Pierre Henry-Labordère

Analysis, Geometry, and Modeling in Finance

Advanced Methods in Option Pricing

Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES

Analysis, Geometry, and Modeling in Finance

Advanced Methods in Option Pricing

CHAPMAN & HALL/CRC
Financial Mathematics Series

Aims and scope:
The field of financial mathematics forms an ever-expanding slice of the financial sector. This series aims to capture new developments and summarize what is known over the whole spectrum of this field. It will include a broad range of textbooks, reference works and handbooks that are meant to appeal to both academics and practitioners. The inclusion of numerical code and concrete real-world examples is highly encouraged.

Published Titles

American-Style Derivatives; Valuation and Computation, *Jerome Detemple*

Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing, *Pierre Henry-Labordère*

Credit Risk: Models, Derivatives, and Management, *Niklas Wagner*

Engineering BGM, *Alan Brace*

Financial Modelling with Jump Processes, *Rama Cont and Peter Tankov*

An Introduction to Credit Risk Modeling, *Christian Bluhm, Ludger Overbeck, and Christoph Wagner*

Introduction to Stochastic Calculus Applied to Finance, Second Edition, *Damien Lamberton and Bernard Lapeyre*

Numerical Methods for Finance, *John A. D. Appleby, David C. Edelman, and John J. H. Miller*

Portfolio Optimization and Performance Analysis, *Jean-Luc Prigent*

Quantitative Fund Management, *M. A. H. Dempster, Georg Pflug, and Gautam Mitra*

Robust Libor Modelling and Pricing of Derivative Products, *John Schoenmakers*

Structured Credit Portfolio Analysis, Baskets & CDOs, *Christian Bluhm and Ludger Overbeck*

Understanding Risk: The Theory and Practice of Financial Risk Management, *David Murphy*

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
4th, Floor, Albert House
1-4 Singer Street
London EC2A 4BQ
UK

Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES

Analysis, Geometry, and Modeling in Finance

Advanced Methods in Option Pricing

Pierre Henry-Labordère

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-8699-7 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Henry-Labordère, Pierre.
Analysis, geometry, and modeling in finance: advanced methods in option pricing / Pierre Henry-Labordère.
p. cm. -- (Chapman & Hall/CRC financial mathematics series ; 13)
Includes bibliographical references and index.
ISBN 978-1-4200-8699-7 (alk. paper)
1. Options (Finance)--Mathematical models. I. Title. II. Series.

HG6024.A3H46 2009
332.64'53--dc22 2008025447

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To Emma and Véronique

List of Tables

List of Figures

Symbol Description

$\mathbb{P}$	Probability measure, usually the risk-neutral measure.
$\mathbb{P}^T$	Forward measure.
$\mathbb{P}^{\alpha\beta}$	Forward swap measure.
$\mathbb{P}^s$	Spot Libor measure.
$\mathbb{E}^{\mathbb{P}}$	Expectation with respect to the measure $\mathbb{P}$.
$\frac{d\mathbb{P}}{d\mathbb{Q}}$	Radon-Nikodyn derivative of the measure $\mathbb{P}$ with respect to $\mathbb{Q}$.
$\mathcal{F}$	Filtration.
W_t	Brownian motion.
$t_1 \wedge t_2$	$\min(t_1, t_2)$.
$\Gamma(E)$	Smooth sections on a vector bundle E.
d	Exterior derivative.
g_{ij}	Metric.
$\mathcal{A}_i$	Abelian connection.
$\mathcal{P}$	Parallel gauge transport.
Q	Potential.
$p(t, x \vert x_0)$	Conditional probability at time t between x and x_0.
$\Delta(x, y)$	VanVleck-Morette determinant.
P_{tT}	Bond between the dates t and T.
D_{tT}	Discount factor between the dates t and T.
$s_{\alpha\beta}$	Swap between the dates T_α and T_β.
$L_i(t)$	Libor at time t between the dates T_{i-1} and T_i.
r_t	Instantaneous interest rate.
f_{tT}	Forward interest curve.
$1(x)$	Heaviside function: 1 if $x \geq 0$ and 0 otherwise.
$\delta(x)$	Dirac function.
$\mathbb{R}_+$	$x \in \mathbb{R}\,, x \geq 0$
$\mathbb{R}_+^*$	$x \in \mathbb{R}\,, x > 0$
δ_i^j	Kronecker symbol: $\delta_i^j = 1$ if $i = j$, zero otherwise.
$< \cdot, \cdot >$	scalar product on $L^2([0,1])$.
csc	$\csc(x) = \frac{1}{\sin(x)}$
$csch$	$csch(x) = \frac{1}{\sinh(x)}$
sec	$\sec(x) = \frac{1}{\cos(x)}$
$sech$	$sech(x) = \frac{1}{\cosh(x)}$
GBM	Geometric Brownian Motion.
LV	Local Volatility.
LVM	Local Volatility Model.
LMM	Libor Market Model.
PDE	Partial Differential Equation.
SDE	Stochastic Differential Equation.
SV	Stochastic Volatility.
SVM	Stochastic Volatility Model.

Contents

Chapter 1

Introduction

> Faire des mathématiques, c'est donner le même nom à des choses différentes.
>
> — Henri Poincaré

With the numerous books on mathematical finance published each year, the usefulness of a new one may be questioned. This is the first book setting out the applications of advanced analytical and geometrical methods used in recent physics and mathematics to the financial field. This means that new results are obtained when only approximate and partial solutions were previously available.

We present powerful tools and methods (such as differential geometry, spectral decomposition, supersymmetry) that can be applied to practical problems in mathematical finance. Although encountered across different domains in theoretical physics and mathematics (for example differential geometry in general relativity, spectral decomposition in quantum mechanics), they remain quite unheard of when applied to finance and allow to obtain new results readily. We introduce these methods through the problem of *option pricing*.

An option is a financial contract that gives the holder the right but not the obligation to enter into a contract at a fixed price in the future. The simplest example is a European call option that gives the right but not the obligation to buy an asset at a fixed price, called *strike*, at a fixed future date, called *maturity date*. Since the work by Black, Scholes [65] and Merton [32] in 1973, a general probabilistic framework has been established to price these options. In this framework, the financial variables involved in the definition of an option are *random variables* and their dynamics follow *stochastic differential equations* (SDEs). For example, in the original Black-Scholes-Merton model, the traded assets are assumed to follow log-normal diffusion processes with constant volatilities. The *volatility* is the standard deviation of a probability density in mathematical finance. The option price satisfies a (parabolic) partial differential equation (PDE), called the Kolmogorov-Black-Scholes pricing equation, depending on the stochastic differential equations introduced to model the market.

The market model depends on unobservable or observable parameters such as the volatility of each asset. They are chosen, we say *calibrated*, in order to reproduce the price of liquid options quoted on the market such as European call options. If liquid options are not available, historical data for the movement of the assets can be used.

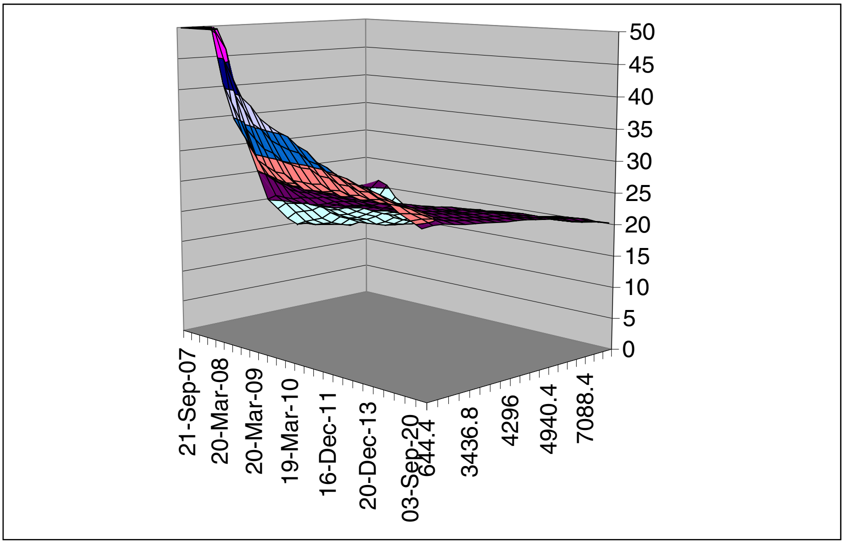

FIGURE 1.1: Implied volatility (multiplied by $\times 100$) for EuroStoxx50 (03-09-2007). The two axes represent the strikes and the maturity dates. Spot $S_0 = 4296$.

This book mainly focuses on the study of the calibration and the dynamics of the implied volatility (commonly called *smile*). The implied volatility is the value of the volatility that, when put in the Black-Scholes formula, reproduces the market price for a European call option. In the original log-normal Black-Scholes model, regarding the hypothesis of constant volatility, the implied volatility is flat as a function of the maturity date and the strike. However, the implied volatility observed on the market is not flat and presents a U-shape (see Fig. 1.1). This is one of the indications that the Black-Scholes model is based on several unrealistic assumptions which are not satisfied under real market conditions. For example, assets do not follow log-normal processes with constant volatilities as they possess "fat tail" probability densities. These unrealistic hypotheses, present in the Black-Scholes model, can be relaxed by introducing more elaborate models called *local volatility models* and *stochastic volatility models* that, in most cases, can calibrate the form of the initially observed market smile and give dynamics for the smile consistent with the fair value of exotic options. To be thorough, we should mention *jump-diffusion models*. Although interesting and mathematically appealing, we will not discuss these models in this book as they are extensively discussed in [9].

On the one hand, local volatility models (LVMs) assume that the volatility $\sigma_{\mathrm{loc}}(t, f)$ of an asset only depends on its price f and on the time t. As shown by Dupire [81], there is a unique diffusion term $\sigma_{\mathrm{loc}}(t, f)$, depending on the derivatives of European call options with respect to the time to maturity and the strike, which can exactly reproduce the current market of European call option prices. Dupire then shows how to get the calibration solution. However, the dynamics is not consistent with what is observed on the market. In this context, a better alternative is to introduce (time-homogeneous) stochastic volatility models.

On the other hand, stochastic volatility models (SVMs) assume that the volatility itself follows a stochastic process. LVMs can be seen as a degenerate example of SVMs as it can be shown that the local volatility function represents some kind of average over all possible instantaneous volatilities in a SVM.

For LVMs or SVMs, the implied volatility can be obtained by various methods. For LVMs (resp. SVMs) European call options satisfy the Kolmogorov and Black-Scholes equation which is a one-dimensional (resp. two-dimensional or more) parabolic PDE. This PDE can be traditionally solved by a finite difference scheme (for example Crank-Nicholson) or by a Monte-Carlo simulation via the Feynman-Kac theorem. These pricing methods turn out to be fairly time-consuming on a personal computer and they are generally not appropriate when one tries to calibrate the model to a large number of (European) options. In this context, it is much better to use *analytical* or *asymptotical methods*. This is what the book is intended for.

For both local and stochastic volatility models, the resulting Black-Scholes PDE is complicated and only a few analytical solutions are known. When exact solutions are not available, singular perturbative techniques have been used to obtain asymptotic expressions for the price of European-style options. We will explain how to obtain these asymptotic expressions, in a unified manner, with any kind of stochastic volatility models using the heat kernel expansion technique on a Riemannian manifold.

In order to guide the reader, here is the general description of the book's chapters:

• A quick introduction to the theory of option pricing: The purpose is to allow the reader to acquaint with the main notions and tools useful for pricing options. In this context, we review the construction of Itô diffusion processes, martingales and change of measures. Problems have been added at the end so that the reader can check his understanding.

• A recall of a few definitions regarding the dynamics of the implied volatility: In particular, we review the different forward starting options (cliquet, Napoleon...) and options on volatility that give a strong hint on the dynamics of the smile (in equity markets). A similar presentation is given for the study of the dynamics of swaption implied volatilities using specific interest rate instruments.

• A review of the heat kernel expansion on a Riemannian manifold endowed with an Abelian connection: thanks to the rewriting of the Kolmogorov-Black-Scholes equation as a heat kernel equation, we give a general asymptotic solution in the short-time limit to the Kolmogorov equation associated to a general multi-dimensional Itô diffusion process. The main notions of differential geometry, useful to grasp the heat kernel expansion, are reviewed carefully. Therefore no prerequisite in geometry is needed.

• A focus on the local and stochastic volatility models: In the geometrical framework introduced in the previous chapter, the stochastic (resp. local) volatility model corresponds to the geometry of complex (resp. real) curves (i.e., Riemann surfaces). For example, the SABR stochastic volatility model, particularly used in the fixed-income market, can be associated to the geometry of the (hyperbolic) Poincaré plane. By using the heat kernel expansion, we obtain a general asymptotic implied volatility for any SVM, in particular the SABR model with a mean-reverting drift.

• Applications to the pricing of multi-asset European options such as equity baskets, Collateralized Commodity Obligations (CCO) and swaptions: In particular, we find an asymptotic implied volatility formula for a European basket which is valid for a general multi-dimensional LVM and an asymptotic swaption implied volatility valid for a stochastic volatility Libor market model. In this chapter, we review the main issues on the construction and calibration of a Libor market model.

• A classification of *solvable* LVMs and SVMs: Solvable means that the price of a European call option can be written in terms of hypergeometric functions. Recasting the Black-Scholes-Kolmogorov PDE in a geometrical setting as described in chapter 4, we show how to reduce the complexity of this equation. The three main ingredients are the group of diffeomorphisms, the group of *gauge transformations* and the *supersymmetry.* For LVMs, these three transformations allow to reduce the backward Kolmogorov equation to a (Euclidean) Schrödinger type equation with a scalar potential for which a classification of solvable potentials is already known. This chapter illustrates the power of the differential geometry approach as it reproduces and enlarges the classification of solvable LVMs and SVMs. In addition, a new useful tool is introduced: the spectral decomposition of unbounded linear operators. A review of the theory of unbounded operators on a Hilbert space will be given in this context.

• Chapter 10 studies the large-strike behavior of the implied volatility for

LVMs and SVMs. We use two-sided Gaussian estimates of Schrödinger equations with scalar potentials belonging to the Kato class.
• In chapter 11, we give a brief overview of the Malliavin calculus. We focus on two applications: Firstly, we obtain probabilistic representations of sensitivities of derivatives products according to model parameters. Secondly, we show how to compute by Monte-Carlo simulation the local volatility function associated to SVMs.
• In the last chapter, our asymptotic methods are applied to non-linear PDEs, mainly the Bellman-Hamilton-Jacobi equation. We focus on the problem of pricing options when the market is not complete or is illiquid. In the latter case, the hedging strategy of a large trader has an impact on the market. The resulting Black-Scholes equation becomes a non-linear PDE.
• Two appendices are included. The first one summarizes the saddle-point method and the second one briefly explains Monte-Carlo methods. In this part, we highlight the Hopf algebra structure of Taylor-Stratonovich expansions of SDEs. This allows to prove the Yamato theorem giving a necessary and sufficient condition to represent asset prices as functionals of Brownian motions.
Throughout this book, we have tried to present not only a list of theoretical results but also as many as possible numerical ones. The numerical implementations have been done with Mathematica© and C++. Also, problems have been added at the end of most chapters. They are based on recently published research papers and allow the reader to check his understanding and identify the main issues arising about the financial industry.

Book audience

In the derivatives finance field, one can mainly identify four types of professionals.

- The structurer who designs the derivatives products.
- The trader who is directly in contact with the market and daily calibrates the computer models. He determines the current prices of the products and the necessary hedging.
- The salespeople who are in charge of selling the products at the conditions decided by the trader.
- The quantitative analyst who is someone with a strong background in mathematics and is in charge of developing the mathematical models & methods and computer programs aiming at the optimum pricing and hedging of the financial risks by the traders.

Among the above mentioned professionals, this book will be useful to quantitative analysts and highly motivated traders to better understand the models they are using.

Also, many Ph.D. students in mathematics and theoretical physics move to finance as quantitative analysts. A purpose of this book is to explain the new applications of some advanced mathematics to better solve practical problems in finance.

It is however not intended to be a full monograph on stochastic differential geometry. Detailed proofs are not included and are replaced by relevant references.

Acknowledgments

I would like to thank my father, Prof. A. Henry-Labordère, for advice in the writing of this book. I would like to thank my colleagues in the Equity Derivatives Quantitative research team at Société Génerale for useful discussions and feedback on the contents of this book.

Finally, this book could not have been written without the support from my wife Véronique and my daughter Emma.

About the author

Dr. Pierre Henry-Labordère works in the Equity Derivatives Quantitative research team at Société Générale as a quantitative analyst. After receiving his Ph.D. at Ecole Normale Supérieure (Paris) in the Theory of Superstrings, he worked in the theoretical physics department at Imperial College (London) before moving to finance in 2004. He also graduated from Ecole Centrale Paris and holds DEAs in theoretical physics and mathematics.

