d)$ or $g \in C_0^\infty(\mathbb{R}^d)$, then (8.18) holds uniformly in $\mathbb{R}^d$.

We remark that the mathematical reason to let $\tau \downarrow 0$ is just to

8.3 The Non-Autonomous Case

We study the non-degenerate, time-inhomogeneous case for the transition density. The key in the extension is the special terminal condition $g(x) = \delta(x-y)$. In fact, we can work formally and exploit the properties of the Dirac's delta distribution. The integrability condition will allow the extension to other terminal conditions $g(x)$ in (8.2).

Asymptotic Expansions of The Transition Density

The first task is to find a formal perturbation series, extending formula (8.7) derived in the previous Chapter. We follow the exposition in [Ris89], Chapter 4, under the only relevant assumption that the elliptic operator (8.1) is uniformly elliptic. We observe that the first (or backward) Kolmogorov equation (see (3.3)), i.e., (8.2) in the form

$$\frac{\partial}{\partial t} p(x,t;y,s) + A(t) p(x,t;y,s) = 0, \tag{8.19}$$

subject to the terminal condition $p(x,t;y,s) \to \delta(x-y)$ as $t \uparrow s$, is formally equivalent to the integral equation[2]

$$p(x,t;y,s) = \delta(x-y) + \int_t^s A(t_1) p(x,t_1;y,s) dt_1, \tag{8.20}$$

that we can iterate, obtaining

$$\begin{aligned} p(x,t;y,s) &= \delta(x-y) + \int_t^s dt_1 A(t_1) \left(\delta(x-y) + \int_{t_1}^s dt_2 A(t_2) p(x,t_2;y,s) \right) \\ &= \delta(x-y) + \int_t^s dt_1 A(t_1) \delta(x-y) + \int_t^s dt_1 \int_{t_1}^s dt_2 A(t_1) A(t_2) p(x,t_2;y,s), \end{aligned}$$

where we can suppose that $A(t_1)$ and $\int_{t_1}^s$ commute since i) $A(t_1)$ does not contain time derivatives and ii) $p(x,t_1;y,s)$ is a diffusion transition function (see [Ris89]). Note that $t \leq t_1 \leq t_2$. Continuing in this fashion we have

$$p(x,t;y,s) = \left[1 + \sum_{n=1}^\infty \int_t^s dt_1 \int_{t_1}^s dt_2 \cdots \int_{t_{n-1}}^s dt_n A(t_1) \cdots A(t_n) \right] \delta(x-y). \tag{8.21}$$

[2]We have, as in Section 10.2, $-\int_t^s \frac{\partial}{\partial t_1} p(x,t_1;y,s) dt_1 = \int_t^s A(t_1) p(x,t_1;y,s) dt_1$.

Now, if the diffusion was time-homogeneous, i.e., $A(t) = A$ for all t, then clearly (8.21) reduces to (8.7) with $g(x) = \delta(x-y)$ and $\tau = s - t$. The presence of time dependent operators complicates the formula. It can be simplified introducing the time-ordering operators $\overrightarrow{T}$ which interchanges the operators $A(t_n)$ in (8.21) such that the the operators with larger times stand to the right. In fact, then we can change the order of integration and get (see [Ris89], page 69 and 85)

$$p(x,t;y,s) = \overrightarrow{T}\left[1 + \sum_{n=1}^{\infty} \frac{1}{n!}\int_t^s dt_1 \int_t^s dt_2 \cdots \int_t^s dt_n A(t_1)\cdots A(t_n)\right]\delta(x-y), \tag{8.22}$$

that admits the just formal writing

$$p(x,t;y,s) = \overrightarrow{T}\exp\left[\int_t^s dt_1 A(t_1)\right]\delta(x-y),$$

which should be interpret as the fact that the solution operator of (8.19) is given by a propagator $U(t,s)$. The point is that if we let $\tau = s - t \downarrow 0$, we find (always formally)

$$p(x,t;y,t+\tau) = \left[\sum_{n=0}^{N} \frac{\tau^n}{n!}\left[A(t)\right]^n + O(\tau^{N+1})\right]\delta(x-y). \tag{8.23}$$

So we have found the asymptotic formal extension to (8.7). Notice that, by the properties of the delta distribution, we can integrate (8.23) in y, as required by the averaging formulas (3.10) or (A.26). In particular, for any $g \in D(A^{N+1})$ we get the extension of (8.7)-(8.16). In fact, for $g \in D(A^{N+1})$ we can integrate termwise (the sum in (8.23) is finite)

$$\begin{aligned}
U(t,t+\tau)g(x) &= \int_{\mathbb{R}^d} g(y)p(x,t;y,t+\tau)dy \\
&= \int_{\mathbb{R}^d}\left[\sum_{n=0}^{N}\frac{\tau^n}{n!}\left[A(t)\right]^n + O(\tau^{N+1})\right] g(y)\delta(x-y)dy \\
&= \sum_{n=0}^{N}\frac{\tau^n}{n!}\left[A(t)\right]^n g(x) + O(\tau^{N+1}),
\end{aligned} \tag{8.24}$$

as $\tau \downarrow 0$. The main difference with our derivation of (8.7)-(8.16) is that here we have worked out the series formally, while we could give a rigorous argument in the previous section, which will be used in the next Chapter.

Actually, we can define also a second coordinate perturbation series starting from the second Cauchy problem satisfied by the transition density $p(x,t;y,s)$. Consider the second (forward) Kolmogorov equation (see (3.7))

$$\frac{\partial}{\partial s}p(x,t;y,s) + A^*(s)p(x,t;y,s) = 0, \tag{8.25}$$

with the initial condition $p(x,t;y,s) \to \delta(y-x) = \delta(x-y)$ as $s \downarrow t$, where $A^*(s)$ is the formal adjoint to (8.1) (see (3.6) and Appendix A.2). Then the very same steps above lead to

$$\begin{aligned}
p(x,t;y,s) &= \left[1 + \sum_{n=1}^{\infty}\int_t^s ds_1 \int_t^{s_1} ds_2 \cdots \int_t^{s_{n-1}} ds_n A^*(s_1)\cdots A^*(s_n)\right]\delta(x-y) \\
&= \overleftarrow{T}\left[1 + \sum_{n=1}^{\infty}\frac{1}{n!}\int_t^s ds_1 \int_t^s ds_2 \cdots \int_t^s ds_n A^*(s_1)\cdots A^*(s_n)\right]\delta(x-y) \\
&= \overleftarrow{T}\exp\left[\int_t^s ds_1 A^*(s_1)\right]\delta(x-y),
\end{aligned}$$

and hence to

$$p(x,t;y,t+\tau) = \left[\sum_{n=0}^{N} \frac{\tau^n}{n!} \left[A^*(t)\right]^n + O(\tau^{N+1})\right] \delta(x-y). \tag{8.26}$$

Note that the perturbation formula (8.26) must hold also in the autonomous case, defining an asymptotic series analogous to (8.7) but with A^* in place of A.

Applications of The Asymptotic Perturbation Series

The asymptotic expansions (8.23) and (8.26) provide an approximation of $p(x,t;y,t+\tau)$ under the condition that each of the two Kolmogorov equation is well posed[3]. Then we have: i) also the other Kolmogorov equation is well posed, so that the two equations (8.19) and (8.25) must be equivalent; ii) the approximation for small τ of $p(x,t;y,t+\tau)$ is good just for (8.23) and (8.26) truncated at $N=1$; as a consequence $p(x,t;y,t+\tau)$, for small τ, is well approximated by a Gaussian density. These two facts are well known, but it is interesting to prove them using the asymptotic perturbations (8.19) and (8.25). The key element is again the very special initial condition used: $g(x) = \delta(x-y)$.

First we prove a Gaussian small time approximation. Suppose, for the sake of simplicity that the diffusion $S(t)$ is scalar. Consider the following first order small-time approximation of $p(x,t;y,t+\tau)$ from (8.26), valid up to corrections of order τ^2, as $\tau \downarrow 0$,

$$\begin{aligned} p(x,t;y,t+\tau) &= \left[1+\tau A^*(t)\right]\delta(x-y) = \exp\left[\tau A^*(t)\right]\delta(x-y) \\ &= \exp\left[\tau\left(-\frac{\partial}{\partial x}\mu(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}a(x,t)\right)\right]\delta(x-y). \end{aligned}$$

Using the Fourier integral representation of the scalar delta function $\delta(x-y) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i(x-y)\eta} d\eta$ and the inverse Fourier transform of derivation ($\frac{\partial^n}{\partial x^n}h(x,t)$ has Fourier transform $(i\eta)^n \frac{\partial^n}{\partial x^n}\widetilde{h}(x,t)$, $\widetilde{h}$ the Fourier transform of h) we have

$$p(x,t;y,t+\tau) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{\tau\left(-i\mu(x,t)\eta - \frac{1}{2}a(x,t)\eta^2\right)} e^{i(x-y)\eta} d\eta, \tag{8.27}$$

which is the inverse characteristic function (the inverse Fourier transform up to a constant) of a Gaussian random variable in y with mean $x+\mu(x,t)\tau$ and variance $a(x,t)\tau$. We make some remarks: i) The correctness of (8.27) is justified also by the observation that it has the right infinitesimal moments as prescribed by conditions (2.3)-(2.4). ii) The Gaussian approximation in (8.27) is not unique. We refer to [Ris89] for other small times Gaussian approximations of the transition density.

The equivalence of the two Kolmogorov equations, in the sense that their solutions coincide, is a consequence of the delta function and that the operators $U(s,t) = \overrightarrow{T}\exp\left[\int_t^s dt_1 A(t_1)\right]$ and $U^*(s,t) = \overleftarrow{T}\exp\left[\int_t^s ds_1 A^*(s_1)\right]$ are adjoint on some suitable Hilbert space, since $A(t)$ and $A^*(s)$ are adjoint and formally $(B_1B_2)^* = B_1^* B_2^*$. Consider now B_x a general real differential operator acting on x, maybe time dependent. We always have $B_x\delta(x-y) = B_x^*\delta(x-y)$, which proves the equality of the formal solutions taking $B_x = U(s,t)$. To prove the claim observe that, for any $\varphi \in D(B_x)$, we have, applying the delta function to φ,

$$B_x\varphi(x) = B_x\int \varphi(y)\delta(x-y)dy = \int \varphi(y)B_x\delta(x-y)dy.$$

On the other hand, we can also apply the delta function to $B_x\varphi(x)$,

[3] A set of restrictive sufficient conditions are given in Theorem 4.

$$B_x\varphi(x) = \int B_y\varphi(y)\delta(x-y)dy = \int \varphi(y)B_y^*\delta(x-y)dy,$$

since B_x, B_x^* are adjoint. Then $\int \varphi(y)\left[B_x\delta(x-y) - B_y^*\delta(x-y)\right]dy = 0$, proving the claim being φ arbitrary.

The Asymptotic Perturbation in [Aze84]

We briefly describe a second, more specific asymptotic approximation of the transition density. [Aze84], extending similar results in [Mol75] for the autonomous case, has proven that, under the assumptions made so far, namely that (8.1) is uniformly elliptic and its coefficients are infinitely differentiable, the expansion (8.23) takes also the form

$$p(x,t;y,t+\tau) = \tau^{-\frac{d}{2}}\exp\left(-\frac{\theta\left(x,t;y\right)}{2\tau}\right)\left[\sum_{n=0}^{N}\alpha_n\tau^n + O(\tau^{N+1})\right], \tag{8.28}$$

as $\tau \downarrow 0$, where $\theta\left(x,t;y\right) := \sum_{ij} a^{ij}(x,t)\left(x_i - y_i\right)\left(x_j - y_j\right)$. The coefficients α_n, $n \geq 0$, are functions of (x,y,t,τ) and depend only on a finite number of derivatives of the coefficients of $A(t)$. This is not a complication here since [Aze84] proves also that there exists a constant, say C, such that $\sum_{n=0}^{N}|\alpha_n| \leq C$ and $\left|O(\tau^{N+1})\right| \leq C(\tau^{N+1})$ as $\tau \downarrow 0$. This implies that (8.28) holds uniformly on compact sets, which is an important fact allowing to integrate (8.28) w.r.t. y. [Aze84] gives also a probabilistic formula for the coefficients α_n and supplies a study of the robustness of (8.28) w.r.t. different kinds of perturbations and distortions. We refer to the original paper for the complex proof of all these properties. We notice that the key tool used is the parametrix method presented in Chapter 10, so that the explicit Gaussian approximation of the transition density in (8.28), actually a refinement of the rough small-time approximation of (8.27), is in accordance to (A.16)-(A.17), that however hold for any $|t-s|$. The asymptotic property of the Gaussian approximation (8.28) is the "cost" from the removal of the boundedness hypothesis of the coefficients in theorem 25.

8.4 Applications

We review some well known applications of the asymptotic expansion (8.7)-(8.16), as $\tau \downarrow 0$. Actually, as already said, the asymptotic property of the perturbation series is not always stressed in the financial literature, despite being the key property, so we discuss it in detail. In particular, for the statistical applications, the asymptotic property of the perturbation series follows by the requirement that the time between observation tends to 0 as the number of observations tends to ∞. This is quite unrealistic, corresponding to the observation of the whole trajectory, and so we try to explain the practical implications of this assumption, also making a theoretical comparison with the results in Chapter 7. We conclude with an application in option valuation theory; we highlight the problems and give some preliminary motivation to the extensions carried out in Chapter 9.

Approximate Maximum Likelihood Estimation

Let $S(t_i)$, $i = 0, \ldots, N$, $0 = t_0 < t_1 < \cdots < t_N = T$, be $N+1$ discrete observations form an autonomous, non-degenerate diffusion described by the SDE $dS(t;\theta) = \mu(x;\theta)dt + \sigma(x;\theta)dW(t)$, where θ is a real vector of parameters, say $\theta \in \Theta \subset \mathbb{R}^m$, and μ, σ are known functions. As done before, we suppress the explicit dependence on the parameter θ. For the sake of easiness we can assume that $S(t)$ is scalar, despite here everything holds in full generality (in contrast to the results in Section 7.4).

The point is that, if $(t_i - t_{i-1})$ is small (for all $i \geq 1$), as shown by (8.27), then the transition density $p(x_{i-1}, x_i; t_i - t_{i-1})$ of $S(t_i)$ given $S(t_{i-1}) = x_{i-1}$ is well approximated by a Gaussian random variable with mean $x_{i-1} + \mu(x_{i-1})(t_i - t_{i-1})$ and variance $\sigma^2(x_{i-1})(t_i - t_{i-1})$, up to terms of order $O\left((t_i - t_{i-1})^2\right)$[4]. Notice that these estimates of the conditional mean and variance can be also carried out using (8.7)-(8.16), applied to the functions $g(x_i) = x_i$ and $g(x_i) = x_i^2$ and truncated at the term of order 1. We remark that the functions $g(x) = x$ and $g(x) = x^2$, and more generally any polynomial (as $g(x) = (x - x_{i-1} - \mu(x_{i-1})(t_i - t_{i-1}))^2$), are actually in $D(A)$, e.g., if the diffusion $S(t)$ is ergodic with invariant density, say $q(x)$, the operator A is defined in $L^2(\mathbb{R}, q(x)dx)$ and the infinitesimal parameters of $S(t)$ are of linear growth. Let us assume this setting, avoiding also the complication to work formally inside the so-called extended domain (as done implicitly, e.g., in [Kes97]).

A Gaussian small-time perturbation of the transition function has been used by [Flo89] (scalar case) and [Yos92] (multivariate case), to compute an AMLE of θ, extending the study in [DaFl86][5] on the loss of efficiency due discretization of the SDE for $S(t)$. Under the assumptions above, other technical details such as that $\theta = (\theta_1, \theta_2)$, where θ_1 enters μ only and θ_2 enters σ only (in particular [Flo89] and [Yos92] take, with abuse of notation, $\sigma(x; \theta_2) = \theta_2 \sigma(x)$) and minimal identification assumptions, the main results in [Flo89], are that: i) If $(t_i - t_{i-1}) = \Delta$, constant for all $i = 0, \ldots, N$, then the approximate likelihood estimator has a bias of order $O(N\Delta^2)$, Δ small (see also [Sør97]). Therefore, if the sample size N is large, the bias can be considerable even for small Δ and, in any case, if Δ is kept fixed, the AMLE cannot be consistent as $N \to \infty$. ii) Let $(t_i - t_{i-1}) = \Delta_N = T_N/N$, for all i, where $\Delta_N \downarrow 0$ and $T_N \to \infty$ as $N \to \infty$ such that $N\Delta_N \to \infty$ as $N \to \infty$. If $N\Delta_N^3 \to 0$ as $N \to \infty$, then the AMLE is consistent and asymptotically normally distributed.

These results are substantially generalized in [Kes97]. Let us suppose the additional properties: a) $\mu, \sigma \in C^K(\mathbb{R})$, for some $K \geq 1$ (even), b) the derivatives of μ, σ are of polynomial growth in x, c) the stationary density $q(x)$ has all moments finite, at least for the true parameter value θ_0, and d) for all $m \geq 0$, $\sup_t E\left|S(t)\right|^m < \infty$. Finally, let Θ be compact, implying the existence of the (unknown) MLE, being involved only continuous functions, and drop the hypothesis that $\sigma(x; \theta_2) = \theta_2 \sigma(x)$. We explain the role of these hypothesis stating two very interesting results form [Kes97].

Assumption a) is the minimal requirement to use (8.7) truncated up to order $K/2$. Hence, for any $k \leq K/2$, any function g of polynomial growth is in $D(A^k)$ and we obtain, along the lines of Section 8.2, an expansion of $T(t_i - t_{i-1})g = E\left[g(S(t_i))|S(t_{i-1}) = x_{i-1}\right]$ in form (8.7)

$$T(t_i - t_{i-1})g(x_{i-1}) = \sum_{n=0}^{k} \frac{(t_i - t_{i-1})^n}{n!} A^n g(x_{i-1}) + R_{k+1}(x_{i-1}, t_i - t_{i-1}), \tag{8.29}$$

where we can write the remainder $R_{k+1}(x_{i-1}, t_i - t_{i-1})$ as (see (8.13))

$$R_{k+1}(x_{i-1}, t_i - t_{i-1}) = \int_0^{t_i - t_{i-1}} \int_0^{s_1} \cdots \int_0^{s_k} E\left[A^{k+1} g(S(t_i))|S(t_{i-1}) = x_{i-1}\right] dt_{k+1} \cdots dt. \tag{8.30}$$

By b) $A^{k+1}g(S(t_i))$ is of polynomial growth, i.e., $\left|A^{k+1}g(S(t_i))\right| \leq \mathit{const.}(1 + |S(t_i)|)^{\mathit{const.}}$. Then, using d) and Lemma 6 in [Kes97] (see also [Fri76], page 102), stating that, for any f of polynomial growth, $E\left[|f(S(t_i))|\,|S(t_{i-1}) = x_{i-1}\right] \leq \mathit{const.}(1 + |x_{i-1}|)^{\mathit{const.}}$, one can conclude that $R_{k+1}(x_{i-1}, t_i - t_{i-1})$ in (8.30) is bounded by

$$R_{k+1}(x_{i-1}, t_i - t_{i-1}) \leq (t_i - t_{i-1})^{k+1} \mathit{const.}(1 + |x_{i-1}|)^{\mathit{const.}}. \tag{8.31}$$

[4] This can be inferred by means of the time-discretization of the SDE using the Euler's approximation. Higher order schemes supply better approximations. See [KlPl99], Chapter 8.

[5] Also [DaFl86] use the asymptotic series (8.7). However, their goal is a specialization of (8.28) in terms of functionals of a Brownian bridge, that, however, are to be computed by simulation.

Notice that (8.31) confirms the more general results in Section 8.2. In particular, the asymptotic property as $(t_i - t_{i-1}) \downarrow 0$ of (8.29)-(8.31), uniformly in compact sets containing x_{i-1}, is plain. However, the point is that if we set additional hypothesis on the generator A in (8.4) and the function $g \in D(A)$, here that the coefficients of $S(t)$ are Lipschitz continuous and of polynomial growth like also g, it can be possible to give a useful representation saying how the remainder behaves w.r.t. the initial condition x_{i-1}.

We return to the pure statistical problem. An important observation is the following (see also [Sør97]). If we truncate (8.29) at $k \leq K/2$, then the bias of the (Gaussian type) AMLE, for $\Delta = (t_i - t_{i-1})$ small, is of order $O(N\Delta^{k+1})$. Now let $\Delta = \Delta_N = T_N/N$, such that $N\Delta_N \to \infty$ as $N \to \infty$. The main result of [Kes97] is that, under the assumptions above, if also there exists any $1 \leq k \leq K/2$ such that $N\Delta_N^{2k+1} \to 0$ as $N \to \infty$, then the AMLE, obtained substituting in a Gaussian likelihood function the approximated conditional moments from the expansion (8.29)-(8.31) truncated at k, is consistent and asymptotically normally distributed as $N \to \infty$.

We make some remarks. i) The techniques in this section work if $\Delta = (t_i - t_{i-1})$ is practically small, but how "small enough" is really small is the curse of any asymptotic perturbation. The answer can be given only case-by-case. ii) Provided $\Delta \ll 1$, then the fact that using $k > 1$ in the small-time approximation (8.29)-(8.31) implies a bias of order $O(N\Delta^{k+1})$ shows that we can always reduce the bias for k relatively large w.r.t. the given Δ. In theoretical terms, this is the meaning of the request $N\Delta_N^{2k+1} \to 0$ as $N \to \infty$, for some k, which is a weaker condition the larger is k. However we need some care. iii) If the latter condition is not met for any $k \leq K/2$, the method breaks down, and this could happen in practice. iv) But the same is true if the condition is met only for a large k (we can generally take $K = \infty$, modulo the other conditions) and the perturbation series (8.29)-(8.31) does not converge. In fact, when the series (8.29) converges for $g(x) = x, x^2$ and a positive ray $\Delta > 0$, then the approximate likelihood converges to the true likelihood as $k \to \infty$. So, for any $\Delta < 1$ inside the ray of convergence, it always pays to use a higher k, with only the obvious trade-off against the exploding complexity of the estimator. When, more likely, the perturbation series does not converges, being only asymptotic, we have to track the actual oscillation of the remainder. So, not only it must be that Δ is really small (for the problem), but also we must check that any added term really improves the approximation of the conditional moments (for the given Δ).

For an excellent discussions of the same issues, also with regard of the other assumptions (e.g., $\theta = (\theta_1, \theta_2)$) and the role of this assumption in finding the optimal rates of convergence, we refer to [Jac04]. Empirical results confirming that already the estimator in [Flo89] work well for Δ small has been performed by [KPSS92]. Another very interesting applications of (8.29) can be found in [BiSo95], in terms of approximate optimal estimating functions. Finally, notice that the Gaussian small-time approximation can be used also if the diffusion $S(t)$ is non-ergodic, and also time inhomogeneous using (8.24). However, in this case things are more complicated and, to our knowledge, in this sample setting, not studied yet.

Approximation of Bond Prices

The asymptotic perturbation (8.7) can be used also to approximate the price of some derivatives, provided some basic conditions are met. First recall that a derivative price, in a SDE-type model, is given by the risk neutral valuation formulas (3.11) or (3.17). So assume that the coefficient of the diffusion $S(t)$ are given under the risk neutral measure $\mathbf{Q}$, i.e., the conditional expectation operator is taken under $\mathbf{Q}$, $E^{\mathbf{Q}}[.|S(t)]$. An interesting application of (8.7) (but completely disregarding the key asymptotic feature of the approximation) to derivative pricing is given in [CLP99][6]. Consider a one-factor term structure autonomous model, i.e., the SDE $dr(t) = \mu(r(t))dt + \sigma(r(t))dW(t)$, where $r(t)$ is a stochastic spot rate. The price $B(r, \tau)$ of a

[6] Actually, [CLP99] apply (8.7) to assess the effects of the use of proxies, instead of the unobservable spot rate, on the estimation of the coefficients of non-affine term-structure PDEs. The potential use in the approximation of derivative prices is noticed in [MeKr06].

default-free zero-coupon bond with time to maturity τ is given by the solution of the Cauchy problem

$$\begin{array}{cc} -\frac{\partial}{\partial\tau}B(x,\tau)+AB(x,\tau)-xB(x,\tau)=0, & \mathbb{R}_{+}\times(0,T] \\ B(x,0)=1, & \mathbb{R}_{+} \end{array}, \tag{8.32}$$

where A is given in (8.4), with $d=1$. In this case it is not convenient to set to 0 the reaction term $c(x)=x$ in (8.32). It is better to define the generator $A_x=A-x$ and recall, from Theorem 14, Section 4.3, that A_x has the same domain of A, $D(A_x)=D(A)$, and generates a (non-unitary) semigroup $R(\tau)$, $\tau\geq 0$, in X given by

$$R(\tau)g := E^{\mathbf{Q}}\left[g(r(\tau))\exp\left[-\int_0^\tau r(s)ds\right]|r(0)\right].$$

Then the bond price, solution of (8.32), is given by $R(\tau)1$. If the SDE driving $r(\tau)$ is not affine, it is almost always the case that we cannot solve (8.32) in closed-form, so that for pricing, hedging, but also estimation ends, it could be necessary to approximate $R(\tau)1$. Notice that $1\in D(A)$ if r is bounded (i.e., r has finite support, recall theorem 6, Section 3.3) or $X=L^p(\mathbb{R}_+,q(x)dx)$, for any probability density $q(x)$; otherwise we could work with the formal extended domain. When the coefficient $\mu(x),\sigma(x)$ are smooth as required, formula (8.7) applies without any difficulty and the asymptotic property as $\tau\downarrow 0$ is clear too. Furthermore, the small-time expansion extends to any autonomous multi-factors term structure (and also generalized BS) model and to any other derivative with smooth payoff.

The problem here with the application of (8.7) is twofold: i) The diffusion $S(t)$ is time-homogeneous; as explained in Sections 2.3-4 and 3.3, this is a limitation, because it is crucial the correct matching of the observed prices. This is not a real problem thanks to the extension (8.24). ii) Most seriously, the payoff function g in (3.11) and (3.17) must be in $D(A^\infty)$, but many usual payoff functions, e.g., the payoff $g(x)=(x-K)^+$ of a call option, are not even derivable, i.e., $g\notin D(A)$. From this two remarks, the actual applicability of (8.7) to pricing problems is very limited. We develop an extension to cope with these issues in the next Chapter.

Chapter 9

More on Small-Time Expansions

We consider a slightly more general problem that can be solved by a small-time expansion. It turns out that the problem is equivalent to the extension of the results of Chapter 8 to the case when the Markov generator is degenerate. The applications are to both the estimation and the pricing problems, in particular we consider the new proposal in [MeKr06]. General references are [EtKu86], [Fri76], [Hin91] and [Øks06].

9.1 The Problem

In this chapter we consider the problem of approximate the following function

$$(U(t,t+\tau)f)(x,t+\tau) := E\left[f(S(t+\tau),t+\tau)|S(t)=x\right], \tag{9.1}$$

where $(S(t), t \geq 0)$ is the possibly time-inhomogeneous diffusion solution of (2.1). Notice that, according to assumption 7. (see formula (3.10)), the time integral of (9.1), $t < s$,

$$v(x,t) = \int_t^s E\left[f(S(\eta),\eta)|S(t)=x\right] d\eta \tag{9.2}$$

is the solution of the Cauchy problem (the interchange of the time-integral and the operator $U(t,t+\tau)$ in (9.2) follows from the assumption of the well posedness of (9.3) in 7. and the Fubini's theorem)

$$\begin{array}{cc} \frac{\partial}{\partial t}u(x,t) + A(t)u(x,t) = -f(x,t), & (x,t) \in \mathbb{R}^d \times [t,s), \\ u(x,s) = 0, & x \in \mathbb{R}^d, \end{array} \tag{9.3}$$

where f satisfies minimal regularity conditions and the operator $A(t)$ is given by the generator of $S(t)$, at least as restricted to functions $C_0^2(\mathbb{R}^d)$, i.e.,

$$A(t) = \sum_{i=1}^d \mu_i(x,t)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{9.4}$$

As done in the previous Chapter, we take $D(A(t)) = D$, independent of $t \in [0,T]$. In particular, if we set the Cauchy problem (9.3) in a suitable time-dependent Banach space and for some integrability conditions on f (see Appendix A.3 and Section 11.4), then the operator $U(t,s)$ in (9.1), which is a Feller propagator in $\widehat{C}(\mathbb{R}^d)$ if we suppose that the Feller property holds (as we always do), is the solution operator of (9.3).

When the transition density of the diffusion $S(t)$ is not known in closed-form, i.e., when the fundamental solution of the parabolic operator $L(t) = \frac{\partial}{\partial t} + A(t)$ is not explicitly known, then it is interesting to give an approximation of $U(t,t+\tau)f$, at least for suitably regular functions f. The simplest approach is to use the Duhamel's principle (see (A.21)-(A.23)) and reduce the problem to that studied in the previous Chapter. However, this requires that the family of auxiliary Cauchy problems (A.22) are well posed, which could be difficult to check. More importantly, the problem to find an approximation to (9.1) arise also independently of the Cauchy problem (9.3), so that again we should artificially introduce it and ask for its well posedness. This approach is well suited when the well posedness follows form the substantial regularity of the coefficients of the operator (9.4), as under the conditions of the parametrix method (see Chapter 10). When the Markov generator $A(t)$ in (9.4) is degenerate or fails the conditions of the parametrix method, it could be a good idea to study the expansion of (9.1) in terms of iterates of (9.4) in abstract terms.

In this chapter we derive, study and apply a small-time expansion of the function (9.1). Without further mention, to make the iterates of $A(t)$ meaningful, we assume that the coefficients $\mu_i(x,t)$ and $a_{ij}(x,t)$, $i,j=1,\ldots,d$, are all in $C^{\infty,\infty}(H)$, where $H := \mathbb{R}^d \times [0,T]$.

9.2 The Asymptotic Expansion

We start with the remark that, form the derivation of formula (4.8) (see (4.7)), the "generator" of $S(t)$ applied to functions $f \in C_0^{2,1}(H)$, takes the form $L(t) = \frac{\partial}{\partial t} + A(t)$, $A(t)$ given in (9.4). This is not, strictly speaking, correct, since the name infinitesimal generator of the Markov process $S(t)$ is reserved to (9.4) only (e.g., see [Dyn65] or [KaSh92]), but this is just formal definitions. Actually, $L(t)$ can be interpreted as the generator of an augmented diffusion, as we will do below. We say that a function $f = f(x,t)$ is in D if $f(.,t) \in D$ for all $t \in [0,T]$. The definition of domain of $L(t)$ is more complicated until we do not interpret it as a genuine Markov generator.

Recall that we have derived the expansion (8.7)-(8.13) iterating the Dynkin's formula (i.e., applying a so-called stochastic Taylor expansion). We can do the same here. For $f \in C^{2,1}(H)$, by Ito's lemma,

$$f(S(t+\tau),t+\tau) = f(S(t),t) + \int_t^{t+\tau} \left(A(\eta) + \frac{\partial}{\partial \eta}\right) f(S(\eta),\eta)d\eta + \sum_{i,k} \int_t^{t+\tau} \sigma_{i,k} \frac{\partial f}{\partial x_i} dW_k(\eta), \tag{9.5}$$

therefore, taking conditional expectation we have

$$E\left[f(S(t+\tau),t+\tau)|S(t)=x\right] = f(x,t) + E\left[\int_t^{t+\tau} \left(A(\eta) + \frac{\partial}{\partial \eta}\right) f(S(\eta),\eta)d\eta | S(t)=x\right], \tag{9.6}$$

provided $E\left[\sum_{i,k}\int_t^{t+\tau} \sigma_{i,k}(S(\eta),\eta) \frac{\partial f(S(\eta),\eta)}{\partial x_i} dW_k(\eta)|S(t)=x\right] = 0$. This is the case if[1]

$$\sum_{i,k} \int_t^{t+\tau} E\left[\sigma_{i,k}(S(\eta),\eta)\frac{\partial}{\partial x_i} f(S(\eta),\eta)\right]^2 d\eta < \infty, \tag{9.7}$$

which in turn is implied by $f \in C_0^{2,1}(H)$. The point is that this latter condition can be very restrictive in the applications. More generally, define $\widehat{D}$ as the set of functions $f \in C^{\infty,\infty}(H)$[2] such that (9.7) is satisfied for any $t \in [0, T-\tau]$, $\tau > 0$. In practice, to say if a specific f is in $\widehat{D}$ one has to take into explicit account the coefficients of (9.4). For instance, if $A(t)$ satisfies the assumptions of the parametrix method (see Section 10.1, in particular the coefficients

[1] This in fact ensures that the last term in the RHS of (9.5) is a true 0-mean martingale.

[2] We can ask much less smoothness for f to get the expansion below truncated at some N, namely $f \in C^{2N+2,N+1}(H)$, but the strong condition is (9.7), not $f \in C^{\infty,\infty}(H)$.

of $A(t)$ are assumed to be bounded) with Holder coefficient $\alpha = 1$, then one can prove that $\sup_{t\in[o,T]} E\left|S(t)\right|^m < \infty$ for all $m \geq 0$, so that (9.7) is satisfied for any $f \in C^{2,1}(H)$ of at most polynomial growth (actually, we could allow f to be of exponential growth and even in the Tychonov class for the specific problem, see (A.18)). This follows from the facts that (in matrix notation, $|.|$ the Euclidean norm) $|\nabla_x f(x,t)\sigma(x,t)| \leq const.(1+|x|^a)$, for some $a > 0$, and that, under the assumptions of the parametrix method, the transition density of $S(t)$ has Gaussian type tails (see theorems 4 and 25, in particular formula (A.17), and most of all [Fri76], pages 147-148).

For $f \in \widehat{D}$, $L(t)f = \left(\frac{\partial}{\partial t} + A(t)\right) f$ is well defined and, applying the Ito's lemma to $L(t+\tau)f(S(t+\tau), t+\tau)$ as in (9.5) and then taking conditional expectation, we find

$$E\left[L(t+\tau)f(S(t+\tau),t+\tau)|S(t)=x\right] = L(t)f(x,t) + E\left[\int_t^{t+\tau} \left[L(\eta)\right]^2 f(S(\eta),\eta)d\eta|S(t)=x\right]. \tag{9.8}$$

Notice that under the hypothesis that the function (9.1) exists (which is equivalent to ask that (9.3) is well posed, as assumed under 7.), by the Fubini's theorem (see [Ash72], page 101), the expectation operator and the time-integrals in (9.6) commute. We end up with (using the compact operator notation set in (9.1) and $s = t + \tau$)

$$U(t,s)f(x,s) = f(x,t) + \tau L(t)f(x,t) + \int_t^s \int_t^{t_1} U(t,t_1)\left[L(t_2)\right]^2 f(x,t_2)dt_2 dt_1. \tag{9.9}$$

Iterating this procedure, for $\left[L(t)\right]^n f \in \widehat{D}$, we find the series expansion

$$U(t,s)f(x,s) = \sum_{n=0}^{N} \frac{\tau^n}{n!}\left[L(t)\right]^n f(x,t) + R_{N+1}(x,\tau), \tag{9.10}$$

where the remainder $R_{N+1}(x,\tau) = R_{N+1}(x, s-t)$ (defined as the difference between $U(t,s)f(x,s)$ and its series expansion (9.10) truncated at N) is given by

$$R_{N+1}(x,\tau) = \int_t^s \int_t^{t_1} \cdots \int_t^{t_N} U(t,t_{N+1})\left[L(t_{N+1})\right]^N f(x,t_{N+1})dt_{N+1}\cdots dt_1, \tag{9.11}$$

for $t \leq t_N \leq \ldots \leq t_1 \leq s = t + \tau$.