Chapter 2

A Brief Course in Financial Mathematics

> The enormous usefulness of Mathematics in the natural sciences is something bordering on the mysterious and that there is no rational explanation for it.[1]
>
> —E. Wigner

Abstract Finance is dominated by stochastic methods. In particular, traded assets are assumed to follow time-continuous Itô diffusion processes. The notions of *(local) martingale*, *Itô diffusion process*, *equivalent measure* appear naturally. In this chapter, we review the main ideas in mathematical finance and motivate the mathematical concepts.

2.1 Derivative products

Instead of giving a general definition of a derivative product, let us present a few classical examples. The first one is a *European call option* on a single asset which is a contract which gives to the holder the opportunity but not the obligation to buy an asset at a given price (called *strike*) at a fixed future date (called *maturity*). Let us suppose a call option is bought with a 100 $ strike and a five year maturity $T = 5$. The initial price of the asset, called the *spot*, is 100 $. If the stock price is 150 $ in five years, the option can be exercised and the buyer is free to purchase the 100 $ stock as the contract allows. The counterpart which has sold the contract must then buy the 150 $ stock on the market (if he does not have it) and sell it at 100 $. When the buyer has received the stock for 100 $, he can sell it directly on the market at 150 $ (we assume that there is no transaction cost). So his net earning is $50 - \mathcal{C}$\$ and the bank has lost $-50 + \mathcal{C}$\$, $\mathcal{C}$ being the price of the contract.
In practice, we don't have a two-player zero sum game as the bank hedges its risk by investing in a portfolio with traded assets and options. The earning

[1] "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" in Communications in Pure and Applied Mathematics, vol. 13, No. I (February 1960).

of the bank at the maturity date T is therefore $-50+\mathcal{C}+\pi_T$ where π_T is the portfolio value at T. As a result, both the option buyer and the seller can have a positive earning at T.

We define the *payoff* representing the potential gain at maturity date T as

$$f(S_T) = \max(S_T - K, 0)$$

with K the strike and S_T the stock price at maturity T. The net earning of the buyer will be

$$f(S_T) - \mathcal{C}$$

whether he calls the option or not because he purchased the contract $\mathcal{C}$. A bank, after defining the characteristics of the option, needs to compute the fair value $\mathcal{C}$ and to build a hedging strategy to cover its own risk.

More generally, each derivative product is characterized by a payoff depending on financial (random) variables such as equity stocks, commodity assets, fixed-income rates and foreign exchange rates between various currencies.

In practice, the derivative products are more complex than a European call option. Several characteristic features can however be outlined. A derivative can be

- *European*: meaning that the option can be exercised by the holder at a specified maturity date. The simplest example is the European call option that we have presented above. Another example most straightforward would be given by a European put option which gives the holder the right but not the obligation to sell an asset at a fixed price (i.e., strike) at the maturity date T. The payoff at T is

$$\max(K - S_T, 0)$$

- *American*: meaning that the option can be exercised either within a specific time frame or at set dates (in the latter case the option is called *Bermudan*). The simplest example is an American put option which gives to the holder the right but not the obligation to sell an asset at a fixed price (i.e., strike) at any time up to the maturity date.

- *Asian*: meaning that the option depends on the path of some assets. For example, a European Asian call option on a single stock with strike K is defined by the following payoff at the maturity date T

$$\max\left(\frac{1}{T-t}\int_t^T S_u du - K, 0\right)$$

with S_u the stock price at a time u. In practice, the integral is understood as a Riemann sum over the business days from $t_1 = t$ to $t_N = T$, $\frac{1}{N}\sum_{i=1}^N S_{t_i}$. The value of the payoff depends therefore on the arithmetic average of the stock price from t up to the maturity date T.

- *Barrier*: meaning that a European option is activated if an asset has not or has reached a barrier level up to the maturity date.

Moreover, an option does not necessarily depend on a single asset but can depend on many financial products. In this case, one says that we have a *multi-asset option*. For example, a classical multi-asset option is a European basket option whose payoff at maturity T with strike K depends on the value of an (equity) basket

$$\max\left(\sum_{i=1} \omega_i S_T^i - K, 0\right)$$

with ω_i the weight normalized by $\sum_{i=1} \omega_i = 1$ and S_T^i the price of the asset i at maturity T. Increasing the complexity level, one can also consider options depending on equity assets, fixed income rates and foreign exchange rates (FX). One says in this case that the option is *hybrid*. For example, let us consider an option that pays a foreign exchange rate FX_{T_i} (resp. a fixed coupon c_i) at some specified date T_i if a stock price S_{T_i} at T_i is greater (resp. lower) than a barrier U. The payoff is therefore[2]

$$FX_{T_i} 1(S_{T_i} - U) + c_i 1(U - S_{T_i})$$

In the following sections, we model the financial random variables that enter in the definition of a derivative product with Itô diffusion processes. For the reader with no familiarity with probability theory, we have included in the next section a reminder of definitions and results in probability. Details and proofs can be found in [26].

2.2 Back to basics

2.2.1 Sigma-algebra

In probability, the space of events that can appear are formalized by the notion of a σ-algebra. If Ω is a given set, then a σ-algebra for Ω is a family $\mathcal{F}$ of subsets of Ω with the following properties:

1. $\varnothing \in \mathcal{F}$
2. $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$ where $F^c = \Omega \setminus F$ is the complement of F in Ω
3. $F_1, F_2, \cdots \in \mathcal{F} \Longrightarrow F = \cup_{i=1}^{\infty} F_i \in \mathcal{F}$

[2] $1(x)$ is the Heaviside function meaning that $1(x) = 1$ (resp. 0) if $x > 0$ (resp. $x \leq 0$).

The pair $(\Omega, \mathcal{F})$ is called a measurable space. In the following, an element of Ω is noted ω. If we replace the third condition by finite union (and intersection), we have an *algebra*.

Example 2.1 Borel σ-algebra
When we have a topological space X, the smallest σ-algebra generated by the open sets is called the Borel σ-algebra that we note below $\mathcal{B}(X)$. $\Box$

We can endow a measurable space with a probability measure.

2.2.2 Probability measure

A probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a function

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$

such that

1. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$
2. If $F_1, F_2, \cdots \in \mathcal{F}$ and $(F_i)_{i=1}^{\infty}$ are disjoint sets then

$$\mathbb{P}(\cup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mathbb{P}(F_i)$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*. It is called a *complete probability space* if $A \subset B$, $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$ then $A \in \mathcal{F}$. The sets B with zero measure are called negligible sets. We say that a property holds *almost surely* (a.s.) if it holds outside a negligible set.
In the following, all probability spaces will be assumed to be complete. It is often simpler to construct a measure on an algebra which generates the σ-algebra. The issue is then to extend this measure to the σ-algebra itself. This extension exists and is unique as stated by the Carathéodory extension theorem.

THEOREM 2.1 Carathéodory extension theorem
Let $\mathcal{A}$ be an algebra and $\mathbb{P} : \mathcal{A} \to [0, 1]$ be a probability measure on $\mathcal{A}$. There exists a unique probability measure $\mathbb{P}' : \sigma(\mathcal{A}) \to [0, 1]$ on the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$ such that $\mathbb{P}'|_{\mathcal{A}} = \mathbb{P}$.

2.2.3 Random variables

A function $X : \Omega \to \mathbb{R}^n$ is called *$\mathcal{F}$-measurable* if

$$X^{-1}(\mathcal{U}) \equiv \{\omega \in \Omega; X(\omega) \in \mathcal{U}\} \in \mathcal{F}$$

for all open sets $\mathcal{U} \in \mathbb{R}^n$.

Example 2.2 Continuous function
Every continuous function f from $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ endowed with its Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ to $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ is measurable. Indeed, according to the definition of a continuous function, $f^{-1}(\mathcal{U})$ with $\mathcal{U} \in \mathcal{B}(\mathbb{R}^m)$ an open set is an open set and therefore belongs to $\mathcal{B}(\mathbb{R}^n)$. $\square$

By definition, a random variable (r.v.) $X : \Omega \to \mathbb{R}^n$ is a $\mathcal{F}$-measurable function. Note that the terminology is not suitable as a r.v. is actually not a variable but a function.
Every r.v. X can induce a probability measure, noted m_X, on $\mathbb{R}^n$ defined by

$$m_X(B) = \mathbb{P}(X^{-1}(B)) \tag{2.1}$$

with B an open set of $\mathbb{R}^n$. Note that this relation is well defined as $X^{-1}(B) \in \mathcal{F}$ by definition of the measurability of X.
At this stage we can introduce the *expectation* of a r.v. First, we restrain the definition to a particular class of r.v., the *simple r.v.s* (also called step functions) which take only a finite number of values and hence can be written as[3]

$$X = \sum_{i=1}^{m} x_i 1_{A_i} \tag{2.2}$$

where $A_i \in \mathcal{F}$. For this class of (measurable) functions, the expectation $\mathbb{E}^{\mathbb{P}}[X]$ is the number

$$\mathbb{E}^{\mathbb{P}}[X] = \sum_{i=1}^{m} x_i \mathbb{P}(A_i)$$

From this definition, it is clear that $\mathbb{E}^{\mathbb{P}}[\cdot]$ is a linear operator acting on the vector space of simple r.v. The operator $\mathbb{E}^{\mathbb{P}}[\cdot]$ can be extended to non-negative r.v. by

$$\mathbb{E}^{\mathbb{P}}[X] = \sup\{\ \mathbb{E}^{\mathbb{P}}[Y]\ ,\ \text{Y simple r.v.}\ ,\ 0 \le Y \le X\ \}$$

Note that the expectation above can be ∞. Finally, for an arbitrary r.v, X, we decompose X into a difference of two non-negative r.v. $X^+ = \max(X, 0)$ and $X^- = -\min(X, 0)$

$$X = X^+ - X^-$$

[3] $1_{A_i}(x) = 0$ if $x \in A_i$, zero otherwise.

and we set

$$\mathbb{E}^{\mathbb{P}}[X] = \mathbb{E}^{\mathbb{P}}[X^+] - \mathbb{E}^{\mathbb{P}}[X^-] \tag{2.3}$$

This expectation is not always defined. When $\mathbb{E}^{\mathbb{P}}[X^+]$ and $\mathbb{E}^{\mathbb{P}}[X^-]$ are finite, the r.v. is called *integrable*. This is equivalent to $\mathbb{E}^{\mathbb{P}}[|X|] < \infty$. $L^1(\Omega, \mathcal{F}, \mathbb{P})$ denotes the set of all integrable r.v. Moreover, the set of all k times integrable r.v. is noted $L^k(\Omega, \mathcal{F}, \mathbb{P})$. The expectation (2.3) of a r.v. is called the first moment and it is also possible to define higher moments:

$$m_k \equiv \mathbb{E}^{\mathbb{P}}[|X|^k] < \infty \ , \ X \in L^k(\Omega, \mathcal{F}, \mathbb{P})$$

For example, the *variance* of a r.v., defined for $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, is

$$\mathrm{Var}[X] = \mathbb{E}^{\mathbb{P}}[X^2] - \mathbb{E}^{\mathbb{P}}[X]^2$$

Note that when the random variable X has a density $p(x)$ on $\mathbb{R}$, it is integrable if and only if $\int_{\mathbb{R}} |x| p(x) dx < \infty$ and the expectation value is

$$\mathbb{E}^{\mathbb{P}}[X] = \int_{\mathbb{R}} x p(x) dx$$

2.2.4 Conditional probability

In probability theory, it may be useful to compute expectations of r.v. conditional to some information that we have. This is formalized by the notion of conditional expectation.

DEFINITION 2.1 Conditional expectation *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then the conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}^{\mathbb{P}}[X|\mathcal{G}]$, is defined as follows:*

1. *$\mathbb{E}^{\mathbb{P}}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable ($\Rightarrow$ it is a r.v.)*
2. *$\mathbb{E}^{\mathbb{P}}[XY] = \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[X|\mathcal{G}]Y]$ holds for any bounded $\mathcal{G}$-measurable r.v. Y.*

It can be shown that the map $X \to \mathbb{E}^{\mathbb{P}}[X|\mathcal{G}]$ is linear.

If we have two r.v. X and Y admitting a probability density, the conditional expectation of $X \in L^1$ conditional to $Y = y$ can be computed as follows:
The probability to have $X \in [x, x + dx]$ and $Y \in [y, y + dy]$ is by definition $p(x, y) dx dy$. Then, we can prove that

$$\mathbb{E}^{\mathbb{P}}[X|Y = y] = \int_{\mathbb{R}} x p(x, y) dx \tag{2.4}$$

Indeed, following the definition, one needs to show that

$$\mathbb{E}^{\mathbb{P}}[f(Y)\mathbb{E}^{\mathbb{P}}[X|Y]] = \mathbb{E}^{\mathbb{P}}[f(Y)X]$$

for all bounded measurable function f. Injecting our guess (2.4) into this equation, we obtain a trivial equality.

2.2.5 Radon-Nikodym derivative

This brief review of probability theory will conclude with the Radon-Nikodym derivative which will be used when we discuss the Girsanov theorem in section 2.8.

Let $\mathbb{P}$ be a probability space on $(\Omega, \mathcal{F})$ and let $\mathbb{Q}$ be a finite measure on $(\Omega, \mathcal{F})$ (i.e., $\mathbb{Q}$ takes its values in $\mathbb{R}_+$ and $\mathbb{Q}(A) < \infty \; \forall A \in \mathcal{F}$).

We say that $\mathbb{Q}$ is equivalent to $\mathbb{P}$ (denoted $\mathbb{Q} \sim \mathbb{P}$) if $\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$ for every $A \in \mathcal{F}$. The Radon-Nikodym theorem states that $\mathbb{Q} \sim \mathbb{P}$ if and only if there exists a non-negative r.v. X such that

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[1_A X] \; , \; \forall A \in \mathcal{F}$$

Moreover X is unique $\mathbb{P}$-almost surely and we note

$$X = \frac{d\mathbb{Q}}{d\mathbb{P}}$$

X is called the *Radon-Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$.

At this stage, we can define a stochastic process. Complements can be found in [34] and [27].

2.3 Stochastic processes

DEFINITION 2.2 Stochastic process *A n-dimensional stochastic process is a family of random variables $\{X_t\}_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $\mathbb{R}^n$.*

In chapter 4, we will consider stochastic processes that take their values in a Riemannian manifold.

In finance, an asset price is modeled by a stochastic process. The past values of the price are completely known (historical data). The information that we have about a stochastic process up to a certain time (usually today) is formalized by the notion of filtration.

DEFINITION 2.3 Filtration *A filtration on $(\Omega, \mathcal{F})$ is a family $\{\mathcal{F}_t\}_{t \geq 0}$ of σ sub-algebras $\mathcal{F}_t \subset \mathcal{F}$ such that for $0 \leq s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$.*

The σ-algebra $\mathcal{F}_t$ represents the information we have up to the time t. As $\mathcal{F}_t$ is a sub σ-algebra of $\mathcal{F}$, we can define the conditional expectation of the r.v. X_s according to the filtration $\{\mathcal{F}_t\}_t$, with $s > t$

$$\mathbb{E}^{\mathbb{P}}[X_s | \mathcal{F}_t]$$

By definition, this r.v. is $\mathcal{F}_t$-measurable. This is not necessarily the case for X_t which should only be $\mathcal{F}$-measurable. When the r.v. X_t is $\mathcal{F}_t$-measurable for every t, we will say that the process X is *adapted* to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

The most important example of stochastic processes is the Brownian motion which will be our building block for elaborating more complex stochastic processes. From this Brownian motion, we can generate a natural filtration.

Example 2.3 Brownian motion
A Brownian motion (or Wiener process) W_t is a continuous adapted process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the properties

- $W_0 = 0$ (with probability one).
- All increments on non-overlapping time intervals are independent: that is $W_{t_n} - W_{t_{n-1}}$, ..., $W_{t_2} - W_{t_1}$ are independent for all $0 \leq t_1 \leq \cdots \leq t_n$.
- Each increment $W_{t_i} - W_{t_{i-1}}$ is normally distributed with zero mean $\mathbb{E}^{\mathbb{P}}[W_{t_i} - W_{t_{i-1}}] = 0$ and variance

$$\mathbb{E}^{\mathbb{P}}[(W_{t_i} - W_{t_{i-1}})^2] = (t_i - t_{i-1}) \tag{2.5}$$

So it means that

$$\mathbb{E}^{\mathbb{P}}[(W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})] = \delta_i^j (t_{i+1} - t_i)$$

The conditional probability of W_{t_i} such as $W_{t_{i-1}} \equiv y$ is

$$p(W_{t_i} \equiv x | W_{t_{i-1}} \equiv y) = \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} e^{-\frac{(x-y)^2}{2(t_i - t_{i-1})}} \tag{2.6}$$

from which we can reproduce the variance (2.5).
Note that $\mathbb{E}^{\mathbb{P}}[W_t W_s] = \min(t, s)$. $\square$

DEFINITION 2.4 Brownian filtration *Let W_t be a Brownian motion. Then, we denote $\mathcal{F}_t^W$ the increasing family of σ-algebras generated by $(W_s)_{s\leq t}$, the information on the Brownian motion up to time t. In other words, $\mathcal{F}_t^W$ is the smallest σ-algebra containing all sets of the form*

$$\{\omega; W_{t_1}(\omega) \in \mathcal{F}_1, \cdots, W_{t_k}(\omega) \in \mathcal{F}_k\}$$

where $t_j \leq t$ and $\mathcal{F}_j \subset \mathbb{R}$ are Borel sets, $j \leq k = 1, 2, \cdots$

Example 2.4
By definition of the filtration $\mathcal{F}_t^W$, the process W_t is $\mathcal{F}_t^W$-adapted. This is not the case for the process W_{2t}. $\square$

In the following, without any specification, we denote $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space where our Brownian motion is defined and the Brownian filtration $\mathcal{F}_t^W$ will be noted $\mathcal{F}_t$.

2.4 Itô process

The price of an asset on a financial market has a deterministic (we say a *drift*) and a fluctuating (we say a *diffusion*) part. We will represent the variation of the asset price $\Delta S_t \equiv S_{t+\Delta t} - S_t$ between $t \geq 0$ and $t + \Delta t > t$ by

$$\Delta S_t = \underbrace{b(t, S_t)\Delta t}_{\text{drift}} + \underbrace{\sigma(t, S_t)(W_{t+\Delta t} - W_t)}_{\text{diffusion}} \tag{2.7}$$

with W_t a Brownian motion. The term $S^{-1}\sigma$ is called the (log-normal) *volatility*.
Here $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are two measurable functions on $\mathbb{R}_+ \times \mathbb{R}$. The Itô diffusion process is formally defined by considering the limit $\Delta t \to 0$ in (2.7). We now give a precise meaning to this limit under the integral form

$$S_t = S_0 + \int_0^t b(s, S_s)ds + \int_0^t \sigma(s, S_s)dW_s \tag{2.8}$$

The first term is a classical Lebesgue integral that should be finite $\mathbb{P}$-almost surely, i.e.,

$$\mathbb{P}[\int_0^t |b(s, S_s(\omega))|ds < \infty] = 1$$

whereas the second term is different as it involves a r.v. $W_{t+\Delta t} - W_t$.

2.4.1 Stochastic integral

As usual in the theory of integration, we will define the integral $\int_0^t \sigma(s, S_s)dW_s$ according to a class of simple functions and then extend the definition to a larger class of functions that can be approximated by these simple functions. In this context, we introduce the class F of simple ($\mathcal{F}_t$-adapted) functions defined on the interval $[0, t]$ by

$$f(s, \omega) = \sum_{j=0}^{n-1} f(t_j, \omega) 1_{t_j \leq s < t_{j+1}} \ , \ \omega \in \Omega \tag{2.9}$$

where we have partitioned the interval $[0, t]$ into n subintervals by means of partitioning points

$$t_0 = 0 \leq t_1 \leq \cdots \leq t_{n-1} \leq t_n = t$$

The *Itô integral* is thus defined as

$$\int_0^t f(s,\omega)dW_s \equiv \sum_{j=0}^{n-1} f(t_j,\omega)\left(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\right) \tag{2.10}$$

Note that we could have introduced simple functions instead

$$f_{\frac{1}{2}}(s,\omega) = \sum_{j=0}^{n-1} \frac{f(t_j,\omega) + f(t_{j+1},\omega)}{2} 1_{t_j \le s < t_{j+1}}$$

and defined the *Stratonovich integral* as

$$\int_0^t f(s,\omega) \diamond dW_s \equiv \sum_{j=0}^{n-1} \frac{f(t_j,\omega) + f(t_{j+1},\omega)}{2} \left(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\right)$$

Now, we introduce Υ to be the class of functions

$$f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$$

such that

1. $(t,\omega) \to f(t,\omega)$ is a $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurable function.
2. $f(t,\omega)$ is $\mathcal{F}_t$-adapted.
3. $\mathbb{P}[\int_S^T f(t,\omega)^2 dt < \infty] = 1 \;\; \forall\, 0 \le S < T < \infty$.

One can show that each $f \in \Upsilon$ can be approximated as a limit of simple functions $\phi_n \in F$ in $L^2([S,T] \times \mathbb{P}) \;\; \forall S < T$ meaning that for each $f \in \Upsilon$, there exists a sequence $\phi_n \in F$ such that

$$\int_S^T \mathbb{E}^{\mathbb{P}}[(\phi_n(t,\cdot) - f(t,\cdot))^2]dt \to_{n\to\infty} 0$$

$\int_0^t f(s,\omega)dW_s$ is then defined as the limit in $L^2(\mathbb{P})$ of $\int_0^t \phi_n(s,\omega)dW_s$ given by (2.10).

DEFINITION 2.5 *Let $f \in \Upsilon$. Then the Itô integral of f (from S to T) is defined by*

$$\int_S^T f(t,\omega)dW_t(\omega) = \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dW_t(\omega)$$

where lim *is the limit in $L^2(\mathbb{P})$ and the ϕ_n is a sequence of simple functions such that*

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}[\int_S^T (f(t,\omega) - \phi_n(t,\omega))^2 dt] = 0 \tag{2.11}$$

The limit does not depend on the actual choice $\{\phi_n\}$ as long as (2.11) holds.

Finally, we have given a meaning to the equation (2.8) which is usually written formally as

$$dS_t = b(t, S_t)dt + \sigma(t, S_t)dW_t \tag{2.12}$$

Following a similar path, it is possible to define a n-dimensional Itô process

$$x_t^i = x_0^i + \int_0^t b^i(s, x_s)ds + \int_0^t \sum_{j=1}^m \sigma_j^i(s, x_s)dW_s^j \ , \ i = 1, \cdots, n$$

that we formally write as

$$dx_t^i = b^i(t, x_t)dt + \sum_{j=1}^m \sigma_j^i(t, x_t)dW_t^j$$

Here W_t is an uncorrelated m-dimensional Brownian motion with zero mean $\mathbb{E}^{\mathbb{P}}[W_t^j] = 0$ and variance: $\mathbb{E}^{\mathbb{P}}[W_t^j W_t^i] = \delta_{ij}t$.

DEFINITION 2.6 Itô process-SDE *Let $W_t(\omega) = (W_t^1(\omega), \cdots, W_t^m(\omega))$ denote an m-dimensional Brownian motion. If each process $\sigma_j^i(t, x)$ belongs to the class Υ and each process $b^i(t, x)$ is $\mathcal{F}_t$-adapted and*[4]

$$\mathbb{P}[\int_0^t |b(s, x_s(\omega))|ds < \infty \ \forall \ t \geq 0] = 1$$

then the process

$$dx_t^i = b^i(t, x_t)dt + \sum_{j=1}^m \sigma_j^i(t, x_t)dW_t^j$$

is an Itô process also called a stochastic differential equation (SDE).

Example 2.5

As an example of a computation of an Itô integral, we calculate $\int_0^t W_s dW_s$. Let us introduce the simple function

$$\phi_n(s, \omega) = \sum_{j=0}^{n-1} W_{t_j}(\omega) 1_{t_j \leq s < t_{j+1}}$$

[4] $|\cdot|$ is the Euclidean norm on $\mathbb{R}^n$: $|x| = \sqrt{\sum_{i=1}^n (x^i)^2}$.

with $t_0 = 0$ and $t_n = t$. This simple function satisfies condition (2.11):

$$\mathbb{E}^{\mathbb{P}}[\int_0^t (W_s(\omega) - \phi_n(s,\omega))^2 \, ds] = \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \mathbb{E}^{\mathbb{P}}[(W_s - W_{t_j})^2] dt$$

By using $\mathbb{E}^{\mathbb{P}}[(W_s - W_{t_j})^2] = |s - t_j|$ according to the definition of a Brownian motion, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[\int_0^t (W_s - \phi_n(s,\cdot))^2 \, ds] &= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} (s - t_j) ds \\ &= \sum_{j=0}^{n-1} \frac{1}{2}(t_{j+1} - t_j)^2 \to 0 \text{ as } \Delta t_j \equiv t_{j+1} - t_j \to 0\end{aligned}$$

Therefore,

$$\int_0^t W_s dW_s = \lim_{\Delta t_j \to 0} \sum_{j=0}^{n-1} W_{t_j}(W_{t_{j+1}} - W_{t_j})$$

As $\sum_{j=0}^{n-1} W_{t_j}(W_{t_{j+1}} - W_{t_j}) = \frac{1}{2}W_t^2 - \frac{1}{2}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^2$ and since

$$\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^2 \to \text{ in } L^2(\mathbb{P})$$

we obtain

$$\int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t)$$

Similarly, we can compute the same integral in the Stratonovich convention and we obtain

$$\int_0^t W_s \diamond dW_s = \frac{1}{2}W_t^2$$

Note that the Stratonovich integral looks like a conventional Riemann integral. In the Itô integral, there is an extra term $-\frac{1}{2}t$. $\Box$

In the same way the computation of a Riemann integral is done without using the Riemann sums, the Itô integral is usually computed without using the definition 2.5. A handier approach relies on the use of Itô's lemma that we explain in the next section.

2.4.2 Itô's lemma

As we will see in the following section, the fair value $\mathcal{C}$ of a European call option on a single asset at time t depends implicitly on the time t and the asset price S_t. This leads to the question of the infinitesimal variation of the fair price of the option between t and $t+\Delta t$. The asset price S_t changes to $S_t + \Delta S_t$ (2.7), during the time Δt, and the variation of $\mathcal{C}$ is

$$\Delta \mathcal{C}_t \equiv \mathcal{C}(t+\Delta t, S_t + \Delta S_t) - \mathcal{C}(t, S_t)$$

By using a Taylor expansion assuming that $\mathcal{C}$ is $C^{1,2}([0,\infty)\times\mathbb{R})$ (i.e., (resp. twice) continuously differentiable on $[0,\infty)$ (resp. $\mathbb{R}$)) then

$$\Delta \mathcal{C}_t = \partial_t \mathcal{C} \Delta t + \partial_S \mathcal{C} \Delta S_t + \frac{1}{2}\partial_S^2 \mathcal{C} (\Delta S_t)^2 + R$$

with R the rest of the Taylor expansion. The arguments of the function $\mathcal{C}$ have not been written so as not to burden the notations. Replacing ΔS_t by its variation (2.7), we obtain

$$\begin{aligned}\Delta \mathcal{C}_t &\equiv \partial_t \mathcal{C} \Delta t + \partial_S \mathcal{C}(b(t,S_t)\Delta t + \sigma(t,S_t)\Delta W_t) \\ &+ \frac{1}{2}\partial_S^2 \mathcal{C}\left(b(t,S_t)\Delta t + \sigma(t,S_t)\Delta W_t\right)^2 + R\end{aligned}$$

where we have noted $\Delta W_t \equiv W_{t+\Delta t} - W_t$. The total variation between $t=0$ up to t is obtained as the sum of the infinitesimal variations $\Delta \mathcal{C}_t$

$$\begin{aligned}\mathcal{C}(t,S_t) - \mathcal{C}(0,S_0) &= \sum_i (\partial_t \mathcal{C}_i + b(t_i, S_{t_i})\partial_S \mathcal{C}_i)\Delta t_i + \sum_i \partial_S \mathcal{C}\sigma(t_i, S_{t_i})\Delta W_{t_i} \\ &+ \sum_i \frac{1}{2}\partial_S^2 \mathcal{C}_i\left(b(t_i,S_{t_i})\Delta t_i + \sigma(t_i,S_{t_i})\Delta W_{t_i}\right)^2 + \sum_i R_i\end{aligned}$$

where we have set $\mathcal{C}_i \equiv \mathcal{C}(t_i, S_{t_i})$. As $\Delta t_i \to 0$, we have

$$\sum_i (\partial_t \mathcal{C}_i + b(t_i,S_{t_i})\partial_S \mathcal{C}_i)\Delta t_i \to \int_0^t (\partial_s \mathcal{C}(s,S_s) + b(s,S_s)\partial_S \mathcal{C}(s,S_s))ds$$

and

$$\sum_i \partial_S \mathcal{C}_i \sigma(t_i,S_{t_i})\Delta W_{t_i} \to \int_0^t \partial_S \mathcal{C}\sigma(s,S_s)dW_s$$

where the last term should be understood as an Itô integral. Moreover, one can show that the sums of the terms in $(\Delta t_i)^2$ and $(\Delta t_i)\Delta W_i$ go to zero as $\Delta t_i \to 0$. Finally, we have in $L^2(\mathbb{P})$

$$\sum_i \sigma(t_i,S_{t_i})^2(\Delta W_{t_i})^2 \to \int_0^t \sigma(t,S_t)^2 dt$$

To prove this, we set $v_i = \sigma(t_i, S_{t_i})^2$ and consider

$$\mathbb{E}^{\mathbb{P}}[\left(\sum_i v_i (\Delta W_{t_i})^2 - \sum_i v_i \Delta t_i\right)^2] = \sum_{ij} \mathbb{E}^{\mathbb{P}}[v_i v_j \left((\Delta W_{t_i})^2 - \Delta t_i\right) \left((\Delta W_{t_j})^2 - \Delta t_j\right)]$$

For $i < j$ (similarly $j > i$), the r.v. (with zero mean) $v_i v_j \left((\Delta W_{t_i})^2 - \Delta t_i\right)$ and $(\Delta W_{t_j})^2 - \Delta t_j$ are independent and therefore the sum cancels. For $i = j$, the sum reduces to

$$\mathbb{E}^{\mathbb{P}}[v_i^2((\Delta W_{t_i})^2 - \Delta t_i)^2] = \mathbb{E}^{\mathbb{P}}[v_i^2]\mathbb{E}^{\mathbb{P}}[(\Delta W_{t_i})^4 - 2\Delta t_i (\Delta W_{t_i})^2 + (\Delta t_i)^2]$$

According to the definition of a Brownian motion, we have $\mathbb{E}^{\mathbb{P}}[(\Delta W_{t_i})^2] = \Delta t_i$, $\mathbb{E}^{\mathbb{P}}[(\Delta W_{t_i})^4] = 3(\Delta t_i)^2$ (see exercise 2) and we obtain

$$\mathbb{E}^{\mathbb{P}}[\left((\Delta W_{t_i})^2 - \Delta t_i\right)^2] = 2(\Delta t_i)^2$$

Finally

$$\mathbb{E}^{\mathbb{P}}[\left(\sum_i v_i (\Delta W_{t_i})^2 - \sum_i v_i \Delta t_i\right)^2] = 2\sum_{ij} \mathbb{E}^{\mathbb{P}}[v_i^2](\Delta t_i)^2 \to 0 \text{ as } \Delta t_i \to 0$$

The argument above also proves that $\sum_i R_i \to 0$ as $\Delta t_i \to 0$. Therefore, putting all our results together, we obtain the *Itô lemma*

$$\mathcal{C}(t, S_t) = \mathcal{C}(0, S_0) + \int_0^t \left(\partial_s \mathcal{C} + \frac{\sigma(s, S_s)^2}{2} \partial_S^2 \mathcal{C} + b(s, S_s)\partial_S \mathcal{C}\right) ds + \int_0^t \sigma(s, S_s)\partial_S \mathcal{C} dW_s$$

This equation can be formally written in a differential form as

$$d\mathcal{C}_t = \left(\partial_t \mathcal{C} + \frac{\sigma(t, S_t)^2}{2} \partial_S^2 \mathcal{C} + b(t, S_t)\partial_S \mathcal{C}\right) dt + \sigma(t, S_t)\partial_S \mathcal{C} dW_t$$

or in the equivalent form

$$d\mathcal{C}_t = \partial_t \mathcal{C} dt + \partial_S \mathcal{C} dS_t + \frac{1}{2}\partial_S^2 \mathcal{C} dS_t^2$$

with dS_t given by (2.12) and dS_t^2 computed using the formal rules

$$dt dW_t = dt.dt = 0$$
$$dW_t dW_t = dt$$

$d\mathcal{C}_t$ is therefore an Itô diffusion process with a drift $(\partial_t \mathcal{C} + \frac{\sigma(t,S_t)^2}{2}\partial_S^2 \mathcal{C} + b(t,S_t)\partial_S \mathcal{C})$ and a diffusion term $\sigma(t,S_t)\partial_S \mathcal{C}$. Similarly for a function of n-dimensional Itô diffusion processes x_t^i (4.4), we can generalize Itô's lemma and we obtain

THEOREM 2.2 The general Itô formula
Let $\{x_t^i\}_{i=1,\cdots,n}$ be an n-dimensional Itô process

$$dx_t^i = b^i(t,x_t)dt + \sum_{j=1}^{m} \sigma_j^i(t,x_t)dW_t^j \tag{2.13}$$

Let $\mathcal{C}(\cdot,\cdot)$ be a $C^{1,2}([0,\infty)\times\mathbb{R}^n)$ real function. Then the process $\mathcal{C}_t \equiv \mathcal{C}(t,x_t)$ is again an Itô process given by

$$d\mathcal{C}_t = \sum_{i=1}^{n}\sum_{j=1}^{m} \sigma_j^i(t,x)\frac{\partial \mathcal{C}(t,x)}{\partial x^i}dW_t^j +$$
$$\left(\partial_t \mathcal{C}(t,x) + \sum_{i=1}^{n} b^i(t,x)\frac{\partial \mathcal{C}(t,x)}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{m}\sigma_k^i(t,x)\sigma_k^j(t,x)\frac{\partial^2 \mathcal{C}(t,x)}{\partial x^i \partial x^j}\right)dt$$

This can be also written as

$$d\mathcal{C}_t = \partial_t \mathcal{C}(t,x)dt + \sum_{i=1}^{n}\frac{\partial \mathcal{C}(t,x)}{\partial x^i}dx^i + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 \mathcal{C}(t,x)}{\partial x^i \partial x^j}dx^i dx^j \tag{2.14}$$

where $dx_t^i dx_t^j$ is computed from (2.13) using the formal rules $dt.dW_t^i = dt.dt = 0$ and $dW_t^i.dW_t^j = \delta_i^j dt$.