So we have found a series expansion generalizing (8.7)-(8.13). We make some remarks. i) First notice the difference between (8.24) and (9.10): in the latter case we use the parabolic operator $L(t)$, instead of the elliptic operator $A(t)$. In the next section we will see that the difference is only in the notation since $L(t)$ can be interpreted as the generator of a degenerate diffusion. ii) In this connection, we notice that stochastic methods are typically used in the solution of PDEs when the problem is degenerate (see [Fre85]). In fact, here we have worked without the assumption that $A(t)$ is uniformly elliptic (see Chapter 8 and the remarks in the following section). iii) The Stochastic Taylor expansion method used here can work also if the condition $f \in \widehat{D}$ is violated. Then the remainder will be considerably more complicated than (9.11). However, with powerful stochastic methods (see [KlPl99], in particular, Chapter 5), it is possible to cope with this more general situation.

Interpretation in Terms of The Dynkin's Formula

We show that $L(t)$ is the generator of a degenerate diffusion when $A(t)$ is the generator of the diffusion $S(t)$. Let us define a new process $Z(t)$ in $\mathbb{R}^{d+1}$ by $Z(t) := (S(t), t)$ (column vector[3]) and

[3] We state the convention, set for ease of notation, that all the vector considered in this subsection are column vectors.

append the time variable t to the space variables x, setting $z := (x,t) \in \mathbb{R}^{d+1}$. Then, with abuse of notation, $f(x,t) = f(z)$ and $E\left[f(S(t+\tau), t+\tau)|S(t)=x\right] = E\left[f(Z(t+\tau))|Z(t)=z\right] =: V(t,t+\tau)f(z)$, $t,\tau \geq 0$. We have to remark two facts: i) Since $S(t)$ is a diffusion, also $Z(t)$ is a diffusion process (e.g., see [Øks06], Chapter 8). In particular, $Z(t)$ has drift $\widetilde{\mu}(z) = (\mu(x,t), 1)$ and diffusion matrix $\widetilde{a}(z) = \widetilde{\sigma}(z)\widetilde{\sigma}^\top(z)$, where $\widetilde{\sigma}_{i,j}(z) = \sigma_{i,j}(x,t)$, $i,j = 1,...,d$[4], and $\widetilde{\sigma}_{d+1,j}(z) = 0$, $j = 1,...,d$. ii) The augmented diffusion matrix $\widetilde{a}$ is singular for all $z \in \mathbb{R}^{d+1}$, i.e., the diffusion $Z(t)$ is degenerate on all the state space since there is not diffusion in the $(d+1)$-th direction. Therefore, we can directly allow the generator $A(t)$ of $S(t)$ to be degenerate. This is a substantial generalization, since we used the uniform ellipticity of $A(t)$ in the derivation, e.g., of (8.23). Actually, we could have allowed $A(t) = A$ to be degenerate in (8.7)-(8.13), but the non-degeneracy has been used (following [Ris89] in the derivation of (8.23)) to carry out the time dependent extension (8.24).

Under the Feller condition on $S(t)$, $V(t,\tau)$ is the propagator associated to $Z(t)$. The generator of $Z(t)$ restricted to $f \in \widehat{D}$ us given by $L(t) = \frac{\partial}{\partial t} + A(t)$. Here f is seen either as a function in $\widehat{C}(\mathbb{R}^{d+1})$, which is clearly true if $f(.,t) \in \widehat{C}(\mathbb{R}^d)$ since $t \in [0,T]$, or a function in some $L^p(\mathbb{R}^{d+1}, dQ(t))$ space, provided the additional condition $\int_0^T |f(x,\eta)|^p \, d\eta < \infty$, x fixed, is satisfied. Then, by the Dynkin formula (4.8), for $[L(t)]^n f \in \widehat{D}$, the conditional expectation $E\left[\int_t^{t+\tau} [L(\eta)]^n f(Z(\eta))d\eta | Z(t) = x\right]$ exists finite and so the expectation and the integral commute (a fact that we had to explicitly suppose for the equivalence of (9.8) and (9.9)). However, the problem is that, when the generator $A(t)$ is degenerate, it is almost impossible to give a characterization of $\widehat{D}$, and, in practice, the condition $[L(t)]^n f \in \widehat{D}$ is difficult to check even on a case by case basis. In this case, an idea, useful also to weaken the requirement $[L(t)]^n f \in \widehat{D}$, could be to exploit the fact that $C_0^\infty(\mathbb{R}^{d+1}) \subset \widehat{D}$ is always a core for the closure of the linear operator $L(t)$ (see theorem 10, Section 4.1, and recall that $C_0^\infty(\mathbb{R}^{d+1})$ is dense in any $L^p(\mathbb{R}^{d+1}, dQ(t))$) and so, for any $f \in C^\infty(\mathbb{R}^{d+1})$, look for a sequence $\{f_k\} \subset C_0^\infty(\mathbb{R}^{d+1})$ such that $f_k \to f$ as $k \to \infty$. Then we could approximate $f \in C^\infty(\mathbb{R}^{d+1})$ with f_k that of course is in $\widehat{D}$.

The Asymptotic Property

To study the behavior of the remainder $R_{N+1}(x,\tau)$, notice that in (9.11) there are only iterates of the operator $L(t_{N+1})$, so we can change the order of integration and get (recall that we set $s = t+\tau$)

$$R_{N+1}(x,\tau) = \int_t^s \frac{\tau^N}{N!} U(t,\eta) \left[L(\eta)\right]^N f(x,\eta) d\eta. \tag{9.12}$$

Letting $\tau \downarrow 0$, we have, by the strong continuity (or, equivalently, by the Feller property) of $U(t,\eta)$ and $\eta \downarrow 0$, $\left\| U(t,\eta) \left[L(\eta)\right]^N f - \left[L(\eta)\right]^N f \right\|_X \to 0$, for all $f \in \widehat{D}$, so that $R_{N+1}(x,\tau) = O(\tau^{N+1})$. Since $f \in \widehat{D}$ is necessarily continuous (but not bounded), again the result holds uniformly in compact sets around x. If also $f \in \widehat{C}(\mathbb{R}^{d+1})$ (e.g., $f \in C_0^2(\mathbb{R}^{d+1})$), the result is uniform in $\mathbb{R}^d$.

9.3 Approximation of The Transition Density

The problem of approximating a function such as (9.1) arise in the fundamental estimation technique in [Aït02]. The only additional assumption in [Aït02] w.r.t. the results presented in the previous section is the necessity that the diffusion $S(t)$ is non-degenerate in the state-space (but see [Aït02] how to recover scalar diffusions such the Feller's square root process (2.12) that are degenerate only at a finite boundary (typically 0) of the state space, at least for a clever

[4] Recall, from section 2.1 that we can suppose w.l.o.g. that the dispersion matrix is a $d \times d$ matrix.

specification of the parameters; however notice that this requires to pose restrictions on the estimation of the parameters).

A Sophisticated Approximation of The Likelihood

Let us assume the usual parametric model. Let $S(t) = S(t;\theta)$ be the autonomous diffusion strong solution of the SDE $dS(t;\theta) = \mu(x;\theta)dt + \sigma(x;\theta)dW(t)$ for each θ in the parameter space $\Theta \subset \mathbb{R}^m$, where μ, σ are known functions. We are given $N+1$ discrete observations $S(t_i;\theta)$, $i = 0, \ldots, N$, $0 = t_0 < t_1 < \cdots < t_N = T$, that w.l.o.g., can be supposed equidistant $\Delta = (t_i - t_{i-1})$.

Following [Aït02] (scalar case) and [Aït04] (general case) we make the following assumptions: a) the diffusion $S(t;\theta)$, possibly multivariate, is autonomous and non-degenerate; b) the coefficients μ, σ are infinitely differentiable; c) the boundaries of the state space are unattainable (as in our assumption 5.). Furthermore, in the scalar case [Aït02] defines a very exhaustive study of the possible behavior of the growth of the coefficients, while in the general case [Aït04] set a more restrictive linear growth condition. We stress that $S(t)$ is not required to be ergodic or stationary. The goal is to approximate the transition density $p(x, y; \Delta; \theta)$ of $S(t;\theta)$ when it is not known in closed form, possibly not only for very small Δ.

The starting point is the simple observation that, when we have enough information to know that p is far from Gaussian for arbitrary Δ, then the small-time approximations of the Section 8.4 can underperform substantially even for small Δ, but fixed. The key idea is to transform $S(t;\theta)$ into a diffusion with unitary dispersion (an identity $d \times d$ dispersion matrix, in general), so that we can of course proficiently approximate that by a Gaussian random variable. This can be done considering the new process (scalar case) $Y(t) = l(S(t); x, \theta)$, where l is the so-called Lamperti transform, given by

$$l(y; x, \theta) = \int_x^y \frac{1}{\sigma(\xi;\theta)} d\xi, \tag{9.13}$$

and $x = S(0)$. For practical ends (the density of $Y(t)$ is very peaked when Δ is small), [Aït02] suggests the further normalization $\Delta^{-1/2}Y(t)$. Notice that: i) $Y(t)$ is a diffusion process with unitary dispersion (and very complicated drift, but it does not matter), as follows from the Ito's lemma. ii) $Z(0) = 0$ $(a.s.)$. iii) The transformation (9.13) depends on the unknown parameter θ, so that it cannot be directly applied to the data, but is invertible. As a consequence, the transition density $q(0, y; \Delta)$ of $\Delta^{-1/2}Y(t)$ is well approximated by a standard random variable for all Δ. We stress that, in the scalar case, the non-degeneracy of $S(t)$ suffices in order to apply (9.13), while in the general case this is not enough, being necessary a special symmetry for $\sigma(x)$ (see [Aït04]). The transition density p can be recovered (at least numerically), by the inverse transformation $p(x, y; \Delta) = \left(\Delta^{1/2}\sigma(x;\theta)\right)^{-1} q(0, l(y; x, \theta); \Delta)$.

The second idea is to exploit the special Gaussian nature of $q(x, y; \Delta)$. [Aït02] devises a series expansion of q in terms of Hermite polynomials (see [Aït02], page 231, extended in [Aït04]), proving that the Hermite series expansion converges to the true density, at least for Δ in a right neighborhood of 0. Therefore, the truncated series, say q_J if stopped at the J-th term, can supply an approximation accurate as desired. However, to compute the coefficients of this truncated expansion q_J one has to evaluate the following expectation

$$V(\Delta)H_j(0) = E\left[H_j(\Delta^{-1/2}Y(t))|Y(0) = 0\right], \tag{9.14}$$

where $H_j(x)$ is the Hermite polynomial of order j, $j = 1, \ldots, J$ (e.g., see [Hoc86]). Since p is not known in closed form, the same will be almost surely the case for q and so (9.14) has to be approximated. This can be done either numerically or by means of (9.10)-(9.12) for Δ small. Notice that this latter approach is particularly suited here, since q_J is very close to a Gaussian and so it is likely that the analytic method works even for mildly large Δ, while it of course works for Δ small. The point is that we have $H_j(\Delta^{-1/2}x) = f(x,\Delta) \in \widehat{D}$. In fact, the tails of $q(x,y;\Delta)$ are Gaussian and $f(x,\Delta)$ is clearly of polynomial growth.

The main result is that i) as the number of observations $N \to \infty$, taking $J = J(N) \to \infty$ as $N \to \infty$, and ii) as $\Delta \to 0$[5], then the approximate likelihood estimator obtained inverting q_J (into p_J, say) is asymptotically equivalent to the true, unknown MLE. Numerical experiments in [Aït02], [Aït04] and also [JePo02] (see also [ELX03] for a non-autonomous scalar extension) and the applications (see [Aït96] and [AïKi05]) show that the technique works extremely well, provided Δ is small enough w.r.t. the specific example, by large offsetting the actual complexity of the method. To our knowledge this is by far the best estimation method for speed/accuracy. Notice also that, whenever it happens that the series expansion for $q(x,y;\Delta)$ converges for some $\Delta > 0$, then the AMLE here is a genuine approximation of the true unknown MLE, but the lack of convergence is not a problem in practice.

9.4 Approximation of Option Prices

We present a recent, very interesting proposal in [MeKr06] which, potentially, allows to approximate the price of any T-claim of the type considered in this thesis (see Section 3.3). The eventual goal is to extend the results of Section 8.4 for the small-time approximation of option prices to those T-contingent claims with non-derivable payoff functions. The idea hinges on the concept of model error (see Chapter 5) and on the fact that the solution of a parabolic Cauchy problem, under natural conditions, enjoys pretty much the same smoothness of the coefficients (for a brief discussion about the so-called regularizing phenomena, see Section 11.2), that here we are already supposing in $C^{\infty,\infty}(H)$ for the well definition of the iterates $[L(t)]^n$, $n \geq 2$. We explain the proposal from the point of view of perturbations, finding a formal series solution using (9.10).

Let $A_0(t)$ be a Markov, non-autonomous generator such that the Cauchy problem[6]

$$\begin{array}{lr} \frac{\partial}{\partial t}u(x,t) + A_0(t)u(x,t) - r(x,t)u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [t,T), \\ u(x,T) = g(x), & x \in \mathbb{R}^d, \end{array} \tag{9.15}$$

is well posed and has an explicit solution, say $u^{A_0(t)}(x,t)$. $A_0(t)$ is the auxiliary model. Then consider a perturbation $A_1(t)$ such that the operator

$$A(t) := A_0(t) + A_1(t), \tag{9.16}$$

is a Markov generator, i.e., it has the form (9.4) for some functional form of the coefficients, and the associated Cauchy problem

$$\begin{array}{lr} \frac{\partial}{\partial t}u(x,t) + A(t)u(x,t) - r(x,t)u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [t,T), \\ u(x,T) = g(x), & x \in \mathbb{R}^d, \end{array} \tag{9.17}$$

is well posed as well, say $u^{A(t)}(x,t)$ is the solution, clearly supposed non-explicit. It is not strictly necessary that (9.15) and (9.17) have the same terminal condition, but it is a natural choice. Notice that, the problem can be faced in the opposite direction, starting from a general problem involving a non-solvable $A(t)$ and then choose the most suitable auxiliary model $A_0(t)$. However, in the way exposed here it is sooner checked the conditions on the perturbation $A_1(t)$ (with this regard observe the setting of the problem in Section 7.3).

[5] In some special univariate, self-adjoint cases [Aït02] proves that this assumption can be removed, i.e., that the small-time perturbation for (9.14) converges for a positive ray of convergence.

[6] We could set also here, as often we do accordingly to Sections 3.3 and 4.3, the reaction terms to 0. However, since the reaction term is necessarily the same for the full and auxiliary problems (9.15) and (9.17), the termwise subtraction to get the PDE for the model error would cancel out automatically the reaction term from (9.18), the PDE for the model error.

The key observation is that the model error $e(x,t) = u^{A(t)}(x,t) - u^{A_0(t)}(x,t)$, from the results of Chapter 5, solves

$$\begin{array}{ll} \frac{\partial}{\partial t}e(x,t) + A(t)e(x,t) = -A_1(t)u^{A_0(t)}(x,t), & (x,t) \in \mathbb{R}^d \times [0,T), \\ e(x,T) = 0, & x \in \mathbb{R}^d. \end{array} \tag{9.18}$$

Clearly, $e(x,t)$ cannot be computed in closed-form, but notice that when the operators $A(t), A_0(t)$ are explicitly given, $A_1(t)u^{A_0(t)}(x,t)$ is explicit as well. In particular, under the mild conditions of proposition 14, the model error can be explicitly bounded, so that we can grossly assess the effects of the perturbation or, equivalently, the performance of the auxiliary solution $u^{A_0(t)}(x,t)$ as a first approximation of $u^{A(t)}(x,t)$. If this approximation is too rough, we can look for a perturbation expansion of the model error $e(x,t)$ to improve the approximation of $u^{A(t)}(x,t)$.

The Series Solution For The Model Error

By the Feynman-Kac theorem (theorem 2, Section 3.2), e.g., if $A_1(t)u^{A_0(t)}(x,t)$ is continuous in its arguments and of polynomial growth, then the solution of (9.18) has representation

$$e(x,t) = E\left[\int_t^T A_1(\eta)u^{A_0(\eta)}(S(\eta),\eta)d\eta|S(t)=x\right] \tag{9.19}$$

$$= \int_t^T E\left[A_1(\eta)u^{A_0(\eta)}(S(\eta),\eta)|S(t)=x\right]d\eta, \tag{9.20}$$

where the equality of (9.19) and (9.20) follows by the minimal requirement that $A_1(t)u^{A_0(t)}(x,t)$ is integrable (despite assumption 7. on the well posedness of (9.15), (9.17), we cannot infer the well posedness of (9.18) without further assumptions). Formula (9.20) is the basis for the approximation scheme for the option price proposed in [MeKr06], which basically involves the use of (9.10) to approximate the conditional expectation in (9.20), i.e.,

$$W(t,\eta)\left(A_1(\eta)u^{A_0(\eta)}(x,\eta)\right) = E\left[A_1(\eta)u^{A_0(\eta)}(S(\eta),\eta)|S(t)=x\right]. \tag{9.21}$$

We can give our theoretical justification of the results in [MeKr06]. Let us take $(T-t)$ small and $\eta < T$. By the results in Section 9.2, if $[L(t)]^n A_1(t)u^{A_0(t)}(x,t) \in \widehat{D}$, for all $n \geq 0$, we can expand (9.21) in the asymptotic series

$$W(t,\eta)\left(A_1(\eta)u^{A_0(\eta)}(x,\eta)\right) \sim \sum_{n=0}^{N} \frac{(\eta-t)^n}{n!}[L(t)]^n A_1(t)u^{A_0(t)}(x,t), \tag{9.22}$$

as $(\eta - t) \downarrow 0$. Then the approximation of (9.19) is obtained, for a very small interval $T-t$, integrating the series (9.22) as in (9.16). A case by case analysis will tell us how many terms is optimal to consider to get the best precision for the given, fixed in practice, $T-t$.

The limitation of our asymptotic justification is just that we should consider only options with a very short time-to-maturity $T-t$, which seems restrictive, because the numerical experiments in [MeKr06] show that, in specific examples, the approximation is good even for not short $T-t$. For instance, the approximation of a call option price under the (explicitly solvable) CEV model ($u^{A(t)}(x,t)$) by means of the price of the same option under the classical BS model (the auxiliary solution $u^{A_0(t)}(x,t)$) requires three additional terms (i.e., $N=2$) from the expansion (9.22), to get a relative error of order less than 0.5%, also for a time-to-maturity as long as 1 year.

It would be optimal to prove the convergence of the perturbation series, at least for special cases, but to our knowledge this is still an open problem. In the next section we study the possible regularity of the perturbation series, giving a very unsatisfactory answer, in terms of the BS example, for which we show that our key condition (stronger than the original condition in

[MeKr06]) $[L(t)]^n A_1(t)u^{A_0(t)}(x,t) \in \widehat{D}$, for all $n \geq 0$, is also satisfied for suitable perturbations $A_1(t)$. We will suppose a lot of boundedness, but we remark that the main pro of the asymptotic expansion (9.22) is that not only it is not necessary to require the uniform ellipticity and the boundedness of the coefficients of $A(t)$ and $A_0(t)$, which are very strong assumptions of the parametrix method (see Chapter 10). In other words, this approximation method has by far more applicability than the other techniques we consider in the thesis and could be considered a natural approach when one needs to use an affine (non-linear) diffusion as full and/or auxiliary models (affine diffusions are singular at least at a boundary and have unbounded coefficients, see Section 2.4).

9.5 A Study of The Black-Scholes Case

We study the application of the method of [MeKr06] to the case that the full model is a generalized BS model, with the obvious choice of the classical BS as auxiliary model. For the sake of simplicity we work out the scalar case but, at least for this special settings, there are not huge complications to work in d dimension (see also Chapter 10). As payoff function we consider two choices $g(x) = (x-K)^+$, i.e., call option, and $g(x) = (K-x)^+$, i.e., put option. In both case there is a singularity (that is, a point of not derivability) at $x = K$, the strike price.

Let the operator $A_0(t)$ in (9.15) be

$$A_0(t) = r(x,t)x\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2}, \tag{9.23}$$

where $x > 0$, $r(x,t)$ is the (stochastic) spot rate and $\sigma > 0$ is just a parameter (the volatility of the classical BS formula). The solution of the Cauchy problem

$$\begin{array}{rr} \frac{\partial}{\partial t}u(x,t) + A_0(t)u(x,t) - r(x,t)u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times [t,T), \\ u(x,T) = (x-K)^+, & x \in \mathbb{R}^d, \end{array} \tag{9.24}$$

is explicit and well known, provided $r(x,t) = r \in \mathbb{R}_+$, and is given, for all $t < T$, by (e.g., see [Duf01], page 89)

$$u^{A_0(t)}(x,t) = x\Phi(z) - e^{-r(T-t)}K\Phi\left(z(x,t) - \sigma\sqrt{T-t}\right), \tag{9.25}$$

where $z(x,t) = \left[\ln(x/K) + \left(r + \sigma^2/2\right)(T-t)\right] / \left[\sigma\sqrt{T-t}\right]$ and Φ is the distribution function of a standard Gaussian random variable. The price of the put option, say $u_{put}^{A_0(t)}(x,t)$, (i.e., (9.24) with the terminal condition $(K-x)^+$) can be recovered form (9.25) by means of the so-called put-call parity (see [Hul06], page 212),

$$u_{put}^{A_0(t)}(x,t) = u^{A_0(t)}(x,t) + Ke^{-r(T-t)} - x, \tag{9.26}$$

where $x = S(t)$, the current value of the underlying. Notice that, for $0 \leq t < T$, (9.25) (and (9.26)) is infinitely derivable in (x,t). The main difference between the put and call option is that $u^{A_0(t)}(x,t)$ is of linear growth in x, while $u_{put}^{A_0(t)}(x,t)$ is bounded on H. This is interesting since the boundedness of the auxiliary model i) implies that the asymptotic property holds uniformly in x; ii) most of all, facilitates the fulfilment of the key condition $f \in \widehat{D}$ in section 9.2.

Now consider the full model, supposing we model the returns as in (9.23) and using the notation of Section 3.3,

$$A(t) = r(x,t)x\frac{\partial}{\partial x} + \frac{1}{2}\left(\Sigma(x,t)\right)^2 x^2 \frac{\partial^2}{\partial x^2}. \tag{9.27}$$

In other words, the diffusion coefficient of the considered generalized BS model is given by $a(x,t) = \left(\Sigma(x,t)\right)^2 x^2$. We stress that here $\Sigma(x,t)$ is not required to be bounded. Suppose that for our specific choice of $\Sigma(x,t)$ (in practice for almost every choice) the solution $u^{A(t)}(x,t)$ of (9.24) with (9.27) in place of $A_0(t)$ is not explicit. The perturbation is given by the operator

$$A_1(t) = \frac{1}{2}\left((\Sigma(x,t))^2 - \sigma^2\right) x^2 \frac{\partial^2}{\partial x^2}, \tag{9.28}$$

and we have to approximate the model error

$$\begin{aligned} e(x,t) &= u^{A(t)}(x,t) - u^{A_0(t)}(x,t) \\ &= \int_t^T E\left[A_1(\eta)u^{A_0(\eta)}(S(\eta),\eta)|S(t)=x\right] d\eta, \end{aligned} \tag{9.29}$$

where we have interchanged expectation and time-integral since the integrand is integrable, as proved in the next section.

The Gamma of The Black-Scholes Model

The function $A_1(t)u^{A_0(t)}(x,t)$ is explicit and is given by

$$A_1(t)u^{A_0(t)}(x,t) = \frac{1}{2}\left((\Sigma(x,t))^2 - \sigma^2\right) x^2 \Gamma^{BS}(x,t), \tag{9.30}$$

where $\Gamma^{BS}(x,t) = \frac{\partial^2}{\partial x^2} u^{A_0(t)}(x,t)$ is the so-called gamma of the BS price. By direct inspection

$$\Gamma^{BS}(x,t) = \frac{1}{\sigma\sqrt{2\pi(T-t)}x} \exp\left[-\frac{1}{2}\left(\frac{\ln(x/K) + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right)^2\right]. \tag{9.31}$$

For $(T-t) > 0$, $\Gamma^{BS}(x,t)$ is a bounded function and, as $x \to 0$ or $x \to \infty$, $\Gamma^{BS}(x,t) \to 0$ (in particular, the higher $\sigma\sqrt{T-t}$ the slower $\Gamma^{BS}(x,t) \to 0$). The problem is what happens as $t \to T$. Let $x > 0$ be fixed and write

$$\begin{aligned} \Gamma^{B\&S}(x,\tau) &= C_1 \frac{1}{\sqrt{T-t}} \exp\left[-C_2 \frac{C_3^2 + C_4(T-t)^2 + 2C_3C_4(T-t)}{(T-t)}\right] \\ &= C_1 \frac{1}{\sqrt{T-t}} \exp\left[-\frac{C_2C_3^2}{(T-t)} - C_2C_4(T-t) - 2C_2C_3C_4\right] \end{aligned} \tag{9.32}$$

where $C_1 = \left(\sigma\sqrt{2\pi}x\right)^{-1} > 0$, $C_2 = \left(2\sigma^2\right)^{-1} > 0$, $C_3 = \ln(x/K) \in \mathbb{R}$ and $C_4 = (r+\sigma^2/2) > 0$. So, if $C_3 = \ln(x/K) \neq 0$, as $t \to T$ the leading term in (9.32) is $\exp\left[-\frac{C_2C_3^2}{\tau}\right]$ and therefore $\Gamma^{BS}(x,t) \to 0$. But if $C_3 = \ln(x/K) = 0$, i.e., $x = K$, the leading term in (9.32) as $t \to T$ is $(T-t)^{-\frac{1}{2}}$ and $\Gamma^{B\&S}(x,\tau) \to +\infty$. Hence, we have found the bad news that $\Gamma^{BS}(x,t)$ has a singularity at $(x,t) = (K,T)$.

The good news is that we can prove that the singularity is integrable. In fact, write (9.32) as

$$\Gamma^{BS}(x,t) = C_5 \frac{1}{\sqrt{T-t}} \exp\left[-\frac{C_2(\ln(x/K))^2}{T-t}\right], \tag{9.33}$$

for some constant C_5, since $\exp\left[-C_2C_4(T-t) - 2C_2C_4\ln(x/K)\right]$ is bounded as $(x,t) \to (K,T)$. Then write (9.33) as

$$\Gamma^{BS}(x,t) = C_5(T-t)^{-\alpha}(\ln(x/K))^{2\alpha-1}\left(\frac{(\ln(x/K))^2}{T-t}\right)^{\frac{1}{2}-\alpha} \exp\left[-\frac{C_2(\ln(x/K))^2}{T-t}\right]$$

for some $0 < \alpha < 1$ and notice that $\ln(x/K) \sim x - K$, as $x \to K$: in other words, $(\ln(x/K))^2 \to 0$ faster than $t \to T$ (i.e., a linear speed), so that also $\left(\frac{(\ln(x/K))^2}{T-t}\right)^{\frac{1}{2}-\alpha}$ and $\exp\left[-\frac{C_2(\ln(x/K))^2}{T-t}\right]$ are bounded as $(x,t) \to (K,T)$. Then, (for the appropriate constant C_6) as $(x,t) \to (K,T)$,

$$\left|\Gamma^{BS}(x,t)\right| \leq C_6 \frac{1}{(T-t)^\alpha} \frac{1}{[(\ln(x/K))^2]^{\alpha-\frac{1}{2}}} \sim C_6 \frac{1}{(T-t)^\alpha} \frac{1}{[(x-K)^2]^{\alpha-\frac{1}{2}}},$$

which proves that $\Gamma^{BS}(x,t)$ is Lebesgue integrable.

In particular, if $t \mapsto \Sigma(x,t)$ in (9.28) is bounded and integrable for $t \to T$ (which is satisfied if $\Sigma(.,t)$ is continuous for $t \in [0,T]$) then also (9.30) is integrable and (9.29) is justified.

The reason of both the singularity and integrability can be understood from the properties of the fundamental solution of (9.24). In fact, one of the essential feature of the fundamental solution $\Gamma(x,t;y,s)$ of any parabolic operator (see Appendix A.2) is that, as $(x,t) \to (y,s)$, $\Gamma(x,t;y,s) \to \delta(y-x)$, the Dirac's delta function. In particular, the singularity of $\Gamma(x,t;y,s)$ is integrable by the very definition of delta function. The same behavior propagates to the gamma $\Gamma^{BS}(x,t)$ by derivation of the solution of the Cauchy problem $u^{A_0(t)}(x,t)$[7], which is actually non-derivable at $(x,t) = (K,T)$, i.e., in correspondence of the non-derivability point of the initial condition $g(x) = (x-K)^+$. This is quite interesting and the same situation happens also at the point where the initial condition $g(x)$ has a jump discontinuity (e.g., for digital payoffs, that, by the way, are problematic for the Monte Carlo simulation methods of solution).

About The Convergence of The Perturbation Series

We conclude with a sketched study of the convergence of the perturbation series in the BS example (see (9.10)). Consider (9.22) where $L(t) = \frac{\partial}{\partial t} + A(t) - r$, and $A(t)$, $A_1(t)$ and $u^{A_0(t)}(x,t)$ are given, respectively, in (9.27), (9.28) and (9.25). Since (9.22) is a power series, we consider its absolute convergence. So one has to estimate the absolute value of the terms $f_n(x,t) = [L(t)]^n A_1(t) u^{A_0(t)}(x,t)$, in the hopes to find that $|f_n(x,t)| \leq |f(x,t)|\,|g_n(x,t)|$, for some functions $f(x,t)$ (independent of n, that hence can be taken out of the sum) and $g_n(x,t)$ such that $\sum_{n=0}^{\infty} \left|\frac{(\eta-t)^n}{n!} g_n(x,t)\right|$ is analytic for a positive ray $(\eta - t) > 0$. The problem is, in theory, explicit, because we know in closed-form $u^{A_0(t)}(x,t)$ and all the other functions involved in (9.22). However, notice that at n fixed, since $A_1(t)$ and $L(t)$ are second order operators, the function $f_n(x,t)$ is the sum of $2 \cdot 3^n$ terms. So to study the remainder (see (9.12)) at N requires to supply an estimate of $2 \cdot 3^{N+1}$ terms and prove that in each case the estimate allows to conclude the convergence. Unfortunately, we could not find any recurrence formula even in this explicit case (one can use any software with symbolic differentiation to get this conclusion) and the problem is really explosive. This makes (perhaps unsatisfactorily) the simple asymptotic result appealing as a justification of the approximation method for option prices (as for any other small-time expansion).

If we simplify at its maximum the problem, it is however possible to get an intuition (not a proof however, we are working on it, but we will see that is unrealistic to hope to generalize the sketched arguments here, without finding a recurrence formula) of why the convergence of (9.22) is not excluded a priori, at least when one uses the BS model as auxiliary model (as in the main numerical experiment in [MeKr06]). Recall that we can find the solution (9.31) applying the transformations described in the formulas (3.14)-(3.16) to the Cauchy problems (9.15) and (9.17) (with $r(x,t) = r = 0$). In particular, the fundamental solution of the transformed BS PDE, i.e., a heat equation (see (3.16)), is given by (here $y = \ln x$)

---

[7]Notice that the solution of the Cauchy problem is given by the averaging formula (A.20), so the derivation in x actually acts on the fundamental solution (see below).

$$\Gamma^{\sigma^2}(y,t;\zeta,s) = (2\pi\sigma^2(s-t))^{-\frac{1}{2}} \exp\left[-\frac{|y-\zeta|^2}{2\sigma^2\sqrt{s-t}}\right], \tag{9.34}$$

for $t<s$, while the terminal condition in (9.24) takes the form $g(x)=g(e^y)=(e^y-K)^+$, that now is of exponential growth. Note that, as should be, nothing has changed from the point of view of the well posedness of the Cauchy problems (e.g., see theorem 3, Section 3.2). In particular the solution $u^{A_0(t)}(x,t)$ of (9.15) can be recovered by

$$w(y,t) = \int_{\mathbb{R}} (e^\zeta - K)^+ \Gamma^{\sigma^2}(y,t;\zeta,T)d\zeta. \tag{9.35}$$

The same transformations has to be applied to all the operators in (9.22). Suppose also that $\Sigma(e^y,t)$ is bounded with bounded derivatives of any order such that $\left|D_y^\alpha D_t^\beta \Sigma(e^y,t)\right| \le C$, C a constant independent of α and β, to simplify at most the situation (e.g., similar assumptions are used, but with C dependent of α,β, in theorem 25 to get the estimate (A.16)). Then, the action of the operators $[L(t)]^n \circ A_1(t)$, n fixed, to the function $u^{A_0(t)}(x,t)$ correspond to the action of some differential operator $D_y^\alpha D_t^\beta$ $(0\le\alpha\le 2n+2,\ 0\le\beta\le n)$ to the function (9.35), since we can "forget" about the coefficients in $A(t)$ thanks to $\left|D_y^\alpha D_t^\beta \Sigma(e^y,t)\right| \le C$. In other words, we have to study $\left|D_y^\alpha D_t^\beta w(y,t)\right|$. The good news is that, since $\Gamma^{\sigma^2}(y,t;\zeta,T)$ is Gaussian, we have

$$D_y^\alpha D_t^\beta \int_{\mathbb{R}} (e^\zeta - K)^+ \Gamma^{\sigma^2}(y,t;\zeta,T)d\zeta = \int_{\mathbb{R}} (e^\zeta - K)^+ D_y^\alpha D_t^\beta \Gamma^{\sigma^2}(y,t;\zeta,T)d\zeta,$$

as could be checked by direct inspection. So, we face the easier problem to study $D_y^\alpha D_t^\beta \Gamma(y,t;\zeta,T)$. For instance, we have to estimate[8]

$$\left|D_y^\alpha \Gamma^{\sigma^2}(y,t;\zeta,T)\right| = \left|\left(-\frac{|y-\zeta|}{\sigma^2\sqrt{T-t}}\right)^\alpha + \text{ other terms}\right| \Gamma^{\sigma^2}(y,t;\zeta,T), \tag{9.36}$$

where the other terms are polynomials in y (to be precise in $|y-\zeta|/\sqrt{T-t}$) but of order strictly less that α, i.e., they allow a lower order estimate and are less problematic. Notice that $D_y^\alpha \Gamma(y,t;\zeta,T)$ corresponds to $D_x^\alpha u^{A_0(t)}(x,t)$ in (9.22), the same α; e.g., at any $n\ge 0$, $\alpha = 2n+2$ (two derivatives from $A_1(t)$ and $2n$ from $[L(t)]^n$). Now, the function

$$\int_{\mathbb{R}} (e^\zeta - K)^+ \left|\frac{|y-\zeta|}{\sigma^2\sqrt{T-t}}\right|^\alpha \Gamma^{\sigma^2}(y,t;\zeta,T)d\zeta, \tag{9.37}$$

is difficult to be evaluated, despite Γ^{σ^2} is Gaussian. However, using a key trick of the parametrix method (see lemma 19, section 10.2), for any $\varepsilon\in(0,1)$, we have (set $c = |y-\zeta|/\sqrt{T-t}$ and $\kappa = (2\pi\sigma^2(T-t))^{-\frac{1}{2}}$)

$$\begin{aligned}
\left(\frac{|y-\zeta|}{\sigma^2\sqrt{T-t}}\right)^\alpha \Gamma^{\sigma^2}(y,t;\zeta,T) &= \kappa\left(\frac{c}{\sigma^2}\right)^\alpha \exp\left[-\frac{c^2}{2\sigma^2}\right] \\
&\le \kappa \exp\left[-\frac{c^2}{2(\sigma^2+\varepsilon)}\right] \max_c \left(\left(\frac{c}{\sigma^2}\right)^\alpha \exp\left[\frac{\varepsilon c^2}{2(\sigma^2+\varepsilon)\sigma^2}\right]\right) \\
&\le \Gamma^{\sigma^2+\varepsilon}(y,t;\zeta,T)\left(\frac{\alpha\sigma^2(\sigma^2+\varepsilon)}{e\varepsilon}\right)^{\alpha/2}.
\end{aligned} \tag{9.38}$$

[8]E.g., these terms correspond to the terms in $[L(t)]^n \circ A_1(t)$ where we perform all the space derivatives and no time derivatives.