This formula is one of the key tools in financial mathematics. In the next paragraph, we present a few examples of resolution of stochastic differential equations (SDE) using Itô's lemma and then explicit under which conditions a SDE admits a unique (strong or weak) solution.

2.4.3 Stochastic differential equations

Example 2.6 Geometric Brownian motion
A Geometric Brownian motion (GBM) is the core of the Black-Scholes market model. A GBM is given by the following SDE

$$dX_t = \mu(t)X_t dt + \sigma(t)X_t dW_t$$

with the initial condition $X_{t=0} = X_0 \in \mathbb{R}$ and $\mu(t)$, $\sigma(t)$ two time-dependent deterministic functions. We observe that X_t is a positive Itô process. By using Itô's lemma on $\ln(X_t)$, we have

$$d\ln(X_t) = \left(\mu(t) - \frac{1}{2}\sigma(t)^2\right) dt + \sigma(t)dW_t$$

and by integration, we deduce the solution

$$X_t = X_0 e^{\int_0^t \mu(s)ds - \frac{1}{2}\int_0^t \sigma^2(s)ds + \int_0^t \sigma(s)dW_s} \tag{2.15}$$

From Itô isometry (see exercise 2.5), $\ln X_t$ is a Gaussian r.v. $\mathcal{N}(m_t, V_t)$ with a mean m_t and a variance V_t equal to

$$m_t = \int_0^t \left(\mu(s) - \frac{1}{2}\sigma(s)^2\right) ds + \ln X_0$$
$$V_t = \int_0^t \sigma(s)^2 ds$$

□

Example 2.7 Ornstein-Uhlenbeck process
An Ornstein-Uhlenbeck process is given by the following SDE

$$dX_t = \gamma X_t dt + \sigma dW_t$$
$$X_{t=0} = X_0 \in \mathbb{R}$$

where γ and σ are two real constants. If $\sigma = 0$, we know that the solution is $X_t = X_0 e^{\gamma t}$. Let us try the ansatz $X_t = e^{\gamma t} Y_t$. Applying Itô's lemma on X_t, we obtain the GBM on Y_t

$$dY_t = \sigma e^{-\gamma t} dW_t$$

and finally we have

$$X_t = e^{\gamma t}\left(X_0 + \sigma \int_0^t e^{-\gamma s} dW_s\right)$$

By using the Itô isometry (see exercise 2.5), the process X_t is a Gaussian r.v. $\mathcal{N}(m_t, V_t)$ with a mean m_t and a variance V_t equal to

$$m_t = e^{\gamma t} X_0$$
$$V_t = \frac{\sigma^2}{2\gamma}(e^{2\gamma t} - 1)$$

□

Strong solution

After these examples, let us present the conditions under which a SDE admits a unique solution. From a classical theorem on ordinary differential equations, $\dot{x} = f(x)$ admits a unique solution if f is a Lipschitz function, meaning that

$$|f(x) - f(y)| \leq K|x - y| \ , \ \forall x, y$$

with K a constant. There is a similar condition for the existence and uniqueness of a SDE:

THEOREM 2.3 Existence and Uniqueness of SDE (Strong solution)

Let $T > 0$ and $b(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable functions satisfying

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|) \ , \ x \in \mathbb{R}^n, t \in [0, T] \tag{2.16}$$

for some constant C (where $|b| = \sqrt{\sum_{i=1}^n |b^i|^2}$ and $|\sigma| = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |\sigma_j^i|^2}$) and such that $\forall \ x, y \in \mathbb{R}^n$, $t \in [0, T]$[5]

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y| \tag{2.17}$$

for some constant D. Then the SDE

$$dX_t^i = b^i(t, X_t)dt + \sum_{j=1}^m \sigma_j^i(t, X_t)dW_t^j \tag{2.18}$$

$$X_{t=0} = X_0 \in \mathbb{R} \tag{2.19}$$

has a unique time-continuous solution X_t for $\forall \ 0 \leq t \leq T$ with the property that X_t is adapted to the filtration $\mathcal{F}_t$ generated by $\{W_s^i\}_{s \leq t, i=1,\cdots m}$. Moreover,

$$\mathbb{E}^{\mathbb{P}}[\int_0^T |X_t|^2 dt] < \infty$$

The Brownian motion is given as an input and X_t is adapted to the filtration generated by the Brownian motion. The solution X_t is called a *strong solution.*

REMARK 2.1 Note that for time-homogeneous SDEs (i.e., $b \equiv b(x)$ and $\sigma \equiv \sigma(x)$), the condition (2.16) is implied by (2.17) and therefore is irrelevant. $\square$

[5] This condition means that the coefficients b and σ are globally Lipschitz. We recall that any function with a bounded first derivative is globally Lipschitz.

Weak solution

The existence and uniqueness of SDEs holds under rather restrictive analytical conditions on the SDE coefficients which are not satisfied in many commonly used market models. In order to circumvent this difficulty, we need to introduce *weak solutions.*
In comparison with strong solutions where the Brownian motion is given as an input, in a weak solution, the Brownian motion W_t is given exogenously: one finds simultaneously W_t and X_t. The only inputs are the drifts and the diffusion parameters. In the pricing of derivative products, one needs to consider weak solutions only as we are solely interested in the law of X_t.

DEFINITION 2.7 Weak solution [34] *A process X_t on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a weak solution to the SDE (2.18) if and only if*

- *X_t is a continuous $\mathcal{F}_t$-adapted process for some complete filtration $\mathcal{F}_t$ which satisfies (2.18).*
- *W_t is a $\mathcal{F}_t$-Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$.*

Equation (2.18) is said to be *unique in law* if any two weak solutions X^1 and X^2 which are defined on two probability spaces $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ and $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2)$ with the filtrations $\mathcal{F}_t^1$ and $\mathcal{F}_t^2$ have the same law.
When we examine local and stochastic volatility models in chapters 5 and 6, we will give some weaker conditions on the coefficients of SDEs which imply existence and uniqueness in law of weak solutions.

Example 2.8
The GBM and the Ornstein-Uhlenbeck processes satisfy the condition (2.17) and therefore have a unique (strong) solution. □

2.5 Market models

The tools to model a market have now been presented. A *market model* will be characterized by a n-dimensional Itô diffusion process

$$dx_t^i = b^i(t, x_t)dt + \sum_{j=1}^{m} \sigma_j^i(t, x_t)dW_t^j \ , \ i = 1 \cdots n \tag{2.20}$$

and by another positive stochastic process B_t satisfying

$$dB_t = r_t B_t dt \ ; \ B_0 = 1 \tag{2.21}$$

The stochastic process x^i can model equity assets, fixed-income rates, foreign exchange currencies or other unobservable r.v. (such as a stochastic volatility that we introduce in chapter 6). B_t is the *money market account* which when you invest 1 at time t gives $1+r_t dt$ at $t+dt$ with r_t the instantaneous interest rate at t. We define the *discount factor* from 0 to t by

$$D_{0t} \equiv \frac{1}{B_t} = e^{-\int_0^t r_s ds} \tag{2.22}$$

and set the discount factor between two dates t and $T > t$ by

$$D_{tT} \equiv \frac{D_{0T}}{D_{0t}} \tag{2.23}$$

The fact that the equation for dB_t does not have a Brownian term means that B_t is a non-risky asset with a fixed return r_t. However r_t is not necessarily a deterministic function and may depend on the risky asset x_t^i or other unobservable Itô processes. The instantaneous rate r_t can therefore be a stochastic process.

2.6 Pricing and no-arbitrage

Arising from previous sections, the natural follow-up question is to find a price for these contracts. It is called the *pricing problem.* The pricing at time t is an operator which associates a number (i.e., price) to a payoff

$$\Pi_t : \mathcal{P} \to \mathbb{R}_+^*$$

Different rules must be imposed on the pricing operator Π_t:
If a trader holds a book of (non-American) options, each one being characterized by a payoff f_i, the total value of the book should be

$$\Pi_t \left(\sum_i f_i \right) = \sum_i \Pi_t (f_i)$$

Therefore each payoff f_i can be priced independently to get the value of the whole portfolio. This implies that Π_t is a linear form on the space of payoff $\mathcal{P}$ which is the space of measurable functions on a measurable space $(\Omega, \mathcal{F})$. If we impose that Π_t is continuous on $\mathcal{P}$, the Riesz representation theorem states that there exists a density μ_t such that

$$\Pi_t(f) \equiv \int f d\mu_t$$

Below, we show that this can be written as a conditional expectation

$$\mathbb{E}^{\mathbb{P}}[f|\mathcal{F}_t] \tag{2.24}$$

with $\mathbb{P}$ a measure on $(\Omega, \mathcal{F})$ and $\mathcal{F}_t$ a filtration of $\mathcal{F}$. In order to characterize the measure $\mathbb{P}$, we need to introduce one of the most important notions in finance: the *no-arbitrage* condition.

2.6.1 Arbitrage

Before getting into the mathematical details, let us present this notion through an example. Let us suppose that the real estate return is greater than the fixed income rate. The trader will borrow money and invest in the real estate. If the real estate return remains the same, the trader earn money. We say that there is an arbitrage situation as money can be earned without any risk. Let us suppose now that as we apply this winning strategy and start earning a lot of money, others who observe our successful strategy will start doing the same thing. According to the offer-demand law, as more and more people will invest in the real estate by borrowing money, the fixed income rate will increase and the real estate return will decrease. The equilibrium will be reached when the real estate return μ will converge to the fixed income return r. So in order to have no-arbitrage, we should impose that $\mu = r$. More generally, we will see in the following that the no-arbitrage condition imposes that the drift of traded assets in our market model is fixed to the instantaneous rate in a well-chosen measure $\mathbb{P}$ (not necessarily unique) called the *risk-neutral measure.* To define precisely the meaning of no-arbitrage, we introduce a class of strategies that could generate arbitrage. In this context, we introduce the concept of a *self-financing portfolio.*

2.6.2 Self-financing portfolio

Let us assume that we have a portfolio consisting of an asset and a money market account B_t. We will try to generate an arbitrage at a maturity date T by selling, buying the asset (we assume that there is no transaction cost) and investing the profit in the market-money account B_t with a return r_t. The portfolio at a time t_i is composed of Δ_i assets (with asset price $x_i = x_{t_i}$) and a money market account $B_i = B_{t_i}$. The portfolio value $\tilde{\pi}_i$ is

$$\tilde{\pi}_i = \Delta_i x_i + B_i \tag{2.25}$$

During our strategy, no cash is injected in the portfolio. This way, we introduce a *self-financing portfolio* by assuming that the variation of the portfolio between dates t_i and t_{i+1} is

$$\tilde{\pi}_{i+1} - \tilde{\pi}_i = \Delta_i \left(x_{i+1} - x_i\right) + r_i \Delta t_i B_i \tag{2.26}$$

with $\Delta t_i = t_{i+1} - t_i$. At time t_i, we hold Δ_i assets with a price x_i and at time t_{i+1}, the new value of the portfolio comes from the variation of the asset price between time t_i and t_{i+1} and the increase of the value of the money market account due to the fixed income return $r_i \Delta t_i$. By using the condition above (2.26), we derive that the money market account at time t_{i+1} is

$$B_{i+1} = B_i(1 + r_i \Delta t_i) - x_{i+1}(\Delta_{i+1} - \Delta_i)$$

Therefore, by recurrence, we have

$$B_n = \prod_{k=0}^{n-1}(1 + r_k \Delta t_k) B_0 - \sum_{k=1}^{n} \tilde{x}_k (\Delta_k - \Delta_{k-1}) \tag{2.27}$$

with $\tilde{x}_k = x_k \prod_{i=k}^{n-1}(1 + r_i \Delta t_i)$, $k = 1, \cdots n-1$ and $\tilde{x}_n = x_n$. By plugging (2.27) into (2.25), the variation of the portfolio between $t_0 = t$ and $t_n = T$ is

$$\tilde{\pi}_T - \prod_{k=0}^{n-1}(1 + r_k \Delta t_k)\tilde{\pi}_t = \sum_{k=0}^{n-1} \Delta_k (\tilde{x}_{k+1} - \tilde{x}_k) \tag{2.28}$$

In the continuous-time limit assuming that the initial value of the portfolio at t is zero, $\tilde{\pi}_t = 0$, the discrete sum (2.28) converges to the Itô integral

$$\pi_T = \int_t^T \Delta(s, \omega) d\bar{x}_s \tag{2.29}$$

with $\pi_T = D_{0T} \tilde{\pi}_T$ and

$$\bar{x}_t = x_t D_{0t} \tag{2.30}$$

From this expression, we formalize the notion of arbitrage:

DEFINITION 2.8 Arbitrage *A self-financing portfolio is called an arbitrage if the corresponding value process π_t satisfies $\pi_0 = 0$ and*

$$\pi_T \geq 0 \; \mathbb{P}^{\text{hist}} \; -\text{almost surely and } \mathbb{P}^{\text{hist}}[\pi_T > 0] > 0$$

with $\mathbb{P}^{\text{hist}}$ the historical (or real) probability measure under which we model our market.

It means that at the maturity date T, the value of the portfolio is non-negative and there is a non-zero probability that the return is positive: there is no risk to lose money and a positive probability to win money. Under which conditions for a specific market model can we build a self-financing portfolio generating arbitrage? To answer this question, we need to define the notion of (local) martingale.

Martingale

A n-dimensional stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a martingale with respect to a filtration $\mathcal{F}_{t\geq 0}$ and with respect to $\mathbb{P}$ if

DEFINITION 2.9 Martingale

- *(a)* M_t *is* $\mathcal{F}_t$*-adapted for all* t.
- *(b)* $\mathbb{E}^{\mathbb{P}}[|M_t|] < \infty \; \forall \, t \geq 0$.
- *(c)* $\mathbb{E}^{\mathbb{P}}[M_t|\mathcal{F}_s] = M_s \;,\; t \geq s$.

Note that condition (c) implies that $\mathbb{E}^{\mathbb{P}}[M_s] = M_0$. However, this condition is much weaker than (c).

Example 2.9
W_t is a martingale. □

Example 2.10
$x_t = W_t^2 - t$ is a martingale with respect to the Brownian filtration $\mathcal{F}_t^W$ because

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[W_t^2 - t|\mathcal{F}_s^W] &= \mathbb{E}^{\mathbb{P}}[(W_t - W_s)^2 + 2(W_t - W_s)W_s + W_s^2 - t|\mathcal{F}_s^W] \\ &= (t-s) + W_s^2 - t = W_s^2 - s\end{aligned}$$

where we have used that $W_t - W_s$ and W_s are independent r.v. □

Local martingale

If we apply Itô's lemma to the martingale $x_t = W_t^2 - t$, we find that x_t satisfies a driftless Itô diffusion process

$$dx_t = 2W_t dW_t$$

More generally, a driftless Itô process such as $\int_0^t a_s dW_s$ is called a *local martingale*.[6] In particular, a martingale is a local martingale. In chapter 5, we explain the Feller criterion which gives a necessary and sufficient condition for which a positive local martingale is a martingale in one dimension. A standard criterion is if $\int_0^T \mathbb{E}^{\mathbb{P}}[a_s^2]ds < \infty$ then the local martingale $\int_0^t a_s dW_s$ is a martingale for $t \in [0, T]$.

[6] A more technical definition can be found in [27].

Fundamental theorem of Asset pricing

All the tools to formulate the fundamental theorem of asset pricing characterizing that a market model does not generate any arbitrage opportunities have now been introduced.
Let us call $\{\mathcal{F}_t\}_{t\geq 0}$ the natural filtration generated by the Brownian motions $\{W_t^i\}_{i=1,\cdots,m}$ that appear in our market model.

DEFINITION 2.10 Asset
An asset means a positive Itô process that can be traded (i.e., sold or bought) on the market.

For example, an equity stock is an asset. The instantaneous interest rate r_t is not a financial instrument that can be sold or bought on the market and therefore is not an asset.

LEMMA 2.1
Let us suppose that there exists a measure $\mathbb{P}$ on $\mathcal{F}_T$ such that $\mathbb{P}^{\text{hist}} \sim \mathbb{P}$ and such that the discounted asset process $\{\bar{x}_t\}_{t\in[0,T]}$ (2.30) is a local martingale with respect to $\mathbb{P}$. Then the market $\{x_t, B_t\}_{t\in[0,T]}$ has no arbitrage.