Similar estimates are possible for all the terms in (9.36), but the fastest growing constant as $\alpha \to \infty$ is, anyway, given by $\left(\frac{\alpha\gamma(\gamma+\varepsilon)}{e\varepsilon}\right)^{\alpha/2}$. Substituting (9.38) into (9.37), we get

$$\begin{aligned}\int_{\mathbb{R}} (e^{\zeta} - K)^+ \left| \frac{|y-\zeta|}{\sigma^2\sqrt{T-t}} \right|^{\alpha} \Gamma^{\sigma^2}(y,t;\zeta,T)d\zeta &\leq \left(\frac{\alpha\sigma^2(\sigma^2+\varepsilon)}{e\varepsilon}\right)^{\alpha/2} \int_{\mathbb{R}} (e^{\zeta} - K)^+ \Gamma^{\sigma^2+\varepsilon}(y,t;\zeta,T)d\zeta \\ &\leq \left(\frac{\alpha\sigma^2(\sigma^2+\varepsilon)}{e\varepsilon}\right)^{\alpha/2} Const.e^{Const.y},\end{aligned}$$

where the second inequality follows from

$$\left| \int_{\mathbb{R}} h(\zeta)\Gamma^{\gamma}(y,t;\zeta,T)d\zeta \right| \leq \int_{\mathbb{R}} |h(\zeta)|\, \Gamma^{\gamma}(y,t;\zeta,T)d\zeta = O(e^{Const.y}),$$

for any $\gamma > 0$ and $h(y) = O(e^{const.y})$, i.e., any function of exponential growth (the constants for h and the integral are typically different). This is a consequence of (A.18) (see also Lemma 9.9, [Fri64]) and the explicit formula (9.34). Again we are using a result from the parametrix method. Notice that $e^{Const.y}$ plays the role of the function f independent of n. On the whole we have, for any $\alpha \geq 0$,

$$\left| D_y^{\alpha} \int_{\mathbb{R}} (e^{\zeta} - K)^+ \Gamma^{\sigma^2}(y,t;\zeta,T)d\zeta \right| \leq \left(\frac{\alpha\sigma^2(\sigma^2+\varepsilon)}{e\varepsilon}\right)^{\alpha/2} Const.e^{Const.y}. \tag{9.39}$$

Therefore, $D_y^{\alpha} w(y,t)$ is of exponential growth and, since $\Gamma^{\sigma^2}(y,t;\zeta,T)$ is Gaussian (and we are working under the conditions of the parametrix method, as in the remark after (9.7)), it follows that the key condition $D_y^{\alpha} w(y,t) \in \widehat{D}$ is satisfied. This can be proved for all the other terms.

The point is that, taking $\alpha = 2n+2$, to hope in the convergence of the series (9.22) we need that the constants in the RHS of (9.39), times $(\eta - t)$ and divided by $n!$, tend to 0 as $n \to \infty$. The result of the quite loose analysis presented is, using the Stirling's formula for n large,

$$\begin{aligned}\left(\frac{(2n+2)\,\sigma^2(\sigma^2+\varepsilon)}{e\varepsilon}\right)^{n+1} \frac{(\eta-t)^n}{n!} &\approx \left(\frac{\sigma^2(\sigma^2+\varepsilon)}{\varepsilon}\right)^{n+1} \frac{2^{n+1}(n+1)^{n+1}}{e^{n+1}} \frac{e^n(\eta-t)^n}{(2\pi)^{1/2}n^{n+1/2}} \\ &= \left(\sigma^2(\sigma^2+\varepsilon)\varepsilon^{-1}\right)^{n+1} n^{1/2}(2\pi)^{-1/2}(\eta-t)^n,\end{aligned}$$

which says that the convergence of the perturbation series is not to so unlikely, at least if σ and/or $(\eta - t)$ are small enough. However, to be sure we should take in explicit account all the constants, their powers, and all the other terms. Therefore, a rigorous proof of the convergence, even for this oversimplified example, seems not feasible and the asymptotic justification is the only real result available. As already said, the appealing of the small-time method for the approximation of the option prices is its intuitive nature and extreme flexibility, allowing for unboundedness in the coefficients of both the Markov auxiliary generator and in the perturbation operator, while here we have asked a lot more. In particular, we have used many assumptions from the parametrix methods (e.g., the boundedness of the coefficients), a technique that always allows to find a regular perturbation expansion. We present this extremely powerful method in the next Chapter, explaining why in that case, and at what costs, we can get the convergence of the perturbation series.

Chapter 10

The Parametrix Method

We present the application of the Levi's parametrix method for the construction of the fundamental solution of PDEs of parabolic type. We stress how to use these results for approximation, with particular regard to financial applications. The references are [Fri64], Chapters 1 and 9, and [IKO62], Section 4. The main result, namely the estimation of the truncation error, is a new result in [CoPa06]. We mention that applications of the parametrix to non-parabolic differential equations and integral problems can be found, e.g., in [Gar64].

10.1 Introduction: Assumptions And Ideas

One of the classical use of the parametrix method is to prove the existence of a fundamental solution of a linear non-degenerate parabolic PDE. The goal of this chapter is to explain why this method can be seen as a perturbation technique useful to approximate the solution of the Cauchy problems considered in the first part of the thesis. We prefer to collect the properties of the fundamental solution constructed in Appendix A.2. In fact, we use those results in other chapters, but not here. Furthermore, their derivation is involved (see [Fri64], Chapter 9) and would not improve the presentation.

The salient idea, not surprisingly, is to convert the differential equation into an integral equation, which can be solved by successive approximations. The problem is that any possible decomposition (e.g., (6.2)) of the diffusion generator considered would produce an unbounded perturbation (see Section 5.1), so that the classical result of regular perturbation (e.g., the perturbation of semigroups, Section 6.3) cannot apply. There are two key steps in the method: first it is considered the special decomposition of the generator into (10.5)-(10.6) and then it is guessed that the fundamental solution can be represented in the form (10.8). We notice here two facts:

1) We can define two formal series solutions, both absolutely and uniformly convergent, that we call first and second parametrix expansions. They are equivalent, as follows from the fact that the first and second Kolmogorov equations for the transition density are equivalent (see Section 11.2). The expansions are also called in the literature forward and backward, but unfortunately in this settings the Cauchy problems are typically posed forward in time (while those for the Markov transition densities are naturally posed backward in time, see Section 3.1) and so the forward (or first, for us) expansion corresponds to the backward (also called first) Kolmogorov equation, and vice versa. The name first and second will avoid any confusion.

2) To prove that exists a unique fundamental solution with a convergent series representation does not mean that we know it in closed-form. Actually, the fundamental solution is known only for few PDEs (in particular, the transition densities of SDE-type models reviewed in Chapter 2), but we can clearly truncate any of the two solution series to get an approximation.

Definitions And Assumptions

Consider the differential operator $L(t) = -\frac{\partial}{\partial t} + A(t)$, where

$$A(t) = \sum_{i=1}^{d} \mu_i(x,t)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j}, \tag{10.1}$$

and the real coefficients μ_i and a_{ij} are defined in the strip $H = \mathbb{R}^d \times [0,T]$. Notice from the signs of $L(t)$ that we are working with the time running forward. The field of matrices $a(x,t)$ is always supposed symmetric and positive semi-definite. From the first part of the thesis we know that (10.1) is the generator of the diffusion $S(t)$ solution of the SDE (2.1). We also know (see Chapters 3 and 4) that there is not loss of generality if we consider the operator $L(t)$ instead of the more general parabolic operator $L_c(t) = L(t) - c(x,t)$, which correspond, if the potential term[1] is such that $c(x,t) \geq 0$ in H, to a subordinated diffusion $S^c(t)$ of $S(t)$. In fact, if we suppose also that $c(x,t)$ is bounded in H, then the results presented here extend without further complication (see [Fri64]).

Let $H_0 = \mathbb{R}^d \times (0,T]$. We recall that a fundamental solution of $L(t)$ is a function $\Gamma(x,t;\xi,\tau)$, $x, \xi \in \mathbb{R}^d$, $0 \leq \tau < t \leq T$, which satisfies the following conditions: i) for fixed (ξ,τ), the function $(x,t) \mapsto \Gamma(x,t;\xi,\tau)$ is a classical solution of $L(t)\Gamma = 0$ for $\tau < t$; ii) for any continuous function $g(x)$ bounded by $O\left[\exp\left(h\left|x\right|^2\right)\right]$, for some $h > 0$ (see Appendix A.2),

$$\lim_{t \downarrow \tau} \int_{\mathbb{R}^d} g(\xi)\Gamma(x,t;\xi,\tau)d\xi = g(x). \tag{10.2}$$

From (10.2) we see that the fundamental solution $\Gamma(x,t;\xi,\tau)$ satisfies the initial condition $\Gamma(x,t;\xi,\tau) \to \delta(x-\xi)$ as $t \downarrow \tau$, so that, as noticed in Section 3.2, if there exists a unique fundamental solution, after the change of variables $(t,\tau) \mapsto (T-t, T-\tau)$, Γ must be the transition density of the diffusion $S(t)$ (supposed existent under the hypotheses 1. − 6. in Chapters 2,3).

We make the following assumptions:

A1 The diffusion matrix $a(x,t)$ is uniformly parabolic in H, with constants $0 < m \leq M < \infty$ (see (A.9)).

A2 The coefficients of $L(t)$ are bounded continuous functions in H and $a_{ij}(x,t)$ are in the Holder space $C^{1,1/2}(H)$, i.e., there exists a constant $\alpha > 0$ such that

$$|a_{ij}(x,t) - a_{ij}(x_0,t_0)| \leq \alpha\left(|x - x_0| + |t - t_0|^{1/2}\right), \tag{10.3}$$

for all $(x,t), (x_0,t_0) \in H$ and $i,j = 1,\ldots,d$.

We stress that the main consequence of **A1** is that the matrix $a(x,t)$ is invertible in H and for the inverse $\left(a^{ij}(x,t)\right) = \left(a(x,t)\right)^{-1}$ holds (A.10), i.e.[2],

$$\frac{|\xi|^2}{M} \leq \sum_{i,j=1}^{d} a^{ij}(x,t)\xi_i\xi_j \leq \frac{|\xi|^2}{m}, \tag{10.4}$$

for all $\xi \in \mathbb{R}^d$ and $(x,t) \in H$. Assumption **A2**, in particular (10.3), is crucial in the proof of the uniform convergence of the perturbation series.

[1] For a description of parabolic operators and PDEs we refer to Appendix A.2.

[2] We use the notation $|x|$ for the Euclidean norm of the vector $x \in \mathbb{R}^d$.

10.2 The First Parametrix Expansion

The parametrix method is characterized by two specific ideas. The first is to decompose the operator (10.1) into $A(t) = A_0 + A_1(t)$, where, for a fixed $(\xi,\tau) \in H$, A_0 is given by the constant coefficients operator

$$A_0 = A_0^{(\xi,\tau)} := \frac{1}{2}\sum_{i,j} a_{ij}(\xi,\tau)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{10.5}$$

Since (ξ,τ) is kept fixed, (10.5) sometimes is called frozen operator at (ξ,τ). Clearly,

$$A_1(t) = A_1^{(\xi,\tau)}(t) = \sum_{i=1}^{d} \mu_i(x,t)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d}\left[a_{ij}(x,t) - a_{ij}(\xi,\tau)\right]\frac{\partial^2}{\partial x_i \partial x_j}. \tag{10.6}$$

We introduce also the notation $L_0^{(\xi,\tau)} = -\partial_t + A_0^{(\xi,\tau)}$, so that $L(t) = L_0^{(\xi,\tau)} + A_1^{(\xi,\tau)}(t)$. The reason to choose such a particular auxiliary model is that the fundamental solution $\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$ of (10.5), for $(\xi,\tau) \in H$ fixed, $\tau < t$, is known in closed-form and is given by (see Appendix A.2)

$$\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau) = (2\pi t)^{-\frac{d}{2}}\left(\det a(\xi,\tau)\right)^{-\frac{1}{2}}\exp\left(-\frac{\theta^{(\xi,\tau)}(x,\xi)}{2(t-\tau)}\right), \tag{10.7}$$

where $\theta^{(\xi,\tau)}(x,\xi) := \sum_{ij} a^{ij}(\xi,\tau)(x_i - \xi_i)(x_j - \xi_j)$. This claim could be checked by direct inspection substituting (10.7) into the equation $L_0^{(\xi,\tau)}u(x,t) = 0$. The function $\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$ is called the parametrix (or Levi's function) of $L(t)$. Parametrix means reference or gauge function for the problem $L(t)\Gamma = 0$ and, actually, it is the simplest possible approximation of Γ. Clearly, $\Gamma_0^{(\xi,\tau)}$ is a multivariate Gaussian probability density function, with the additional property $\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau) = \Gamma_0^{(\xi,\tau)}(x-\xi,t-\tau)$.

Now we aim to convert the differential problem into an integral problem. The second step of the parametrix method is to try to find a fundamental solution Γ of the full problem in the form

$$\Gamma(x,t;\xi,\tau) = \Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau) + \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(x,t;y,s)\Phi(y,s;\xi,\tau)dyds, \tag{10.8}$$

where Φ is to be determined by means of the condition $L(t)\Gamma = 0$. In fact, if we apply $L(t)$ to both sides of (10.8), we have

$$0 = L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau) + L(t)\int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(x,t;y,s)\Phi(y,s;\xi,\tau)dyds. \tag{10.9}$$

Note that, if $\Phi(y,s;\xi,\tau)$ is continuous and bounded as required, the double integral in RHS of (10.8) is an improper integral, called volume potential of Φ, since $\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$ has a singularity at $(x,t) = (\xi,\tau)$. However, this singularity is absolutely and separately integrable (see below or Appendix A.2). The bulk of the work of the parametrix method is to prove the following theorem (see [Fri64], Theorems 1.1-1.6 and Section 6) about a general volume potential

$$V(x,t) = \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(x,t;y,s)f(y,s)dyds.$$

Theorem 18 *Let $f(x,t)$ be continuous in (x,t) and Holder continuous in $x \in \mathbb{R}^d$, uniformly in t. Then $L(t)V(x,t)$ is well defined and continuous in (x,t) and we have*

$$L(t)V(x,t) = \int_\tau^t \int_{\mathbb{R}^d} \left(L(t)\Gamma_0^{(y,s)}\right)(x,t;y,s)f(y,s)dyds - f(x,t). \tag{10.10}$$

We remark that in the proof of this theorem one has to show that the improper integral in the RHS of (10.10) converges. This is true since it turns out that the action of $L(t)$ on $\Gamma_0^{(y,s)}$ smooths out the singularity of $\Gamma_0^{(y,s)}$ and so also the singularity of $L(t)\Gamma_0^{(y,s)}$ is integrable. We return on this important fact below (here we refer to formula (4.3), Chapter 1, [Fri64]).

The application of the theorem to $f(y,s) = \Phi(y,s;\xi,\tau)$, supposing for the moment that Φ is jointly continuous in (x,t) and Holder continuous in $x \in \mathbb{R}^d$, uniformly in t, implies that (10.9) is equivalent to

$$\Phi(x,t;\xi,\tau) = L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau) + \int_\tau^t \int_{\mathbb{R}^d} \left(L(t)\Gamma_0^{(y,s)}(x,t;y,s) \right) \Phi(y,s;\xi,\tau) dyds, \qquad (10.11)$$

i.e., an integral equation of Volterra-type for $\Phi(x,t;\xi,\tau)$. It is natural to apply the method of successive approximations. Iterating (10.11) we find the formal series solution

$$\Phi(x,t;\xi,\tau) = \sum_{n=1}^{\infty} \Phi_n(x,t;\xi,\tau), \qquad (10.12)$$

where $\Phi_1(x,t;\xi,\tau) = L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$ and, for $n \geq 1$,

$$\Phi_{n+1}(x,t;\xi,\tau) = \int_\tau^t \int_{\mathbb{R}^d} \left(L(t)\Gamma_0^{(y,s)}(x,t;y,s) \right) \Phi_n(y,s;\xi,\tau) dyds. \qquad (10.13)$$

One has to prove: i) That the improper integrals in (10.13), $n \geq 1$, converge. E.g., for $n = 2$, the integrand in (10.13) is the product of two singular functions (for $(x,t) = (y,s) = (\xi,\tau)$). ii) That the series solution (10.12) converges uniformly, so that it is a genuine solution of (10.11) (see Section 6.3). iii) That Φ, as given in (10.12), is actually jointly continuous in (x,t) and Holder continuous in $x \in \mathbb{R}^d$, uniformly in t.

The classical approach of the proof of i)-iii) above (see [Fri64] and [IKO62]) is to use the fact that $\Gamma_0^{(\xi,\tau)}$ is a multivariate Gaussian density function and therefore it has the following simple estimate:

$$\left| \Gamma_0^{(\xi,\tau)} \right| \leq C_1 (t-\tau)^{-\frac{d}{2}} \exp\left[-\frac{c_1 \left| x - \xi \right|^2}{(t-\tau)} \right], \qquad (10.14)$$

C_1, c_1 positive constant. Then, by means of the elementary inequality

$$z^c \exp(-c_1 z) \leq C \exp(-c_2 z), \qquad (10.15)$$

true for any $z \in [0, +\infty), c \geq 0, 0 < c_2 < c_1$, C a positive constant dependent on c, one can study the terms Φ_n in (10.12) bounding any derivative of $\Gamma_0^{(\xi,\tau)}$ (which, notice, is clearly infinitely derivable in each variable x_i, $i = 1, \ldots, d$, and t) by

$$\left| D_x^{n_1} D_t^{n_2} \Gamma_0^{(\xi,\tau)} \right| \leq const. (t-\tau)^{-\frac{d}{2} - |n_1| - 2n_2} \exp\left[-\frac{const. \left| x - \xi \right|^2}{(t-\tau)} \right]. \qquad (10.16)$$

These estimates (10.16) of the parametrix $\Gamma_0^{(\xi,\tau)}$ are powerful and allow to conclude the absolute and uniform convergence of (10.12), the other properties required and also (A.16)-(A.17) (from (10.8); see [Fri64] or [IKO62]). However, (10.16) is too rough (i.e., the constants are not explicit) to allow to study the remainder of (10.12).

Estimation of The First Term of The Series Solution

The idea in [CoPa06] to solve the problem explained at the end of the previous section is to still approximate $\Gamma_0^{(\xi,\tau)}$ by a simpler Gaussian density, but allowing to keep explicit the constants in (10.14). Let $\Gamma_0^\gamma(x,t;\xi,\tau)$, for $(\xi,\tau)\in H$ fixed, $\tau<t$, be the fundamental solution of the heat operator $\frac{\gamma}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i\partial x_j}-\partial_t$, $\gamma>0$. $\Gamma_0^\gamma(x,t;\xi,\tau)$ is given by (10.7) with $a(x,t)=\gamma I_{d\times d}$, $I_{d\times d}$ the identity matrix. From (10.4) we have the inequality

$$\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\le(2\pi m(t-\tau))^{-\frac{d}{2}}\exp\left[-\frac{|x-\xi|^2}{2M(t-\tau)}\right]=\left(\frac{M}{m}\right)^{\frac{d}{2}}\Gamma_0^M(x,t;\xi,\tau). \tag{10.17}$$

To estimate $\Phi_1(x,t;\xi,\tau)=L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$, we have to compute the derivatives of $\Gamma_0^{(\xi,\tau)}$. Thanks to the explicit formulas

$$\partial_{x_i}\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)=-\frac{\sum_j a^{ij}(\xi,\tau)(x_j-\xi_j)}{(t-\tau)}\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau),$$

$$\partial_{x_i}\partial_{x_j}\Gamma_0^{(\xi,\tau)}=\frac{-a^{ij}(\xi,\tau)(t-\tau)+\sum_h a^{ih}(\xi,\tau)(x_h-\xi_h)\sum_k a^{jk}(\xi,\tau)(x_k-\xi_k)}{(t-\tau)^2}\Gamma_0^{(\xi,\tau)},$$

applying (10.17) and again (10.4), we have

$$\left|\partial_{x_i}\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\right|\le\frac{|x-\xi|}{m(t-\tau)}\left(\frac{M}{m}\right)^{\frac{d}{2}}\Gamma_0^M(x,t;\xi,\tau), \tag{10.18}$$

$$\left|\partial_{x_i}\partial_{x_j}\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\right|\le\frac{1}{(t-\tau)}\left(\frac{1}{m}+\frac{|x-\xi|^2}{m^2(t-\tau)}\right)\left(\frac{M}{m}\right)^{\frac{d}{2}}\Gamma_0^M(x,t;\xi,\tau). \tag{10.19}$$

It would be ideal if we could get rid of the terms $|x-\xi|$ and $|x-\xi|^2$ in (10.18)-(10.19). It holds the following lemma

Lemma 19 *For every $\gamma>0$, $\varepsilon\in(0,1)$ and $n=0,1,2,\ldots$,*

$$\left(\frac{|x-\xi|}{(t-\tau)^{1/2}}\right)^n\Gamma_0^\gamma(x,t;\xi,\tau)\le\left(\frac{n}{\varepsilon}\right)^{\frac{n}{2}}(\gamma+\varepsilon)^n\left(\frac{\gamma+\varepsilon}{\gamma}\right)^{\frac{d}{2}}\Gamma_0^{\gamma+\varepsilon}(x,t;\xi,\tau), \tag{10.20}$$

for all $(x,t),(\xi,\tau)\in H$, $\tau<t$.

To see that this Lemma is true, notice that $\Gamma_0^\gamma(x,t;\xi,\tau)=\Gamma_0^\gamma(x-\xi,t-\tau)$, so that we can assume, w.l.o.g., that $(\xi,\tau)=(0,0)$. Then observe that $|x|/\sqrt{t}=:y>0$. We can write explicitly

$$\left(|x|/\sqrt{t}\right)^n\Gamma_0^\gamma(x,t;0,0)=y^n(2\pi\gamma t)^{-d/2}\exp\left[-y^2/(2\gamma)\right]$$

Now let $-\frac{1}{\gamma}=-\frac{1}{\gamma+\varepsilon}+\left(\frac{1}{\gamma+\varepsilon}-\frac{1}{\gamma}\right)$ and $\overline{y}:=\max\left(y^n\exp\left[\frac{y^2}{2}\left(\frac{1}{\gamma+\varepsilon}-\frac{1}{\gamma}\right)\right]\right)$. We have

$$\left(|x|/\sqrt{t}\right)^n\Gamma_0^\gamma(x,t;0,0)\le\overline{y}\left(\frac{\gamma+\varepsilon}{\gamma}\right)^{\frac{d}{2}}\Gamma_0^{\gamma+\varepsilon}(x,t;0,0).$$

(10.20) follows since $\overline{y}=(n\gamma(\gamma+\varepsilon)/(e\varepsilon))^{n/2}\le\left(\frac{n}{\varepsilon}\right)^n(\gamma+\varepsilon)^n$. We remark that (10.20) is just (10.15) with explicit constants. Moreover, the estimate in (10.20) is the sharpest possible if one wants an estimate such as (10.16) (i.e., with the exponential tails of a Gaussian) but with explicit constants.

We are ready for the estimation of $\Phi_1(x,t;\xi,\tau) = L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$. By the definition of fundamental solution, $L_0^{(\xi,\tau)}\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau) = 0$, so

$$\begin{aligned}\left|L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\right| &= \left|\left(L(t) - L_0^{(\xi,\tau)}\right)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\right| \\ &= \left|A_1^{(\xi,\tau)}(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\right|.\end{aligned}$$

By (10.6) we have

$$\left|L(t)\Gamma_0^{(\xi,\tau)}\right| \leq \sum_{i=1}^{d}|\mu_i(x,t)|\left|\partial_{x_i}\Gamma_0^{(\xi,\tau)}\right| + \frac{1}{2}\sum_{i,j=1}^{d}|a_{ij}(x,t)-a_{ij}(\xi,\tau)|\left|\partial_{x_i}\partial_{x_j}\Gamma_0^{(\xi,\tau)}\right|,$$

applying (10.18), (10.19), (10.20) and (10.3), we get

$$\left|L(t)\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)\right| \leq \frac{\eta_\varepsilon}{(t-\tau)^{1/2}}\Gamma_0^{M+\varepsilon}(x,t;\xi,\tau), \tag{10.21}$$

where $\varepsilon \in (0,1)$ and η_ε is a positive explicit constant given by

$$\eta_\varepsilon := \alpha d^2\left(\frac{2}{\varepsilon}\right)^{\frac{3}{2}}\left(\frac{M+\varepsilon}{m}\right)^{\frac{d}{2}+2}\left(M+\varepsilon+\sqrt{\varepsilon/2}\right) + \beta\frac{d}{2\sqrt{\varepsilon}}\left(\frac{M+\varepsilon}{m}\right)^{\frac{d}{2}+1}, \tag{10.22}$$

with $\beta := \sup_{i,(x,t)}|\mu_i(x,t)| < \infty$. Note that (10.21) confirms that the singularity of $L(t)\Gamma_0^{(\xi,\tau)}$ is integrable.

Convergence of The Series

To estimate the higher order terms in (10.12), suppose that for every $\varepsilon \in (0,1)$ and $n \geq 1$,

$$|\Phi_n(x,t;\xi,\tau)| \leq \frac{[\Gamma_E(1/2)]^n}{\Gamma_E(n/2)}\frac{\eta_\varepsilon^n}{(t-\tau)^{1-n/2}}\Gamma_0^{M+\varepsilon}(x,t;\xi,\tau), \tag{10.23}$$

for all $(x,t),(\xi,\tau) \in H$, $\tau < t$, where Γ_E is the gamma function.

Φ_1 satisfies (10.23) by (10.21). Assume (10.23) holds for $n > 1$, then, by (10.21) and the inductive hypothesis

$$\begin{aligned}|\Phi_n(x,t;\xi,\tau)| &\leq \int_\tau^t\int_{\mathbb{R}^d}|\Phi_1(x,t;y,s)|\,|\Phi_n(y,s;\xi,\tau)|\,dyds \\ &\leq \frac{[\Gamma_E(1/2)]^n}{\Gamma_E(n/2)}\eta_\varepsilon^{n+1}\Gamma_0^{M+\varepsilon}(x,t;\xi,\tau)\int_\tau^t(t-s)^{-\frac{1}{2}}(s-\tau)^{\frac{n}{2}-1},\end{aligned}$$

since $\Gamma_0^{M+\varepsilon}$ is a Gaussian density and so, by convolution[3],

$$\Gamma_0^{M+\varepsilon}(x,t;\xi,\tau) = \int_{\mathbb{R}^d}\Gamma_0^{M+\varepsilon}(x,t;y,s)\Gamma_0^{M+\varepsilon}(y,s;\xi,\tau)dy. \tag{10.24}$$

By the change of variable $z = (s-\tau)/(t-\tau)$, since $\int_0^1(1-z)^{\alpha-1}z^{\beta-1}dz = \Gamma_E(\alpha)\Gamma_E(\beta)/\Gamma_E(\alpha+\beta)$, (10.23) is proved.

Because[4] $\frac{[\Gamma_E(1/2)]^n}{\Gamma_E(n/2)} = \frac{(2\pi)^{\frac{n-1}{2}}}{(n-2)!!}$, estimate (10.23) implies that the series (10.12) converges

[3]Note that this property extends to any fundamental solution (see (A.15)), which is anything but the Chapman-Kolmogorov equation for the transition density.

[4]The double factorial is recursively defined as $k!! := 1$ if $k = -1,0,1$ and $k!! := k(k-2)!!$ if $k \geq 2$.

absolutely and uniformly in the strip $\mathbb{R}^d \times [\tau, T]$, for any $\tau < T$, proving that (10.12) is a solution of (10.11) and that the function $\Phi(x, t; \xi, \tau)$ is jointly continuous. Furthermore, from (10.23) we can see that the singularity of $\Phi_n(x, t; \xi, \tau)$ is not only integrable, but also weaker that the singularity of $\Phi_{n-1}(x, t; \xi, \tau)$, for all $n \geq 1$. The proof that Φ, as given in (10.12), is Holder continuous in $x \in \mathbb{R}^d$, uniformly in t, can be found in [Fri64], Theorem 1.7.

Some remarks on The Convergence of The Perturbation Series

We remark now why the parametrix methods allows to prove the uniform convergence of the perturbation series in such generality (recall the difficult proof for an autonomous scalar problem in Section 7.3 and the huge problems in Section 9.5 for more or less the same problem), highlighting the analogies and differences with the other perturbation techniques described in Chapters 7-9.

The key feature of any particular perturbation method is the decomposition of the full problem used (i.e., the rearrangement of the PDE problem): this implies the structure of the iterative procedure that one can exploit to get the perturbation series. The parametrix methods has not flexibility in this choice, but this is not a con, since the chosen decomposition (10.5) is about the simplest model available, probably the most intuitive choice. The regular approach is quite similar, with much more flexibility, but one has to introduce artificially the perturbative parameter ϵ to control and bound (ϵ is small), when possible, the action of the unbounded perturbation (7.3). No decomposition is considered in the small-time expansions (but see (9.16)). As a result, the iteration in the latter case is given by (8.8)-(8.20), that is the most natural way to get a series solution of the PDE, but this complicates the study of the possible convergence because we use in the iterations the operators (the generator and the semigroup) of the full, not solvable problem. In fact, we are not simplifying the problems unless the time-span is shrunk to 0, the asymptotic condition, which allows (thanks to the strong continuity) to get rid of the semigroup of the full problem that appears in the remainder of the small-time expansion. On the other hand, the other two methods use a similar strategy, despite the clear difference in the way one defines the iteration: the simplest way (7.8) in the regular case, the clever choice (10.11), exploiting the special properties of the operators involved. Then the regular perturbation ends up with the recursion (7.9), that is quite similar to (10.13), being a time integral of an expectation taken under the auxiliary model, that is the main difference with the small-time expansions. The advantage, and the limitation, of the parametrix method is that the auxiliary, explicit transition density is Gaussian. Moreover, the action of $L(t)$ on $\Gamma_0^{(y,s)}$ in (10.13) smooths out the singularity of the transition density. In the same way, the integrand of (10.13) (some volume potentials), though more complicated, is more regular than that in (7.9). This facilitates the convergence of the series solution.

Let us analyze the peculiarities of the parametrix expansion and the actual two reasons of the convergence. While in the regular perturbation case one has to carry out a specific analysis on the operators and then, hopefully, the integrals do not mess up the result, in the parametrix technique it the presence of the integrals what ensures the convergence. The expectation is easily computed thanks to the convolution property (10.24) enjoyed by any parametrix $\Gamma_0^{M+\varepsilon}$, while in (7.9) one has to explicitly evaluate it. But most of all, the crucial fact is the presence of the beta integral $\int_0^1 (1-z)^{\alpha-1} z^{\beta-1} dz$ that implies the term $\Gamma_E(n/2)$ in the denominator of (10.23); in fact, the method iterates the full operator $L(t)$ just to get that fundamental time integral. Note that, unfortunately, despite also the small-time perturbations in Chapter 9 iterate the parabolic operator, one has to supply a case by case analysis since neither the expectation in the iteration can be computed (as in (7.9)), given the use of the operator of the full problem, nor there is a time-integral to enhance the possible convergence.

Approximation of The Fundamental Solution

From (10.8) and (10.12) we have the following absolutely and uniformly convergent expansion of the fundamental solution of $L(t)$

$$\Gamma(x,t;\xi,\tau) = \sum_{n=0}^{\infty} \Gamma_n(x,t;\xi,\tau), \tag{10.25}$$

where $\Gamma_0(x,t;\xi,\tau) = \Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$ and, for $n \geq 1$,

$$\Gamma_n(x,t;\xi,\tau) = \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(x,t;y,s)\Phi_n(y,s;\xi,\tau)dyds. \tag{10.26}$$

The remainder of the series (10.25) truncated at $N \geq 0$ is given by

$$\begin{aligned}|R_{N+1}(x,t;\xi,\tau)| &= \left| \sum_{n=N+1}^{\infty} \Gamma_n(x,t;\xi,\tau) \right| \\ &\leq \sum_{n=N+1}^{\infty} \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(x,t;y,s) \left|\Phi_n(y,s;\xi,\tau)\right| dyds.\end{aligned}$$

By (10.17), (10.23) and (10.24) we have

$$\begin{aligned}|R_{N+1}(x,t;\xi,\tau)| &\leq \left(\frac{M+\varepsilon}{m}\right)^{\frac{d}{2}} \Gamma_0^{M+\varepsilon}(x,t;\xi,\tau) \sum_{n=N+1}^{\infty} \frac{(2\pi)^{\frac{n-1}{2}}}{(n-2)!!}\eta_\varepsilon^n \frac{(t-\tau)^{\frac{n}{2}}}{n/2} \\ &= \left(\frac{M+\varepsilon}{m}\right)^{\frac{d}{2}} \Gamma_0^{M+\varepsilon}(x,t;\xi,\tau)\sqrt{\frac{2}{\pi}} \sum_{n=N+1}^{\infty} \frac{\left(\eta_\varepsilon \sqrt{2\pi(t-\tau)}\right)^n}{n!!}.\end{aligned}$$

Call $\eta := \left(\eta_\varepsilon \sqrt{2\pi(t-\tau)}\right)$. Suppose $N+1$ is even (otherwise take the first term out of the series and $N+2$ is even). To deal with the double factorial, for which $(2n)!! = 2^n n!$, it is convenient to perform the following decomposition

$$\begin{aligned}\sum_{n=N+1}^{\infty} \frac{\eta^n}{n!!} &= \sum_{n=\frac{N+1}{2}}^{\infty} \frac{\eta^{2n}}{(2n)!!} + \sum_{n=\frac{N+1}{2}+1}^{\infty} \frac{\eta^{2n-1}}{(2n-1)!!} \\ &\leq \sum_{n=\frac{N+1}{2}}^{\infty} \frac{\eta^{2n}}{(2n)!!} + \sum_{n=\frac{N+1}{2}+1}^{\infty} \frac{\eta^{2n-1}}{(2n-2)!!} \\ &= \sum_{n=\frac{N+1}{2}}^{\infty} \frac{\left(\eta^2/2\right)^n}{n!} + \sum_{n=\frac{N+1}{2}}^{\infty} \frac{\eta^{2n+1}}{2^n n!}.\end{aligned}$$

Since $\sum_{n=N+1}^{\infty} \frac{\eta^n}{n!!} \leq e^\eta \eta^{N+1}/(N+1)!$, we have (for $N+1$ even)

$$|R_{N+1}(x,t;\xi,\tau)| \leq \left(\frac{M+\varepsilon}{m}\right)^{\frac{d}{2}} \Gamma_0^{M+\varepsilon}(x,t;\xi,\tau)\sqrt{\frac{2}{\pi}} e^{\frac{\eta^2}{2}} \left(\frac{\eta^2}{2}\right)^{\frac{N+1}{2}} \left(\frac{N+1}{2}!\right)^{-1}, \tag{10.27}$$

which is actually very complicated, but is explicit. We notice that, when $\eta \ll 1$ (e.g., if $(t-\tau)$ is very small), the rate of convergence of the series (10.12) and (10.25) is very fast.

10.3 The Second Parametrix Expansion

We develop now the dual versions of the series (10.12) and (10.25). We need the following additional assumption

A3 The functions $\partial_{x_i} a_{ij}(x,t)$, $\partial_{x_i}\partial_{x_j} a_{ij}(x,t)$ and $\partial_{x_i}\mu_i(x,t)$ are well defined (i.e., $a(x,t)$ is derivable as required), continuous and bounded functions for all $i,j = 1,\ldots,d$.

Under **A3** the formal adjoint operators of $A(t)$ and $L(t)$ are well defined (see Appendix A.2 and Chapters 3 and 4) and given by

$$A^*(t)u(x,t) = -\sum_i \frac{\partial}{\partial x_i}(\mu_i(x,t)u(x,t)) + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}(x,t)u(x,t))$$

and $L^*(t) = \partial_t + A^*(t)$. We remark that the formal adjoint operator $L^*(t)$ is obtained from $L(t)$ by formally inverting the time direction (we can see this from the signs of $L(t)$ and $L^*(t)$) and extending the action of the partial differential operators to the coefficients. A fundamental solution of the adjoint equation $L^*(t)u(x,t) = 0$ is given by a function $\Gamma^*(x,t;\xi,\tau)$, $x,\xi \in \mathbb{R}^d$, $0 \le t < \tau \le T$, which solves $L^*(t)u(x,t) = 0$ (notice in (x,t), but for $t < \tau$, i.e., in the backward variables) and (10.2), but now as $t \uparrow \tau$. Given the symmetry between $x,\xi \in \mathbb{R}^d$, it is convenient to retain $\tau < t$ and write $\Gamma^*(\xi,\tau;x,t)$, so that $(\xi,\tau) \mapsto \Gamma^*(\xi,\tau;x,t)$ solves $L^*(\tau)u = 0$. All this complication are due to the fact that in PDE theory (see [Fri64] or [IKO62]), when a parabolic operator acts on a function of four variables, as $\Gamma^*(\xi,\tau;x,t)$, it is always supposed to act on the first two variables (the other variables are kept fixed and treated as parameters).

Now the decomposition (10.5)-(10.6) takes the form $A^*(\tau) = A_0^{*(x,t)} + A_1^{*(x,t)}(\tau)$, where the frozen operator $A_0^{*(x,t)}$ at (x,t), $\tau < t$, is given by

$$A_0^{*(x,t)}u(\xi,\tau) = \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial \xi_i \partial \xi_j}\left(a_{ij}(x,t)u(\xi,\tau)\right).$$

Let $L_0^{*(x,t)}(\tau) = A_0^{*(x,t)} + \partial_\tau$. Notice that the function $\Gamma_0^{*(x,t)}(\xi,\tau;x,t)$, given by (10.7) with the matrix $a(x,t)$ in place of $a(\xi,\tau)$, is a fundamental solution of $L_0^{*(x,t)}(\tau)u(\xi,\tau) = 0$. Finally, let $\Gamma_0^{*\gamma}(\xi,\tau;x,t)$ be the fundamental solution of the adjoint heat operator.

This time the method guesses that the fundamental solution of the adjoint problem $L_0^{*(x,t)}(\tau)u(\xi,\tau) = 0$ is given by

$$\Gamma^*(\xi,\tau;x,t) = \Gamma_0^{*(x,t)}(\xi,\tau;x,t) + \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{*(y,s)}(\xi,\tau;y,s)\Phi^*(y,s;x,t)dyds, \tag{10.28}$$

where Φ^*, by the condition $L^*(\tau)\Gamma^* = 0$ and the analogue of (10.10), satisfies

$$\Phi^*(\xi,\tau;x,t) = L^*(\tau)\Gamma_0^{(x,t)}(x,t;\xi,\tau) + \int_\tau^t \int_{\mathbb{R}^d} \left(L^*(\tau)\Gamma_0^{*(y,s)}(\xi,\tau;y,s)\right)\Phi^*(y,s;x,t)dyds,$$

Again, the integral equation of Volterra-type for $\Phi^*(x,t;\xi,\tau)$ can be solved by the method of successive approximations. By iteration we find the formal series solution

$$\Phi^*(\xi,\tau;x,t) = \sum_{n=1}^\infty \Phi_n^*(\xi,\tau;x,t), \tag{10.29}$$

where $\Phi_1^*(\xi,\tau;x,t) = L^*(\tau)\Gamma_0^{*(x,t)}(\xi,\tau;x,t)$ and, for $n \ge 1$,

$$\Phi_{n+1}^*(\xi,\tau;x,t)=\int_\tau^t\int_{\mathbb{R}^d}\left(L^*(\tau)\Gamma_0^{*(y,s)}(\xi,\tau;y,s)\right)\Phi_n^*(y,s;x,t)dyds. \tag{10.30}$$

Given the perfect symmetry of $\Gamma_0^{*(x,t)}$ in $x,\xi\in\mathbb{R}^d$ (e.g., $\Gamma_0^{*(x,t)}(\xi,\tau;x,t)=\Gamma_0^{*(x,t)}(\xi-x,t-\tau)$) and noticing that $\Gamma_0^{*\gamma}(x,t)=\Gamma_0^{\gamma}(x,t)$, for any $\gamma>0$ and $(x,t)\in H_0$, it not difficult to see that the estimates (10.18)-(10.20) hold (with the obvious modifications, e.g., the derivatives are taken w.r.t. ξ). In particular one can prove

$$|\Phi_n^*(\xi,\tau;x,t)|\le\frac{[\Gamma_E(1/2)]^n}{\Gamma_E(n/2)}\frac{(\eta_\varepsilon^*)^n}{(t-\tau)^{1-n/2}}\Gamma_0^{M+\varepsilon}(\xi,\tau;x,t), \tag{10.31}$$

for all $n\ge1$, where η_ε^* is identical to η_ε in (10.22) but with

$$\beta^*:=\sup_{i,j,(x,t)}\left|\partial_{x_i}a_{ij}(x,t)+\partial_{x_i}\partial_{x_j}a_{ij}(x,t)+\partial_{x_i}\mu_i(x,t)\right|<\infty$$

in place of β. From (10.31) follows the absolute and uniform convergence of (10.29). By (10.28) we have also the convergent expansion for the adjoint fundamental solution

$$\Gamma^*(\xi,\tau;x,t)=\sum_{n=0}^{\infty}\Gamma_n^*(\xi,\tau;x,t), \tag{10.32}$$

where $\Gamma_0^*(\xi,\tau;x,t)=\Gamma_0^{*(x,t)}(\xi,\tau;x,t)$ and, for $n\ge1$,

$$\Gamma_n^*(\xi,\tau;x,t)=\int_\tau^t\int_{\mathbb{R}^d}\Gamma_0^{*(y,s)}(\xi,\tau;y,s)\Phi_n^*(y,s;x,t)dyds. \tag{10.33}$$

In turn, the truncation error is estimated by[5]

$$\left|R_{N+1}^*(\xi,\tau;x,t)\right|\le\left(\frac{M+\varepsilon}{m}\right)^{\frac{d}{2}}\Gamma_0^{M+\varepsilon}(\xi,\tau;x,t)\sqrt{\frac{2}{\pi}}e^{\frac{\eta^2}{2}}\left(\frac{(\eta^*)^2}{2}\right)^{\frac{N+1}{2}}\left(\frac{N+1}{2}!\right)^{-1},$$

where $\eta^*:=\left(\eta_\varepsilon^*\sqrt{2\pi(t-\tau)}\right)$.

An Equivalent Representation of The Second Expansion

In practice it could be sometimes cumbersome to apply the operator $L^*(\tau)$, as prescribed by (10.30). First of all notice that $\Gamma_0^{*(x,t)}(\xi,\tau;x,t)=\Gamma_0^{(x,t)}(x,t;\xi,\tau)$ (where (x,t) is fixed): this is obvious from (10.17). An important classical result of the parametrix method is that this property extends to the fundamental solutions of the full problem $L(t)u=0$ and of its adjoint problem $L^*(\tau)u=0$ (see Theorem 1.15, [Fri64]), i.e., for $\tau<t$,

$$\Gamma^*(\xi,\tau;x,t)=\Gamma(x,t;\xi,\tau).$$

As a consequence, the two expansions (10.25) and (10.32) are equivalent, in the sense that they have the same sum. However, the two expansions (10.25) and (10.32) will generally differ in the leading terms, so that the truncation at the same small N will give different approximations). In particular, (10.32) appears more complex (since we need to perform more derivatives), but we could provide an easier equivalent representation.

As noticed, at $n=0$, $\Gamma_0^{*(x,t)}=\Gamma_0^{(x,t)}$. Using this fact

[5] Note that $R_N^*(\xi,\tau;x,t)=R_N^*(x,t;\xi,\tau)$.

$$\Gamma_1^*(\xi,\tau;x,t) = \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(y,s;\xi,\tau) L^*(s) \Gamma_0^{*(x,t)}(y,s;x,t) dyds. \tag{10.34}$$

Since the singularities of $\Gamma_0^{(y,s)}$ and $L^*(s)\Gamma_0^{*(x,t)}$ (at $(x,t) = (y,s) = (\xi,\tau)$) are absolutely and separately integrable, the order of integration in (10.34) is immaterial. Now, decompose $L^*(s) = \partial_s + A^*(s)$ into its elementary differential operators. We have $\partial_s \Gamma_0^{*(x,t)}(y,s;x,t) = \partial_s \Gamma_0^{(x,t)}(x,t;y,s)$ and, integrating by parts formally,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \int_\tau^t \Gamma_0^{(y,s)}(y,s;\xi,\tau) \partial_s \Gamma_0^{(x,t)}(x,t;y,s) dsdy \\
&= \left[\int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(y,s;\xi,\tau) \Gamma_0^{(x,t)}(x,t;y,s) \right]_{s=\tau}^t - \int_{\mathbb{R}^d} \int_\tau^t \partial_s \Gamma_0^{(y,s)}(y,s;\xi,\tau) \Gamma_0^{(x,t)}(x,t;y,s) dsdy \\
&= - \int_{\mathbb{R}^d} \int_\tau^t \partial_s \Gamma_0^{(y,s)}(y,s;\xi,\tau) \Gamma_0^{(x,t)}(x,t;y,s) dsdy.
\end{aligned}$$

In fact, since $\Gamma_0^{(y,t)}(x,t;y,t) = \delta(y-x)$ and $\Gamma_0^{(y,\tau)}(y,\tau;\xi,\tau) = \delta(y-\xi)$, we have that

$$\left[\int_{\mathbb{R}^d} \Gamma_0^{(y,s)}(y,s;\xi,\tau) \Gamma_0^{(x,t)}(x,t;y,s) \right]_{s=\tau}^t = 0.$$

The same conclusion holds for all the space partial-derivatives as well, finding

$$\Gamma_1^*(\xi,\tau;x,t) = \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(x,t)}(x,t;y,s) L(s) \Gamma_0^{(y,s)}(y,s;\xi,\tau) dyds.$$

One can prove by induction (see [CoPa06]) that

$$\Gamma_n^*(\xi,\tau;x,t) = \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(x,t)}(x,t;y,s) \Psi_n(y,s;\xi,\tau) dyds,$$

where $\Psi_1(x,t;\xi,\tau) = L(t)\Gamma_0^{(x,t)}(x,t;\xi,\tau)$ and, for $n > 1$,

$$\Psi_{n+1}(x,t;\xi,\tau) = \int_\tau^t \int_{\mathbb{R}^d} L(t) \Gamma_0^{(x,t)}(x,t;y,s) \Psi_n(y,s;\xi,\tau) dyds.$$

All in all we have found the expansion (termwise equal to (10.32))

$$\Gamma(x,t;\xi,\tau) = \sum_{n=0}^\infty \Lambda_n(\xi,\tau;x,t), \tag{10.35}$$

where $\Lambda_0(\xi,\tau;x,t) = \Gamma_0^{(x,t)}(x,t;\xi,\tau)$ and, for $n \geq 1$,

$$\Lambda_n(\xi,\tau;x,t) = \int_\tau^t \int_{\mathbb{R}^d} \Gamma_0^{(x,t)}(x,t;y,s) \Lambda_n(y,s;\xi,\tau) dyds. \tag{10.36}$$

In conclusion, we summarize the two results of this equivalent representation: 1) we can use the same operator $L(t)$ in both the first and second expansions; 2) in (10.36) we can integrate not w.r.t. the freezing point, in contrast to (10.26) and (10.33), simplifying a lot the formula. Notice that it is also possible to perform the very same steps and find a (fourth) perturbation series, equivalent to the first parametrix expansion.

10.4 Applications

We discuss some possible applications of the parametrix method. The authors in [CoPa06] suggest to use the method to approximate the option price (see the second subsection). We extend the potential applications to approximated likelihood estimation for discretely observed diffusion processes. According to the theoretical interest in the properties of the perturbation methods, we want to stress the pros and cons w.r.t. the other perturbation techniques presented. Both the proposals appear to be quite new, at least to our knowledge. However, no results are available in the direction of the performances of the AMLE based on the method, while only minimal evidence has been gained for the option price approximation, which however seem to be extremely good.

Approximation of The Transition Density

The perturbation series (10.25) and (10.35) are directly applicable to find the AMLE of the unknown parameters of the transition density, provided assumption $\mathbf{A1}-\mathbf{A3}$ hold. The interesting problem, still an open question, is under what conditions is possible to obtain the optimal asymptotic properties of the true unknown MLE. We make some general (we hope not too vague) remarks on the method of approximation in terms of this possible use, then we discuss the specific hypotheses and problems faced by the available techniques.

1) $\mathbf{A1}-\mathbf{A3}$ are quite strong requirements. The uniform ellipticity condition is a standard requirement and is typically satisfied by the transition density of the most used models, ruling out, e.g., the Feller square root process (2.12). In this case, most of the AMLE, especially the numerical (see Chapter 1 for a review of the literature), are in trouble as well. Only the method of [Aït02] (scalar case), can face this problem, despite the condition to recover this case are not always possible to check before the estimation of the parameters (see Section 9.3). The boundedness requirement is instead a major issue and not always can be achieved by some straightforward transformation. Anyway, we notice that $\mathbf{A1}-\mathbf{A3}$ hold true for the generalized BS model, where the element of the dispersion matrix are given by $\Sigma_{ij}(x,t)x_i$, provided the matrix $\Sigma(x,t)\Sigma^\top(x,t)$ is bounded (which was a very problematic case in Section 7.4). The generalized BS model is by far the most relevant model in Finance and it is a very interesting fact is that it is not really more complex to apply the results in more than 1 dimension. This is not true for other regular perturbations (see Chapter 7), while it is possible for the small-time techniques (Chapters 8,9) but only in terms of singular perturbations.

2) Another cost and limitation of the method is that there is not flexibility in the choice of the auxiliary model, but just the choice amongst the two expansions, choosing the expansion performing better in the case at hand. As a matter of fact, this is a great advantage in the generalized BS model, where the parametrixes $\Gamma_0^{(\xi,\tau)}$ and $\Gamma_0^{(x,t)}$ are transition densities of the classical solvable BS model (with different state-dependent volatilities), the natural auxiliary model to choose. So the parametrix is expected to outperform the other techniques in this setting, but also in the approximation of any model its constant coefficient version is a well known solvable model. Otherwise, the use of a different perturbation seems better. In particular, again, in theoretical terms the small-time technique of [Aït02] (and its extension [Aït04]) is by far the best possible choice.

3) The perturbation series converges. This fact is very rare in perturbations of differential problems. For the parametrix is an issue since the method has been devised to prove the existence of a fundamental solution. The advantage is twofold. First we saw that, for $(t-\tau)$ small, the rate of convergence is quite fast, but the convergence is uniform in all H. This implies, in terms of an estimation process, that the method does not require the unrealistic assumption of all the small-time techniques that the distance between the observations tends to 0. This is the real reason to try to apply the parametrix method to AMLE. Second, we can in principle get a good-as-desired approximation, just adding more terms, pretty much in the same philosophy of the

Hermite expansion of [Aït02], which however in general is only asymptotic. Actually, maybe the only weak point of [Aït02], is that the method requires many terms of the Hermite expansion to get a good approximation. In the experiments in [CoPa06] seems that this is not necessary also for not very small $(t-\tau)$.

We give now the interpretation of the two approximation and their terms. We have seen in Section 8.3 that for a small time $(t-\tau)$ any diffusion transition density is well approximated by a Gaussian density. The Gaussian approximation does not need to be good if the time-span $(t-\tau)$ is large. However, under $\mathbf{A1}-\mathbf{A3}$, after the necessary transformations to get the boundedness of the coefficients, the parametrix approximation says the following two things: i) The transition density is upper-bounded by a Gaussian (e.g., see (A.17)). But more is true. ii) The transition density is essentially a sum of Gaussian-type terms (which in turn implies i)).

To explain the second property we examine (10.25) and (10.35), highlighting their analogies and differences. The main term $(n=0)$ of both the approximation series is a Gaussian density, the parametrixes $\Gamma_0^{(\xi,\tau)}(x,t;\xi,\tau)$ and $\Gamma_0^{(x,t)}(x,t;\xi,\tau)$. The difference is the point in which one freezes the diffusion matrix and is connected with the action of the operators $L(t)$ and $L^*(\tau)$. The other terms $(n\geq 1)$ are time-integrals of expected values w.r.t. a Gaussian density (see formulas (10.26) and (10.36)): but, while the second parametrix $\Gamma_0^{(x,t)}(x,t;y,s)$ is the same Gaussian for each value of the integrating variables (y,s), the first parametrix $\Gamma_0^{(y,s)}(x,t;y,s)$ is a different Gaussian for each (y,s). More precisely, as (y,s) varies, also the variance of $\Gamma_0^{(y,s)}(x,t;y,s)$ varies (see (10.7)), so that we are not integrating w.r.t. an exact density, in the sense that it will not generally integrate to 1. As far the higher order terms are concerned, from (10.26) (and analogously (10.36)), recalling (3.10) and, in particular, (A.26), we see that $\Gamma_n(x,t;\xi,\tau)$ is a solution of the inhomogeneous equation $L_0^{(\xi,\tau)}v(x,t;\xi,\tau)=-L(t)\Phi_n(x,t;\xi,\tau)$, with an homogeneous initial condition. The source term is interpretable as the error committed at the previous stage of the approximation[6]. As a consequence, due to the convolution property of Gaussian kernel and the fact that $\Gamma_n(x,t;\xi,\tau)$ is a time-integral of a Gaussian-type function, the truncated series are in turn mixtures of Gaussian-type functions. This Gaussian feature has the practical advantage to render the numerical integration of the terms of order $n\geq 1$ easy (typically easier than those in (7.9), where the transition density does not need to be Gaussian), despite the discussed integrable singularity.

We conclude with the analysis of the assumptions in [Aït02] to get the equivalence of the AMLE to the true MLE. In fact, the method here is very similar to those of [Aït02] in that the goal is just to approximate the true likelihood with a sequence of convergent approximations. By means of the Hermite polynomial [Aït02], through the classical parametrix method here. Therefore it seems reasonable (but to study at all), that the very same identification assumptions (quite strong, as from [BaSc83], in the non-ergodic case, see Assumption 4 and proposition 3 in [Aït02]) should work, since the bulk of the effort in [Aït02] is to prove that the approximation expansion converges to the true transition density.

Approximation of The Option Price

A natural application of the parametrix approximation is the pricing of derivatives when the SDE-type model is not explicitly known. This is how [CoPa06] propose to use the method[7], in particular for hedging ends, which means to compute approximate Greeks of the unsolvable model (in fact, the price of any liquid derivative is directly observed in the market and the pricing is necessary only for hedging the position in the derivative). We summarize the pros and cons of the approach.

[6]By the estimates (10.31) and (10.23), we note that this errors decreases for all $n\geq 1$ (not only definitely). The more the diffusion matrix is near to that of the frozen auxiliary diffusions, the faster will be the convergence.

[7]See also [BoIs99] and related works for a similar application in the study of the implied local volatility.

We start recalling (see Chapter 3, in particular (3.13)) that the option price in the generalized BS model is given by the solution of the Cauchy problem (set here forward in time)

$$\begin{cases} -\frac{\partial}{\partial t}u(x,t) + \frac{1}{2}\sum_{i,j} a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j}u(x,t) = 0, & (x,t) \in \mathbb{R}^d \times (0,T], \\ u(x,0) = g(x), & x \in \mathbb{R}^d, \end{cases} \tag{10.37}$$

where we suppose that $a(x,t) := \Sigma(x,t)\Sigma^\top(x,t)$ is bounded (and, w.l.o.g., we have set the potential term $r(x,t)$ in (3.13) equal to 0). Then we exploit the representation formula (A.20) (which is a special case of (3.10))

$$u(x,t) = \int_{\mathbb{R}^d} g(y)\Gamma(x,t;y,0)dy. \tag{10.38}$$

Substituting the first expansion (10.25) into (10.38) we find the perturbation series for the option price

$$u(x,t) = \sum_{n=0}^{\infty} u_n(x,t), \tag{10.39}$$

where

$$u_0(x,t) = \int_{\mathbb{R}^d} g(y)\Gamma_0^{(y,0)}(x,t;y,0)dy, \tag{10.40}$$

$$\begin{aligned} u_1(x,t) &= \int_{\mathbb{R}^d} g(y)\Gamma_1^{(y,0)}(x,t;y,0)dy \\ &= \int_{\mathbb{R}^d} g(y)\int_0^t \int_{\mathbb{R}^d} \Gamma_0^{(y_0,s_0)}(x,t;y_0,s_0)L(s_0)\Gamma_0^{(y_0,s_0)}(y_0,s_0;y,0)dy_0 ds_0 dy \\ &= \int_0^t \int_{\mathbb{R}^d} dy_0 ds_0 \Gamma_0^{(y_0,s_0)}(x,t;y_0,s_0)L(s_0)\int_{\mathbb{R}^d} g(y)\Gamma_0^{(y_0,s_0)}(y_0,s_0;y,0)dy \\ &= \int_0^t \int_{\mathbb{R}^d} \Gamma_0^{(y_0,s_0)}(x,t;y_0,s_0)L(s_0)u_0(y_0,s_0)dy_0 ds_0, \end{aligned}$$

with the third equality that holds since $\Gamma_0^{(y_0,s_0)}$ is a Gaussian density (so $\int_{\mathbb{R}^d}$ and $L(s_0)$ commute) and all the singularities are integrable (so the order of integration is immaterial). At $n = 2$, recalling the definition of $\Gamma_2^{(y,0)}(x,t;y,0)$ and formula (10.10) applied to $f(x,t) = L(t)u_0(x,t)$, the same reasoning as above leads to

$$\begin{aligned} u_2(x,t) &= \int_{\mathbb{R}^d} g(y)\Gamma_2^{(y,0)}(x,t;y,0)dy \\ &= \int_0^t \int_{\mathbb{R}^d} \Gamma_0^{(y_0,s_0)}(x,t;y_0,s_0)L(s_0)\left[u_0(y_0,s_0) + u_1(y_0,s_0)\right]dy_0 ds_0. \end{aligned}$$

In general, we have

$$u_n(x,t) = \int_0^t \int_{\mathbb{R}^d} \Gamma_0^{(y_0,s_0)}(x,t;y_0,s_0)L(s_0)U_{n-1}(y_0,s_0)dy_0 ds_0, \tag{10.41}$$

where $U_{n-1}(x,t) := \sum_{k=0}^{n-1} u_k(x,t)$.

From the second parametrix perturbation we have a second option price perturbation

$$u(x,t) = \sum_{n=0}^{\infty} u_n^*(x,t), \tag{10.42}$$

where

$$u_0^*(x,t) = \int_{\mathbb{R}^d} g(y)\Gamma_0^{(x,t)}(x,t;y,0)dy, \tag{10.43}$$

and, for $n \geq 1$ (the proof is analogous to that of (10.41)),

$$u_n^*(x,t) = \int_0^t \int_{\mathbb{R}^d} \Gamma_0^{(x,t)}(x,t;y_0,s_0)L(s_0)U_{n-1}^*(y_0,s_0)dy_0ds_0, \tag{10.44}$$

with $U_{n-1}^*(x,t) := \sum_{k=0}^{n-1} u_k^*(x,t)$.

The two convergent series (10.39) and (10.42), exactly as (10.25) and (10.35), are clearly equivalent, in the sense that they have the same sum, but they are different termwise. Let's interpret the terms of the two perturbation series in a generalized BS world. Recall that the classical BS price (in d dimensions) is the solution of the Cauchy problem (10.37) associated to the heat operator $-\frac{\partial}{\partial t} + \frac{\gamma}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}u(x,t)$, $\gamma > 0$. Then the freezing of the operator $\frac{1}{2}\sum_{i,j} a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j}u(x,t)$ reduces the first term of both expansions to an easy computable slight extension of the BS option price.

We start from the first term. The second expansion turns out to be more easily interpretable, thanks to (10.36). The first term (10.43) is just a solvable BS price (with volatility, i.e., the diagonal elements of the frozen dispersion matrix Σ at (x,t)) dependent on the current value of the underlying. This is in accordance to the market practice to use almost always the classical BS model, even if patently wrong, just because it is analytically tractable. On the other hand, the first term of the first parametrix (10.40) is also very similar to the BS price, with the complication that we integrate w.r.t. y, where we freeze the diffusion operator. In terms of the option price it means that the volatility now depends on the terminal value y of the underlying, w.r.t. we are averaging. Furthermore, since $\Gamma_0^{(y,0)}$ does not need to be a density, the option price $u_0(x,t)$ is itself an approximation of the BS price, but is however again analytically solvable.

We turn to the higher order terms. The point is that both expansions tell us how to correct the error committed approximating the option value with a BS price. The terms (10.44) and (10.41) of the series are time-integrals of fictitious BS option prices with instantaneous payoffs (or transaction costs) $L(s_0)U_{n-1}^*(y_0,s_0)$ and $L(s_0)U_{n-1}(y_0,s_0)$, respectively. Again, these somehow complicated payoffs are in fact the errors committed at the previous step[8]. The point is that all these quantities are in principle computable in closed-form (or, anyway, by means of standard numerical or Monte Carlo procedures to evaluate Gaussian integrals).

For other interesting financial interpretations of the two expansions, but outside the scope of the thesis, we refer to [CoPa06] and [Cor06b]. Actually, an additional nice feature of this application of the parametrix to option pricing is that the two expansions allow different interpretations and also to justify some standard practitioner behaviors, that despite working well in practice, has not been theoretically justified so far.

[8]To be more explicit, e.g., $u_n(x,t)$, for any $n \geq 1$, solves the equation $L_0^{(\xi,\tau)}v(x,t) = -L(t)U_{n-1}(x,t)$ with homogeneous initial datum $v(x,0) = 0$.

Chapter 11

Some Abstract Results

We present some abstract results, among which an extension of the parametrix method that allows (in principle) to weaken the assumptions of boundedness of the coefficients and non-degeneracy of the diffusion matrix. The main assumption is that the Markov generator is a sectorial operator. The eventual goal is the study of the well posedness of the Cauchy problems considered in the first part of the thesis. References are [EnNa99], [ItKa02] and [Lun95]. The abstract parametrix method can be found in [Fri69] and [Paz83].

11.1 Introduction: Why Abstract Results?

We could have concluded the previous chapter with a brief analysis of the numerical experiments in [CoPa06], but it is more striking to do it now, to see that, despite the parametrix method presented in Chapter 10 has a broad applicability in financial problems, the assumptions of non-degeneracy and boundedness of the coefficients are restrictive.

The authors in [CoPa06] perform three numerical experiments considering the CEV model (see Section 3.3), a local volatility model, i.e., a scalar generalized BS model with dispersion parameter $\sigma(x,t) = \theta\left(1 + \frac{x^2}{1+t}\right)$, and the Hobson-Rogers model [HoRo98] for stochastic volatility. The option price under the CEV model is known in closed-form (see [Epp00]), while in the other two cases a Monte Carlo technique has been used to find the derivative price. Specified the free parameters, for a somehow short time-to-maturity of .5 years, the empirical results show that any of the two parametrix expansions truncated at 1 (i.e., considering u_1 or u_1^*) gives a stunning agreement result: in all three cases the approximation is very good, with a relative error of the order of 1%. There is only a slight difference for deep in or out of the money values for the Monte Carlo prices, confirming the well known non-complete reliability of that numerical technique in that range of the underlying[1].

Surprisingly enough, given the results achieved, all the examples proposed violate one of the key assumptions of the method in Chapter 10 (namely assumptions **A1** and **A2**). In particular, after all the suitable transformations have been applied, the CEV and local volatility models still have an unbounded diffusion coefficient, while the diffusion matrix of the Hobson-Rogers model (chosen bounded in the example) is singular in all $H = \mathbb{R}^2 \times [0,T]$, since there is not diffusion in one of the two space variables. Can we explain why?

Actually, the unboundedness in the underlying can be easily dealt with supposing that the underlying price is restricted to some compact, e.g., taking $\widetilde{\sigma}(x,t) = \max\{\sigma(x,t), C\}$, C a positive constant. This is typically done in the Finite difference approach for the numerical solution of PDEs, where one has to reduce the problem to a finite dimensional space truncating practically

[1]Note that this is exactly the kind of information one wants to obtain from a perturbation.

large enough values of the space variables (typically unbounded in financial problems). In fact, on a compact any continuous function is necessarily bounded, and the parametrix method applies.

Alternatively we could resort to some powerful extensions of the parametrix method allowing either for unbounded coefficients (e.g., see [Bes79] or [DeKr02]) or for possibly degenerate diffusions (see [DiPa05] and, in particular, [DiPa04]), explaining the empirical results. In particular, we choose to consider the abstract version of the problem, an approach that allows to deal with both the unboundedness and the possible degeneracy of the diffusion matrix at once. However, we stress that we will need to make other restrictive assumptions, as we will clarify. In practice, on a case by case basis, it should be more convenient to use the extension results cited above, as for the specific results about the Hobson-Rogers model. The main advantage in our presentation is that the connections with the regular perturbations of semigroups (Section 6.3) are manifest.

In the next section we collect (without proof, as in the appendixes, but explaining in detail) some properties of the so-called analytic semigroups, which are used in the final section for the study of the well posedness of non-autonomous abstract Cauchy problems.

11.2 Analytic Semigroups

We know from Chapter 4 that the conditional expectation of a diffusion process, under the Feller property, defines a strongly continuous evolution family of contractions $U(t,s)$, $0 \le t < s \le T$, in the space $\widehat{C}(\mathbb{R}^d)$[2]. In this section we are interested in the case the diffusion is time-homogeneous, such that the evolution family is a one-parameter semigroup $T(\tau) = U(t,s)$, $\tau = (s-t) > 0$, and the generator A of $T(\tau)$ is time-independent, having the form

$$A = \sum_{i=1}^{d} \mu_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}. \tag{11.1}$$

Let $D(A) \subset \widehat{C}(\mathbb{R}^d)$ be the domain of the unbounded operator A. By the Hille-Yoshida theorem (see Appendix A.2), we have that the closure of A is a dissipative, densely defined linear operator in $\widehat{C}(\mathbb{R}^d)$ and its resolvent set includes the positive complex plane. In Appendix A.3 it is also reviewed the fact that the abstract Cauchy problem

$$\begin{array}{c} -\frac{d}{dt}U(t) + AU(t) = 0, \quad t > 0 \\ U(0) = g, \end{array} \tag{11.2}$$

where $U : \widehat{C}(\mathbb{R}^d) \to D(A)$, A given in (11.1), is well posed for each $g \in D(A)$ and the solution is given by $U(t) = T(t)g$, i.e., the solution operator of (11.2) is the semigroup $T(t)$, $t \ge 0$, generated by A.

In the application of the abstract results to PDEs it is standard to require more regularity to the spectra of the generator A involved, asking A to be a sectorial operator (see [Lun95], [Paz83], Chapters 4,5,7 or [EnNa99], Section 6.5). The goals are: i) To extend the class of initial datum for which (11.2) is well posed. ii) To infer more derivability of the solution $U(t)$ from that of the solution semigroup operator $T(t)$.

Sectorial Operators

Let Σ_θ be a *sector* of angle θ in the complex plane, i.e.,

$$\Sigma_\theta = \{\lambda \in \mathbb{C} : |\arg \lambda| < \theta\} \setminus \{0\},$$

[2]For the extension of semigroup and generator to $L^p(\mathbb{R}^d, dQ)$, $p \ge 1$, see Section 4.1.

where $\theta \in [0, \pi]$, with the convention that $\Sigma_0 = \emptyset$. Note that $\Sigma_\pi = \mathbb{C}\backslash\{0\}$, while if $\theta \in (0, \pi/2]$, the closure $\overline{\Sigma}_\theta$ is a symmetric cone (i.e., a convex sector with boundaries and origin included) in the positive complex plane; in particular, $\overline{\Sigma}_\theta \supset \mathbb{R}_+ \cup \{0\}$.

A densely defined linear operator A in a Banach space X is called *sectorial* (of angle δ) if there exists $0 < \delta \leq \pi/2$ such that the sector $\Sigma_{\pi/2+\delta}$ is contained in the resolvent set $\rho(A)$ and, for each $\varepsilon \in (0, \delta)$, there exists a constant $b_\varepsilon \geq 1$ such that $\|R(\lambda, A)\| \leq b_\varepsilon / |\lambda|$, for all $0 \neq \lambda \in \overline{\Sigma}_{\delta-\varepsilon}$.

We make some important remarks. i) Since the resolvent set $\rho(A)$ is not empty, the operator A must be closed. So, w.l.o.g., we will always consider directly the closure of the Markov generator (11.1). ii) Form the definition it follows that the spectrum of a sectorial operator is contained in the symmetric cone $\mathbb{C}\backslash\Sigma_{\pi/2+\delta}$ (notice, origin included) in the negative complex plane. iii) Observe that the resolvent operator of a sectorial operator is characterized by a single estimate. In particular, if we can take $b_\varepsilon = 1$ for all $\lambda > 0$, then the sectorial operator is also dissipative. This must be the case if A is the generator of a contraction semigroup.

Now we explain why we are naturally interested in sectorial operators. Consider the Laplacian $\Delta := \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}$. It is well known (see [KaTa81], Vol.2) that $\frac{1}{2}\Delta$ is the generator in $\widehat{C}(\mathbb{R}^d)$ of the d-dimensional Brownian motion. In fact, the Laplacian is also sectorial in $\widehat{C}(\mathbb{R}^d)$. This can be proved by the explicit representation of the transition density of the arithmetic Brownian motion (for the long proof, see, e.g., [LLMP04]). In particular, the Laplace operator is a special case of the following result (see [Ste74] and [Ste80]):

Theorem 20 *Consider the operator A in (11.1). If the matrix $a(x) = (a_{ij}(x))$ is uniformly elliptic and all the coefficients $\mu_i(x)$ and $a_{ij}(x)$ are bounded and continuous, $a_{ij}(x)$ uniformly continuous, then A is sectorial in $\widehat{C}(\mathbb{R}^d)$.*

This theorem tells us that possible sufficient conditions such that the Markov generator (11.1) in $\widehat{C}(\mathbb{R}^d)$ is sectorial are the very same conditions such that we can apply the parametrix method to approximate the solution of the autonomous problem (11.2). It turns out that we need that a general time-dependent Markov generator $A(t)$ is sectorial for all $t \geq 0$, in order to extend the parametrix method to a general non-autonomous abstract Cauchy problems (see Section 11.4 below).