PROOF (Sketch) Let us suppose that we have an arbitrage (see definition 2.8). From (2.29), we have that π_t is a local martingale under $\mathbb{P}$. Assuming that π_t is in fact a martingale, we obtain that

$$\mathbb{E}^{\mathbb{P}}[\pi_T] = \pi_0 = 0$$

As $\pi_T \geq 0$, we have $\pi_T = 0$ $\mathbb{P}$-almost surely, which contradicts the fact that $\mathbb{P}[\pi_T > 0] > 0$ (and therefore $\mathbb{P}^{\text{hist}}[\pi_T > 0] > 0$).
Note that the martingale condition can be relaxed by introducing the class of *admissible* portfolio [34]. ∎

A measure $\mathbb{P} \sim \mathbb{P}^{\text{hist}}$ such that the normalized process $\{\bar{x}_t\}_{t\in[0,T]}$ is a local martingale with respect to $\mathbb{P}$ is called an *equivalent local martingale measure.* Conversely, one can show that the no-arbitrage condition implies the existence of an equivalent local martingale measure [11].

THEOREM 2.4 Fundamental theorem of Asset pricing
The market model defined by $(\Omega, \mathcal{F}, \mathbb{P}^{\text{hist}})$ and asset prices $x_{t\in[0,T]}$ is arbitrage-free if and only if there exists a probability measure $\mathbb{P} \sim \mathbb{P}^{\text{hist}}$ such that the discounted assets $\{\bar{x}_t\}_{t\in[0,T]}$ (defined by (2.30)) are local martingales with respect to $\mathbb{P}$.

In practice, arbitrage situation may exist. *Statistical arbitrage traders* try to detect them to generate their profits. On a long scale period, the market can be considered at equilibrium and thus there is no arbitrage.
In the following, we assume that our market model is arbitrage free, meaning that the assets $\bar{x}_t^i = D_{0t} x_t^i$ are local martingales under $\mathbb{P}$ and therefore driftless processes. Therefore, there exists a diffusion function $\bar{\sigma}_j^i(t,\omega)$ such that the processes $\bar{x}_t^i$ satisfy the SDE under $\mathbb{P}$

$$d\bar{x}_t^i = \sum_{j=1}^m \bar{\sigma}_j^i(t,\omega) dW_t^j$$

As $x_t^i = D_{0t}^{-1}\bar{x}_t^i$, we have under $\mathbb{P}$

$$dx_t^i = r_t x_t^i dt + \sum_{j=1}^m \sigma_j^i(t,\omega) dW_t^j$$

with $\sigma_j^i(t,\omega) = D_{0t}^{-1}\bar{\sigma}_j^i(t,\omega)$ and W_t a Brownian motion under $\mathbb{P}$. A measure $\mathbb{P}$ for which the drift of the traded assets is fixed to the instantaneous rate r_t is called a *risk-neutral measure*.
Note that a derivative product $\mathcal{C}_t$ can be considered as an asset as it can be bought and sold on the market. Therefore according to the theorem above, $D_{0t}\mathcal{C}_t$ should be a local martingale in an arbitrage-free market model under a risk-neutral measure $\mathbb{P}$. Assuming the integrability condition to ensure that $D_{0t}\mathcal{C}_t$ is not only a local martingale but also a martingale, we obtain the **pricing formula**

$$D_{0t}\mathcal{C}_t = \mathbb{E}^{\mathbb{P}}[D_{0T}\mathcal{C}_T|\mathcal{F}_t]$$

and using (2.22)

$$\mathcal{C}_t = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds}\mathcal{C}_T|\mathcal{F}_t] \tag{2.31}$$

Using the Markov property, it can be shown that $\mathcal{C}_t = \mathcal{C}(t, x_t)$ [34].
A fair price is given by the mean value of the discounted payoff under a risk-neutral measure $\mathbb{P}$. This is the main formula in option pricing theory. It is an equilibrium price fair to both the buyer and the seller. The seller will add a premium to the option price.
A pricing operator Π_t (2.24) is therefore the mean value according to a risk-neutral measure for which traded assets have a drift fixed to the instantaneous fixed income rate. Note that in some market models such as Libor market models (see section 2.10.3.3), the instantaneous rate is not defined. We therefore require that the discounted traded assets be local martingales under another measure, the forward measure, that we explain in example (2.13). Below, we introduce the classical Black-Scholes market model and price a European call option using pricing formula (2.31).

Example 2.11 Black-Scholes market model and Call option price
The market model consists of one asset S_t and a deterministic money market account with constant interest rate r. Therefore, under a risk-neutral measure $\mathbb{P}$, the process $f_t^T \equiv \frac{S_t}{D_{tT}} = S_t e^{r(T-t)}$, called the forward of maturity T and denoted $\bar{x}_t$ above, is a local martingale and therefore driftless. We model it by a GBM

$$df_t^T = \sigma f_t^T dW_t$$

with the constant volatility σ and the initial condition $f_0^T = S_0 e^{rT}$. The solution is (see example 2.6)

$$f_t^T = f_0^T e^{-\frac{\sigma^2}{2}t + \sigma W_t} \tag{2.32}$$

We want to find in this framework the fair price of a European call option with payoff $\max(S_T - K, 0)$. First, we remark that as $S_T = f_T^T$, the payoff (at maturity) can be written as

$$\max(f_T^T - K, 0)$$

Moreover as r is constant, the discount factor is $D_{tT} = e^{-r(T-t)}$ and the pricing formula (2.31) gives at $t = 0$

$$\mathcal{C}(0, S_0) = e^{-rT}\mathbb{E}^{\mathbb{P}}[\max(f_T^T - K, 0)]$$

By plugging the value for f_T^T in (2.32), we have

$$\mathcal{C}(0, S_0) = e^{-rT}\mathbb{E}^{\mathbb{P}}[(f_0^T e^{-\frac{\sigma^2}{2}T + \sigma W_T} - K)1\left(W_T - \frac{\ln\frac{K}{f_0^T}}{\sigma} + \frac{\sigma T}{2}\right)]$$

By using the probability density (2.6) for W_T, we obtain

$$\mathcal{C}(0, S_0) = e^{-rT}\int_{\frac{\ln\frac{K}{f_0^T}}{\sigma} - \frac{\sigma T}{2}}^{\infty} (f_0^T e^{-\frac{\sigma^2}{2}T + \sigma x} - K)\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx$$

The integration over x does not present any difficulty and we obtain

$$\mathcal{C}(0, S_0) = S_0 N(d_+) - K e^{-rT} N(d_-) \tag{2.33}$$

with

$$d_\pm = -\frac{\ln(\frac{Ke^{-rT}}{S_0})}{\sigma\sqrt{T}} \pm \frac{\sigma\sqrt{T}}{2}$$

and with $N(x) = \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\frac{dy}{\sqrt{2\pi}}$ the *cumulative normal distribution*. This is the Black-Scholes formula for a European call option. ∎

2.7 Feynman-Kac's theorem

According to the previous example (2.11), the fair price can be obtained by integrating the payoff according to the probability density. Another approach to get the same result shows that $\mathcal{C}_t$ satisfies a parabolic partial differential equation (PDE). This is the Feynman-Kac theorem.
First, we assume that under a risk-neutral measure $\mathbb{P}$ our market model is described by (2.20, 2.21). The fair value $\mathcal{C}(t,x)$ depends on the n-dimensional Itô diffusion processes $\{x_t^i\}$ characterizing our market model plus the money market account. Using the general Itô formula (2.14) we obtain that the drift of $D_{0t}\mathcal{C}$ is

$$D(t,x_t) \equiv D_{0t}^{-1}\mathrm{Drift}[d(D_{0t}\mathcal{C})] = \partial_t\mathcal{C}(t,x_t) - r_t\mathcal{C}(t,x_t) \tag{2.34}$$
$$+\sum_{i=1}^n b^i(t,x_t)\frac{\partial\mathcal{C}(t,x_t)}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^n\sum_{k=1}^m \sigma_k^i(t,x_t)\sigma_k^j(t,x_t)\frac{\partial^2\mathcal{C}(t,x_t)}{\partial x^i\partial x^j}$$

As $\mathcal{C}$ is a traded asset under a risk-neutral measure $\mathbb{P}$, $D_{0t}\mathcal{C}$ is a local martingale and its drift should cancel. Then one can show under restrictive smoothness assumption on $\mathcal{C}$, $D(t,x_t)=0$ implies that $D(t,x)=0$ for all x in the support of the diffusion. We obtain the PDE

$$\partial_t\mathcal{C}(t,x) + \sum_{i=1}^n b^i(t,x)\frac{\partial\mathcal{C}(t,x)}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^n\sum_{k=1}^m \sigma_k^i(t,x)\sigma_k^j(t,x)\frac{\partial^2\mathcal{C}(t,x)}{\partial x^i\partial x^j}$$
$$-r_t\mathcal{C}(t,x) = 0$$

More precisely, the Feynman-Kac theorem stated as below is valid under the following hypothesis [12]
Hypothesis A: The functions $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ satisfy the Lipschitz condition (2.17). The function $\mathcal{C}$ also satisfies the Lipschitz condition in $[0,T]\times\mathbb{R}^n$; moreover, the functions b, σ, $\mathcal{C}$, $\partial_x b$, $\partial_x\sigma$, $\partial_x\mathcal{C}$, $\partial_{xx}b$, $\partial_{xx}\sigma$, $\partial_{xx}\mathcal{C}$ exist, are continuous and satisfy the growth condition (2.16).

THEOREM 2.5 Feynman-Kac
Let $f\in C^2(\mathbb{R}^n)$, $r\in C(\mathbb{R}^n)$ and r be lower bounded.
Given $\mathcal{C}(t,x) = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds} f(x_T)|\mathcal{F}_t]$, we have under hypothesis A

$$-\partial_t\mathcal{C}(t,x) = D\mathcal{C}(t,x) - r_t\mathcal{C}(t,x)\ ,\ t>0,\ x\in\mathbb{R}^n$$

with the terminal condition $\mathcal{C}(T,x)=f(x)$ and with the second-order differential operator D defined as

$$D\mathcal{C}(t,x) = \sum_{i=1}^n b^i(t,x)\frac{\partial\mathcal{C}(t,x)}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^n\sum_{k=1}^m \sigma_k^i(t,x)\sigma_k^j(t,x)\frac{\partial^2\mathcal{C}(t,x)}{\partial x^i\partial x^j}$$

This theorem is fundamental as we can convert the mean-value (2.31) into a PDE. We will apply this theorem to the case of the Black-Scholes market model and reproduce the fair value of a European call option (2.33) by solving explicitly this PDE.

Example 2.12 Black-Scholes PDE and call option price rederived
Starting from the hypothesis that the forward $f_t^T = S_t e^{r(T-t)}$ follows a log-normal diffusion process in the Black-Scholes model

$$df_t^T = \sigma f_t^T dW_t$$

the resulting Black-Scholes PDE for a European call option, derived from the Feynman-Kac theorem, is

$$-\partial_t \mathcal{C}(t,f) = \frac{1}{2}\sigma^2 f^2 \partial_f^2 \mathcal{C}(t,f) - r\mathcal{C}(t,f)$$

with the terminal condition $\mathcal{C}(T,f) = \max(f-K,0)$. The solution is given by

$$\mathcal{C}(t,f) = e^{-r(T-t)} \int_0^\infty \max(f'-K,0) p(T,f'|t,f) df' \tag{2.35}$$

$p(T,f'|t,f)$ is the fundamental solution of the PDE

$$-\partial_t p(T,f'|t,f) = \frac{1}{2}\sigma^2 f^2 \partial_f^2 p(T,f'|t,f)$$

with the initial condition $\lim_{t\to T} p(T,f'|t,f) = \delta(f'-f)$. Doing the change of variable $s = \frac{\sqrt{2}}{\sigma}\ln(\frac{f}{f'})$ and setting $\tau = T-t$, one can easily show that the new function $p'(\tau,s)$ defined by

$$p(T,f'|t,f) = e^{\frac{1}{2}\ln(\frac{f}{f'}) - \frac{\sigma^2}{8}\tau} p'(\tau,s)\frac{\sqrt{2}}{\sigma f'} \tag{2.36}$$

satisfies the *heat kernel* equation on $\mathbb{R}$

$$\partial_\tau p'(\tau,s) = \partial_s^2 p'(\tau,s)$$

with the initial condition $\lim_{\tau\to 0} p'(\tau,s) = \delta(s)$. The solution is the Gaussian heat kernel

$$p'(\tau,s) = \frac{1}{\sqrt{4\pi\tau}} e^{-\frac{s^2}{4\tau}}$$

From (2.36) we obtain the conditional probability for a log-normal process

$$p(T,f'|t,f) = \frac{1}{f'\sqrt{2\pi\sigma^2(T-t)}} e^{-\frac{(\ln(\frac{f'}{f}) + \frac{\sigma^2(T-t)}{2})^2}{2\sigma^2(T-t)}}$$

Doing the integration over f' in (2.35), we obtain our previous result (2.33). $\square$

The Feynman-Kac theorem allows to price European options using a PDE approach. For Barrier options we need the following extension of the Feynman-Kac theorem.

THEOREM 2.6 Localized Feynman-Kac
Let τ be the stopping time the process leaves the domain $\mathcal{D}$:

$$\tau = \inf\{t \ : \ X_t \notin \mathcal{D}\}$$

Given $\mathcal{C}(t,x) = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds} f(x_T) 1_{\tau > T} | \mathcal{F}_t]$, we have

$$-\partial_t \mathcal{C}(t,x) = D\mathcal{C}(t,x) - r_t \mathcal{C}(t,x) \ , \ t > 0, \ x \in \mathcal{D} \tag{2.37}$$

with the terminal condition $\mathcal{C}(T,x) = f(x)$ and the Dirichlet boundary condition

$$\mathcal{C}(x) = 0 \ , \ x \notin \mathcal{D} \ , \ \forall \, t \in [0,T]$$

REMARK 2.2 This theorem can be formally derived from the original FK theorem by taking

$$\begin{aligned} r(x) &= 0 \ , \ x \in \mathcal{D} \\ &= \infty \ , \ x \notin \mathcal{D} \end{aligned}$$

With this choice, the mean value is localized to the domain of $\mathcal{D}$. $\square$

2.8 Change of numéraire

The Girsanov theorem is a key tool in finance as it permits to considerably simplify the analytical or numerical computation of (2.31) by choosing an appropriate numéraire. By definition,

DEFINITION 2.11 *A numéraire is any positive continuous asset.*

Under the no-arbitrage condition, the drift of a numéraire, being an asset, is constrained to be the instantaneous interest rate r_t under a risk-neutral measure $\mathbb{P}$.

Before stating the Girsanov theorem, we motivate it with a simple problem consisting in the pricing of a European option depending on two assets S^1 and S^2 whose payoff at time T is the spread option

$$\max(S_T^2 - S_T^1, 0)$$

From (2.31), we know that the fair price at time $t \leq T$ is

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}}[\frac{D_{0T}}{D_{0t}} \max(S_T^2 - S_T^1, 0)|\mathcal{F}_t]$$

For our market model, we assume that the two assets S^1 and S^2 follow the diffusion processes under a risk-neutral measure $\mathbb{P}$

$$\frac{dS_t^1}{S_t^1} = r_t dt + \sigma^1(t, S^1).dW_t$$
$$\frac{dS_t^2}{S_t^2} = r_t dt + \sigma^2(t, S^2).dW_t$$

where W_t is a m-dimensional Brownian motion.[7]
Let us suppose that the two assets are valued in euros. As an alternative, the second asset could be valued according to the first one. In this case, the numéraire is the first asset and we can consider the dimensionless asset $X_t = \frac{S_t^2}{S_t^1}$. By using Itô formula, the dynamics of X_t is then

$$\frac{dX_t}{X_t} = (\sigma^2(t, S_2) - \sigma^1(t, S_1)).(dW_t - \sigma^1(t, S_1)dt)$$

Note that to get this result easily, it is better to apply the Itô formula on $\ln X_t$ beforehand. We observe that X_t has a non-trivial drift under $\mathbb{P}$.
Let us define

$$d\hat{W}_t = dW_t - \sigma^1(t, S_1)dt$$

and the dynamics of X_t can be written as

$$\frac{dX_t}{X_t} = (\sigma^2(t, S_2) - \sigma^1(t, S_1)).d\hat{W}_t$$

This rewriting is completely formal at this stage as $\hat{W}_t$ is not a Brownian motion according to $\mathbb{P}$. However, according to the Girsanov theorem, we can define a new measure $\hat{\mathbb{P}}$ equivalent to $\mathbb{P}$ such that $\hat{W}_t$ is a Brownian motion under $\hat{\mathbb{P}}$. Therefore, X_t is a local martingale under $\hat{\mathbb{P}}$. More precisely, we have

[7] $\sigma^1(t, S^1).dW_t$ means $\sum_{j=1}^m \sigma_j^1(t, S^1) dW_t^j$ with $\{W_t^j\}_j = 1, \cdots, m$ a m-dimensional uncorrelated Brownian motion.

THEOREM 2.7 Girsanov
Let $X_t \in \mathbb{R}^m$ be an Itô diffusion process of the form

$$dX_t = \lambda_t dt + dW_t \ , \ t \leq T \ , \ X_0 = 0$$

where $T \leq \infty$ a given constant, λ_t a $\mathcal{F}_t$-adapted process and $(W_t, \mathcal{F}_t, \mathbb{P})$ is a m-dimensional Brownian motion. Then M_t given by

$$M_t = e^{-\int_0^t \lambda_s dW_s - \frac{1}{2}\int_0^t \lambda_s^2 ds} \tag{2.38}$$

is a (positive) local martingale with respect to $\mathcal{F}_t$.
Let us define the measure $\hat{\mathbb{P}}$ on $\mathcal{F}_T$ by the Radon-Nikodym derivative

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}|_{\mathcal{F}_T} = M_T$$

If M_t is a $\mathcal{F}_t$-martingale (see remark 2.3 below) then $\hat{\mathbb{P}}$ is a probability measure on $\mathcal{F}_T$ and X_t is a m-dimensional Brownian motion according to $\hat{\mathbb{P}}$ for $0 \leq t \leq T$.

REMARK 2.3 Under the Novikov condition

$$\mathbb{E}^{\mathbb{P}}[e^{\frac{1}{2}\int_0^T \lambda_s^2 ds}] < \infty \tag{2.39}$$

the positive local martingale M_t given by (2.38) is a martingale. This condition can be weakened by

$$\mathbb{E}^{\mathbb{P}}[e^{\frac{1}{2}\int_0^T \lambda_s dW_s}] < \infty \tag{2.40}$$

□

This theorem is standard and the proof can be found in [34] for example (see also exercise 2.7). Note that the fact that M_t is a local martingale can be easily proved by observing that

$$dM_t = -\lambda_t M_t dt$$

as an application of Itô's lemma.
When dealing with conditional expectation, we have that for a $\mathcal{F}_T$-measurable r.v. X satisfying $\mathbb{E}^{\hat{\mathbb{P}}}[|X|] < \infty$,

$$\mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[X\frac{M_T}{M_t}|\mathcal{F}_t] \tag{2.41}$$

PROOF By using in the following order the definition of $\mathbb{E}^{\hat{\mathbb{P}}}$, the definition of the conditional expectation and the martingale property of M_t, we have

$\forall\ A \in \mathcal{F}_{t<T}$

$$\mathbb{E}^{\hat{\mathbb{P}}}[1_A \frac{1}{M_t}\mathbb{E}^{\mathbb{P}}[XM_T|\mathcal{F}_t]] = \mathbb{E}^{\mathbb{P}}[1_A \mathbb{E}^{\mathbb{P}}[XM_T|\mathcal{F}_t]]$$
$$= \mathbb{E}^{\mathbb{P}}[1_A X M_T]$$
$$= \mathbb{E}^{\hat{\mathbb{P}}}[1_A X]$$

$\square$

Coming back to our example, applying the Girsanov theorem, the change of measure from $\mathbb{P}$ to $\hat{\mathbb{P}}$ is (here $\lambda_t = -\sigma^1(t, S_t^1)$)

$$M_T \equiv \frac{d\hat{\mathbb{P}}}{d\mathbb{P}}|\mathcal{F}_T = e^{\int_0^T \sigma^1 dW_s - \frac{1}{2}\int_0^T (\sigma^1)^2 ds}$$
$$= \frac{S_T^1}{S_0^1 B_T} = \frac{S_T^1}{S_0^1} D_{0T} \tag{2.42}$$

Therefore the fair value is

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}}[\frac{D_{0T}}{D_{0t}} \max(S_T^2 - S_T^1, 0)|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\frac{D_{0T}}{D_{0t}} S_T^1 \max(X_T - 1, 0)|\mathcal{F}_t]$$
$$= \mathbb{E}^{\hat{\mathbb{P}}}[\frac{D_{0T}M_t}{D_{0t}M_T} S_T^1 \max(X_T - 1, 0)|\mathcal{F}_t]$$
$$= S_t^1 \mathbb{E}^{\hat{\mathbb{P}}}[\max(X_T - 1, 0)|\mathcal{F}_t]$$

In the second line, we have used (2.41) and in the last equation (2.42).
Before closing this section, let us present the general formula enabling to transform an Itô diffusion process under a measure $\mathbb{P}^1$ (associated to a numéraire N_1) into a new Itô process under the measure $\mathbb{P}^2$ (associated to a numéraire N_2).

Change of numéraire: General formula

Let us consider a n-dimensional Itô process X_t given under $\mathbb{P}^1$ by

$$\frac{dX_t}{X_t} = \mu_X^{N_1} dt + \sigma_X . dW_t^1$$

and under the numéraire $\mathbb{P}^2$ by

$$\frac{dX_t}{X_t} = \mu_X^{N_2} dt + \sigma_X . dW_t^2$$

Note that the diffusion terms are the same in the two equivalent measures $\mathbb{P}^1$ and $\mathbb{P}^2$ as the change of measure only affects the drift terms. Moreover, the writing of the dynamics for X_t using a "log-normal" form $\frac{dX_t}{X_t}$ is solely for convenience. Under the Girsanov theorem, the two measures differ by

$$dW_t^2 = dW_t^1 + \lambda_t dt$$

with

$$\mu_X^{N_1} = \mu_X^{N_2} + \sigma_X.\lambda_t \tag{2.43}$$

Here $\sigma_X.\lambda_t$ is a n-dimensional vector with components $\sum_{j=1}^m \sigma_{X\ j}^i \lambda_t^j$. The change of measure is

$$\frac{d\mathbb{P}^2}{d\mathbb{P}^1}|_{\mathcal{F}_T} \equiv M_T = e^{-\int_0^T \lambda_s dW_s^1 - \frac{1}{2}\int_0^T \lambda_s^2 ds}$$

If we price a European contract (with maturity T) and payoff f_T, the fair value can be written in two different manners using the measure $\mathbb{P}^1$ and $\mathbb{P}^2$

$$\begin{aligned}\mathcal{C} &= \mathbb{E}^{\mathbb{P}^1}[\frac{N_t^1}{N_T^1} f_T|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}^2}[\frac{N_t^2}{N_T^2} f_T|\mathcal{F}_t] \\ &= \mathbb{E}^{\mathbb{P}^2}[\frac{M_t}{M_T}\frac{N_t^1}{N_T^1} f_T|\mathcal{F}_t]\end{aligned}$$

Therefore,

$$\frac{M_t}{M_T} = \frac{N_t^2 N_T^1}{N_T^2 N_t^1}$$

Modulo an irrelevant multiplicative factor, we deduce that

$$M_t = \frac{N_t^2}{N_t^1}$$

Moreover, we assume that the two numéraires under the measure $\mathbb{P}^1$ follow the Itô diffusion processes

$$\begin{aligned}\frac{dN_1}{N_1} &= \mu^{N_1} dt + \sigma^1.dW_t^1 \\ \frac{dN_2}{N_2} &= \mu^{N_2} dt + \sigma^2.dW_t^1\end{aligned}$$

As we have $dM_t = -\lambda_t M_t dW_t^1$, we obtain the identification

$$\lambda_t = \sigma^1 - \sigma^2$$

and therefore from (2.43), we have

$$\mu_X^{N_2} = \mu_X^{N_1} + \sigma_X.(\sigma^2 - \sigma^1)$$

The result is summarized in the following proposition

PROPOSITION 2.1
Consider two numéraires N_1 and N_2 with Itô diffusion processes under $\mathbb{P}^1$ (associated to the numéraire N_1)

$$\frac{dN_1}{N_1} = \mu_1 dt + \sigma^1.dW_t^1$$
$$\frac{dN_2}{N_2} = \mu_2 dt + \sigma^2.dW_t^1$$

Then if we consider an Itô diffusion process given under $\mathbb{P}^1$ by

$$\frac{dX_t}{X_t} = \mu_X^{N_1} dt + \sigma_X.dW_t^1$$

then the process X_t under the measure $\mathbb{P}^2$ (associated to the numéraire N_2) is

$$\frac{dX_t}{X_t} = (\mu_X^{N_1} + \sigma_X.(\sigma^2 - \sigma^1))dt + \sigma_X.dW_t^2$$

In our previous example, the two processes S_t^1 and S_t^2 are written under the measure $\mathbb{P}^1$ associated to the numéraire S_t^1

$$\frac{dS_1}{S_1} = (r_t + (\sigma^1)^2)dt + \sigma^1.dW_t^1$$
$$\frac{dS_2}{S_2} = (r_t + \sigma^2.\sigma^1)dt + \sigma^2.dW_t^1$$

as the volatility associated to S_1 is σ^1 and the volatility associated to the money market account B_t is zero. From Itô's lemma, we can check that $\frac{S_2}{S_1}$ is a local martingale under $\mathbb{P}^1$ as expected

$$d\left(\frac{S_2}{S_1}\right) = \left(\frac{S_2}{S_1}\right)(\sigma^2 - \sigma^1).dW_t^1$$

In order to illustrate the power of the change of measure technique, we present a new measure, the *forward measure*, particularly useful to compute the fair value of an option (2.31) when the discount factor D_{0t} is a stochastic process.

Example 2.13 Forward and Forward measure
We want to price a European call option (with a maturity T and a strike K) and we assume that the discount factor D_{0t} is a stochastic process. The fair price is under a risk-neutral measure $\mathbb{P}$

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}}[\frac{D_{0T}}{D_{0t}} \max(S_T - K, 0)|\mathcal{F}_t] \tag{2.44}$$

with $D_{0t} = e^{-\int_0^t r_s ds}$. We set

$$P_{tT} \equiv \mathbb{E}^{\mathbb{P}}[\frac{D_{0T}}{D_{0t}}|\mathcal{F}_t]$$
$$= \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds}|\mathcal{F}_t]$$

the fair value of a contract paying 1 at the maturity T. It is called a *bond*. Since P_{tT} is a traded asset, its dynamics under the measure $\mathbb{P}$ is

$$\frac{dP_{tT}}{P_{tT}} = r_t dt + \sigma^P . dW_t$$

with σ^P the volatility of the bond P_{tT} (left unspecified). Moreover, we assume that under $\mathbb{P}$

$$\frac{dS_t}{S_t} = r_t dt + \sigma^S . dW_t$$

Doing a change of measure from $\mathbb{P}$ to $\mathbb{P}^T$ associated to the numéraire P_{tT} (called the forward measure), the dynamics of P_{tT} and S_t under $\mathbb{P}^T$ are

$$\frac{dP_{tT}}{P_{tT}} = (r_t + \sigma^P . \sigma^P) dt + \sigma^P . dW_t^T$$
$$\frac{dS_t}{S_t} = (r_t + \sigma^S . \sigma^P) dt + \sigma^S . dW_t^T$$

The Radon-Nikodym derivative is

$$M_T \equiv \frac{d\mathbb{P}^T}{d\mathbb{P}}|_{\mathcal{F}_T} = \frac{P_{TT}}{B_T} = D_{0T} \tag{2.45}$$

Moreover, we define the *forward* $f_t^T = \frac{S_t}{P_{tT}}$ ($f_T^T = S_T$ as $P_{TT} = 1$). As being the ratio of a traded asset S_t and a bond P_{tT}, f_t^T is a local martingale under the forward measure $\mathbb{P}^T$. Indeed, it can be checked using Itô's lemma that f_t^T is a local martingale

$$df_t^T = f_t^T \left(\sigma^S . dW_t^T - \sigma^P . dW_t^T \right)$$

The European call option can be written as

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}^T}[\frac{D_{0T}}{D_{0t}} \frac{d\mathbb{P}}{d\mathbb{P}^T} \max(f_T^T - K, 0)] = \mathbb{E}^{\mathbb{P}^T}[\frac{D_{0T}}{D_{0t}} \frac{M_t}{M_T} \max(f_T^T - K, 0)]$$

and finally using (2.45)

$$\mathcal{C} = P_{tT} \mathbb{E}^{\mathbb{P}^T}[\max(f_T^T - K, 0)] \tag{2.46}$$

Using the forward measure, the (stochastic) discount factor has disappeared in the pricing formula (2.44). ∎

We conclude this section with a PDE interpretation of the Girsanov transform for Itô diffusion processes.

PDE interpretation of the Girsanov transform

We restrict our discussion to one-dimensional Itô diffusion processes although everything we discuss below can be trivially extended to higher-dimensional processes.
We assume that under a measure $\mathbb{P}$ (not necessarily being a risk-neutral measure), the spot process S_t satisfies the SDE

$$dS_t = \mu(S_t)dt + \sigma(S_t)dW_t$$

and under a new measure $\hat{\mathbb{P}}$

$$dS_t = r(S_t)dt + \sigma(S_t)d\hat{W}_t$$

Let us introduce the process M_t defined by the SDE

$$dM_t = \left(\frac{\mu(S_t) - r(S_t)}{\sigma(S_t)}\right) M_t d\hat{W}_t$$

According to the Feynman-Kac theorem, $\hat{\mathcal{C}}(t, S, M) \equiv \mathbb{E}^{\hat{\mathbb{P}}}[M_T \Phi(S_T)|\mathcal{F}_t]$ satisfies the PDE

$$\partial_t \hat{\mathcal{C}}(t, S, M) + r(S)\partial_S \hat{\mathcal{C}}(t, S, M) + \frac{1}{2}\sigma(S)^2 \partial_S^2 \hat{\mathcal{C}}(t, S, M)$$
$$+\frac{1}{2}M^2 \left(\frac{\mu(S) - r(S)}{\sigma(S)}\right)^2 \partial_M^2 \hat{\mathcal{C}}(t, S, M) + M\left(\mu(S) - r(S)\right)\partial_{MS}\hat{\mathcal{C}}(t, S, M) = 0$$

with the terminal condition $\hat{\mathcal{C}}(T, S, M) = M\Phi(S)$. Let us try the ansatz $\hat{\mathcal{C}}(t, S, M) = M\mathcal{C}(t, S)$. We obtain that the function $\mathcal{C}(t, S)$ satisfies the PDE

$$(\partial_t + \mu(S)\partial_S + \frac{1}{2}\sigma(S)^2\partial_S^2)\mathcal{C}(t, S) = 0$$

with the terminal condition $\mathcal{C}(T, S) = \Phi(S)$. By the Feynman-Kac theorem, this is equivalent to

$$M_t \mathbb{E}^{\mathbb{P}}[\Phi(S_T)] = \mathbb{E}^{\hat{\mathbb{P}}}[M_T \Phi(S_T)]$$

This is the Girsanov transform.

2.9 Hedging portfolio

Up to this stage, the pricing problem has been our main focus. However, in order to reduce the risk, a specific strategy called a **hedging strategy** should also be used. For example, should we sell a European option characterized

by an unbounded payoff Φ_T at the maturity T, the loss could potentially be unlimited if we did nothing.
Let us assume that our market model is described by a money market account B_t, n traded assets $\{x_t^i\}_{i=1\cdots n}$ and m unobservable Itô processes $\{x_t^\alpha\}$ with $\alpha = n+1, \cdots, n+m$. The (hedging) strategy consists in buying and selling the assets x_t^i from $t=0$ to $t=T$ and investing the profit in the money market account. The amount of asset i that we hold at time t is noted Δ_t^i. From section 2.6, the value of a self-financing portfolio at the maturity T is

$$\pi_T = \mathbb{E}^{\mathbb{P}}[D_{tT}\Phi_T|\mathcal{F}_t] - \Phi_T + \int_t^T \sum_{i=1}^n \Delta^i(t,x)d\bar{x}_t^i \tag{2.47}$$

with $\bar{x}_t^i = x_t^i D_{0t}$. The first term is the fair value of the contract with payoff Φ_T as given by the pricing formula (2.31). The second term is the payoff Φ_T exercised at maturity T. The last term is the value of a self-financing portfolio (2.29). From (2.47), we introduce the notion of a complete market.