An even more interesting reason for which sectorial operators are crucial for us is that any self-adjoint Markov generator in a Hilbert space X is sectorial. Actually, the spectrum of a self-adjoint operator is included in the real axis (see Appendix A.1), while the spectra of any Markov generator is included in the left-half complex plane (possibly imaginary axis included). By the intersection of the two sets, we have that the spectra of the self-adjoint generator is included in $\mathbb{R}_- \cup \{0\}$. This fact and the single estimate of the Hille-Yoshida theorem for contraction semigroups (see Appendix A.3) imply that any self-adjoint Markov generator must be sectorial. This result is important since in many estimation techniques, reviewed in Chapter 7,8, it is required that the diffusion is stationary and ergodic, conditions that are ensured by the self-adjointness of the Markov generator in $L^2(\mathbb{R}^d, dQ)$, Q the stationary measure of the diffusion process. In turn (see Section 4.1), this extension can be recovered by the study of the sectorial operators in $\widehat{C}(\mathbb{R}^d)$.

Generation of Analytic Semigroups

We recall (see Appendix A.3) that, if A was the generator of a strongly continuous semigroup $T(t)$, $t \geq 0$, i.e., $Ag = \frac{d}{dt}T(t)g|_{t=0^+}$ for all $g \in D(A)$, then the resolvent would have the integral representation (A.36), i.e.,

$$R(\lambda, A)g = \int_0^\infty e^{-\lambda s} T(s) g ds, \tag{11.3}$$

where the improper integral converges for all $g \in X$ and $\lambda \in \rho(A)$. In other words, the resolvent operator is also the Laplace transform of the semigroup $T(t)$. This is correct since there is a one-to-one correspondence between a densely defined closed operator A and its resolvent operator $R(\lambda, A)$ (see [EnNa99], Chapter 2). However, for a general strongly continuous semigroup, formula (11.3) cannot be inverted since, by the Hille-Yoshida theorem, the spectrum of A is in general too large. But the inverse transform of (11.3) exists if A is also sectorial.

In fact, because the spectrum of a sectorial operator A is included in the cone $\mathbb{C}\backslash\Sigma_{\pi/2+\delta}$, one can define a family of operators $T(z)$, $z \in \Sigma_\delta$, by means of the formula

$$T(z) := \frac{1}{2\pi i}\int_\gamma e^{\mu z} R(\mu, A) d\mu, \tag{11.4}$$

and $T(0) := I$, where γ is a piecewise smooth curve going from $\infty e^{-i\vartheta}$ to $\infty e^{+i\vartheta}$, $\vartheta \in (\pi/2, \pi/2+\delta)$. The integration of the contour integral (11.4) must be performed for increasing $\operatorname{Im}\mu$ along γ, i.e., γ is a counterclockwise oriented contour around $\sigma(A)$. It turns out (see [EnNa99], Section 2.4.a) that: i) The integral (11.4) converges in the uniform operator topology. ii) The map $z \mapsto T(z)$ is bounded and strongly continuous in $\Sigma_{\delta'} \cup \{0\}$, for any $\delta' \in (0, \delta)$. iii) The map $z \mapsto T(z)$ is analytic in Σ_δ and satisfies the semigroup property in Σ_δ, i.e., $T(z+y) = T(z)T(y)$ for all $z, y \in \Sigma_\delta$. iv) The restriction of the map $z \mapsto T(z)$ to $z \in \mathbb{R}_+ \cup \{0\}$ is the strongly continuous semigroup generated by A.

All these properties follow from the single estimate of the resolvent of the sectorial operator and, most of all, the fact that the resolvent operator $R(\mu, A)$ is always analytic in $\rho(A) \supset \Sigma_\delta$ (e.g., see [Paz83], Section 1.5), so that the integrand in (11.4) is analytic and we can apply the Cauchy integral theorem (for operator valued functions). So, if the generator is sectorial, the generated strongly continuous semigroup $T(t)$, $t \geq 0$, is analytic in $t > 0$ and can be also extended analytically to the sector Σ_δ, $0 < \delta \leq \pi/2$. An operator valued function $T : \Sigma_\delta \cup \{0\} \to \mathcal{L}(X)$ with properties ii) and iii) above is called *analytic semigroup* (of angle δ). Note that the properties of an analytic semigroup, in particular the convergent representation (11.4), allow to write (not only formally) $T(z) = e^{Az}$, $z \in \Sigma_\delta \cup \{0\}$, where A is the generator of $T(z)$[3]. Note that if A is bounded, we do not need to use definition (11.4) (which actually holds), but we already see from the $\mathcal{L}(X)$-convergent series representation (A.33) of the semigroup that any (and only any) uniformly continuous semigroup can be extended analytically to all $\mathbb{C}$ (i.e., the boundedness of the generator is the only way to resolve completely the non-analyticity (i.e., singularity) at $\tau = 0$ of the mapping $\tau \mapsto T(\tau)$).

Analytic semigroups are characterized by the following properties (see [EnNa99], Theorem 4.4.6)

Theorem 21 *For an operator A in the Banach space X, the following facts are equivalent:*

- *A is sectorial.*
- *A generates an analytic semigroup $T(z)$, $z \in \Sigma_\delta \cup \{0\}$, in X.*
- *A generates a strongly continuous semigroup $T(t)$, $t \geq 0$, in X with the additional properties*

$$Rg(T(t)) \subset D(A), \quad \textit{for all } t > 0, \tag{11.5}$$

$$b := \sup_{t>0} \|tAT(t)\| < \infty. \tag{11.6}$$

Properties (11.5) and (11.6) are the key to see why the function $T(z)$, $z \in \Sigma_\delta$, is analytic by means of elementary facts. We need the following result (see [ItKa02], Proposition 3.2)

[3] We recall that the notation $T(t) = e^{At}$ for a strongly continuous semigroup with generator A is only formal and, therefore, generally avoided.

Proposition 22 *Let $T(t)$, $t \geq 0$, be a strongly continuous semigroup in X with generator A. Then $T(t)X \subset D(A)$ for all $t > 0$ if, and only if, the mapping $t \mapsto T(t)$ is differentiable on $t > 0$. In this case, if A^k, $k \geq 2$, are well defined, then $T(t)$ has derivative of any order in $t > 0$ and $\frac{d^n}{dt^n}T(t) = A^n T(t)$ for all $t > 0$.*

Suppose now that A^k, $k \geq 2$, is well defined (as we suppose, e.g., in Chapters 8,9). Then $T(t)$ has derivative of any order in $t > 0$ and hence we can apply the Taylor's theorem to $T(t)$ around $t_0 > 0$ obtaining

$$T(t) = \sum_{n=0}^{N-1} \frac{(t-t_0)^n}{n!} A^n T(t_0) + R_N(t), \tag{11.7}$$

for $t > 0$, where $R_N(t) = \frac{1}{(N-1)!}\int_{t_0}^t (t-\tau)^{N-1} A^N T(\tau) d\tau$. By the semigroup property $(A^n T(t)) = (AT(t/n))^n$, for $t > 0$. Then, using $\frac{n^n}{n!} \leq e^n$ for n large enough (e.g., from the Stirling's formula) and property (11.6), we have, for $\tau \in (t_0, t)$, $\frac{1}{N!}\left\|A^N T(\tau)\right\| \leq \frac{1}{N!}\left\|AT(\tau/n)\right\|^N \leq (eb/t_0)^N$. This implies that $\|R_N(t)\| \to 0$, as $N \to \infty$, for all $0 \leq (t-t_0) < \frac{t_0}{eb}$. In other words, the power series (11.7) converges for all $0 \leq (t-t_0) < \frac{t_0}{eb}$. By the arbitrary of t_0 we see that $t \mapsto T(t)$ is analytic for all $t > 0$. Furthermore, $T(t)$ in (11.7) extends analytically to the sector Σ_δ (where $\delta = \arctan(1/(eb))$).

Notice also that, by standard fact about analytic operator-valued functions (see [HiPh57] or, better, [ItKa02], page 98), the function $t \mapsto T(t)$ is analytic in $t > 0$ if, and only if, for any $g \in X$ the mapping $t \mapsto T(t)g$ is analytic in $t > 0$.

Interpretation in Terms of Cauchy Problems

Here we make some remarks about the effects of the sectorial property of the Markov generator A to the Cauchy problem (11.2). Assume that one of theorems 9 or 10, Section 4.1, holds, so that A in (11.1) generates a Feller semigroup.

The main point is that, if A in (11.1) is the generator of a analytic semigroup in a Banach space X, then the Cauchy problem (11.2) is well posed for each $g \in X$ (see theorem 4.1.5 of [Paz83]). This essentially follows by the property (11.5), which ensures that $AT(t)g$ is defined for all $g \in X$. If the Cauchy problem is inhomogeneous, i.e.,

$$\begin{array}{c} -\frac{d}{dt}U(t) + AU(t) = f(t), \quad t > 0 \\ U(0) = g, \end{array} \tag{11.8}$$

if the Markov generator A is sectorial and $f \in C(0,T;X) \cap L^1(0,T;X)$, then also (11.8) has a unique classical solution for every $g \in X$ (e.g., see [Lun95], Chapter 4). The solution is given by the variation of constants formula

$$U(t) = T(t)g - \int_0^t T(t-\tau)f(\tau)d\tau.$$

The second remark is about the differentiability of the solution $U(t)$ of (11.2). The request that A^k, $k \geq 2$, is well defined is equivalent to ask that the coefficients of (11.1) are infinitely differentiable. This is an obvious requirement (in the same direction see, e.g., (A.16)). Then, if A is sectorial, $t \mapsto T(t)$ is analytic in $t > 0$, but has a singularity at $t = 0$, as already noticed. At $t = 0$ we recover the initial condition $U(0) = T(0)g = g$. In fact, for $g \in X \backslash D(A)$, still $\frac{d}{dt}T(t)g = AT(t)g$ (for a strongly continuous semigroup it must be that $g \in D(A)$), but $AT(t)g \to Ag$ as $t \downarrow 0$ by strong continuity, so that it is manifest that $t \mapsto T(t)g$ cannot be even differentiable in t, at $t = 0$, for $g \notin D(A)$.

From these two remarks we see that the effect of a sectorial operator is that of smooth out the initial condition. E.g., if we set the Cauchy problem in the Banach space $X = L^p(\mathbb{R}^d, dQ)$,

$p \geq 1$, than $g \in X$ does not need to be a continuous function. But also, if we take $X = \widehat{C}(\mathbb{R}^d)$, then $g \in \widehat{C}(\mathbb{R}^d)$ is continuous but not necessarily differentiable and in $D(A)$. However, if A is sectorial, $T(t)g = e^{At}g$ is the unique solution of (11.2) for all $g \in X$, so that $AT(t)g$ and $\frac{d}{dt}T(t)g$ are well defined and belong to X. Moreover, if we require more regularity on the coefficients of A, then also the solution $T(t)g$ enjoys pretty much the same regularity of the coefficients for $t > 0$. In other words, sectorial operators are associated to *regularizing phenomena*, typically described by parabolic PDEs (for a general discussion, but without reference to analytic semigroups, see, e.g., [Sal04] and [Duf06]).

The sectorial property of the generator is a useful property but, as it is always the case for broadly applicable abstract result, it is sometimes a "restrictive" assumption, i.e., something more deep can be said in specific cases. E.g., the Laplace operator $\frac{1}{2}\Delta$ is sectorial in $\widehat{C}(\mathbb{R}^d)$, but not in $B(\mathbb{R}^d)$, since $\widehat{C}(\mathbb{R}^d)$ is not dense in $B(\mathbb{R}^d)$. So we cannot apply the results above for $g \in B(\mathbb{R}^d)$. However, $T(t)g(x) = \int_{\mathbb{R}^d} g(y)p(x,y;t)dy$, $p(x,y;t)$ the transition density of the Brownian motion, is well defined for all $g \in B(\mathbb{R}^d)$. Furthermore, by direct inspection, since $p(x,y;t)$ is a Gaussian, we see that $T(t)g$, $g \in B(\mathbb{R}^d)$, not only solves the Cauchy problem (11.2) with $A = \frac{1}{2}\Delta$ (this is more generally true under the conditions for the first Kolmogorov equation, see Sections 3.1 and 4.1), but it is also a function in $C^\infty(\mathbb{R}^d \times (0,T])$[4].

11.3 Analytic Vectors

We study possible sufficient (but extremely restrictive) conditions ensuring the convergence of the asymptotic perturbation series (8.7)-(8.16), at least for specific given functions g[5]. The main result presented is from [Sch04]. However, the result is in the same direction of what we have done in Section 9.5: the only way to prove that the perturbation series (8.7) converges for some specific g, fact that is not excluded, one has to specify the functional form of the generator A and find a good estimate of Ag. This is typically difficult by means of elementary calculus (see Section 9.5), but it is not simple, in the few feasible cases, even if we tackle the problem with powerful abstract methods. We conclude with an idea, object of current research, about the possibility to extend the results presented.

First of all we need some notation. Let $T(\tau)$ be a strongly continuous semigroup in the Banach space X (equal to either $\widehat{C}(\mathbb{R}^d)$ or $L^p(\mathbb{R}^d, dQ)$) and A be the associated generator given in (8.4). So we explicitly require that $\mu_i(x), a_{ij}(x) \in C^\infty(\mathbb{R}^d)$, ensuring, as already discussed, that A^n, $n \geq 0$, are well defined operators. Following [Nel59] (see also [ReSi80], Section 10.6, Vol.2, and the Notes of that Section for some remarks and extensions), we define an element $g \in D(A^\infty) = \cap_{n=0}^{\infty} D(A^n)$ ($D(A)$ the domain of A) an *analytic vector for A* if the power series (8.7) converges in the strong operator topology with a strictly positive ray of convergence, i.e., if

$$\sum_{n=0}^{N} \frac{\tau^n}{n!} \|A^n g\|_X < \infty, \tag{11.9}$$

for some $\tau > 0$. The main result in [Nel59], at least for our sake, is the following

Theorem 23 *Let X be an Hilbert space and A be a closed symmetric operator in X. A is self-adjoint if, and only if, $D(A)$ contains a dense set of analytic vectors for A.*

We remark that, in this theorem, A is not required to be a semigroup generator. This extra requirement is good, implying that $D(A^\infty)$ is dense in X (see Section 8.1). In fact, for a general symmetric operator A it could be that $D(A^\infty) = \emptyset$. The strong conclusion is that any self-adjoint

[4]Note that $t = 0$ is excluded. Obviously $\left(\frac{1}{2}\Delta\right)^n$, $n \geq 1$, makes sense.

[5]Another set of (quite restrictive) sufficient conditions for the convergence of the perturbation series (8.7)-(8.16) is given in [Aït02].

Markov generator has a dense set of analytic vectors. However, there is not any general criterion to see if an element $g \in D(A)$, A a self-adjoint Markov generator, is an analytic vector for A. Again, given A, one has to try to compute some estimates of $\|A^n g\|_X$ and show that, for a particular g, the series (11.9) converges for some $\tau > 0$.

We turn to the results in [Sch04]. The main result there (again, to our ends) is the remark that, if the Markov generator A is sectorial, then there must exists a non-empty subset of analytic vectors for the generator A. This is important since many diffusion generators are sectorial (see Section 11.2), but not necessarily self-adjoint: e.g., the generator of the Brownian motion, associated to the heat equation and, by the log-transform, to the classical BS model and the geometric Brownian motion.

The proof of the claim is based on two special properties of the analytic semigroup $T(\tau)$ generated by a sectorial Markov generator A (see theorem 19, Section 8.2): i) $Rg(T(\tau)) \subset D(A)$, for all $\tau > 0$; ii) $AT(\tau)$ is a bounded operator for all $\tau > 0$: in particular, $\|AT(\tau)\| \leq b/\tau$ (see (11.6) for the definition of b). From i) and the relation $AT(\tau)g = T(\tau)Ag$, valid for all $\tau \geq 0$ and $g \in D(A)$, one can verify (e.g., see [ItKa02], Proposition 3.3.12) that actually $Rg(T(\tau)) \subset D(A^\infty)$, for all $\tau > 0$. Then, at least for $g \in Rg(T(\tau_0))$, $\tau_0 > 0$, using the estimate in ii), [Sch04] proves that (11.9) converges for all $0 \leq \tau < \tau_0/(be)$.

Actually, this fact can be more easily seen invoking the fact that, for an analytic semigroup $T(\tau)$, $\tau \geq 0$, the mapping $\tau \mapsto T(\tau)$ is analytic for all $\tau > 0$, i.e., the power series representation

$$T(\tau) = \sum_{n=0}^{\infty} \frac{(\tau - \tau_0)^n}{n!} A^n T(\tau_0),$$

is convergent for all $0 \leq (\tau - \tau_0) < \frac{\tau_0}{eb}$ (see (11.7)). Now, consider $g = T(\tau_0)f$, $f \in D(A^\infty)$, for $\tau_0 > 0$, i.e., $g \in Rg(T(\tau_0))$. Then, using the semigroup property, for $\tau > \tau_0$, we have $T(\tau)f = T(\tau - \tau_0 + \tau_0)f = T(\tau - \tau_0)T(\tau_0)f = T(\tau - \tau_0)g$, and hence

$$\begin{aligned} T(\tau - \tau_0)g &= \sum_{n=0}^{\infty} \frac{(\tau - \tau_0)^n}{n!} A^n T(\tau_0) f \\ &= \sum_{n=0}^{\infty} \frac{(\tau - \tau_0)^n}{n!} A^n g, \end{aligned}$$

convergent for all $0 \leq (\tau - \tau_0) < \frac{\tau_0}{eb}$. Call $\tau_1 = (\tau - \tau_0)$ and we get again the claim in [Sch04] (with τ_1 in place of τ).

We notice that, chosen a particular $g \in D(A^\infty)$, the condition $g = T(\tau_0)f$, for $\tau_0 > 0$, in general cannot be inverted (it defines an ill-posed integral equation). However, a second important result in [Sch04] is that the equation $g = T(\tau_0)f$ can be inverted in at least a very special case. For the sake of clarity we make some steps.

1) *Assumptions.* Let X be an Hilbert space and suppose the Markov generator is self-adjoint. In our case, we have to take $X = L^2(\mathbb{R}^d, dQ)$, Q the invariant distribution associated to A (see Section 4.2). The first two consequences are that: i) as stated in the theorem above, the set of analytic vectors for A is dense in $D(A)$; ii) the spectrum of A is included in $(-\infty, 0]$. Suppose also that the spectrum of A is purely discrete and countable, say $\{\lambda_n\}_{n\geq 0}$, with $0 = \lambda_0 > \lambda_1 > \cdots > \lambda_n \to -\infty$ and that the set of eigenfunctions is known in closed form, say $\{\varphi_n\}_{n\geq 0}$. This last assumption is very restrictive, but satisfied by some commonly used SDE-type models (see Section 9.2 and [KaTa81], Section 15.13). Then the sequence of eigenfunction form an orthonormal basis for the Hilbert space X (see Appendix A.1).

2) *Representation of g.* The function $g \in D(A^\infty)$ has the representation

$$g = \sum_{n=0}^{\infty} \langle g, \varphi_n \rangle_X \varphi_n, \tag{11.10}$$

convergent in the norm of X (i.e., the norm induced by the inner product $\langle .,.\rangle_X$). Recall that (11.10) is convergent if, and only if, $\sum_{n=0}^{\infty} |\langle g,\varphi_n\rangle_X|^2 < \infty$ (see (A.5)).

3) *Representation of* $T(\tau_0)f$. The mapping $\tau \mapsto T(\tau_0)f = e^{A\tau_0}f$, for $f \in D(A^\infty)$ and $\tau_0 > 0$, is analytic. But the generator A is unbounded and so we cannot use the analytic functional calculus (see Appendix A.1). However, by the assumptions in 1) we can use the more powerful measurable functional calculus (see (A.6)) and get the convergent series representation (since the spectra of A is discrete)

$$T(\tau_0)f = e^{A\tau_0}f = \sum_{n=0}^{\infty} e^{\lambda_n \tau_0} \langle f,\varphi_n\rangle_X \varphi_n. \tag{11.11}$$

This fact can be maybe more easily seen, for $f \in D(A^\infty)$, by: i) $\frac{d}{dt}T(\tau_0)f = AT(\tau_0)f$, since $T(\tau_0)$ is a semigroup; ii) $AT(\tau_0)f = \lambda T(\tau_0)f$, which is the singular Sturm-Liouville problem used to find the eigenvalues $\{\lambda_n\}_{n\geq 0}$; then $\frac{d}{dt}T(\tau_0)f = \lambda T(\tau_0)f$, and hence $T(\tau_0)f = e^{\lambda\tau_0}f$. Clearly, we can represent $e^{\lambda\tau_0}f$ as (11.11). We remark that these two derivations of (11.11) hold under much weaker conditions: for the first see, e.g., [HaSc95], for the second see [KeSo99].

4) *The condition for* $g \in Rg(T(\tau_0))$. By (11.10) and (11.11), the condition $g = T(\tau_0)f$, for $f \in D(A^\infty)$ and $\tau_0 > 0$, is equivalent to

$$\sum_{n=0}^{\infty} \langle g,\varphi_n\rangle_X \varphi_n = \sum_{n=0}^{\infty} e^{\lambda_n\tau_0} \langle f,\varphi_n\rangle_X \varphi_n.$$

By the uniqueness of the Fourier series representations (11.10) and (11.11) given the basis $\{\varphi_n\}_{n\geq 0}$ of X, we see that $g = T(\tau_0)f$ if, and only if, the Fourier coefficients are such that $\langle g,\varphi_n\rangle_X = e^{\lambda_n\tau_0}\langle f,\varphi_n\rangle_X$ for all $n \geq 0$. So, given a specific function $g \in D(A^\infty)$, we have that $g \in Rg(T(\tau_0))$ for all $\tau_0 > 0$ such that

$$\sum_{n=0}^{\infty} \left| e^{-\lambda_n\tau_0} \langle g,\varphi_n\rangle_X \right|^2 < \infty. \tag{11.12}$$

The condition (11.12) is very restrictive, since the sequence $\{-\lambda_n\}_{n\geq 0}$ diverges to ∞, typically as or faster than $O(n)$, and so $e^{-\lambda_n\tau_0}$ grows at least as $O(e^n)$. This means that the coefficients $\langle g,\varphi_n\rangle_X$ must tend to 0 extremely fast, at least definitely, to satisfy (11.12) for a strictly positive τ_0. Due to this fact, it is not rare that, when (11.12) is satisfied, the ray of convergence τ_0 is very small. In this case, the advantages from the convergence of the perturbation series are minimized. Clearly the behavior of $\langle g,\varphi_n\rangle_X$ depends not only on the generator A (i.e., the eigenfunctions φ_n), but most of all on the specific function g.

In [Sch04] is considered the following very interesting example. Let $A = -\mu x\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}$, i.e., the generator of the Ornstein-Uhlenbeck process (2.10), defined in $X = L^2(\mathbb{R}, q(x)dx)$, where $q(x)$ is the invariant measure of the process given by a Gaussian random variable with 0 mean and variance equal to $\sigma^2/(2\mu)$. The spectra of A is purely discrete, given by the eigenvalues $\lambda_n = -\mu n$, $n \geq 0$, while the associated eigenfunctions are given by $\varphi_n(x) = H_n(x\sqrt{\mu}/\sigma)/\left(2^{n/2}\pi^{1/4}\sqrt{n!}\right)$, $n \geq 0$, where $H_n(x)$ is the Hermite polynomial of order n (see [KaTa81], page 333).

The inner product of X is given by $\langle g,\varphi_n\rangle_X = \int_{\mathbb{R}} g(x)\varphi_n(x)q(x)dx$. So the Fourier coefficients $\langle g,\varphi_n\rangle_X$ will die out the faster the thinner are the tails of the function g. For this reason, [Sch04] considers $g(x) = e^{-x^2/c}$, $c > 0$. Notice that this function is extremely regular, $g \in C^\infty(\mathbb{R}) \cap \widehat{C}(\mathbb{R})$, with very thin tails, and can be perfectly approximated by functions in $C_0^\infty(\mathbb{R})$. By long calculations one can arrive to say (we refer to [Sch04] for the details) that $g(x) = e^{-x^2/c} \in Rg(T(\tau_0))$, for all $0 < \tau_0 \leq \ln\left(1 + \mu c/\sigma^2\right)/(2\mu)$, which is a very good result.

The problem is that there are only few self-adjoint diffusions for which i) the spectra is purely discrete and ii) the eigenfunctions are known in closed form. Three diffusions with these properties play a special role: the Ornstein-Uhlenbeck process (2.10), the Feller Square-root process (2.12) and the Jacobi process (e.g., see [KaTa81], page 355), that, having a compact support, has received very few interest in Finance. The associated eigenfunctions are given by the classical Hermite, Laguerre and Jacobi polynomials. Many other diffusions with the same feature can be seen as a one-to-one transformation of one of these (see, e.g., [AlKu05] and [AlLa05] and the reference therein). In all this cases however the transition density is clearly known, so that better techniques can be used rather than of an asymptotic series expansion, even if convergent.

An Approximation Scheme

We conclude the section sketching an idea, currently under development, that could prove to be useful in the quest for the conditions for saying if a function in $g \in D(A^\infty)$ is actually an analytic vector for the generator A. The idea is to approximate the semigroup $T(\tau)$ generated by A by means of a sequence of bounded generators A_h and try to understand when a function $g \in D(A^\infty)$, which of course is an analytic vector of A_h[6], is also an analytic vector for A. The idea of reduce the study of a strongly continuous semigroup to that of a sequence of uniformly continuous semigroup is typical (e.g., see the proof of the Hille-Yoshida theorem, [EnNa99], Section 2.3). This is also naturally connected with the finite difference approximation of the Cauchy problems involving the differential operator A (see [ItKa96]) and necessary for the numerical implementation of any small-time expansion. To be more precise we consider, in the scalar case, the following steps[7]:

a) *Space discretization.* Truncate the state space $\mathbb{R}$ taking a large interval I and define the lattices $I_h = h\mathbb{Z} \cap I$ with $N := card(h\mathbb{Z} \cap I)$. We need that the lattices are nested as $h \downarrow 0$, so that we could take $h = \{h_n\}$, $h_n = 2^{-n}$. Consider a finite difference approximation of $A = \mu(x)\nabla + \frac{1}{2}\sigma^2(x)\Delta$ (∇ the gradient and Δ the Laplacian operators in the scalar case) given by $A_h = \mu(x)\nabla_h^+ + \frac{1}{2}\sigma^2(x)\Delta_h$, where ∇_h^+ is the forward first differences operator and Δ_h is the central second differences operator. E.g., $\nabla_h^+ f(x) = \left(f(x+h) - f(x)\right)/h$. Notice that A_h well approximates the action of A for any function $g \in C_0^2(\mathbb{R})$.

b) *The generator A_h.* Notice that A_h is a continuous operator, that can be represented by an $N \times N$ matrix. Basic condition on the matrix representation of A_h such that it is a Markov generator are given in [GrSt01] (see Theorem 6.10.10): these conditions can be stated in terms of the original diffusion coefficients asking $\sigma(x) > 0$ and -$\sigma(x) < h\mu(x) < \sigma(x)$, for all x. Then A_h on the discrete state space I_h is the generator of a birth and death process. In fact, the associated Markov process can jump only to the nearest state on both directions, i.e., the matrix representation of A_h is tridiagonal (a very nice matrix to be diagonalized).

c) *The semigroups.* Since A_h is bounded, it generates an uniformly continuous semigroup $T_h(\tau)$. We can infer the convergence of the semigroups $T_h(\tau)g \to T(\tau)g$ for all $g \in X$, by a suitable convergence of the generators (see [Paz83], Section 3.6), under the conditions of the Trotter-Kato theorem (a special version of this theorem used in finite difference approximations is given in [ItKa96] and [ItKa02]). Note that the theorem says that we can approximate, at least numerically, the solution $T(\tau)g$ of (8.6) by $T_h(\tau)g$.

d) *Convergence of the series.* This is the point: is it possible, probably by means of a sieve method, to infer if a given function $g \in D(A^\infty)$ is an analytic vector for A, just studying its behavior w.r.t. A_h? To be sure that A actually has analytic vectors, we are working under the hypothesis that the generator A is self-adjoint. But this problem is extremely involved.

[6] Clearly, the series (11.9) converges for all $g \in D(A) = X$ if A is bounded, so that its set of analytic vectors is all $D(A) = X$.

[7] Steps a,b) are used, e.g., by [AlMi06a],[AlMi06b] to approximate option prices by means of spectral methods.

11.4 Abstract Non-autonomous Cauchy Problems

The eventual goal of the Chapter is to study from an abstract point of view the well posedness of the Cauchy problems considered in the first part of the thesis. Consider the evolution equation

$$-\tfrac{d}{dt}u(t) + A(t)u(t) = 0, \quad t > s, \tag{11.13}$$

in the Banach space X, where, for $s \in [0, T)$, $A(t)$ is a Markov generator, i.e.,

$$A(t) = \sum_{i=1}^{d} \mu_i(x,t)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x,t)\frac{\partial^2}{\partial x_i \partial x_j}, \tag{11.14}$$

$a(x, t)$ symmetric and positive semi-definite. We look for classical solutions of (11.13), i.e., functions $u : [s, T] \to D(A(t))$ in $C(s, T; D(A(t))) \cap C^1(s, T; X)$, for $t \in [s, T]$. Given a suitable initial value $u(s) = g$, it is reasonable to suppose that the solution operator of (11.13), i.e., an $\mathcal{L}(X)$-operator valued function $U(t, s)$, $0 \le s \le t \le T$, such that $u(t) = U(t, s)g$, is a strongly continuous evolution system (see Chapter 4 and Appendix A.3)[8]. We make the following assumptions

B1 The domain D of $A(t)$, $t \in [0, T]$, is dense in X and independent of t.

B2 For each $t \in [0, T]$, the resolvent $R(\lambda, A(t))$ of $A(t)$ exists for all λ such that $\mathrm{Re}\,\lambda \ge 0$, and $\|R(\lambda, A(t))\| \le const.(|\lambda| + 1)^{-1}$.

B3 For any $t, s \in [0, T]$, there is $\alpha \in (0, 1]$ such that $\left\|[A(t) - A(s)]\, A(s)^{-1}\right\| \le const.\, |t - s|^{\alpha}$.

We remark that **B2** implies that, for all $\eta \in [0, T]$, the operator $A(\eta)$ is sectorial. Therefore, for each $\eta \in [0, T]$, $A(\eta)$ generates an analytic semigroup $T_\eta(t) = e^{tA(\eta)}$, $t \ge 0$. In particular, $T_\eta(t)$, η fixed, is strongly continuous for $t \ge 0$, $\frac{d}{dt}T_\eta(t) = A(\eta)T_\eta(t)$ is a bounded operator, for $t > 0$, and (see (11.6))

$$\|A(\eta)T_\eta(t)\| \le const.t^{-1}, \quad t > 0. \tag{11.15}$$

Furthermore, for η fixed and $g \in X$, the Cauchy problem $-\frac{d}{dt}u(t) + A(\eta)u(t) = 0$, $t > s$, and initial condition $u(s) = g$, has a unique solution $T_\eta(t - s)g$, $t \ge 0$.

The Abstract Parametrix

In the classic parametrix method the first step is to freeze the coefficients of (11.14) at the values of the initial condition. In the abstract formulation the PDE is regarded as an ODE in time. So we freeze the operator $A(t)$ at the initial time s: $A(s)$, s fixed, is used as auxiliary model and generates a one-parameter C_0-semigroup $T_s(t-s)$; the perturbation of the auxiliary model $A(s)$ is given by the unbounded operator $(A(t) - A(s))$. The idea is to solve the problem in terms of the solution operators as in Section 6.3, allowing to transform the differential problem into an integral problem. The conditions $\mathbf{B1} - \mathbf{B3}$ are sufficient to bound the effect of $(A(t) - A(s))$ on the perturbation expansion in terms of the solution operators. We call $T_s(t-s)$ the parametrix (at the initial time s) of the evolution equation (11.13) and we use it as a centre of the perturbation expansion of $U(t, s)$, the solution operator of (11.13).

The second step of the abstract parametrix method, pretty much as in the classic parametrix, is to suppose that, under $\mathbf{B1} - \mathbf{B3}$, the solution operator $U(t, s)$, $0 \le s \le t \le T$, is an evolution system of the form

[8] Notice that, since we are setting here the Cauchy problems forward in time, we write the solution operator as $U(t, s)$, i.e., with larger times on the left. This is necessary for the correct ordering of the composition products: e.g., see (11.17) or (11.19). As usual, the change of variables $(t, s) \mapsto (T - t, T - s)$ does not affect the validity of the results, but only the sign of the time-derivatives, etc.

$$U(t,s) = T_s(t-s) + \int_s^t T_\tau(t-\tau)R(\tau,s)d\tau, \tag{11.16}$$

where the operators $R(t,s)$ has to be found by (from (11.13))

$$-\frac{\partial}{\partial t}U(t,s) + A(t)U(t,s) = 0. \tag{11.17}$$

Differentiating (formally, we need to validate this) (11.16), we have

$$\frac{\partial}{\partial t}U(t,s) = A(s)T_s(t-s) + R(t,s) + \int_s^t A(\tau)T_\tau(t-\tau)R(\tau,s)d\tau,$$

since $T_t(0) = I$. Define

$$R_1(t,s) := (A(t) - A(s))T_s(t-s). \tag{11.18}$$

By (11.16),

$$\frac{\partial}{\partial t}U(t,s) - A(t)U(t,s) = R(t,s) - R_1(t,s) - \int_s^t R_1(t,\tau)R(\tau,s)d\tau,$$

so that, by (11.17), we find the Volterra integral equation

$$R(t,s) = R_1(t,s) + \int_s^t R_1(t,\tau)R(\tau,s)d\tau, \tag{11.19}$$

which defines the perturbation iteration. The variation of constant formula (11.19) is analogous to (6.15) with $T_s(t-s)$ in place of $T(t-s)$ and the perturbation A_1 given by $(A(t) - A(s))$. By the discussion in Section 6.3, it is clear that we need that $R_1(t,s)$ in (11.18) is bounded. By **B3** and (11.15), we can estimate (11.18) by

$$\begin{aligned}
\|R_1(t,s)\| &= \left\|(A(t)-A(s))A(s)^{-1}A(s)T_s(t-s)\right\| \\
&\le \left\|(A(t)-A(s))A(s)^{-1}\right\| \|A(s)T_s(t-s)\| \\
&\le const.\, |t-s|^{\alpha}\, |t-s|^{-1} \le const.\, |t-s|^{\alpha-1}\,,
\end{aligned} \tag{11.20}$$

so that $R_1(t,s)$ is actually a continuous operator in t and s.