DEFINITION 2.12 Complete market *The payoff Φ_T is attainable (in the market $(x,B)_{t\in[0,T]}$) if there exists a self-financing portfolio π such that $\pi_T = 0$ $\mathbb{P}$-almost surely. If each payoff Φ_T is attainable, then the market is called complete. Otherwise, the market is called incomplete.*

Let us see under which conditions a market is complete. As the fair price $f(t,x_t)$ of an option with payoff Φ_T is a local martingale, we have that under a risk-neutral measure $\mathbb{P}$

$$df(t,x_t) = \sum_{i=1}^n \frac{\partial f(t,x_t)}{\partial x_i} d\bar{x}_t^i + \sum_{\alpha=n+1}^{n+m} \frac{\partial f(t,x_t)}{\partial x_\alpha}[dx_t^\alpha]$$

with $[dx_t^\alpha]$ the local martingale part of dx_t^α (i.e., we disregard the drift part). Integrating this equation between t and T and using that $f(t,x_t) = \mathbb{E}^{\mathbb{P}}[D_{tT}\Phi_T|\mathcal{F}_t]$, we obtain

$$\Phi_T \equiv f(T,x_T) = \mathbb{E}^{\mathbb{P}}[D_{tT}\Phi_T|\mathcal{F}_t] + \int_t^T \sum_{i=1}^n \frac{\partial f}{\partial x_i} d\bar{x}_t^i + \int_t^T \sum_{\alpha=n+1}^{n+m} \frac{\partial f}{\partial x_\alpha}[dx_t^\alpha]$$

By plugging this expression in (2.47), we obtain

$$\pi_T = \int_t^T \sum_{i=1}^n \left(\Delta^i(t,x_t) - \frac{\partial f(t,x_t)}{\partial x_i}\right) d\bar{x}_t^i - \int_t^T \sum_{\alpha=n+1}^{n+m} \frac{\partial f(t,x_t)}{\partial x_\alpha}[dx_t^\alpha]$$

Therefore if we choose

$$\Delta^i(t,x) = \frac{\partial f(t,x)}{\partial x_i}$$

and the market model is composed of traded assets only (plus a money market-account), then Φ_T is attainable and the market is complete. We say that we have a *dynamic Delta hedging* strategy which consists in holding $\Delta^i(t,x)$ asset i at time t. The resulting risk at maturity T cancels as the option Φ_T is attainable.
On the contrary, if we have unobservable Itô processes $\{x^\alpha\}$ such as a stochastic volatility (i.e., $[dx_t^\alpha] \neq 0$), the model is incomplete. More generally, we can prove that

THEOREM 2.8 Second theorem of asset pricing
A market defined by the assets $(S_t^0, S_t^1, \cdots, S_t^d)_{t\in[0,T]}$ (plus a money market account), described as Itô processes on $(\Omega, \mathcal{F}, \mathbb{P})$, is complete if and only if there is a **unique** *locale martingale measure $\mathbb{Q}$ equivalent to $\mathbb{P}$.*

2.10 Building market models in practice

The building of a market model is to be a two-step process: The first step consists in introducing the right financial traded assets relevant to our pricing. One could model equity assets, foreign exchange (Fx) rates, commodity assets .... In the second step, we model the fixed income rate which enter the definition of the money market account. Once the dynamics of the interest rate has been fixed, we should specify an Itô diffusion process for traded assets. In this context, a choice of measure should be done. Usually, we use the risk-neutral or the forward measure. Let us now describe the modeling in different cases.

2.10.1 Equity asset case

For an equity asset, we know that under a risk-neutral measure (associated to the money market account as a numéraire) the drift is constrained to be the instantaneous interest rate r_t.
The forward $f_t^T = \frac{S_t}{P_{tT}}$ is a strictly positive local martingale in the forward measure $\mathbb{P}^T$ (associated to the bond P_{tT} as numéraire). We can use the following proposition 2.2 which characterizes a strictly positive local martingale. Previous to that, we recall the definition of a quadratic variation process:

DEFINITION 2.13 Quadratic variation *For X_t a continuous stochastic process, the quadratic variation is defined by*

$$< X, X >_t (\omega) = \lim \sup_{\Delta t_i \to 0} \sum_{i=1}^{n-1} |X_{t_{i+1}}(\omega) - X_{t_i}(\omega)|^2$$

where $0 = t_1 < t_2 < \cdots < t_n = t$ *and* $\Delta t_i = t_{i+1} - t_i$. *If* X_t *and* Y_t *are two continuous stochastic processes, we define*

$$< X, Y >_t = \frac{1}{2} \left(< X + Y, X + Y >_t - < X, X >_t - < Y, Y >_t \right)$$

Note that for $X_t = \int_0^t \sigma_s dW_s$ *with* σ_t *an* $\mathcal{F}_t$*-adapted process, we have*

$$< X, X >_t = \int_0^t \sigma_s^2 ds$$

PROPOSITION 2.2

If f_t *is a strictly positive local martingale then there exists a local martingale* X_t *such that*

$$f_t = f_0 e^{\left(X_t - \frac{1}{2} < X, X >_t \right)}$$

Moreover if $< X, X >_t$ *is absolutely continuous meaning that there exists a* $L^2(d\mathbb{P} \times [0, t])$ *r.v.,* $\sigma.$*, such that*

$$< X, X >_t = \int_0^t \sigma_s^2 ds$$

then

$$X_t = \int_0^t \sigma_s . dW_s$$

with W_t *a Brownian motion.* σ_s *is called the (stochastic) volatility.*

In particular, if we assume that the forward and the bond are Itô processes, the dynamics of f_t^T under $\mathbb{P}^T$ is

$$df_t^T = (\sigma_t - \sigma_t^P).dW_t^T$$

with σ_t^P the volatility of the bond P_{tT}. In the risk-neutral measure $\mathbb{P}$, we have

$$\frac{dS_t}{S_t} = r_t dt + \sigma_t . dW_t$$

At this stage, the form of σ and σ^P and the number of Brownians we use to drive the dynamics of f_t^T are unspecified. We will come back to this point when we discuss fixed income rate modeling.

2.10.2 Foreign exchange rate case

For Fx rates, the specification of the dynamics is slightly more evolved. Let the exchange rate between two currencies (domestic and foreign) be $S_t^{d/f}$. By convention, one unit in the foreign currency corresponds to $S_t^{d/f}$ in the domestic currency. The domestic (resp. foreign) money market account is noted D_t^d (resp. D_t^f) and the domestic (resp. foreign) bond between t and T is P_{tT}^d (resp. P_{tT}^f). We also denote $\mathbb{P}^d$ (resp. $\mathbb{P}^f$) the domestic (resp. foreign) risk-neutral measure.

Let us consider two different strategies:

The first strategy consists in investing one unit in foreign currency at t up to the maturity T and then convert the value in the domestic currency using the exchange rate $S_T^{d/f}$ at T. The undiscounted portfolio valued in the domestic currency at T is $\frac{D_t^f}{D_T^f} S_T^{d/f}$.

In the second strategy, we directly convert our initial unit investment in the domestic currency at t and then invest in the domestic money market account. The value of this new portfolio, still valued in the domestic currency, is $\frac{D_t^d}{D_T^d} S_t^{d/f}$.

In order to avoid arbitrage, we should impose that the expectations of these two different discounted portfolios are equal

$$\mathbb{E}^{\mathbb{P}^d}[\frac{D_T^d D_t^f}{D_t^d D_T^f} S_T^{d/f} | \mathcal{F}_t] = S_t^{d/f}$$

equivalent to

$$\mathbb{E}^{\mathbb{P}^d}[\frac{D_T^d}{D_T^f} S_T^{d/f} | \mathcal{F}_t] = \frac{D_t^d}{D_t^f} S_t^{d/f}$$

Therefore, $\frac{D_t^d}{D_t^f} S_t^{d/f}$ should be a (local) martingale under the domestic risk-neutral measure $\mathbb{P}^d$. As $\frac{D_t^d}{D_t^f} = e^{-\int_0^t (r_d - r_f) ds}$, the drift of the exchange rate $S_t^{d/f}$ is constrained to be the difference between the domestic and foreign rates:

$$\frac{dS_t^{d/f}}{S_t^{d/f}} = (r_d - r_f)dt + \sigma_{d/f}.dW_t^d \tag{2.48}$$

The product of the Fx forward (defined as $f_t^{d/f} = \frac{S_t^{d/f} P_{tT}^f}{P_{tT}^d}$) with the domestic bond P_{tT}^d is a contract paying 1 in foreign currency valued in the domestic currency. Therefore $f_t^{d/f}$ is a local martingale under the domestic forward measure $\mathbb{P}_d^T$. Its dynamics under $\mathbb{P}_d^T$ is

$$\frac{df_t^{d/f}}{f_t^{d/f}} = \sigma_{d/f}.dW_t^d - \sigma_{P^d}.dW_t^d + \sigma_{P^f}.dW_t^d$$

with σ_S the volatility of the Fx rate and σ_{P^d} (resp. σ_{P^f}) the volatility of the domestic (resp. foreign) bond.
To conclude this section, we consider the dynamics of an asset, valued in the foreign currency, under the domestic risk-neutral measure $\mathbb{P}^d$. In $\mathbb{P}^f$, we have

$$\frac{dS_t^f}{S_t^f} = r_f dt + \sigma_S . dW_t^f$$

By definition, the process $S_t^{d/f} S_t^f$ is the foreign asset valued in the domestic currency and therefore should be driven under $\mathbb{P}^d$ by

$$\frac{dS_t^{d/f} S_t^f}{S_t^{d/f} S_t^f} = r_d dt + \sigma_S . dW_t^d + \sigma_{d/f} . dW_t^d$$

for which we deduce that

$$d \ln \left(S_t^{d/f} S_t^f \right) = \left(r_d - \frac{1}{2} \sigma_S . \sigma_S - \frac{1}{2} \sigma_{d/f} . \sigma_{d/f} - \sigma_S . \sigma_{d/f} \right) dt \\ + \sigma_S . dW_t^d + \sigma_{d/f} . dW_t^d$$

Then, by using (2.48), we deduce that the dynamics of S_t^f under $\mathbb{P}^d$ is

$$\frac{dS_t^f}{S_t^f} = \left(r_f - \sigma_{d/f} . \sigma_S \right) dt + \sigma_S . dW_t^d$$

2.10.3 Fixed income rate case

As previously discussed, the modeling of equity assets, Fx rates or market models in general depends on the modeling of the money market account (resp. bond) if we work in a risk-neutral measure (resp. forward measure). The modeling of a fixed-income rate is slightly more complex as there is no standard way of doing this. In the following section, we quickly review the main models that have been introduced in this purpose: short-rate models, HJM model and Libor Market Models. Details and extensive references can be found in [7].

2.10.3.1 Short rate models

In the framework of short-rate models, one decides to impose the dynamics of the instantaneous interest rate r_t. As r_t is not a traded financial contract, there is no restriction under the no-arbitrage condition on its dynamics. For example, we can impose the time-independent one-dimensional Itô diffusion process

$$dr_t = a(r_t) dt + b(r_t) dW_t$$

TABLE 2.1: Example of one-factor short-rate models.

one-factor short-rate model	SDE
Vasicek-Hull-White	$dr_t = k(\theta - r_t)dt + \sigma dW_t$
CIR	$dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t$
Dooleans	$dr_t = kr_t dt + \sigma r_t dW$
EV-BK	$dr_t = r_t(\eta - \alpha \ln(r_t))dt + \sigma r_t dW_t$

See table (2.1) for a list of functions a and b that are commonly used. By definition, the discount factor is

$$D_t = e^{-\int_0^t r_s ds}$$

and the fair pricing formula for a payoff Φ_T in a risk-neutral measure $\mathbb{P}$ is

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds}\Phi_T]$$

Moreover, the value of a bond quoted at t maturing at T is

$$P_{tT} = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds}]$$

Note that using the Feynman-Kac theorem, one gets under $\mathbb{P}$

$$\frac{dP_{tT}}{P_{tT}} = r_t dt + \sigma^P.dW_t \tag{2.49}$$

2.10.3.2 HJM model

The main drawback of short-rate models is that it is slightly suspicious to assume that the forward curve defined by

$$f_{tT} = -\frac{\partial \ln(P_{tT})}{\partial T}$$

is driven by the short rate $r_t = \lim_{T\to t} f_{tT}$. This situation can be improved by directly modeling the forward curve f_{tT}. This is the purpose of the HJM framework. A priori, f_{tT} is not a financial contract and therefore there should be no restriction under the no-arbitrage condition on its dynamics. However, we should have (2.49) and we will see that this constraint imposes the drift of the forward as a function of its volatility. Without loss of generality, the forward f_{tT} follows the SDE

$$df_{tT} = \mu(t,T)dt + \sigma(t,T)d\hat{W}_t$$

We will now derive the SDE followed by the bond P_{tT}. Starting from the definition, we have

$$\begin{aligned}
\ln(P_{tT}) &= -\int_t^T f_{ts} ds \\
&= -\int_t^T ds(f_{0s} + \int_0^t \mu(u,s)du + \int_0^t \sigma(u,s)d\hat{W}_u)
\end{aligned}$$

Itô's lemma gives

$$d\ln(P_{tT}) = \left(f_{0t} + \int_0^t \mu(u,t)du + \int_0^t \sigma(u,t)d\hat{W}_u\right) - \int_t^T ds\left(\mu(t,s)dt + \sigma(t,s)d\hat{W}_t\right)$$

Recognizing the first bracket as $r_t \equiv f_{tt}$, we obtain

$$d\ln(P_{tT}) = \left(r_t - \int_t^T \mu(t,s)ds\right)dt - \int_t^T \sigma(t,s)dsd\hat{W}_t$$

and

$$\frac{dP_{tT}}{P_{tT}} = \left(r_t dt - \int_t^T \mu(t,s)ds + \frac{1}{2}\int_t^T\int_t^T \sigma(t,u)\sigma(t,s)dsdu\right)dt - \int_t^T \sigma(t,s)dsd\hat{W}_t$$

Therefore, there is no-arbitrage if there exists a $\mathcal{F}_t$-adapted function λ_t such that

$$\int_t^T \mu(t,s)ds = \frac{1}{2}\int_t^T\int_t^T \sigma(t,u)\sigma(t,s)dsdu - \lambda_t\int_t^T \sigma(t,s)ds$$

Differentiating with respect to T, we obtain

$$\mu(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du - \lambda_t\sigma(t,T)$$

Finally, under the risk-neutral measure $\mathbb{P}$, the forward curve follows the SDE

$$df_{tT} = \sigma(t,T)\int_t^T \sigma(t,u)du + \sigma(t,T)dW_t$$

where we have set $dW_t = d\hat{W}_t - \lambda_t dt$.
Note that from a HJM model, we can derive a short rate model $r_t = \lim_{t\to T} f_{tT}$ given by

$$r_t = f_{0t} + \int_0^t \sigma(s,t)ds\int_s^t \sigma(s,u)du + \int_0^t \sigma(s,t)dW_s$$

We should note that this process is not necessarily Markovian.

2.10.3.3 Libor market models

The main drawback of the HJM model is that the (continuous) forward curve is an idealization and is not directly observable on the market. The forward rates observed on the market are the *Libor rates* corresponding to the rates between two future dates, usually spaced out three or six months apart. We note $L_i(t)$ the value of the Libor forward rate at time t between two future dates T_{i-1} and T_i (we say the *tenor* structure). The Libor Market Model (LMM) corresponds to the modeling of the forward rate curve via the Libor variables. The forward bond at t between T_{i-1} and T_i is by definition

$$P(t, T_{i-1}, T_i) \equiv \frac{1}{1 + \tau L_i(t)}$$

Besides the no-arbitrage condition requires that

$$P(t, T_{i-1}, T_i) = \frac{P_{tT_i}}{P_{tT_{i-1}}}$$

which gives

$$L_i(t) = \frac{1}{\tau_i}\left(\frac{P_{tT_{i-1}}}{P_{tT_i}} - 1\right)$$

As the product of the bond P_{tT_i} with the forward rates $L_i(t)$ is a difference of two bonds with maturities T_{i-1} and T_i, $\frac{1}{\tau_i}(P_{tT_{i-1}} - P_{tT_i})$, therefore a traded asset, $L_i(t)$ is a (local) martingale under $\mathbb{P}^i$, the (forward) measure associated with the numéraire P_{tT_i}. Therefore we assume the following driftless dynamics under $\mathbb{P}^i$

$$dL_i(t) = L_i \sigma_i(t) dW_t^i$$

$\sigma_i(t)$ can be a deterministic function; in this case we get the *BGM model*, or a stochastic process (which can depend on the Libors). In chapter 8, we will discuss in details the BGM model and its extension including stochastic volatility processes. In particular, we will explain how the asymptotic methods presented in this book can be used to calibrate such a model.

2.10.4 Commodity asset case

Commodity assets are particular as their dynamics does not depend on the modeling of interest rates as it is the case for equity assets and Fx rates because it is impossible to build a self-financing portfolio with commodity assets (which can not be stored). On the commodity market, one can trade commodity future f_{tT} which is the contract paying the spot $S_T \equiv f_{TT}$ at the maturity T. Being a traded asset, f_{tT} should be a local martingale under the forward measure $\mathbb{P}^T$

$$df_{tT} = \sigma . dW_t$$

where σ is an unspecified volatility.
Note that the spot defined as $S_t = \lim_{t\to T} f_{tT}$ does not necessarily follow a Markovian dynamics. This phenomenon is analog to what is observed in the HJM model for the instantaneous interest rate.

2.11 Problems

Exercises 2.1 Central limit theorem
We have N independent r.v. $x_i = \{+1, -1\}$ such that $p(+1) = p(-1) = \frac{1}{2}$. We define the sum $X = \sum_{i=1}^{N} x_i$

1. Compute the mean-value $\mathbb{E}[X]$.
2. Compute the variance $\mathbb{E}[X^2] - \mathbb{E}[X]^2$.
3. Compute the probability $P(M)$ such that $X = M$.
4. Take the limit $N \to \infty$ of $\frac{\sqrt{N}}{2} P(\sqrt{N}x)$ with $x \in \mathbb{R}$.
5. Deduce the result above using the central limit theorem [26] (Hint: use the Stirling formula).

Exercises 2.2
Prove that (2.1) satisfies the definition of a measure.

Exercises 2.3 Wick's identity
Let W_t be a Brownian motion.