By successive substitution we define the formal series solution

$$R(t,s) = \sum_{n=1}^{\infty} R_n(t,s), \tag{11.21}$$

where, for $n \ge 1$,

$$R_{n+1}(t,s) = \int_s^t R_1(t,\tau)R_n(\tau,s)d\tau. \tag{11.22}$$

To solve (11.19) by successive approximations we must prove that the series solution $R(t,s)$, $0 \le s \le t \le T$: i) is continuous in t and s; ii) converge in $\mathcal{L}(X)$ (see Section 6.3); iii) and that the improper integral (11.22) is convergent and continuous in $\mathcal{L}(X)$. We claim that, for $n \ge 1$, it holds the estimate

$$\|R_n(t,s)\| \le \frac{(const.\Gamma_E(\alpha))^n}{\Gamma_E(n\alpha)}(t-s)^{n\alpha-1}, \tag{11.23}$$

which is enough to prove i)-iii). One can prove (11.23) by induction. From (11.20), (11.23) is true for $n = 1$. Let's assume that (11.23) holds for $n > 1$. We have[9]

$$\begin{aligned}
\|R_{n+1}(t,s)\| &\leq \int_s^t \|R_1(t,\tau)\| \, \|R_n(\tau,s)\| \, d\tau \\
&\leq \frac{(const.\Gamma(\alpha))^n}{\Gamma(n\alpha)} const. \int_s^t (t-\tau)^{\alpha-1} (\tau-s)^{n\alpha-1} d\tau \\
&\leq \frac{(const.\Gamma(\alpha))^n}{\Gamma(n\alpha)} const. (t-s)^{(n+1)\alpha-1} \frac{\Gamma(n\alpha)\Gamma(\alpha)}{\Gamma((n+1)\alpha)} \\
&= \frac{(const.\Gamma(\alpha))^{n+1}}{\Gamma((n+1)\alpha)} (t-s)^{(n+1)\alpha-1},
\end{aligned}$$

proving the inductive hypothesis (11.23).

Finally, since $(s-t)^{k\alpha-1} \leq T^{\alpha(k-1)}(s-t)^{\alpha-1}$, by Stirling's formula,

$$\|R(s,t)\| \leq \sum_{n=1}^{\infty} \frac{(const.\Gamma(\alpha))^n}{\Gamma(n\alpha)} T^{\alpha(n-1)} (t-s)^{\alpha-1} \leq const. (t-s)^{\alpha-1}.$$

Thereby, the operator $U(s,t)$ defined in (11.16) is strongly continuous in t and s, $0 \leq s \leq t \leq T$; it also belongs to $\mathcal{L}(X)$, since

$$\begin{aligned}
\|U(t,s)\| &\leq \|T_s(t-s)\| + \int_s^t \|T_\tau(t-\tau)\| \, \|R(\tau,s)\| \, d\tau \\
&\leq const. + const. \int_s^t (\tau-s)^{\alpha-1} d\tau \leq const.
\end{aligned}$$

We refer to [Paz83], theorem 5.6.1, for the (long) completion of the proof of

Theorem 24 *Under* $\mathbf{B1}-\mathbf{B3}$ *there exists a unique propagator* $U(t,s)$, $0 \leq s \leq t \leq T$, *satisfying, for all* $s \leq t$: *i)* $\|U(t,s)\| \leq const.$; *ii)* $U(t,s) : X \to D$, $t \mapsto U(t,s)$ *is strongly differentiable; the derivative* $\frac{\partial}{\partial t}U(t,s) \in \mathcal{L}(X)$ *and it is strongly continuous in* t, *for* $s < t$; *moreover,* $-\frac{\partial}{\partial t}U(t,s) + A(t)U(t,s) = 0$, *for* $s < t$; *iii) for every* $g \in D$, $t \in (0,T]$, $s \mapsto U(t,s)g$ *is differentiable and* $\frac{\partial}{\partial s}U(t,s) + U(t,s)A(s) = 0$, *for* $s \leq t$.

By analogy to the analytic case, [Fri69] calls the solution operator $U(t,s)$ fundamental solution of (11.13). Then we see that, by (11.16) and (11.21), the fundamental solution has the expansion

$$U(t,s) = \sum_{n=0}^{\infty} U_n(t,s), \tag{11.24}$$

where, $U_0(t,s) = T_s(t-s)$ and, for $n \geq 1$,

$$U_{n+1}(t,s) = \int_s^t T_\tau(t-\tau) U_n(\tau,s) d\tau.$$

We remark also that there exists a second parametrix expansion, equivalent to (11.24), formally derived by the adjoint equation to (11.13), i.e., $\frac{\partial}{\partial s}U(t,s)g + U(t,s)A(s)g = 0$, that clearly holds only for $g \in D$ (see also property iii) in the theorem). We refer to [Paz83], page 160-162 for this. We apply and interpret the parametrix expansion (11.24) in the next Subsection.

[9] *const.* denotes a general positive constant, which can be appropriately chosen to satisfy the recursion.

The Cauchy Problem And Interpretation

Consider the abstract Cauchy problem in the Banach space X

$$\begin{array}{c} -\frac{d}{dt}u(t) + A(t)u(t) = -f(t), \quad t > 0 \\ u(0) = g, \end{array} \tag{11.25}$$

where $g \in X$. Under $\mathbf{B1} - \mathbf{B3}$, if f is Holder continuous in $[0, T]$, then (see theorem 2.3.2, [Fri69]) there exists a unique solution of (11.25) given by

$$u(t) = U(t,s)g + \int_s^t U(t,\tau)f(\tau)d\tau,$$

where $U(t,s)$ is the fundamental solution of (11.13) given by (11.24). Since D is dense in X and the solution operator $U(t,s)$ is continuous, the Cauchy problem (11.25) is well posed (see Appendix A.3). This is the most important result of the chapter which proves that our assumptions in the first part of the thesis are largely justified.

The use of the abstract parametrix as a perturbation method is straightforward, but its scope is clearly more limited than the results presented in Chapter 10. Suppose we know in closed-form the operator $T_s(t-s)$, which means that we know in closed-form the transition density of the diffusion associated to $A(s)$, s fixed. Then we can approximate $u(t)$ by $\sum_{n=0}^N u_n(t)$, N fixed, where

$$u_0(t) = T_s(t-s)g + \int_s^t T_\tau(t-\tau)f(\tau)d\tau,$$

and so on, according to the terms in (11.24). The abstract parametrix allows to approximate a time-dependent problem with an autonomous problem: this is also the limit of the method, at least with respect to the classical parametrix, which uses a much simpler auxiliary model. In other words, it is more likely that, if we cannot solve in closed-form (11.25), we could for a completely constant coefficient operator (as in the classic parametrix), rather than if we freeze only the time parameters. Furthermore, given the generality of the abstract result, it is impossible to give an explicit estimate of the truncated error as in (10.27). The advantage of the abstract method is that the assumptions are much weaker, which explains why, in practice, the classical parametrix method works in much greater generality that the assumptions in Chapter 10 suggest.

A possible practical use of the method presented here is in a two stage approximation: in fact often a perturbation technique for PDEs works well for autonomous problems but not for time-dependent problems (e.g., see the results and the problems in Chapters 7, where in order to use the separation of variables method, the generator must be time-independent); so we could first simplify the problem with an autonomous operator, which in turn could be solve by means of a second perturbation technique. This proposal deserves further research.

Differentiability of The Fundamental Solution

We conclude stating a result that can be found in [Fri69], Chapter 2, Part 3. Under $\mathbf{B1} - \mathbf{B3}$ and other strong conditions (which can be relaxed substantially if X is an Hilbert space), then the fundamental solution $U(t,s)$, as given in (11.24), can be extended as an analytic function in (t,s), for $s < t$ (actually the extension can be to a sector of the complex plane, but we are not interested in this). This extension can be defined at $t = s$ such that it is strongly continuous for $0 \leq s \leq t \leq T$. Note that, as in the autonomous case, the sectorial property of the generator allows to say (under other technical conditions) that the propagator is analytic in its arguments, for $s < t$, with the same singularity discussed in Section 11.1 at $t = s$.

Appendix A

Operator Theory And Cauchy Problems

We review and collect here some basic definitions and results that we use in the thesis about: (A.1) Bounded and unbounded linear operators; (A.2) Parabolic PDEs; (A.3) Semigroups of operators.

A.1 Linear Operators

We supply a brief introduction to some elements of functional analysis and operator theory. Some more advanced results are developed and explained in detail in the main text where needed. The main references are [Fri70], [Kat80], [ReSi80], [Yos80].

Operators on Banach Spaces: Definitions

Let X_1, X_2 be Banach spaces with norms $\|.\|_1, \|.\|_2$. An operator is a mapping $B : X_1 \to X_2$. B is called *linear* if $B(\alpha x + \beta y) = \alpha Bx + \beta By \in X_2$, for any $x, y \in X_1$ and $\alpha, \beta \in \mathbb{R}$ or $\mathbb{C}$. All the operators that we consider in the thesis are linear, so we often suppress the adjective. If $X_1 = X_2$, we say that B is an operator in X_1.

The linear space $D(B) \subset X_1$ where B is defined is called the *domain* of B, the subspace of all $x \in D(B)$ such that $Bx = 0$ is called the null space of the operator, while the set $Rg(B) = B(D(B))$ is called the range of B. Note that the domain of a linear operator B is always a linear space, but $D(B)$ need not to be closed and hence a Banach space (with $\|.\|_1$). Often the domain is a dense subset of the given Banach space; in this case the operator is said to be densely defined. Two transformations B_1, B_2 that share the same domain D are said equal if, for any $x \in D$, $B_1 x = B_2 x$. Notice that the same mapping on different domains defines different operators. In particular, if $B_1 x = B_2 x$ on $D(B_1)$ and $D(B_1) \subset D(B_2)$, we say that B_2 extends B_1 (or that B_1 is a restriction of B_2), writing $B_1 \subset B_2$. An operator that does not admit extensions on a given Banach space is said maximal.

As an example, consider $X_1 = X_2 = C[\alpha, \beta]$ with the sup-norm and $B = \frac{d}{dx}$. The domain of B is the subset of all the differentiable functions on $[\alpha, \beta]$, which is only a dense linear subset of $C[\alpha, \beta]$. Therefore $B = \frac{d}{dx}$ is densely defined and maximal in $C[\alpha, \beta]$. Let D_0 be the subset of all the functions $u \in D(B)$ satisfying the boundary condition $u(\alpha) = 0$. The operator $B_0 = \frac{d}{dx}$ in $C[\alpha, \beta]$ with domain D_0 is a restriction of B (but note that D_0 is no longer dense in $C[\alpha, \beta]$). To conclude the example, we notice that the action of the operator and the space of definition are strongly related: e.g., the operator $\frac{d}{dx}$ in $L^p[\alpha, \beta]$, $1 \le p < \infty$, on the domain of all the absolutely continuous functions, acts as a weak derivative.

We stress that when we speak about an operator we should always define the Banach space in which it is defined and specify its domain. We say that a property concerning an operator holds *formally* if we miss to specify the domain and/or the action of the operator. Then we must check if there exists a version of the operator and a domain, if any, such that the property holds.

An operator B is said *bounded* if, for some $c \geq 0$, $\|Bx\|_2 \leq c\|x\|_1$ for all $x \in X_1$. The norm of B, written $\|B\|$, is the smallest $c \in \mathbb{R}_+$. If we cannot find this bound, i.e., for some $x \in X_1$, $c = +\infty$, the operator is said *unbounded*. We have that

$$\|B\| = \sup_{\|x\|_1=1} \|Bx\|_2 = \sup_{x\neq 0} \frac{\|Bx\|_2}{\|x\|_1}. \tag{A.1}$$

The differential operator $\frac{d}{dx}$ in $C[\alpha,\beta]$ is clearly unbounded, for $\|u'\|_1$ could be arbitrarily large for $\|u\|_1 = 1$. The same is true for its restriction B_0. Actually, we could make $\frac{d}{dx}$ bounded regarding it as a map $C^1[\alpha,\beta] \to C[\alpha,\beta]$ (with the natural norms), but this could be a good idea only if we are given a particular operator. When we consider a broad class of operators *in* a given Banach space, it turns out that differential operators are typically unbounded. However, there are large classes of bounded operators. All the operators in finite-dimensional spaces are bounded. Other examples are projections operators or unitary operators in Hilbert spaces (for a complete picture, see [Kat80]). For our sake, it is important to notice that integral operators are often bounded. For instance, the inverse of B_0 on $D(B_0^{-1}) = Rg(B_0) = C[\alpha,\beta]$, given by $B_0^{-1}u(x) = \int_\alpha^x u(y)dy$, is an everywhere defined bounded integral operator[1].

Some Theorems For Bounded Operators

Let X_1, X_2 be Banach spaces as in the previous sections. Since X_1, X_2 are complete metric spaces, we can wonder when the mapping B is continuous at the point $x \in D(B)$, i.e., $\|x_n - x\|_1 \to 0$ and $\{x_n\} \subset D(B)$, only if $\|Bx_n - Bx\|_2 \to 0$. The answer is that the following are equivalent:

- B is continuous at one point;
- B is continuous on its domain;
- B is bounded.

Bounded operators not only are the only continuous operators, but also enjoy special nice properties (for all the theorems about bounded operators see, e.g., [Yos80]). For our ends, the most important theorems are:

- Bounded linear transformation (B.L.T.) theorem. *Let B be a bounded mapping on a normed linear space $(D(B) \subset X_1, \|.\|_1)$ to a Banach space X_2. It is always possible to extended B to a bounded transformation $\widetilde{B}$, with the same norm c, on the completion of $(D(B), \|.\|_1)$, i.e., such that $D(\widetilde{B}) = X_1$*[2].
- Principle of uniform boundedness (Banach-Steinhaus theorem). *Let X_1 be a Banach space and X_2 be a normed linear space. Let $\mathcal{B}$ be a family of bounded operators from X_1 to X_2. If for each $x \in X_1$ the set $\{\|Bx\|_2 : B \in \mathcal{B}\}$ is bounded, then $\{\|B\| : B \in \mathcal{B}\}$ is bounded.*

Another important theorem for bounded operators is the Closed graph theorem. An operator $B : D(B) \subset X_1 \to X_2$ is said closed if its graph, i.e., the set of points $G_B := (x, Bx) \subset X_1 \times X_2$, is a closed subset of the Cartesian product $X_1 \times X_2$. First of all note that if B is one-to-one, its inverse B^{-1} has closed graph as well, therefore, B^{-1} is closed if, and only if, B is closed. Notice

[1] Note that the maximal operator $\frac{d}{dx}$ is not invertible, for $\frac{d}{dx}u = 0$ if $u = const.$.

[2] Hence, if B is bounded, we can directly consider its extension as defined on the whole Banach space. The B.L.T. theorem can be seen as a particular case of the Hahn-Banach theorem (see [Roy88]).

also that the graph of an operator is always a linear subspace of $X_1 \times X_2$, so that an operator is closed if, and only if, its graph is a closed linear subspace. Hence, B is closed if, and only if, $\{x_n\} \subset D(B)$, $\|x_n - x\|_1 \to 0$ and $\|Bx_n - y\|_2 \to 0$, $y \in X_2$, imply $x \in D(B)$ and $Bx = y$. In other words, limits of Cauchy sequences in $D(B)$ are in $D(B)$ and the convergence is preserved under the operator.

The natural question is about the relations between continuity and closedness. We have

- *A continuous operator B is closed if, and only if, $D(B)$ is closed. In particular, every everywhere defined continuous operator is closed.*

- Closed graph theorem. *A closed operator B with domain $D(B) = X_1$ is continuous.*

Note that the maximal operator $\frac{d}{dx}$ in $C[\alpha, \beta]$ is closed (because the limit of a sequence of derivable functions is derivable). From the Closed graph theorem $\frac{d}{dx} : C^1[\alpha, \beta] \to C[\alpha, \beta]$ is continuous. This trick works for any closed operator which domain can be made a Banach space. We draw the additional conclusion that, when the Banach space X is given, a closed unbounded operator in X is typically at most densely defined.

Topologies on Bounded Operators

In this section we review some formalism to study the convergence of sequences of bounded operators and the continuity of operator valued functions.

The Banach spaces X_1, X_2 are linear spaces, so we can perform addition and scalar multiplication. Let B_1, B_2 be not necessarily bounded operators on a common domain D dense in X_1 to X_2 and $\alpha \in \mathbb{C}$. Define $(B_1 + B_2)\, x := B_1 x + B_2 x$ and $(\alpha B_1)\, x := \alpha B_1 x$. Then the space of all linear operators on D is a linear space. However, to have more structure on this linear space, mainly due the fact that unbounded operators are only densely defined, we have to restrict ourselves to continuous operators (which we suppose everywhere defined). Let $\mathcal{L}(X_1, X_2)$ ($\mathcal{L}(X_1)$ if $X_1 = X_2$) be the linear subspace of all bounded operators on X_1 to X_2 endowed with the operator norm $\|.\|$ in (A.1). The main result is that, provided (as here) X_2 is complete, $\mathcal{L}(X_1, X_2)$ is a Banach space.

We can introduce a natural topology on $\mathcal{L}(X_1, X_2)$. A sequence of bounded operators $\{B_n\}$ is said *uniformly convergent* if there is a bounded operator B such that $\|B_n - B\| \to 0$. The topology induced on $\mathcal{L}(X_1, X_2)$ by this convergence is called *uniform (operator) topology* and in this topology the composition product mapping (let X_3 be another Banach space and B_1 be onto X_2) $(B_1, B_2) : \mathcal{L}(X_1, X_2) \times \mathcal{L}(X_2, X_3) \to \mathcal{L}(X_1, X_3)$, given by $(B_1, B_2) := B_2\,(B_1 x) = B_2 B_1 x \in X_3$, $x \in X_1$, is jointly continuous. If $X_i = X$, $i = 1, 2, 3$, then the composition product makes $\mathcal{L}(X)$ a Banach algebra (e.g., if $X = \mathbb{R}^d$, then $\mathcal{L}(X)$ can be identified with the algebra of the $(d \times d)$ matrices). On this algebra we can define the so called *analytic functional calculus.* Let $f(x) = \sum_{n=0}^{\infty} \alpha_n x^n$ be a complex analytic function with ray of convergence r. For any $B \in \mathcal{L}(X)$ with $\|B\| < r$, the power series $\sum_{n=0}^{\infty} \alpha_n B^n$ converges in $\mathcal{L}(X)$[3], so it is natural to set $f(B) := \sum_{n=0}^{\infty} \alpha_n B^n$.

There are other two topologies on $\mathcal{L}(X_1, X_2)$ weaker than the uniform one. The *strong operator topology* is the weakest topology on $\mathcal{L}(X_1, X_2)$ such that all the maps $C_x : \mathcal{L}(X_1, X_2) \to X_2$, given by $C_x(B) := Bx$, are continuous for all $x \in X_1$. In this topology a sequence of bounded operators $\{B_n\}$ converges to $B \in \mathcal{L}(X_1, X_2)$ if, and only if, $\|B_n x - Bx\|_2 \to 0$ for all $x \in X_1$. The uniform convergence implies the strong convergence, but the composition product mapping is only separately continuous in the strong operator topology. This notion of convergence is useful since any strongly convergent sequence of operators $\{B_n\}$ has a strong limit.

The third operator topology needs the concept of dual (or adjoint) space. The dual space of the Banach space X is defined as $X^* = \mathcal{L}(X, \mathbb{C})$. The Hahn-Banach theorem assures that

[3] We have that $\|\sum B_n\| \leq \sum \|B_n\|$ and $\left\|B^{n+m}\right\| \leq \|B^n\| \, \|B^m\|$, so $\|B^n\| \leq \|B\|^n$.

the adjoint space always contains "sufficiently many" functionals, in the sense that if $x, y \in X$ are such that $x \neq y$, then there exists an $l \in X^*$ such that $l(x) \neq l(y)$. In other words, in X^* there are enough elements to distinguish the elements of X. Therefore, one can define a topology on $\mathcal{L}(X_1, X_2)$, called weak operator topology, as the weakest topology such that the maps $C_{x,l} : \mathcal{L}(X_1, X_2) \to \mathbb{C}$, given by $C_{x,l}(B) := l\,(Bx)$, are continuous for all $x \in X_1$, $l \in X_2^*$. This topology is weaker than the strong operator topology, but the composition product is still separately continuous. In this topology a sequence of bounded operators $\{B_n\}$ converges to B if, and only if, $|l\,(B_n x) - l\,(Bx)| \to 0$ for all $x \in X_1$, $l \in X_2^*$.

Commutativity

On the algebra $\mathcal{L}(X)$ is naturally defined the notion of commutativity. We say that $B_1, B_2 \in \mathcal{L}(X)$ commute if $B_1 B_2 = B_2 B_1$. It is very difficult to extend this definition to unbounded operators (e.g., see [ReSi80], vol.1, chapter 8), because generally unbounded operators are not everywhere defined and the intersection of their domains could be 0 alone; but this cannot happen if one of the operators is bounded (and everywhere defined). Let $B \in \mathcal{L}(X)$ and A be an operator in X. We say that A and B commute if $BA \subset AB$, i.e., for each $x \in D(A)$, we have $Bx \in D(A)$ and $BAx = ABx$.

Closable And Adjoint Operators

An operator B from X_1 to X_2 is said to be *closable* if B has a closed extension. The natural characterization of closedness is in terms of sequences

$$(x_n) \subset D(B), x_n \to 0, Bx_n \to x \quad \Rightarrow \quad x = 0.$$

The closure of B, denoted $\overline{B}$, is the smallest closed extension of B. The graph of the closure is $G_{\overline{B}} = \overline{G_B}$ in $X_1 \times X_2$. If B is closed, the domain $D(A)$ of any closable operator A with closure $\overline{A} = B$ is called a *core* for B. We have that the set of elements $(x, Bx) \subset X_1 \times X_2$, $x \in D(A)$, is dense in G_B. So any closed operator could always be recovered from a restriction on a core, i.e., to study a closed operator it is enough to find a core and study the restriction.

A second criterion requires the concept of adjoint operator. Let $B : X_1 \to X_2$ be densely defined. The (Banach) adjoint B^* of B is the (unique) operator $X_2^* \to X_1^*$, with domain $D(B^*) := \{l \in X_2^* : \text{there is a } g \in X_2^* \text{ such that } l\,(Bx) = g(x), \text{ for all } x \in D(B)\}$, defined by $(B^* l)\,(x) := g(x)$, $x \in D(B)$. The adjoint is always closed, but could be trivial (i.e., $D(B^*) = 0$). We have:

Let B and B^* be adjoint to each other (i.e., $(B)^* = B^*$ and $(B^*)^* = B$): if one is densely defined, the other is closable.

In particular, if X_1, X_2 are reflexive (i.e., a Banach space that can be identified with its second adjoint space), for a closable and densely defined operator B, its adjoint B^* is closed, densely defined and $B^{**} = \overline{B}$.

Spectra And Resolvent

Let B be a closed operator in a Banach space X with domain $D(B)$. The spectral theory studies the properties of the operator $(\lambda I - B)$ as λ varies in $\mathbb{C}$, where I is the identity operator on X. Note that the domain of $(\lambda I - B)$ is $D(B)$. The *resolvent set* of B is given by $\rho(B) := \{\lambda \in \mathbb{C} : (\lambda I - B) \text{ is bijective}\}$. For $\lambda \in \rho(B)$, it is defined the operator

$$R(\lambda, B) := (\lambda I - B)^{-1}, \tag{A.2}$$

which is, by the Closed graph theorem ((A.2) is closed and everywhere defined), continuous. The operator valued function $\lambda \mapsto R(\lambda, B)$ is called the *resolvent* of B at $\lambda \in \rho(B)$. We refer to [Yos80] or [EnNa99] for the properties of the resolvent. Here we notice that $\rho(B)$ is always an open subset of $\mathbb{C}$ (while if $\rho(B) \neq \emptyset$, then B is closed, but not vice versa). For all $\lambda, \mu \in \rho(B)$ it holds the important *resolvent equation*

$$R(\lambda, B) - R(\mu, B) = (\lambda - \mu) R(\lambda, B) R(\mu, B). \tag{A.3}$$

If $\rho(B)$ is not empty we have a useful criterion for the commutativity: in order that B and $A \in \mathcal{L}(X)$ commute, it is necessary that $R(\lambda, B)A = AR(\lambda, B)$ for all $\lambda \in \rho(B)$ and it is sufficient that the same condition holds for some $\lambda \in \rho(B)$.

The resolvent operator is an extremely powerful theoretical tool, but it is rare to know it in closed-form. To study the qualitative feature of the closed operator B it is easier to consider the complement in $\mathbb{C}$ of the resolvent set, i.e., $\sigma(B) := \mathbb{C} \backslash \rho(B)$, the so-called *spectrum*. $\sigma(B)$ is a closed set, but to say more, if B is an unbounded operator, one has to consider the particular operator. For example, the operator $B = \frac{d}{dx}$ in $C[\alpha, \beta]$ has spectrum $\sigma(B) = \mathbb{C}$, because $(\lambda I - B)$ is not invertible for all $\lambda \in \mathbb{C}$. On the other hand, its restriction B_0 is such that $(\lambda I - B_0)$ is bijective for all λ and so B_0 has an empty spectrum. If instead $B \in \mathcal{L}(X)$, we can say that $\sigma(B)$ is compact and non-empty; in particular, the spectral radius (i.e., $\sup\{|\lambda| : \lambda \in \sigma(B)\}$) is finite and less than $\|B\|$.

For a general closed operator B, we notice that the spectrum is possibly quite heterogeneous, in function of the behavior of $(\lambda I - B)$ at $\lambda \in \sigma(B)$: many possible decompositions of the spectrum are possible, here we limit ourselves to the most important subset. The totally of points $\lambda \in \sigma(B)$ such that $(\lambda I - B)$ is not invertible is called the *point spectrum* $P\sigma(B)$ of B. Note that if $P\sigma(B) \neq \emptyset$, it does not need to be a discrete (possibly infinite) subset of $\mathbb{C}$; for instance, the maximal operator $\frac{d}{dx}$ in $C[\alpha, \beta]$ has $P\sigma\left(\frac{d}{dx}\right) = \mathbb{C}$. This assertion follows from the fact that, for a closed operator B in X, $\lambda_0 \in P\sigma(B)$ if, and only if, the eigenproblem $Bx = \lambda_0 x$ has a solution $0 \neq x \in X$. In this case λ_0 is called an eigenvalue of B and x the corresponding eigenelement. The null space of $(\lambda_0 I - B)$ is the eigenspace of (B, λ_0) and its dimension is the multiplicity of the eigenvalue λ_0. Back to the example, we have $P\sigma(d/dx) = \mathbb{C}$, because the equation $u'(x) = \lambda_0 u(x)$ has a non-trivial solution $u(x) = const. \exp(\lambda_0 x)$ for all λ_0.

The best situation is when the spectrum is composed only of isolated eigenvalues (that of course are in $P\sigma(B)$), but even self-adjoint operators in Hilbert spaces need not to have only a purely discrete (infinite) spectrum of eigenvalues. To account the fact that it is much easier to deal with isolated eigenvalues rather than with a general element of the spectrum, we use sometimes the following disjoint decomposition of the spectra[4]: $\sigma(B) = \sigma_{disc}(B) \cup \sigma_{ess}(B)$, where $\sigma_{disc}(B)$ is called discrete spectrum and is the set of all isolated eigenvalues of finite multiplicity of B; $\sigma_{ess}(B)$, called essential spectra, is defined by $\sigma_{ess}(B) = \sigma(B) \backslash \sigma_{disc}(B)$. We use a "large" definition of $\sigma_{ess}(B)$, so that any "continuous" (this term should be made precise) part of the spectrum is in $\sigma_{ess}(B)$; in particular any accumulation point of $P\sigma(B)$ (e.g., eigenvalues of infinite multiplicity or a limit point of $P\sigma(B)$).

Hilbert Spaces

Let X be a Hilbert space (usually on $\mathbb{R}$, but sometimes is necessary to use the field $\mathbb{C}$) and $\langle ., . \rangle$ be its inner (or scalar) product. For simplicity, we consider here only operators in X, everything extends if the domain and the range are in different Hilbert spaces. Any Hilbert space is a Banach space with norm $\|x\| := \langle x, x \rangle^{1/2}$. Every Banach space such that its norm $\|.\|$ satisfies the parallelogram law, $\|x + y\| + \|x - y\| = 2(\|x\| + \|y\|)$, are Hilbert spaces with inner product given by $\langle x, x \rangle := \|x\|^2$. For instance, $L^2[\alpha, \beta]$ with $\|f\| = \left(\int_\alpha^\beta |f(x)|^2 \, dx\right)^{1/2}$ is a Hilbert space with $\langle f, g \rangle = \int_\alpha^\beta f(x)\overline{g(x)} dx$.

Hilbert spaces enjoy a geometric structure that comes from the inner product. $x, y \in X$ are said orthogonal if $\langle x, y \rangle = 0$. Two sets $M, N \subset X$ are *orthogonal* if for each pair $(x, y) \in M \times N$,

[4]Notice that there is not uniformity in the literature in the use of the following terms (for a discussion, see [Kat80]). The definition given here is the one used in the Sturm-Liouville theory, see [Zet05].

x and y are orthogonal. We write $M \perp N$. The set $M^\perp$ of all $x \in X$ orthogonal to $M \subset X$ is called orthogonal complement. If $M \perp N$, then $N \subset M^\perp$. Note that if M is a closed subspace of X, then M and $M^\perp$ are Hilbert spaces with the original scalar product. In this case the Projection theorem says that any $x \in X$ can be uniquely decomposed in $x = y + z$, where $y \in M$ and $z \in M^\perp$. With this theorem one can prove that any Hilbert space X is self-adjoint and reflexive, i.e., $H^{**} = H^* = H$.

A second special feature of Hilbert spaces is that any complete orthonormal family plays pretty much the role of a basis in an Euclidean space. A collection $\{x_i\} = K \subset X$ is said *orthonormal* if $\langle x_i, x_i \rangle = 1$ for all i, and $\langle x_i, x_j \rangle = 0$ for $i \neq j$. For any $x \in X$, the scalars $\langle x, x_i \rangle$ are called the (Fourier) coefficients of x with respect to $\{x_i\}$. The family $\{x_i\}$ is said complete if it spans X (equivalently if $K^\perp = 0$). A complete orthonormal family in X is called an orthonormal basis of X. We have the following remarkable facts:

- *Let K be an orthonormal basis of X. For any $x \in X$, $\langle x, x_i \rangle = 0$ for all but a countable number of elements in K. Call $K_x = \{x_i \in K : \langle x, x_i \rangle \neq 0\}$. We have*

$$x = \sum_{y \in K_x} \langle x, y \rangle \, y, \tag{A.4}$$

  and the series (A.4) converges independently of the order of summation.

- *Every Hilbert space has an orthonormal basis.*

- *Any orthonormal basis in a separable Hilbert space is countable.*

We note also that if $\{x_n\} \subset X$ is an orthonormal sequence and $\{\alpha_n\}$ a sequence of scalars (in $\mathbb{R}$ or $\mathbb{C}$), then the series $\sum_n \alpha_n x_n$ is convergent if, and only if, $\sum_n |\alpha_n|^2 < \infty$, and in that case

$$\left\| \sum_n \alpha_n x_n \right\| = \left(\sum_n |\alpha_n|^2 \right)^{1/2} \tag{A.5}$$

and the sum is independent of the order of summation. When α_n are Fourier coefficients of a function $x \in X$, equation (A.5) is called Parseval's formula.

Operators in Hilbert Spaces

Operators in Hilbert spaces have special features. As always the main distinction is between bounded and unbounded operators, but here we present directly the results about unbounded operators.

Let B be an operator in the Hilbert space X. If B is densely defined we can define the Hilbert space adjoint B^* of B. Let $D(B^*)$ the set of $y \in X$ for which there exists a $z \in X$ with $\langle Bx, y \rangle = \langle x, z \rangle$, for all $x \in D(B)$. For each such $y \in D(B^*)$ define $B^* y = z$. The density of $D(B)$ ensures that z is uniquely defined. Note that the Banach and Hilbert adjoint are related, but the two notions are different (see [Kat80] for a discussion). If $D(B)$ is dense in X we have

- B^* is closed.
- B is closable if, and only if, $D(B^*)$ is dense, in which case $\overline{B} = B^{**}$.
- If B is closable, $\left(\overline{B}\right)^* = B^*$.

A densely defined operator B in X is called *symmetric* if $B \subset B^*$. From the definition of Hilbert adjoint, B is symmetric if, and only if, $\langle Bx, y \rangle = \langle x, By \rangle$, for all $x, y \in D(B)$. From the definition follows that a symmetric operator is such that $\langle Bx, x \rangle$ is real.

If B is symmetric and also $B^* \subset B$, i.e., $B = B^*$, B is said *self-adjoint*. There are some important remarks: 1) a symmetric operator B is closable, because $D(B^*)$ is dense by definition; 2) if B is symmetric, B^* is closed and therefore is a closed extension of B and of the closure of B: $B \subset B^{**} \subset B^*$; 3) if B is closed, $B = B^{**} \subset B^*$, i.e., any closed operator (in a Hilbert space) is symmetric; 4) for a self-adjoint operator $B = B^{**} = B^*$, i.e., a closed symmetric operator is self-adjoint if, and only if, B^* is symmetric. The distinction between symmetric and self-adjoint operators is fundamental since only for the latter the spectral theorem holds. For the criteria to check if a symmetric operator is self-adjoint, we defer to [ReSi80] Vols.1,2. Other conditions are given in [Kat80] (e.g., see section 5.3.7).

Spectral Theory

Self-adjoint operators enjoy good spectral properties. An example of self-adjoint operator is given by any (orthogonal) projection operator: an operator $P \in \mathcal{L}(X)$ such that $P^2 = P$ and $P = P^*$. P acts as the identity on its range and as the 0-operator on $Rg(P)^\perp$. If, by the projection theorem, $x = y + z$, with $y \in Rg(P)$ and $z \in Rg(P)^\perp$, then $Px = y$. For instance, if $\{x_i\}$ is an orthonormal family, then the operator $Ex = \sum_{y \in K_x} \langle x, y\rangle\, y$ is a projection on the closed linear subspace spanned by $\{x_i\}$.

The version of the spectral theorem we present uses the so called spectral projections (or resolution of the identity). Recall that the spectrum of a self-adjoint operator is a subset of the real axis (a closed interval if the operator is bounded, an unbounded subset if unbounded). Each isolated point is an eigenvalue and eigenfunctions associated to different eigenvalues are orthogonal (so if the spectrum of a self-adjoint operator on a separable Hilbert space is a countable point spectrum, we have found a basis for the Hilbert space), but the eigenvalues need not to be isolated points of the spectrum and the spectrum to be discrete.

The spectral projections of B, self-adjoint, are a family of projections $P_\Omega(B)$, for Ω any Borel set of $\mathbb{R}$, defined by $P_{(\lambda-\epsilon,\lambda+\epsilon)}(B) \neq 0$ if, and only if, $\lambda \in \sigma(B)$. $\{P_\Omega\}$ has the properties: 1) each P_Ω is a projection (note that if in $\Omega \cap \sigma(B)$ there are only a finite number of eigenvalues, then P_Ω spans a finite dimensional space); 2) $P_\emptyset = 0$, $P_\mathbb{R} = I$ and $P_{\Omega_1} P_{\Omega_2} = P_{\Omega_1 \cap \Omega_2}$; 3) if $\Omega = \cup_n \Omega_n$, $\Omega_n \cap \Omega_m = \emptyset$, $n \neq m$, then $\sum_{n=1}^N P_{\Omega_n} \overset{N\to\infty}{\to} P_\Omega$ in the strong operator topology.

Let B be self-adjoint and $\{P_\Omega\}$ be its spectral projections. Denote $dP_\lambda = \langle x, P_\Omega x\rangle$ for any $x \in X$. dP_λ is a Borel measure on $\mathbb{R}$. For a real-valued Borel measurable function g define the *measurable functional calculus*

$$g(B) = \int_\mathbb{R} g(\lambda) dP_\lambda, \tag{A.6}$$

where $g(B)$ means $\langle x, g(B)x\rangle$. The spectral theorem says that, if X is separable, there is a one-to-one correspondence between self-adjoint operators B and spectral projections $\{P_\Omega\}$ given by $B = \int_\mathbb{R} \lambda dP_\lambda$. From the spectral theorem it follows that $\int_\mathbb{R} \lambda dP_\lambda = \int_{\sigma(B)} \lambda dP_\lambda$ and $D(B) = \left\{x \in X : \int_\mathbb{R} \lambda^2 dP_\lambda < \infty\right\}$. Furthermore, if g is a real-valued Borel measurable function, then $g(B)$ is self-adjoint and the spectrum of $g(B)$ is given by $g(\sigma(B))$. The best reference for the spectral theorem, in its many guises, is [ReSi80], vol.1.

A.2 PDEs of Parabolic Type

We introduce the concepts of second order linear parabolic PDEs, Cauchy problems and well posedness. General references are [Eva98], [Pol02], [Sal04] and [Wei65]. We give some intuition by means of the heat equation, which main reference is [Wid75]. We present also some result about fundamental solutions of Cauchy problems, from [Fri64] and [IKO62].

Elliptic And Parabolic Differential Operators

Second order (from the order of the highest derivative) linear PDEs are classified in three major subclasses, parabolic (e.g., in two variables, the heat equation $-u_t + u_{xx} = 0$), hyperbolic (e.g., the wave equation $-u_{tt} + u_{xx} = 0$) and elliptic (e.g., the Laplace equation $u_{xx} + u_{yy} = 0$). We consider the first and third classes, that are strongly linked together. The important fact in these simple examples are the signs, that tell us that in the first two cases the role of t and x are different, while in the third x and y have the same role. t has the role of temporal variable (running forward) and it is confined in a bounded interval $[0, T]$, $T > 0$. x, y are spatial variables and their number depends on the particular problem. So we pose a general problem in the strip $H = \mathbb{R}^d \times [0, T]$.

The general second order linear parabolic differential operator defined on the strip H, which means that all the coefficients appearing in the operators are real functions on H, is given by the sum of three operators with peculiar roles. Let us write a parabolic operator as

$$L_c(t) := -\frac{\partial}{\partial t} + A(t) - c(x,t). \tag{A.7}$$

- $-\frac{\partial}{\partial t}$, a first order differential operator, tells us that $L_c(t)$ describes a system in evolution. ∂_t and $A(t)$ have opposite signs[5] since we are assuming that the information comes from $t = 0$.

- The second operator is given by

$$A(t) := \sum_{i=1}^{d} \mu_i(x,t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{A.8}$$

  so that $A(t)$ is a second order linear differential operator. $A(t)$, that acts only on the space variables, depends on t because its coefficients $\mu_i(x,t)$ (called *drift* or convection) and $a_{ij}(x,t)$ (*diffusion coefficients*, here the leading coefficients) are function of t. If the coefficients are independent of t, $A(t) = A$ is said autonomous and A describes a system in steady state. The classification of $A(t)$ depends on the behavior of the leading coefficients. Let the field of matrices $a(x,t) := (a_{ij}(x,t))$ in H be symmetric. For fixed t, $A(t)$ is said *elliptic* at the point x if, for (x,t) fixed, $\sum_{i,j=1}^{d} a_{ij}\xi_i\xi_j > 0$, for any $\xi \in \mathbb{R}^d$, $\xi \neq 0$. In this case $L_c(t)$ is *parabolic* at (x,t). If, for any $t \in [0,T]$, $A(t)$ is elliptic at all points $x \in \mathbb{R}^d$ (i.e., elliptic in $\mathbb{R}^d$), we say that $L_c(t)$ is parabolic in H.

- $c(x,t)$ is a multiplicative operator and is called *reaction* (or potential term). It represents a source (if $c(x,t) < 0$, a sink otherwise) proportional to the state of the system at (x,t).

In the applications the class of elliptic (and parabolic) operators could be too vast or too narrow. We can ask $A(t)$ to be *uniformly* elliptic in H: suppose there exist two constants $0 < m < M < \infty$ (independent of (x,t)) such that, for any $\xi \in \mathbb{R}^d$, $\xi \neq 0$,

$$m\,|\xi|^2 \leq \sum_{i,j=1}^{d} a_{ij}(x,t)\xi_i\xi_j \leq M\,|\xi|^2\,, \tag{A.9}$$

for all $(x,t) \in H$, where $|\xi|$ is the Euclidean norm. This ensures that $a(x,t)$ is invertible for all $(x,t) \in H$. Let $(a^{ij}(x,t))$ be the inverse of $a(x,t)$. If the minimum and the maximum of the Euclidean norms of the field of matrices $a(x,t)$ (i.e., the min and the max over H of the eigenvalues of $a(x,t)$) are finite, we can take these as m and M. Under this technical assumption we have that, for any $\xi \in \mathbb{R}^d$, $\xi \neq 0$,

[5] Nothing changes if we change *all* the signs in the presentation.

$$\frac{|\xi|^2}{M} \leq \sum_{i,j=1}^{d} a^{ij}(x,t)\xi_i\xi_j \leq \frac{|\xi|^2}{m}. \tag{A.10}$$

(A.10) is the reason for considering an uniformly elliptic operator. On the other hand, we can extend the class of elliptic operators allowing $a(x,t)$ to be only semi-definite positive for some (x,t). Then the operators $A(t)$ and $L_c(t)$ are called *degenerate*. In this case the advantage is that we no longer need to keep the roles of x and t separate. In other words, we can append t to the space variables. Therefore if $A(t)$ in (A.7) is elliptic or degenerate, we can always see the operator $L_c(t)$ as a degenerate operator on some space of functions from $\mathbb{R}^{d+1}$ to $\mathbb{R}$.

We sometimes use other two operators related to $L_c(t)$ and $A(t)$ in (A.7). Let $a_{ij} \in C^{2,0}(H)$ and $\mu_i \in C^{1,0}(H)$, then (following [Fri64]) the formal *adjoint* to $L_c(t)$ is the unique second order linear parabolic operator $L_c^*(t)$ that solves the equation $\int_H [vL_c(t)u - uL_c^*(t)v]\, dxdt = 0$, for $u, v \in C_0^\infty(H)$. It turns out that $L_c^*(t) = \frac{\partial}{\partial t} + A^*(t) - c(x,t)$, where $A^*(t)$ is given by $A^*(t)u = -\sum_i \frac{\partial(\mu_i u)}{\partial x_i} + \frac{1}{2}\sum_{i,j}\frac{\partial^2(a_{ij}u)}{\partial x_i \partial x_j}$. From the definition (inverting the roles of $L_c(t), L_c^*(t)$) we have $L_c^{**}(t) = L_c(t)$. Notice that, for the heat operator (in d space variables) $-\partial_t + \alpha\sum_i \partial_{x_i x_i}$, $\alpha > 0$, the change of variable $t \mapsto -t$ (but also $t \mapsto T - t$) produces the adjoint operator. This is clearly not true for a general $L_c(t)$.

Homogeneous PDEs And Classical Solutions

A *homogeneous* parabolic PDE is given by the evolution equation[6]

$$(L_c(t)u)(x,t) = 0, \tag{A.11}$$

for $(x,t) \in H$. For (A.11) to make sense, we must specify the class of function u belongs to and the action of $L_c(t)$. We say that u is a *classical* solution of (A.11) if $u \in C^{2,1}(H)$ and satisfies the equation; we defer to the next Appendix any consideration on the domain of $L_c(t)$. As is well known, a differential equation could admit an infinite number of classical particular solutions (in particular, (A.11) has always the trivial solution $u \equiv 0$). Since $L_c(t)$ is linear, we can combine the particular solutions of the homogeneous equation by means of the so-called *superposition principle*: if u_0 and u_1 solve (A.11) in the classical sense, then also $\alpha u_0 + \beta u_1$, $\alpha, \beta \in \mathbb{R}$, solves the same equation. This is obvious, but this principle extends (formally) not only to series, but also to linear combination of any number of particular solutions: for instance, if $(u(x,t;\theta), \theta \in (\alpha,\beta))$ is a family of solution, then for any suitable function $\varphi(\theta)$, the function

$$\int_\alpha^\beta u(x,t;\theta)\varphi(\theta)d\theta, \tag{A.12}$$

is a formal solution of (A.11). To exploit fully the possibility of build solutions from available ones, we first look for solutions with special properties.

Fundamental Solutions

To understand the definition of fundamental solution of an uniformly parabolic operator (A.7) in H, it is better if we first consider the heat equation $-u_t + u_{xx} = 0$ and look for a similarity solution (i.e., a solution suggested by the symmetry of the problem). Note that if $u(x,t)$ solves the heat equation, then also the parabolically dilated function $w(x,t) = \alpha u(\alpha x, \alpha^2 t)$ solves the equation for any $\alpha > 0$. If we choose $\alpha = t^{-\frac{1}{2}}$ and we set $y = x/t^{\frac{1}{2}}$, we have that $w(x,t) = t^{-\frac{1}{2}}v(y)$. Inserting this into the heat equation we find $-w_t + w_{xx} = t^{-\frac{3}{2}}[v'' + yv'/2 + v/2]$, so that it must be $v'' + yv'/2 + v/2 = 0$. If we prescribe the behavior of v on the infinite boundaries by

[6] Any PDE involving time derivatives is called evolution equation.

$v(\pm\infty) = 0$, then the solution is $v(y) = Const. \exp(-y^2/4)$. To specify the constant we need an extra condition and we require that $\int_{\mathbb{R}} v(y)dy = 1$: so $const. = (4\pi)^{-\frac{1}{2}}$. All in all, we have that $w(x,t) = (4\pi t)^{-\frac{1}{2}} \exp(-x^2/(4t))$ is a solution of the heat equation. These reasonings extend to any uniformly elliptic constant coefficients operator $-\partial_t + \frac{1}{2}\sum_{i,j} a_{ij}\partial_{x_i x_j}$, provided the matrix $a = (a_{ij}) \in \mathbb{R}^{d\times d}$ is positive define. Let $\left(a^{ij}\right) = a^{-1}$. We have that the similarity solution for $-\partial_t + \frac{1}{2}\sum_{i,j} a_{ij}\partial_{x_i x_j}$ is also explicit and given by

$$w(x,t) = (2\pi t)^{-\frac{d}{2}} \left(\det a\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2t}\sum_{ij} a^{ij} x_i x_j\right). \tag{A.13}$$

The solution (A.13) has the following properties:

- Invariance with respect to translations, i.e., for any $\xi \in \mathbb{R}^d$, $0 < \tau < t$ fixed, $w(x-\xi, t-\tau)$ is a solution in the (forward) variables (x,t). The symmetry is perfect for the space variable (called radial symmetry). Furthermore, $w(x-\xi, t-\tau)$ is a solution of the adjoint equation in the (backward) variables (ξ, τ).
- $w(x,t)$ is non-negative for all (x,t), bounded for all $t \in (0,T]$ and have an integrable singularity (see [Fri64]) at $(0,0)$. Actually, we have that $w(x,t) \to 0$ as $|x| \to \infty$ ($t > 0$ fixed) and (formally, actually δ acts on the test functions $C_0^\infty(\mathbb{R}^d)$) $w(x,t) \to \delta(x)$ as $t \to 0$. Hence, in particular[7], $\int_{\mathbb{R}^d} w(x,t)dx = 1$.

The results about the similarity solution of the heat equation suggest the following definition under the hypothesis that $L_c(t)$ is uniformly parabolic in H (so that the diffusion matrix is invertible in H). A fundamental solution[8] of (A.11) in H is a function $\Gamma(x,t;y,s)$, for $0 \le s < t \le T$ and $x, y \in \mathbb{R}^d$, that solves $L_c(t)u = 0$ in the variables (x,t) (i.e., it solves the equation on $H_0 := \mathbb{R}^d \times (0,T]$) and, for any continuous function $\varphi(x)$ bounded by $O\left[\exp\left(h\left|x\right|^2\right)\right]$, where $0 < h < (2MT)^{-1}$, Γ is such that

$$\lim_{t\downarrow s} \int_{\mathbb{R}^d} \Gamma(x,t;y,s)\varphi(y)dy = \varphi(x). \tag{A.14}$$

We sketch the involved construction of the fundamental solution for $L_c(t)$ when we present the parametrix method (see Chapter 10). We collect here for reference the main properties of $\Gamma(x,t;y,s)$ (see [Fri64], Chapters 1 and 9, and [IKO62]):

Theorem 25 *If $L_c(t)$ is uniformly parabolic in H and the coefficients are continuous bounded functions in H, such that $a_{ij} \in C^{\alpha,\alpha/2}(H)$, $b_i, c \in C^{\alpha,0}(H)$ (uniformly w.r.t. t), $0 < \alpha \le 1$, for all i, j, then there exists a fundamental solution $\Gamma(x,t;y,s)$ of $L_c(t)$. $\Gamma(x,t;y,s)$ is non-negative, is bounded in any domain $|x-y| + (t-s) > 0$ and has an integrable singularity at $(x,t) = (y,s)$. $\Gamma(x,t;y,s)$ satisfies the reproduction property ($s < \tau < t$)*

$$\Gamma(x,t;y,s) = \int_{\mathbb{R}^d} \Gamma(x,t;\xi,\tau)\Gamma(\xi,\tau;y,s)d\xi. \tag{A.15}$$

If $L_c(t) = L_c$, then any fundamental solution depends only on $x, y, \tau = (t-s)$. If $L = -\partial_t + \frac{1}{2}\sum_{i,j} a_{ij}\partial_{x_i x_j}$, with constant coefficients, then $\Gamma(x,t;y,s) = \Gamma(\left|x-y\right|, t-s)$ is a Gaussian probability density function with mean y and covariance matrix $(a_{i,j})(t-s)$.

If also $a_{ij} \in C^{2+\alpha,\alpha/2}(H)$, $\mu_i \in C^{1+\alpha,0}$ (uniformly w.r.t. t) and the derivatives are bounded functions, then there exists a fundamental solution $(y,s) \mapsto \Gamma^*(y,s;x,t)$, $s < t$, of the adjoint equation $L_c^*(s)u = 0$ and $\Gamma(x,t;y,s) = \Gamma^*(y,s;x,t)$.

[7] Note that in the scalar case this is in accordance to the conditions on $v(y)$.

[8] More generally, a fundamental solution of a partial differential operator B is any generalized function E satisfying in the weak sense (see below) the inhomogeneous equation $BE = \delta$.

$(x,t) \mapsto \Gamma(x,t;y,s) \in C^{2,1}(H_0)$. Furthermore, if all the coefficients of $L_c(t)$ are in $C^{\infty,\infty}(H)$ and all the derivatives are bounded functions, then $\Gamma(x,t;y,s)$ is infinitely differentiable in all the variables and

$$\left| D_x^\alpha D_t^\beta D_y^a D_s^b \Gamma(x,t;y,s) \right| \leq const. \frac{\exp\left[-const.\left(\frac{|x-y|^2}{t-s}\right)\right]}{(t-s)^{(d+|\alpha|+2\beta+|a|+2b)/2}}, \tag{A.16}$$

where the positive constants depend only on $\alpha, \beta, a, b \in \mathbb{Z}_+$ (α, a are multi-indexes). In particular,

$$|\Gamma(x,t;y,s)| \leq const.(t-s)^{-\frac{d}{2}} \exp\left[-\left(\frac{|x-y|^2}{2(t-s)M_0}\right)\right], \quad M_0 > M. \tag{A.17}$$

(A.14) asks that $\Gamma(x,t;y,s)$, for fixed (y,s), acts as the Dirac's delta distribution $\delta(x-y)$ on the class of continuous functions bounded by $O\left[\exp\left(h|x|^2\right)\right]$, called Tychonov class (for $L_c(t)$)[9]. From (A.17), if $\varphi(x)$ is in the *Tychonov class*, the integral in (A.14) converges. In fact, for any $0 < \alpha < \beta$, we have (e.g., see Lemma 9.9, [Fri64])

$$\int_{\mathbb{R}^d} \exp\left[-\alpha|x-y|^2\right] \exp\left[\beta|y|^2\right] dy \leq const. \exp\left[const.|x|^2\right]. \tag{A.18}$$

The same argument is used to prove the well posedness of the Cauchy problem for $L_c(t)$ (see [Fri64], Theorems 1.12 and 1.16). A strong consequence of (A.17) is that any fundamental solution from the parametrix method is upper-bounded by a Gaussian probability distribution with variance-covariance matrix $(t-s)M_0 I_{d\times d}$, where $I_{d\times d}$ is the identity matrix. This is not necessarily true for fundamental solutions (if any exists) of an operator $L_c(t)$ for which the conditions of the theorem above do not hold.

Cauchy Problems; Well Posedness

A Cauchy (or pure initial value) problem for the homogeneous equation (A.11) is the problem to find a solution of (A.11) on H_0 that satisfies also the initial condition

$$u(x,0) = g(x), \quad x \in \mathbb{R}^d \ . \tag{A.19}$$

This problem is said to be *well* (or properly) *posed* if there exists a unique solution of (A.11) - (A.19) and the solution depends continuously on the initial datum $g(x)$, which means that if we change slightly the function g, then the solution should change slightly[10]. We always suppose that our problems are well posed, despite we could not be able to compute in closed-form the solution. Here we review some sufficient conditions for existence and uniqueness. We specialize to the heat equation to say something on the continuity on the datum, sometimes called stability.

We can get the existence of the solution and a representation formula using the fundamental solution and the superposition principle as in (A.12). First notice that the fundamental solution $\Gamma(x,t;y,0)$ solves the Cauchy problem with the initial condition $\delta(x-y)$, since $\Gamma(x,t;y,0) \to \delta(x-y)$ as $t \downarrow 0$. Then we get a formal solution of (A.11) - (A.19) by

$$u(x,t) = \int_{\mathbb{R}^d} \Gamma(x,t;y,0) g(y) dy. \tag{A.20}$$

Note that if g is in the Tychonov class, then the integral converges and, by (A.14), the initial condition is satisfied. A rather complicated analysis ([Fri64], Chapters 1,9) proves that (A.20) solves the given Cauchy problem and it satisfies $|u(x,t)| \leq const. \exp\left[const.|x|^2\right]$ (called exponential

[9] This interpretation is correct because $C_0(\mathbb{R}^d)$, on which δ acts, is a subset of the Tychonov class.

[10] To assess this continuity we need some metric on the space of functions for the initial datum and so we defer this issue to the next chapter.

growth condition). (A.20) is called a representation formula, because it holds independently of the knowledge in closed-form of the fundamental solution.

As for ODEs or SDEs, uniqueness requires much less. However, since we do not have spacial boundary conditions, again we must restrict the growth of the coefficients of $A(t)$ and of the initial datum[11]. Suppose that the coefficients of $A(t)$ are continuous and satisfy: $a(x,t)$ is semi-definite positive in H, $|a_{i,j}(x,t)| \leq const. \left(|x|^2+1\right)$, $|b_i(x,t)| \leq const. (|x|+1)$ and $c(x,t) \geq 0$. It is proved in [Fri76], that under these conditions, there exists at most one solution satisfying the polynomial growth condition $|u(x,t)| \leq const. [|x|^\alpha + 1]$, for some $\alpha \geq 0$. If the diffusion coefficients are bounded, the solution is unique in the wider class of function satisfying the exponential growth condition. This uniqueness result and the representation formula (A.20) imply that the fundamental solution Γ of $L_c(t)$ is unique. In fact, suppose that Γ_1 is another fundamental solution of $L_c(t)$, then

$$\int_{\mathbb{R}^d} \left(\Gamma(x,t;y,0) - \Gamma_1(x,t;y,0)\right) g(y) dy = 0,$$

and uniqueness follows for g is arbitrary and Γ, Γ_1 are continuous.

The result about uniqueness follows from the weak maximum principle for parabolic equation on unbounded domains. It is a rather technical result for a general operator (A.7), but it takes a nice form for a constant coefficient operator, allowing also to prove stability. So consider the heat equation on H. The maximum principle for the Cauchy problems says that if $u \in C^{2,1}(H_0) \cap C(H)$ solves the heat equation with initial condition (A.19) and satisfies an exponential growth condition, then $\sup_H u(x,t) = \sup_{\mathbb{R}^d} g(x)$. By the existence result above, the estimate is satisfied. Therefore if u_0, u_1 satisfy the heat equation with initial data g_0, g_1, then $(u_0 - u_1)$ solves the heat equation with datum $(g_0 - g_1)$ and we have

$$\sup_H |u_0 - u_1| = \sup_{\mathbb{R}^d} |g_0 - g_1| .$$

Inhomogeneous Problems And The Duhamel's Principle

An *inhomogeneous* equation is given by $\left(L_c(t)u\right)(x,t) = f(x,t)$, where f is called source (or datum). For classical solutions the general principle from ODEs holds: the general solution of the nonhomogeneous equation is given by the sum of the general solution of (A.11) plus any particular solution of the inhomogeneous equation. The Duhamel's Principle extends this fact to the Cauchy problems for the inhomogeneous equation with homogeneous initial datum, i.e., the problem

$$\begin{array}{rl} \left(L_c(t)u\right)(x,t) = f(x,t), & H_0 \\ u(x,0) = 0, & \mathbb{R}^d \end{array} . \tag{A.21}$$

The principle tells us that we can solve the problem (A.21) using the results of the previous section and again (A.12). For a fixed $s \in [0,T]$, consider the auxiliary Cauchy problem for the homogeneous equation[12]

$$\begin{array}{rl} \left(L_c(t)w\right)(x,t;s) = 0, & \mathbb{R}^d \times (s,T] \\ w(x,s;s) = -f(x,s), & \mathbb{R}^d \times \{t = s\} \end{array} . \tag{A.22}$$

If, for any $s \in [0,T]$, the problem (A.22) is well posed, we have a family of solutions $\{w(x,t;s), s \in [0,T]\}$. Then (e.g., see [KGO95]) also problem (A.21) is well posed and its solution has representation

[11] E.g., the heat equation with initial condition $u(x,0) = 0$ admits infinite solutions other than the trivial (but physically correct) $u \equiv 0$, see [Joh82]. All the other solutions grow very fast.

[12] The $-$ is due to the signs of the operator $L_c(t)$. For the intuition see Evans, Chapter 2, for a detailed discussion see Fri64, Theorem 1.9.

$$u(x,t) = \int_0^t w(x,t;s)ds. \tag{A.23}$$

When the hypotheses of the existence result of the previous section are satisfied and $f(x,t)$ is in the Tychonov class, for any fixed value of the parameter s there exists a unique solution of (A.22) and it has representation (A.20). Therefore, (A.23) has representation

$$-\int_0^t \int_{\mathbb{R}^d} f(y,s)\Gamma(x,t;y,s)dyds. \tag{A.24}$$

By linearity, under the assumptions of the existence result and that $f(x,t)$ is continuous, Holder continuous (exponent α) uniformly in t and in the Tychonov class, there exists a unique solution of the general Cauchy problem

$$\begin{array}{rl} \left(L_c(t)u\right)(x,t) = f(x,t), & H_0 \\ u(x,0) = g(x), & \mathbb{R}^d \end{array}, \tag{A.25}$$

and the solution has representation

$$u(x,t) = \int_{\mathbb{R}^d} \Gamma(x,t;y,0)g(y)dy - \int_0^t \int_{\mathbb{R}^d} f(y,s)\Gamma(x,t;y,s)dyds. \tag{A.26}$$

We stress that the change of variable $t \mapsto T - t$ does not change the well posedness (at least, existence and uniqueness) of the problem (A.25). The effects are that the time runs backward, we have to take $0 \le t < s \le T$ and we deal with a terminal condition $u(x,T) = g(x)$. On the other hand, the change of variable $t \mapsto -t$ in general destroys the well posedness of the problem (A.25), because we retain an initial condition and so we cannot apply with backward time neither the uniqueness conditions nor the existence results.

Note that from (A.24) we have $|u(x,t)| \le T \sup |f(x,t)|$. Since the maximum principle for constant coefficient elliptic operators extends to the problem (A.25), also the Cauchy problem for an inhomogeneous heat equation is stable.

Other topics

We collect here some issues that either we treat elsewhere or we do not consider at all. We give some references.

- Separation of variable method. It is the most powerful method to solve in closed-form one-space variable problems, with specified boundary condition (also in the case $|x| \to \infty$). It is necessary that $A(t) = A$ in (A.7). It is based on the ansatz (i.e., guess) that the solution factorize somehow, for linear problem as $u(x,t) = v(x)w(t)$. Then the problem reduces to an eigenproblem (a Sturm-Liouville problem if on $\mathbb{R}$) and to the building of a series solution. We apply this technique in Chapter 7. See [Wei65], [Pol02].

- Integral transform methods. The idea is to apply a transform (for instance the Fourier transform to the space variables) to reduce the Cauchy problem to an ODE or an algebraic equation. Then the solution is recovered inverting the transform. Usually these methods work for constant coefficient equations and require a lot of regularities. The Laplace transform is one of the key tools in Semigroups theory. See [Eva98], [Pol02] and [Kry96].

- Weak (or generalized) solutions. A function that admits the representation (A.26) (or any other integral representation formula) does not need necessarily to be in $C^{2,1}(H)$. From this fact, the weak formulation of the Cauchy problem looks for a solution in some Sobolev space (depending on the chosen space we have weak, mild and strong solutions). The derivatives are meant in the weak sense, i.e., as Schwartz' distributions (or generalized functions). It is clear that any fundamental solution is a weak solution. Since the fundamental solution of a Cauchy problem is also a function, in the thesis we prefer to work (sometimes formally) with classical solutions, but a possible extension is to consider a weak version of the problem and then check when the required smoothness holds. Note that if a problem is well posed in classical sense, the weak solutions must agree with the classical solution. See [Eva98], [Fri63], [Sal04].

- A priori estimates. An a priori estimate is an estimate of the solution obtained without the knowledge that the solution actually exists. In fact, these estimates are often used to prove the existence of a solution. For classical solutions estimates in norm on Banach spaces (Schauder's and Bernstein's estimates, for instance) are typically used. The estimates in term of the L^2 norm, fundamental in the weak formulation, are called energy estimates (integral estimates, for general L^p). An energy estimate is considered in Chapter 5. See [Eva98], [Fri64], [Fri76] and [IKO62].
- Numerical methods. The two main numerical methods of solution are the Finite difference and Finite elements methods. The idea is to reduce (project) the infinite dimensional problem to some finite dimensional problem. The former approximates the derivatives on a mesh (lattice) with finite difference operators and studies the well posedness of the discretized problem (see Section 10.5). The latter uses functional analytical methods to find a solution in a finite dimensional subspace (e.g., a space of bounded piecewise polynomials) that approximates in some norm the true solution. It is strongly connected with the separation of variable method. See [RiMo67], [Duf06], [Smi85], [QuVa97].

A.3 Semigroup Theory

This is an introduction to semigroups of bounded operators, presenting the basic definitions and properties. References are [HiPh57] and [Yos80]. A more modern treatment can be find in [EnNa99], [ItKa02] and [Paz83]. An important reference for Abstract Cauchy problems is [RiMo67].

Abstract Cauchy Problems

Let A be a second order elliptic autonomous operator and consider the homogeneous Cauchy problem

$$\begin{array}{cc} \frac{\partial}{\partial t}u(x,t) - (Au)(x,t) = 0, & \mathbb{R}^d \times (0,T] \\ u(x,0) = g(x), & \mathbb{R}^d \end{array} . \tag{A.27}$$

We can study this problem from a functional analytical point of view exploiting the fact that A is a linear operator. The main advantage of this approach is that we can reduce (A.27) to a Cauchy problem for an ODE. The mathematical complexity is traded-off by some calculus simplification (e.g., see (A.29) below).

The first step is to separate the roles of time and space. Let X be some Banach space of real functions on $\mathbb{R}^d$; since usually in our applications $g(x) \in C(\mathbb{R}^d)$, we can set, e.g., $X = C_b(\mathbb{R}^d)$ (with the sup-norm). Then we have to see $u(x,t)$ in the following way. For each $t \in [0,T]$, let $u(x,t) \in X$, so that we can define $u(.,t) = U(t)$, $U : [0,T] \to X$. The space of continuous function $[0,T] \to X$ is denoted $C(0,T;X)$ and is endowed with the sup-norm $\|U\|_{C(0,T;X)} = \sup_{t\in[0,T]} \|U(t)\|_X$. Here we see that the choice of X is crucial. For instance, if $X = C_b(\mathbb{R}^d)$, $C(0,T;X) = C_b(\mathbb{R}^d \times [0,T])$. Note that, for U to solve (A.27) we must ask $U \in C^1(0,T;X)$ and $U : [0,T] \to D(A)$. We can represent (A.27) by

$$\begin{array}{cc} \frac{d}{dt}U(t) - AU(t) = 0, & t \in (0,T] \\ U(0) = x, & x \in X \end{array} , \tag{A.28}$$

which is called Abstract Cauchy problem. "Abstract" comes from the fact that it is required that A is linear, but not necessarily a differential operator. The manifest advantage is not (only) this broad generality, but (also) the fact that there is a lot of theory about linear operators and (A.28) is, as anticipated, an ODE Cauchy problem.

To study the notions of continuity at a point, differentiability at a point, etc., of the function $U : [0,T] \to X$ one has to take into account a topology on X. In particular, the Riemann's integration theory extends, also for improper integrals. For instance, the Fundamental theorem of calculus holds: if $f \in C(0,T;X)$, then the integral function $F(t) = \int_0^t f(s)ds$ is differentiable and $F'(t) = f(t)$. A useful result is: if $A : D(A) \to X$ is a linear closed operator and $f : [0,T] \to D(A)$ is such that $t \mapsto f(t)$ and $t \mapsto Af(t)$ are integrable, then $\int_0^T f(t)dt \in D(A)$ and

$$A \int_0^T f(t)dt = \int_0^T Af(t)dt. \tag{A.29}$$

Semigroup property

The semigroup theory deals with the study of the exponential function in an infinite dimensional setting. It better if we start from the property of the scalar exponential function in which we are mostly interested. It is well known that the real-valued exponential function $t \mapsto e^{at}$, $a \in \mathbb{R}$, satisfies the Cauchy problem

$$\begin{array}{c} \frac{d}{dt}T(t) - aT(t) = 0, \quad t \geq 0, \\ T(0) = 1 \end{array} . \tag{A.30}$$

Conversely, for any $a \in \mathbb{R}$, $T(t) = e^{at}$ is the only differentiable solution of (A.30). Note that $a = \frac{d}{dt}T(t)|_{t=0}$. (A.30) is exactly (A.28) in the scalar case, i.e., with $X = \mathbb{R}$ and $A = a$. So we expect that also the solution of (A.28) with the initial value $x = I$ is some extension of the exponential function. This is true and the key is to use a second characterization of the exponential function. $t \mapsto e^{at}$, $a \in \mathbb{R}$, is the only *continuous* function $T : \mathbb{R}_+ \to X$, $X = \mathbb{R}$, satisfying

$$T(t+s) = T(t)T(s), \quad \forall t,s \geq 0, \quad T(0) = I. \tag{A.31}$$

The functional equation (A.31) is called *semigroup property* and its study goes back to Cauchy who first studied the scalar-valued case, $X = \mathbb{R}$ ($X = \mathbb{C}$ to be correct). We can summarize in this way: the unique continuous solution $t \mapsto e^{at}$, $a \in \mathbb{R}$, of (A.31) is actually differentiable for all $t \geq 0$ and also solves $\frac{d}{dt}T(t) - aT(t) = 0$ with $a = \frac{d}{dt}T(t)|_{t=0}$. This explains the approach of semigroup theory: assume that $T : \mathbb{R}_+ \to X$, X a Banach space, satisfies the semigroup property (A.31) and study the function T (in particular its differentiability) and its relation with A such that the Cauchy problem (A.28) is well posed for a large enough class of initial data.

Let X be a Banach space. A family of bounded linear operators $\{T(t), t \geq 0\}$ in X (i.e., the operator-valued function $T : \mathbb{R}_+ \to \mathcal{L}(X)$) satisfying (A.31) is called a *one-parameter semigroup of operators*. The condition $T(t) \in \mathcal{L}(X)$, for all $t \geq 0$, to ensures that the composition product in (A.31) is well defined. We will see that the boundedness of $T(t)$ is also required by the well posedness of the initial value problem.

Uniformly Continuous Semigroups

The semigroup $T : \mathbb{R}_+ \to \mathcal{L}(X)$ is assumed to be continuous. We must put a topology on $\mathcal{L}(X)$ and the most natural is the uniform operator topology. In this case, $\|T(t) - I\| \to 0$ as $t \downarrow 0$ and $T(t)$ is said uniformly continuous. From $T(t)$ we can derive the so-called *infinitesimal generator* of the semigroup given by the operator

$$Ax = \lim_{t\downarrow 0} \frac{T(t)x - x}{t} = \frac{d^+}{dt} T(t)x|_{t=0}, \tag{A.32}$$

where the limit is in the uniform sense and the derivative is from the right. The maximal domain of the linear operator A[13] is

$$D(A) = \left\{ x \in X : \frac{d^+}{dt} T(t)x|_{t=0} \text{ exists} \right\}.$$

It is clear that each uniformly continuous semigroup has a unique generator; on the other hand if two uniformly continuous semigroups have the same generator they coincide. The important fact is:

A linear operator A is the generator of a uniformly continuous semigroup if, and only if, A is a bounded, everywhere defined, linear operator.

The sufficiency follows from the analytic functional calculus. If A is bounded, since the exponential function is analytic in $\mathbb{R}$, we can define

$$T(t) := e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n, \tag{A.33}$$

which converges in the operator norm for each $t \geq 0$. It is also clear that (A.33) satisfies (A.31). The functional calculus (A.33) does not extends to unbounded operators and the characterization tells us that the solution operator (see below) of a Cauchy problem (A.28) in which A is an unbounded linear operator cannot be uniformly continuous semigroup.

Strongly Continuous Semigroups

Uniform continuity is too strong a requirement to applications. If we endow $\mathcal{L}(X)$ with the strong operator topology, we have $\|T(t)x - x\|_X \to 0$ as $t \downarrow 0$ for all $x \in X$ and we call $T(t)$ a strongly continuous (or a C_0) semigroup. The mapping $t \mapsto T(t)x$, $x \in X$, is continuous for all $t > 0$, while for $x \in D(A)$ is continuous for all $t \geq 0$. The notion of infinitesimal generator extends (the limit in (A.32) is taken in the strong sense) and we have the following results that tells us that the strong continuity is the right continuity.

Let $T(t)$ be a C_0 semigroup and A its unique generator. Then

- *There exists $\omega \in \mathbb{R}$ and $M \geq 1$ such that $\|T(t)\| \leq Me^{t\omega}$, for all $t \geq 0$.*
- *Two C_0 semigroups with the same generator are equal.*
- *$D(A)$ is ($\|.\|_X$) dense in X and A is a closed linear operator.*
- *A semigroup is strongly continuous if, and only if, is weakly continuous.*

We collect here also some calculus facts. If $x \in X$, then $\int_0^t T(s)xds \in D(A)$ and

$$A \int_0^t T(s)xds = T(t)x - x. \tag{A.34}$$

If $x \in D(A)$, (A.34) extends to $\int_s^t AT(\tau)xd\tau = T(t)x - T(s)x$. Finally, if $x \in D(A)$, then $T(t)x \in D(A)$ and, for all $t \geq 0$, since $\frac{T(h)-I}{h} T(t) = T(t) \frac{T(h)-I}{h}$,

$$\frac{d}{dt} T(t)x = AT(t)x = T(t)Ax. \tag{A.35}$$

[13] Note that A is defined as a differential operator. From this follows that A is linear.

Generation Theorems

The unique generator of a C_0 semigroup is a closed densely defined linear operator. When does a closed densely defined linear operator A generate a C_0 semigroup? The answer is quite involved when A is not bounded. In fact, for A unbounded, it is unrealistic to expect the convergence (in the strong sense) of the series $\sum_{n=0}^{\infty} \frac{t^n}{n!} A^n x$ for a dense set of elements in X (see, e.g., [EnNa99] and [Kry96]). It can be that for a given A and x the series converges for some $t > 0$, but this is not enough.

Since we assume A closed, the resolvent of A, $R(\lambda, A) = (\lambda I - A)^{-1}$, $\lambda \in \rho(A)$, is a well defined bounded operator. Note that we can recover the operator A from its resolvent, by $A = \lambda I - R(\lambda, A)^{-1}$. In other words, the domain of A coincides with the range of its resolvent. Yet, this definition is only of theoretical interest because the computation of the resolvent is almost always impossible (at least if A is a second order differential operator). Another important property of the resolvent of a semigroup is its integral representation for $x \in X$,

$$R(\lambda, A)x = \int_0^{\infty} e^{-\lambda s} T(s) x ds, \tag{A.36}$$

i.e., $R(\lambda, A)$ is the Laplace transform of $T(t)$. Can we invert the transform to get $T(t)$? In general it is not possible, because of the shape of the spectrum of $T(t)$ (too large, as proved in the generation theorems), but it is possible for a special class of semigroups, called analytic semigroups, that we study in Chapter 11.

The idea behind the generation theorems is to try to use some other exponential formulas. Recalling that, for $a \in \mathbb{R}$, $e^{at} = \lim_{n\to\infty} \left(1 - \frac{ta}{n}\right)^{-n}$, one can use the operator-valued version of this formula (called Hille's approximation)

$$\lim_{n\to\infty} \left(\frac{n}{t} R(n/t, A)\right)^n \tag{A.37}$$

in the hope that the limit defines a C_0 semigroup. Another (theoretical) approximation formula (the Yoshida's approximation) is given by the family e^{tA_λ}, $\lambda \in \rho(A)$, of uniformly continuous semigroups generated by

$$A_\lambda := \lambda A R(\lambda, A), \tag{A.38}$$

in the hope again that the limit of e^{tA_λ} as $\lambda \to \infty$ generates a C_0 semigroup. Both formulas (A.37) and (A.38) work in the proof of the Hille-Yoshida theorem. One of the crucial steps is to prove that not only $\rho(A) \neq \emptyset$, but also that $\rho(A)$ is a quite nice set.

Theorem 26 *Let A be a linear operator in the Banach space X and let $\omega \in \mathbb{R}$ and $M \geq 1$. The following are equivalent.*

1. *A generates a C_0 semigroup such that $\|T(t)\| \leq M e^{t\omega}$, for all $t \geq 0$.*
2. *A is closed, densely defined and for any $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > \omega$, λ $\lambda \in \rho(A)$ (i.e., $\{\lambda : \operatorname{Re}\lambda > \omega\} \subset \rho(A)$) and $(\operatorname{Re}\lambda - \omega)^n \|R(\lambda, A)^n\| \leq M$.*

This theorem actually extends the Hille-Yoshida theorem, which holds for contraction C_0 semigroups, i.e., C_0 semigroup such that $\omega = 0$ and $M = 1$. In this case one needs only one estimate $\operatorname{Re}\lambda \|R(\lambda, A)\| \leq 1$ in 2.. In the applications the practical difficulty to compute these estimates of the resolvent (which usually is not known in closed form) is tackled using dissipative operators.

Let X^* be the dual space of the Banach space X. Denote $x^*(x)$ the value of $x^* \in X^*$ at $x \in X$. For any $x \in X$, the duality set of x is given by $F(x) = \{x^* \in X^* : x^*(x) = \|x\|_X^2 = \|x^*\|_{X^*}^2\}$. The Hahn-Banach theorem ensures that $F(x) \neq \emptyset$. A linear operator is said *dissipative* if for every $x \in D(A)$ there is a $x^* \in F(x)$ such that $\operatorname{Re} x^*(Ax) \leq 0$. In such a case, $-A$ is said accretive. In the case X is a Hilbert space, the condition becomes $\langle Ax, x \rangle_X$. We have

- *A densely defined dissipative operator is closable.*
- *The closure of a closable dissipative operator is dissipative.*
- *A linear operator A is dissipative if, and only if, $\|(\lambda I - A)x\|_X \geq \lambda \|x\|_X$ for all $x \in D(A)$ and $\lambda > 0$.*
- *For all $\lambda > 0$, $(\lambda I - A)$ is injective and, for any $z \in R(\lambda I - A)$, $\left\|(\lambda I - A)^{-1}z\right\|_X \leq \lambda^{-1}\|z\|_X$.*
- *The generator of any contraction semigroup is dissipative.*

The Lumer-Phillips theorem states

Theorem 27 *Let A be a densely defined dissipative operator in X. The following are equivalent.*

1. *$\overline{A}$ generates a contraction semigroup.*
2. *$R(\lambda I - A)$ is dense in X for all $\lambda > 0$.*

As a corollary we have that if A is a densely defined operator and A, A^* are dissipative, then $\overline{A}$ generates a contraction semigroup. An exhaustive treatment of dissipative operators can be found in [ItKa02].

Some Spectral Theory

The Hille-Yoshida theorem tells us that, for a C_0 semigroup such that $\|T(t)\| \leq Me^{t\omega}$, $\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda > \omega\} \subset \rho(A)$, so that the spectrum of the generator A must lie in a left half of the complex plane. In particular, if $T(t)$ is a contraction, $\omega = 0$ and $\sigma(A)$ lies in the closed left half of $\mathbb{C}$. What can we infer about the spectrum $\sigma(T(t))$? Can $\sigma(T(t))$ be described by $\sigma(A)$?

If X is a Hilbert space and the generator A is self-adjoint, then the measurable functional calculus says that $T(t)$ is self-adjoint and that $\sigma(T(t)) = e^{t\sigma(A)}$. We see that in general this description cannot hold, because the Spectral theorem gives a characterization of self-adjoint operators. However, the generator of a C_0 semigroup has a lot of nice properties, so it is not surprising that the following relations hold.

- $\sigma(T(t)) \supset e^{t\sigma(A)}$, for $t \geq 0$.
- $e^{tP\sigma(A)} \subset P\sigma(T(t)) \subset e^{tP\sigma(A)} \cup \{0\}$, for $t \geq 0$.

For a complete spectral theory, see [EnNa99], [Kat80] and [Paz83].

Well Posedness of Abstract Cauchy Problems

Consider again the homogeneous Cauchy problem (A.28)

$$\begin{aligned} \tfrac{d}{dt}U(t) - AU(t) = 0, &\quad t > 0 \\ U(0) = x, & \end{aligned} \tag{A.39}$$

where X is a Banach space (chosen as suggested by the problem) and $A : D(A) \subset X \to X$ is a closed densely defined (possibly unbounded) linear operator. We look for solutions $U : \mathbb{R}_+ \to D(A)$ of the ODE $\frac{d}{dt}U(t) - AU(t) = 0$ such that $U \in C^1(0, \infty; X)$, i.e., *classical solutions.* It is clear that the derivative of a classical solution $U(t)$ at each point $t > 0$, which is given by $\frac{d}{dt}U(t) = \lim_{h\downarrow 0} \frac{U(t+h)-U(t)}{h}$ in the norm of X, must satisfy

$$\left\| \frac{U(t+h) - U(t)}{h} - AU(t) \right\|_X \overset{h\downarrow 0}{\to} 0, \tag{A.40}$$

uniformly in t. Suppose that A satisfies the conditions of a generation theorem above and let $T(t)$ be the C_0 semigroup generated by A. Recalling (A.35), for any $x \in D(A)$, $T(t)x \in D(A)$ and satisfies (A.40) so that $\frac{d}{dt}U(t) - AU(t) = 0$ for all $t \geq 0$. The uniqueness of $T(t)$ implies[14]:

For every $x \in D(A)$ the function $U : t \mapsto U(t) := T(t)x$ is the unique classical solution of the Abstract Cauchy problem (A.39).

A family of operators $B(t)$, $t \geq 0$, such that, for a large enough class of initial data $x \in X$, $U(t) = B(t)x$ is a solution of (A.39), is called a *solution operator*. By this definition, $B(t)$ is linear and $B(0) = I$. The semigroup property of the solution operator $T(t)$ not only ensures existence, but also the well posedness of (A.39). In the abstract setting the requirement of continuous dependence on the initial datum for a solution operator $B(t)$ translates into

- the domain of the solution operator $B(t)$, i.e., the set D of $x \in X$ such that $B(t)x$ is a classical solution of (A.39), is independent of t and is dense in X;
- $\{B(t), t \geq 0\}$ is uniformly bounded.

The latter condition, which follows from the Uniform boundedness principle if $B(t)$ is a family of bounded operators, says that, if K is the uniform bound and $x, y \in D$, then $\|B(t)x - B(t)y\|_X \leq K\|x - y\|_X$. The former condition ensures that for each $(x_n) \subset D$ and $\|x_n - x\|_X \to 0$, then $\|B(t)x_n - B(t)x\|_X \to 0$. Therefore, the solution $U(t) = B(t)x$ depends continuously on the data. We have

If $x \in D(A)$, (A.40) is well posed for $B(t) = T(t)$ and $D = D(A)$.

We can now give the interpretation of the semigroup property (A.31). The Cauchy problem (A.39) for a second order differential operator A describes the deterministic evolution of an autonomous system and its unique solution operator is the law of evolution of the system. The fact that the solution operator is a semigroup means that, given an initial value $x \in D(A)$, we can compute the solution in two different but equal ways. We can compute the solution at $t+s$, $U(t+s) = T(t+s)x$, either directly or by first computing the solution at t, $U(t) = T(t)x = y \in D(A)$, and then solve the same problem (A.39) at $s+t$ with initial condition $U(t) = y$, obtaining $U(s+t) = T(s)y = T(s)T(t)x = T(s+t)x$. Therefore, the semigroup property is a rigorous mathematical formulation of the so-called law of determinism (for a thorough treatment of this issue, see [EnNa99]).

Extensions

A first possible generalization stems by noticing that by the B.L.T. theorem we could extend the solution operator to X. This avenue leads to the weak formulations of the problem that we do not consider. Instead, we consider the extension to inhomogeneous Cauchy problems

$$\begin{array}{c} \frac{d}{dt}U(t) - AU(t) = -f(t), \quad t > s \\ U(s) = x, \end{array} \tag{A.41}$$

where $f : [s, \infty) \to X$ and $s \geq 0$ is fixed. The result is

For every $x \in D(A)$ and $f \in C(s, \infty; X) \cap L^1(s, \infty; X)$ the Abstract Cauchy problem (A.41) is well posed and for each $t \geq s$ the unique solution is given by the variation of constants formula

$$U(t) = T(t-s)x - \int_s^t T(t-\tau)f(\tau)d\tau. \tag{A.42}$$

Now consider the non-autonomous Cauchy problems

$$\begin{array}{c} \frac{d}{dt}U(t) - A(t)U(t) = -f(t), \quad t > s \\ U(s) = x, \end{array} \tag{A.43}$$

[14] If A is only closable, consider its closure. Actually, uniqueness follows under weaker assumptions (e.g, see [Paz83], page 101).

which is called an evolution problem. $s \geq 0$ is fixed. We will always assume that $D(A(t)) = D$, independent of t, and dense in X. It turns out, but it is intuitive, that any solution operator of (A.43) is a two-parameters family of operators, $V(t,s)$, $0 \leq s < t < \infty$. The semigroup property extends to

$$V(t,s) = V(t,r)V(r,s), \quad 0 \leq s \leq r \leq t < \infty, \quad V(s,s) = I. \tag{A.44}$$

If (A.44) holds, $V(t,s)$ is called an *evolution system* (or *propagator* or two-parameters semigroup). We require that $(t,s) \to V(t,s)$ is strongly continuous for all $0 \leq s \leq t < \infty$. The generator of the propagator at $t > 0$ is given by the strong limit

$$\lim_{h \downarrow 0} \frac{V(t+h,t)x - x}{h} = A(t)x,$$

for $x \in D(A(t))$. The fact that $A(t)$ is time-dependent makes the study of the well posedness of (A.43) quite involved. We sketch how to do it with the parametrix method (see [Fri69] and Chapter 11).

Bibliography

[Aït96] Aït-Sahalia, Y., (1996), Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance*, **54**, 1361-1395.

[Aït02] Aït-Sahalia, Y., (2002), Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach, *Econometrica*, **70**, 223-262.

[Aït04] Aït-Sahalia, Y., (2004), Closed-form likelihood expansions for multivariate diffusions, *Working paper.*

[AHS04] Aït-Sahalia, Y., L.H. Hansen, J. Scheinkman, (2004), Operator methods for continuous-time Markov processes. *Handbook of Financial Econometrics*, North-Holland, Amsterdam. Forthcoming.

[AïKi05] Aït-Sahalia, Y., R. Kimmel, (2005), Estimating affine multifactor term structure models using closed-form likelihood expansions, *Working paper.*

[AïMy03] Aït-Sahalia, Y., P. Mykland, (2003), The effects of random and discrete sampling when estimating continuous-time diffusions, *Econometrica*, **71** (2).

[AlKu05] Albanese, C., A. Kuznetsov, (2003), Transformations of Markov processes and classification scheme for solvable driftless diffusions, *Working paper.*

[AlLa05] Albanese, C., S. Lawi, (2005), Laplace transforms for integrals of Markov processes, *Working paper.*

[AlMi06a] Albanese, C., A. Mijatovic, (2006), Convergence rates for diffusions on continuous-time lattices, *Working paper.*

[AlMi06b] Albanese, C., A. Mijatovic, (2006), Spectral methods for volatility derivatives, *Working paper.*

[AlBo99] Aliprantis, C.D., K.C. Border, (1999), *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag, Berlin.

[AHS03] Anderson, E.W., L.P. Hansen, T.J. Sargent, (2003), A quartet of semigroups for model specification, detection, robustness, and the price of risk. *Journal of the European Economic association*, **1.1**, 68-123.

[Aro68] Aronson, D.G., (1968), Non-negative solutions of linear parabolic equations. *Ann. Scuola Normale Superiore Pisa*, **22**, 607–694.

[Ash72] Ash, R.B., (1972), *Real Analysis and Probability*. Academic Press, New York.

[Aze84] Azencott, R., (1984), Densité des diffusions en temps petit: développements asymptotiques (part I). *Séminaire de probabilités de Strasbourg*, **18**, 402-498.

[Bal00] Baldi, P. (2000), *Equazioni Differenziali Stocastiche e Applicazioni*. Pitagora, Bologna.

[BaSo94] Barndorff-Nielsen, O. E., M. Sorensen, (1994), A review of some aspects of asymptotic likelihood theory for stochastic processes, *International Statistical Review*, **62**, 133-165.

[BaSc83] Basawa, I.V., D.J. Scott, (1983), *Asymptotic Optimal Inference for Non-ergodic Models*, Lecture Notes in Statistics, New York. Springer-Verlag.

[Bel64] Bellman, R.P., (1964), *Perturbation Techniques in Mathematics, Physics and Engineering*. Holt, Rinehart and Winston, New York.

[Bes79] Besala, P., (1979), Fundamental solution and Cauchy problem for a parabolic system with unbounded coefficients, *J. Differential Equations*, **33**, 26–38.

[BJS04] Bibby, B.M., M. Jacobsen, M. Sorensen, (2004), Estimating functions for discretely sampled diffusion-type models. *Handbook of Financial Econometrics*, North-Holland, Amsterdam. Forthcoming.

[BiSo95] Bibby, B.M., M. Sorensen, (1995), Martingale estimation functions for discretely observed diffusion processes, *Bernoulli*, **1**, 17-39.

[Bjö03] Björk, T., (2003), *Arbitrage Theory in Continuous Time*. Oxford University Press.

[Bla76] Black, F. (1976), The pricing of commodity contracts, *Journal of Financial Economics*, **3**, 167–179.

[BlSc73] Black, F., M. Scholes, (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637-659.

[BlGe68] Blumenthal, R.M., R.K. Getoor, (1968), *Markov Processes and Potential Theory.* Academic Press, New York.

[BoIs99] Bouchouev, I., V. Isakov, (1999), Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets. *Inverse Problems*, **15**, R95–R116.

[CaCo06] Carta, P., F. Corielli, (2006), On analytic perturbations of the Black&Scholes PDE, *Working Paper.*

[CLP99] Chapman, D.A., J.B. Long, N.D. Pearson, (1999), Using proxies for the short-term rate: when are three months like an instant?, *Review of Financial Studies*, **12**, 763-806.

[CiPe98] Cifarelli, D.M., L. Peccati, (1998), *Equazioni Differenziali Stocastiche con Apllicazioni Economiche e Finanziarie*. Egea, Milan.

[Cla93] Clark, S., (1993), The valuation problem in Arbitrage pricing theory. *Journal of Mathematical Economics*, **22**, 463-478.

[Cor04] Corielli, F., (2004), Eigenfunctions based estimating martingales for perturbed diffusions, *Working Paper.*

[Cor06a] Corielli, F., (2006), Hedging with energy, *Mathematical Finance*, **16** (3), 495-517.

[Cor06b] Corielli, F., (2006), Model error and analytical approximations, *Working Paper.*

[CoPa06] Corielli, F., A. Pascucci, (2006), Parametrix approximations for option prices, *Working Paper*.

[Cox75] Cox, J., (1975), Notes on option pricing I: Constant elasticity of variance diffusions. *Working paper*, Stanford University.

[CIR85] Cox, J.C., J.E. Ingersoll, S.E. Ross, (1985), A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.

[CoRo76] Cox, J., S.E. Ross, (1976), The valuation of options for alternative stochastic processes. *J. Financ. Econ.*, **3**, 145–166.

[DaFl86] Dacunha-Castelle, D., D. Florens-Zmirou, (1986), Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

[DaGu01] Darolles, S., C. Gourieroux, (2001), Truncated dynamics and estimation of diffusion equations, *Journal of Econometrics*, **102**, 1-22.

[DaLi03] Davydov, D., V. Linetsky, (2003), Pricing options on scalar diffusions: an eigenfunction approach. *Operations Research*, **51**, 185-209.

[DeKr02] Deck, T., S. Kruse, (2002), Parabolic differential equations with unbounded coefficients: A generalization of the parametrix method. *Acta Applicandae Mathematicae*, **74**, 71–91.

[DeSc94] Delbaen, F., W. Schachermayer, (1994), A general version of the fundamental theorem of asset pricing, *Math. Annalen*, **300**, 463-520.

[DiPa04] DiFrancesco, M., A. Pascucci, (2004), On the complete model with stochastic volatility by Hobson and Rogers. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, **460**, 3327–3338.

[DiPa05] DiFrancesco, M., A. Pascucci, (2005), On a class of degenerate parabolic equations of Kolmogorov type. *AMRX Appl. Math. Res. Express*, 77–116.

[Duf01] Duffie, D., (2001). *Dynamic asset pricing theory*. Princeton University Press.

[DFS03] Duffie, D., D. Filipovic, W. Schachermayer, (2003), Affine processes and applications in Finance. *Annals of Applied Probability*, **13** (3), 984–1053.

[DuGl01] Duffie, D., P. Glynn, (2004), Estimation of continuous-time Markov processes sampled at random time intervals. *Econometrica*, **72** (6), 1773-1808.

[DuKa96] Duffie, D., R. Kan, (1996), A yield-factor model of interest rates. *Mathematical Finance*, **6**, 379-406.

[Duf06] Duffy, D.J., (2006), *Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach*. Wiley, New York.

[Dyn65] Dynkin, E.B., (1965), *Markov Processes, Vols. 1-2*. Springer-Verlag, Berlin.

[ELX03] Egorov, A.V., H. Li, Y. Xu, (2003), Maximum likelihood estimation of time-inhomogeneous diffusions, *Journal of Econometrics*, **114**, 107-139.

[ECS01] Elerian, O., S. Chib, N. Shephard, (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, **69**, 959–993.

[EnNa99] Engel, K., R. Nagel, (1999), *One-Parameter Semigroups for Linear Evolution Equations*. Springer-Verlag, New York.

[Epp00] Epps, T.W., (2000), *Pricing derivative securities*. World Scientific, Singapore.

[Era01] Eraker, B., (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. and Econom. Statist.*, **19**, 177–191.

[EtKu86] Ethier, S.N., T.G. Kurtz, (1986), *Markov Processes: Characterization and Convergence*. Wiley, New York.

[Eva98] Evans, L.C., (1998), *Partial Differential Equations*. A.M.S. Pubblications.

[Fel51] Feller, W., (1951), Two singular diffusion problems, *The Annals of Mathematics*, 2nd Ser., **54**, No. 1., 173-182.

[Flo89] Florens-Zmirou, D., (1989), Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, **20** (4), 547–557.

[FPS00] Fouque, J.P., G. Papanicolaou, K.R. Sircar, (2000), *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press.

[FPSS03] Fouque, J.P., G. Papanicolaou, K.R. Sircar, K. Solna, (2003), Singular perturbations in option pricing. *SIAM J. Appl. Math.*, **63**.

[Fre85] Freidlin, M., (1985), *Functional Integration and Partial Differential Equations*. Ann. of Math. Studies 109, Princeton University Press.

[Fri63] Friedman, A., (1963), *Generalized Functions and Partial Differential Equations*. Prentice-Hall, New York.

[Fri64] Friedman, A., (1964), *Partial Differential Equations of Parabolic Type*. Prentice-Hall, New York.

[Fri69] Friedman, A., (1969), *Partial Differential Equations*. Holt, Rinehart and Winston, New York.

[Fri70] Friedman, A., (1970), *Foundations of Modern Analysis*. Holt, Rinehart, and Winston, New York.

[Fri71] Friedman, A., (1971), *Advanced Calculus*. Holt, Rinehart, and Winston, New York.

[Fri76] Friedman, A., (1976), *Stochastic Differential Equations and Applications, Vols. 1-2*. Academic Press, New York.

[Gar64] Garabedian, P.R., (1964), *Partial differential equations*. Wiley, New York.

[Geo95] Georgescu, A., (1995), *Asymptotic Treatment Of Differential Equations*, Chapman-Hall.

[GeJa93] Genon-Catalot, V. J. Jacod, (1993), On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. H. Poincare´ Probab. Statist.*, **29**, 119–151.

[GiSk69] Gikman, I., A.V. Skorokhod, (1969), *Introduction to the Theory of Random Processes*. Saunders, Philadelphia.

[GiSk74] Gikman, I., A.V. Skorokhod, (1974-75-79), *The Theory of Stochastic Processes, Vols. 1-2-3*. Springer-Verlag, Berlin.

[Gla04] Glasserman, P., (2004), *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

[GoHe87] Godambe, V.P., C.C. Heyde, (1987), Quasi likelihood and optimal estimation, *International Statistical Review*, **55**, 231-244.

[GrSt01] Grimmet, G., D. Stirzaker, (2001), *Probability and Random Processes*. University Press, Oxford.

[HKLW02] Hagan, P.S., D. Kumar, A.S. Lesniewski, D.E. Woodward, (2002), Managing smile risk. *Wilmott Magazine*, **11**.

[HaSc95] Hansen, L.H., J.A. Scheinkman, (1995), Back to the future: generating moment implications for continuous-time Markov processes, *Econometrica*, **63** (4), 767-804.

[HST98] Hansen, L.H., J.A. Scheinkman, N. Tuozi, (1998), Spectral methods for identifying scalar diffusions, *Journal of Econometrics*, **86**, 1-32.

[HaKr79] Harrison, J. M., D. M. Kreps, (1979), Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20, 381–408.

[HaPl81] Harrison, J. M., S. R. Pliska, (1981), Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes Applications*, **11**, 215–260.

[HaPl83] Harrison, J. M., S. R. Pliska, (1983), A stochastic calculus model of continuous trading: complete markets. *Stochastic Processes Applications*, **15**, 313–316.

[Hes93] Heston, S.L., (1993), A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies*, **6**, 327-343.

[HiPh57] Hille, E., S. Phillips, (1957), *Functional Analysis And Semi-Groups*. A.M.S. Colloquium Pubblications.

[Hin91] Hinch, E.J., (1991), *Perturbation Methods*. Cambridge University Press.

[HiSh86] Hinton, D.B., J.K. Shaw, (1986), Absolutely continuous spectra of second-order differential operator with long and short range potentials, *SIAM J. Math. Anal.*, **17**, 182-196.

[HoRo98] Hobson, D.G., L.C.G. Rogers, (1998), Complete models with stochastic volatility. *Math. Finance*, **8**, 27–48.

[Hoc86] Hochstadt, H., (1986), *The Functions of Mathematical Physics*. Dover, New York.

[Hul06] Hull, J., (2006), *Options, Futures and Other Derivatives*. Prentice-Hall, New York.

[HuWh87] Hull, J., A. White, (1987), The pricing of options on assets with stochastic volatilities, *The Journal of Finance*, **42**, No. 2, 281-300.

[HuWh90] Hull, J., A. White, (1990), Pricing interest-rate-derivative securities, *Review of Financial Studies*, **3**, 573-592.

[IKO62] Il'in, A.M., A.S. Kalashnikov, O.A. Oleinik, (1962), Linear equations of the second order of parabolic type, *Russian Math. Surveys*, **17** (3), 1–143.

[ItKa96] Ito, K., F. Kappel, (1996), The Trotter-Kato theorem and approximation of PDEs. *Mathematical Computation*, **67**, 21–44.

[ItKa02] Ito, K., F. Kappel, (2002), *Evolution Equations and Approximations*. World Scientific, Singapore.

[Jac04] Jacod, J., (2004), Inference for stochastic processes. *Handbook of Financial Econometrics*, North-Holland, Amsterdam. Forthcoming.

[Jac06] Jacod, J., (2006), Parametric inference for discretely observed non-ergodic diffusions, *Bernoulli*, **12**(3) 383–401.

[JePo02] Jensen, B., Poulsen, R., (2002), A comparison of approximation techniques for transition densities of diffusion processes. *Journal of Derivatives*, **22**, 55-71.

[JoPo04] Johannes, M., N. Polson, (2004), MCMC methods for continuous-time financial econometrics. *Handbook of Financial Econometrics*, North-Holland, Amsterdam. Forthcoming.

[Joh82] John, F., (1982), *Partial Differential Eqations.* Springer-Verlag, New York.

[Kal97] Kallenberg, O., (1997), *Foundations of Modern Probability.* Springer-Verlag, New York.

[KaSh92] Karatzas, I., S.E. Shreve, (1992), *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

[KaTa81] Karlin, S., H.M. Taylor, (1981), *A Second Course in Stochastic Processes.* Academic Press, New York.

[Kat80] Kato, T., (1980), *Perturbation Theory for Linear Operators.* Springer-Verlag, Berlin.

[Kes97] Kessler, M., (1997), Estimation of an ergodic diffusion from discrete observations, *Scand. J. Statist.*, **24**, 211-229.

[KeSo99] Kessler, M., M. Sorensen, (1999), Estimating equations based on eigenfunctions for a discretely observed diffusion process, *Bernoulli*, **5**, 299-314.

[Kev00] Kevorkian, J., (2000), *Partial Differential Equations: Analytical Solution Techniques*, Springer-Verlag, New York.

[KeCo96] Kevorkian, J., J.D. Cole, (1996), *Multiple Scale and Singular Perturbation Methods*, Springer-Verlag, New York.

[KlPl99] Kloeden, P.E., E. Platen (1999), *Numerical Solution of Stochastic Differential Equations*, Springer Verlag, New York.

[KPSS92] Kloeden, P.E., E. Platen, H. Schurz, M. Sorensen, (1992), On effects of discretization on estimators of drift parameters for diffusion processes, *J. Appl. Prob.*, **33**, 1061-1076.

[KGO95] Kreiss, H., B. Gustafsson, J. Oliger, (1995), *Time Dependent Problems and Difference Methods.* Wiley, New York.

[Kry80] Krylov, N.V., (1980), *Controlled Diffusion Processes*, Springer-Verlag, New York.

[Kry96] Krylov, N.V., (1996), *Lectures on Elliptic And Parabolic Equations in Hölder Spaces*, American Mathematical Society, Providence.

[Lam77] Lamperti, J., (1977), *Stochastic Processes.* Springer-Verlag, New York.

[Lo88] Lo, A.W., (1988). Maximum likelihood estimation of generalized Ito processes with discretely sampled data. *Econometric Theory*, **4**, 231–247.

[LLMP04] Lorenzi, L., A. Lunardi, G. Metafune, D. Pallara, (2004), Analytic semigroups and reaction-diffusion problems, *Internet Seminar*, http://www.math.unipr.it/~lunardi/LectureNotes/I-Sem2005.pdf

[Lun95] Lunardi, A., (1995), *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birkhäuser Verlag, Basel.

[Man68] Mandl, P., (1968), *Analytical Treatment of One-Dimensional Markov Processes*. Springer-Verlag, Berlin.

[MeKr06] Mele, A., D. Kristensen, (2006), Adding and subtracting Black-Scholes: a new approach to approximating derivative prices in continuous-time asset prices models, *Working Paper*.

[Mol75] Molchanov, S.A., (1975), Diffusion processes and Riemannian geometry, *Russian Math. Surveys*, **30** (1), 1-63.

[Nay73] Nayfeh, A.H., (1973), *Perturbation Methods*. Wiley, New York.

[Nel59] Nelson, E., (1959), Analytic vectors. *Annals of Mathematics*, **70**, 572-615.

[Øks06] Øksendal, B., (2006), *Stochastic Differential Equations. An Introduction with Applications*. Springer-Verlag, New York.

[PaSa90] Pagani, C.D., S. Salsa (1990), *Analisi Matematica, Vols. 1-2*. Masson, Milan.

[Paz83] Pazy, A. (1983), *Semigroup of Linear Operators and Applications to Partial Differential Equations*. Springer-Verlag, New York.

[Ped95] Pedersen, A.R., (1995), A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**, 55–71.

[Pia04] Piazzesi, M., (2004), Affine term structure models. *Handbook of Financial Econometrics*, North-Holland, Amsterdam. Forthcoming.

[Pol02] Polyanin, A.D., (2002), *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton.

[Pou99] Poulsen, R., (1999), Approximate maximum likelihood estimation of discretely observed diffusion processes, *Working paper 29*, Centre for Analytical Finance, Aarhus.

[Pra88] Prakasa Rao, B.L.S., (1988), Statistical inference from sampled data for stochastic processes, *Contemporary Mathematics*, **80**, 249-284.

[Pra99a] Prakasa Rao, B.L.S., (1999a), *Semimartingales and Their Statistical Inference*. Chapman & Hall, Boca Raton.

[Pra99b] Prakasa Rao, B.L.S., (1999b), *Statistical Inference for Diffusion Type Processes*. Arnold, London.

[QuVa97] Quarteroni, A., A.Valli, (1997), *Numerical Approximation of Partial Differential Equations*. Springer-Verlag, Berlin.

[Reb04] Rebonato, R., (2004), *Volatility and Correlation*. Wiley, New York.

[ReSi80] Reed, M., B. Simon, (1980-75), *Methods of Modern Mathematical Physics, Vols. 1-2-4*. Academic Press, New York.

[RiMo67] Richtmyer, R.D., K.W. Morton, (1967), *Difference Methods For Initial-value Problems.* Wiley, New York.

[Ris89] Risken, H., (1989), *The Fokker-Planck Equation. Methods of Solution And Applications.* Springer-Verlag, Berlin.

[Roy88] Royden, H.L., (1988), *Real Analysis.* Prentice-Hall, New York.

[Sal04] Salsa, S., (2004), *Equazioni a Derivate Parziali.* Springer-Verlag, Milan.

[Sch04] Schaumburg, E., (2004), Estimation of Markov processes with Levy-type generators, *Working Paper.*

[Shr04] Shreve, S.E., (2004), *Stochastic Calculus for Finance, Vols. 1-2.* Springer-Verlag, New York.

[Smi85] Smith, D.G., (1985), *Numerical Solution of Partial Differential Equations: Finite Difference Methods.* Clarendon Press, Oxford.

[Sør01] Sørensen, H., (2001), Discretely observed diffusions: Approximation of the continuous-time score function, *Scand. J. Statist.*, **28**, 113-121.

[Sør04] Sørensen, H., (2004), Parametric inference for diffusion processes observed at discrete points in time: A survey, *International Statistical Review*, **72** (3), 337–354,

[Sør97] Sørensen M., (1997), Estimating functions for discretely observed diffusions: A review, *Working Paper, Aarhus University.*

[Sør98] Sørensen, M., (1998), Statistical inference for discretely observed diffusions, *Lectures at Berliner Graduiertenkolleg*, Univerity of Copenhagen.

[Ste74] Stewart, H.B., (1974), Generation of analytic semigroups by strongly elliptic operators, *Trans. Amer. Math. Soc.*, **199**, 141-162.

[Ste80] Stewart, H.B., (1980), Generation of analytic semigroups by strongly elliptic operators under general boundary conditions, *Trans. Amer. Math. Soc.*, **259**, 299-310.

[Vas77] Vasicek, O., (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, 177-188.

[Wei65] Weinberger H.F., (1965), *A First Course in Partial Differential Equations*, Dover, New York.

[Wid75] Widder, D.V., (1975), *The Heat Equation.* Academic Press, New York.

[WDAN05] Widdicks, M., P.W. Duck, A.D. Andricopoulos, D.P. Newton, (2005), The Black-Scholes equation revisited: asymptotic expansions and singular perturbations. *Mathematical Finance*, **15** (2), 373-391.

[WHD95] Wilmott, P, S. Howison, J. Dewynne, (1995), *The mathematics of financial derivatives: A student introduction.* Cambridge University Press.

[YaWa71] Yamada, T., S. Watanabe, (1971), On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, **11**, 155-167.

[Yos80] Yosida, K., (1980), *Functional Analysis.* Springer-Verlag, New York.

[Yos92] Yoshida, N., (1992), Estimation for diffusion processes from discrete observation, *J. Multivariate Analysis*, **41**, 220-242.

[Zet97] Zettl, A., (1997), *Sturm-Liouville Problems. Spectral Theory And Computational Methods of Sturm Liouville Problems.* Hinton-Shaeffer eds. http://www.math.niu.edu/ zettl/SL2/

[Zet05] Zettl, A., (2005), *Sturm-Liouville Theory.* American Mathematical Society, Providence.

[Zwi98] Zwillinger, D., (1998), *Handbook of Differential Equations*, Academic Press, New York.