1. Prove that

$$\mathbb{E}^{\mathbb{P}}[e^{iuW_t}] = e^{-\frac{u^2 t}{2}}$$

2. Deduce the following formula, called the Wick identity, $\forall n \in \mathbb{N}$

$$\mathbb{E}^{\mathbb{P}}[(W_t)^{2n}] = \frac{(2n-1)!}{2^{n-1}(n-1)!} t^n$$
$$\mathbb{E}^{\mathbb{P}}[(W_t)^{2n+1}] = 0$$

Exercises 2.4 Exponential martingale
Let us define $M_t = \exp(W_t - \frac{t}{2})$ with W_t a Brownian motion. Prove that M_t is a martingale using the definition 2.9. Then prove that M_t is a (local) martingale using Itô's lemma.

Exercises 2.5 Itô's isometry
Let us consider a deterministic function $f : \mathbb{R} \to \mathbb{R}$.

1. Prove that

$$\mathbb{E}^{\mathbb{P}}[\int_0^t f(s)dW_s] = 0$$
$$\mathbb{E}^{\mathbb{P}}[\left(\int_0^t f(s)dW_s\right)^2] = \int_0^t f(s)^2 ds$$

   Hint: Assume that $f(s)$ can be approximated by simple functions $\chi(s) = \chi_i 1_{s_i<s<s_{i+1}}$ (for the L^2 norm).

   These expressions can be extended to the case where $f(s,\omega)$ is an adapted process. We have the Itô isometry formula:

$$\mathbb{E}^{\mathbb{P}}[\int_0^t f(s,\omega)dW_s] = 0$$
$$\mathbb{E}^{\mathbb{P}}[\left(\int_0^t f(s,\omega)dW_s\right)^2] = \int_0^t \mathbb{E}^{\mathbb{P}}[f(s,\omega)^2]ds$$

2. Let $Z \in N(0,\sigma^2)$ be a normal r.v. with zero mean and variance equal to σ^2. Prove that

$$\mathbb{E}[e^Z] = e^{\frac{\sigma^2}{2}}$$

3. Deduce that

$$\mathbb{E}^{\mathbb{P}}[e^{\int_0^t f(s)dW_s}] = e^{\frac{1}{2}\int_0^t f(s)^2 ds}$$

Exercises 2.6 Ornstein-Uhlenbeck SDE
The Ornstein-Uhlenbeck process is defined by the following SDE

$$dX_t = -\alpha X_t dt + \sigma dW_t$$

with the initial condition $X_0 = x_0 \in \mathbb{R}$.

1. Is there a unique solution?

   Hint: Verify the Lipschitz condition.

2. Compute the mean-value $\mathbb{E}^{\mathbb{P}}[X_t]$.
3. Compute the variance $\mathbb{E}^{\mathbb{P}}[X_t^2] - \mathbb{E}^{\mathbb{P}}[X_t]^2$.
   Hint: You should need the result 1 in problem (2.5).
4. Deduce the distribution for X_t.

Exercises 2.7 Simplified Girsanov's theorem
Let us suppose that $\hat{W}_t$ satisfies the SDE

$$d\hat{W}_t = dW_t + \lambda(t)dt$$

with $\lambda(\cdot)$ a deterministic function and W_t a Brownian with respect to a measure $\mathbb{P}$. We want to prove that $\hat{W}_t$ is a Brownian motion with respect to the measure $\hat{\mathbb{P}}$ such that

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}|_{\mathcal{F}_t} = e^{-\int_0^t \lambda(s)dW_s - \frac{1}{2}\int_0^t \lambda(s)^2 ds}$$

1. Prove that

$$\mathbb{E}^{\hat{\mathbb{P}}}[e^{iu\hat{W}_t}] = \mathbb{E}^{\mathbb{P}}[\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}e^{iu\hat{W}_t}]$$

2. By completing the square, prove that

$$\mathbb{E}^{\hat{\mathbb{P}}}[e^{iu\hat{W}_t}] = \mathbb{E}^{\mathbb{P}}[e^{-\int_0^t \lambda(s)dW_s - \frac{1}{2}\int_0^t \lambda(s)^2 ds}e^{iu\hat{W}_t}] = e^{-\frac{u^2 t}{2}}$$

3. Deduce that $\hat{W}_t$ is a Brownian motion according to $\hat{\mathbb{P}}$.

Exercises 2.8 Barrier option in Bachelier's model

In the Bachelier model, a forward f_t follows a normal process

$$df_t = dW_t$$

Here we have taken a unit volatility. Note that this model has a severe drawback as the forward can become negative.
In this exercise, we want to price a barrier option (i.e., down-and-out call) which pays a call with strike K at the maturity T if the forward has not reached the barrier level B. We take $K > B$.

1. By using the localized Feynman-Kac theorem, prove that the fair value $\mathcal{C}$ of the barrier option satisfies the PDE

$$\partial_t \mathcal{C}(t,f) + \frac{1}{2}\partial_f^2 \mathcal{C}(t,f) = 0 \ , \forall\ f > B$$
$$\mathcal{C}(T,f) = \max(f-K,0)1_{f>B}$$
$$\mathcal{C}(t,B) = 0 \ , \forall\ t \in [0,T]$$

2. Show that the solution can be written as

$$\mathcal{C}(t, f) = u(t, f) - u(t, 2B - f)$$

with u a call option with strike K and maturity T. This is called the *image's method.*

Exercises 2.9 P&L Theta-Gamma

We study the variation of a trader's self-financing portfolio composed of an asset with price S_t at time t and a European option with payoff $\phi(S_T)$ at the maturity date T. The fair value of the option is $\mathcal{C}(t, S_t)$ at time t. The portfolio's value is at time t:

$$\pi_t = \mathcal{C}(t, S_t) - \Delta_t S_t$$

We take zero interest rate.

1. We assume that on the market the asset follows a Black-Scholes log-normal process with a volatility σ_{real}. By using Itô's formula, prove that the infinitesimal variation of the self-financing *delta-hedge* portfolio between t and $t + dt$ is

$$d\pi_t = \partial_t \mathcal{C} dt + \frac{1}{2} S^2 \partial_S^2 \mathcal{C} \sigma_{\text{real}}^2 dt$$

The term $\partial_t \mathcal{C}$ (resp. $\partial_S^2 \mathcal{C}$) is called the *Theta* (resp. the *Gamma*) of the option.

2. We assume that the option was priced using a Black-Scholes model with a volatility σ_{model}. Prove that

$$d\pi_t = \frac{1}{2} S^2 \partial_S^2 \mathcal{C} \left(\sigma_{\text{real}}^2 - \sigma_{\text{model}}^2 \right) dt$$

3. Deduce that the expectation value $\mathbb{E}^{\mathbb{P}}[\Delta \pi_T]$ of the variation of the portfolio between $t = 0$ to $t = T$ is

$$\mathbb{E}^{\mathbb{P}}[\Delta \pi_T] = \frac{1}{2} \int_0^T \mathbb{E}^{\mathbb{P}}[S^2 \partial_S^2 \mathcal{C}] \left(\sigma_{\text{real}}^2 - \sigma_{\text{model}}^2 \right) dt$$

This expression is called the profit and loss (P&L) Theta-Gamma. For a European call option, the Gamma is positive and therefore, if $\sigma_{\text{real}} > \sigma_{\text{model}}$ (resp. $\sigma_{\text{real}} < \sigma_{\text{model}}$) , the P&L is positive (resp. negative). When the market realized volatility σ_{real} equals the model volatility σ_{model}, the P&L vanishes. In this context, σ_{model} is called the *break-even volatility.*

Chapter 3

Smile Dynamics and Pricing of Exotic Options

Abstract We review the definition of the implied volatility in equity markets and present examples of exotic options which give an insight into the dynamics of the implied volatility. Similar definitions and examples are also presented for fixed-income markets.

3.1 Implied volatility

The Black-Scholes *implied volatility* $\sigma_{\mathrm{BS}\,t}(K,T)$, starting at t, also called *smile*, is defined as the "wrong number" which, when put in the Black-Scholes formula $\mathcal{C}^{\mathrm{BS}}$ for a European call option with strike K and maturity T quoted at t, reproduces the fair price on the market $\mathcal{C}^{\mathrm{mkt}}$. More precisely, we have the following definition

DEFINITION 3.1 *The implied volatility $\sigma_{\mathrm{BS}\,t}(K,T)$ is such that*

$$\mathcal{C}^{\mathrm{mkt}}(K,T|t) = \mathcal{C}^{\mathrm{BS}}(K,T,\sigma_{\mathrm{BS}t}(K,T)|S,t)$$

where

$$\mathcal{C}^{\mathrm{BS}}(K,T,\sigma|S,t) = SN(d_+) - KP_{tT}N(d_-) \tag{3.1}$$

with

$$d_\pm = -\frac{\ln(\frac{KP_{tT}}{S})}{\sigma_{\mathrm{BS}\,t}(K,T)\sqrt{T-t}} \pm \frac{\sigma_{\mathrm{BS}\,t}(K,T)\sqrt{T-t}}{2}$$

Here $N(x) = \int_{-\infty}^{x} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$ is the cumulative normal distribution, P_{tT} the bond value expiring at T and S is the spot.

The solution is unique because $\mathcal{C}^{\mathrm{BS}}$ is strictly increasing in σ. In practice, the numerical solution can be found using a dichotomy algorithm. As a function of the strike K and the maturity T, $\sigma_{\mathrm{BS}\,t}(K,T)$ describes a surface or a membrane if the time t is added. For a fixed maturity T (and time t), the form of $\sigma_{\mathrm{BS}\,t}$ as a function of the strike can vary according to the market considered. In equity markets (resp. Fx markets), the smile is a decreasing (resp. parabolic) function of the strike. An important characteristic of the smile is its skew defined as the derivative of the smile according to the log of the strike at-the-money (i.e., the strike is fixed to the spot):

DEFINITION 3.2 *Volatility Skew:*

$$\mathcal{S}(T) = K\frac{\partial \sigma_{\mathrm{BS}}}{\partial K}|_{K=S}$$

We can also define the skew at-the-money forward (i.e., the strike is fixed to the forward f_t^T):

$$\mathcal{S}_F(T) = K\frac{\partial \sigma_{\mathrm{BS}}}{\partial K}|_{K=f_t^T}$$

We will now describe several useful properties satisfied by the smile which restrain the form that the implied volatility surface can take.

- Let us consider a time-homogeneous stochastic volatility model given by the following SDE under the forward measure $\mathbb{P}^T$

$$\frac{df_t^T}{f_t^T} = \sigma_t dW_t$$

with σ_t an adapted process and f_t^T the forward. We assume that conditional on the filtration generated by the process σ_t, the forward is log-normal. Then the implied volatility is a function of the moneyness $m = K/f_t$, the time-to-maturity $\tau = T - t$ and the volatility σ_t

$$\sigma_{\mathrm{BS}\,t}(K,T) = \sigma_t \Phi(m, \sigma_t^2 \tau) \tag{3.2}$$

PROOF This can be derived using the fact that the parameters $[\sigma_t^2 \tau]$ and $m = K/f_t$ are dimensionless. Thus, if we write the implied volatility

$$\sigma_{\mathrm{BS}\,t}(K,T) = \sigma_t \Phi(\tau, f_t, K, \sigma_t)$$

the general function $\Phi(\cdot)$ must be dimensionless and therefore depends on dimensionless parameters only, i.e., m and $\sigma_t \tau^2$ $\square$

- Given the inequality $-1 < P_{tT}^{-1} \frac{\partial \mathcal{C}^{\text{mkt}}}{\partial K} < 0$,[1] we derive the following (rough) bounds for the skew at-the-money forward

$$-\frac{N(d_-)}{\sqrt{T-t}n(d_-)}|_{K=f_t^T} \leq -\mathcal{S}_F(T) \leq \frac{N(d_+)}{\sqrt{T-t}n(d_+)}|_{K=f_t^T}$$

with $n(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$.

PROOF From the Black-Scholes formula (3.1), we have

$$-P_{tT}^{-1} \frac{\partial \mathcal{C}^{\text{mkt}}}{\partial K}|_{K=f_t^T} = N(d_-)|_{K=f_t^T} - \mathcal{S}_F(T) n(d_-)|_{K=f_t^T} \sqrt{T-t}$$

From the inequality $0 < -P_{tT}^{-1} \frac{\partial \mathcal{C}^{\text{mkt}}}{\partial K} < 1$ and $(N(d_+) + N(d_-))\,|_{K=f_t^T} = 1$, we obtain the bounds above. ∎

Example 3.1 Merton model
The Merton model is given by the SDE on the forward f_t^T

$$df_t^T = \sigma(t) f_t^T dW_t$$

Thanks to a change of local time $t' = \int_0^t \sigma(s)^2 ds$, we obtain the GBM

$$df_{t'}^T = f_{t'}^T dW_{t'}$$

The implied volatility at the maturity T is therefore

$$\sigma_{\text{BS}\ t}(T)^2 (T-t) = \int_t^T \sigma(s)^2 ds$$

∎

By definition, European call-put options are sensitive to the implied volatility given a specific strike and maturity. In the following section, we will see that a general European option quoted at time t with a maturity date T on a single asset depends on the full implied volatility at T: $K \to \sigma_{\text{BS}\ t}(K, T)$.

3.2 Static replication and pricing of European option

A European option on a single stock or index is characterized by a payoff $\phi(S_T)$ depending only on the price S_T of a single stock or an index at the

[1] $\frac{\partial \mathcal{C}^{\text{mkt}}}{\partial K} = -P_{tT} \mathbb{E}^{\mathbb{P}^T}[1(S_T - K)]$.

maturity date T. This payoff $\phi(S_T)$ can be decomposed (we say *replicated*) over an infinite sum of calls and puts of different strikes as written below. Firstly, we have the following identity for $S_T > 0$

$$\phi(S_T) = \phi(S_t) + \phi'(S_t)(S_T - S_t) + \int_0^{S_t} \phi''(K) \max(K - S_T, 0) dK + \int_{S_t}^{\infty} \phi''(K) \max(S_T - K, 0) dK \quad (3.3)$$

The prime indicates a derivative with respect to S. The derivation is done according to the distribution theory where $\phi(S_T)$ is considered as a distribution belonging to the space $\mathcal{D}'$. For example, the first derivative of a call payoff $\max(x - K, 0)$ is the Heaviside function $1(x - K)$ and the second derivative is the Dirac function $\delta(x - K)$.

PROOF We take $x > 0$ and start from the representation of the function $f(\cdot)$ as

$$\begin{aligned} f(x) &= \int_0^{\infty} f(K)\delta(K - x) dK \\ &= \int_0^{x^*} f(K) \frac{\partial^2}{\partial K^2} \max(K - x, 0) dK + \int_{x^*}^{\infty} f(K) \frac{\partial^2}{\partial K^2} \max(x - K, 0) dK \end{aligned}$$

Integrating by parts twice gives (3.3) with $x \equiv S_T$ and $x^* \equiv S_t$ ∎

As the fair value of an option is given in the forward measure $\mathbb{P}^T$ by

$$P_{tT} \mathbb{E}^{\mathbb{P}^T}[\phi(S_T)|\mathcal{F}_t]$$

multiplying both sides of (3.3) by the bond value P_{tT} and taking the pricing operator $\mathbb{E}^{\mathbb{P}^T}[\cdot|\mathcal{F}_t]$, we obtain the fair price of a European option with payoff $\phi(S_T)$

$$P_{tT} \mathbb{E}^{\mathbb{P}^T}[\phi(S_T)|\mathcal{F}_t] = P_{tT}\phi(S_t) + \phi'(S_t)S_t(1 - P_{tT}) + \int_0^{S_t} \phi''(K)\mathcal{P}^{\mathrm{BS}}(t, K, T) + \int_{S_t}^{\infty} \phi''(K)\mathcal{C}^{\mathrm{BS}}(t, K, T) \quad (3.4)$$

with

$$\mathcal{C}^{\mathrm{BS}}(t, K, T) = P_{tT} \mathbb{E}^{\mathbb{P}^T}[\max(S_T - K, 0)|\mathcal{F}_t]$$

and

$$\mathcal{P}^{\mathrm{BS}}(t, K, T) = P_{tT} \mathbb{E}^{\mathbb{P}^T}[\max(K - S_T, 0)|\mathcal{F}_t]$$

the fair price of European call and put options with strike K and maturity T quoted at $t \leq T$. We have used that

$$P_{tT}\mathbb{E}^{\mathbb{P}}[S_T] = P_{tT}\mathbb{E}^{\mathbb{P}}[f_T^T] = P_{tT}f_t^T = S_t$$

We observe that the fair value of the payoff $\phi(S_T)$ is completely model independent as it can be decomposed over an infinite sum of European call-put options that are quoted on the market. Note that in practice only a finite number of strikes are quoted on the market and the pricing of (3.4) requires the extrapolation of the implied volatility outside the known strikes (the so-called *wings*). The payoff ϕ is therefore sensitive to the wings of the implied volatility.

A common (European) derivative product sensitive to the wings of the implied volatility is a variance swap option.

Example 3.2 Variance swap

A variance swap option is a contract which pays at the maturity date T the (daily) realized variance of the stock from t to T minus a strike. The strike is called the variance swap fair value, noted VS below. The payoff at T is

$$\frac{1}{N}\sum_{i=0}^{N-1} \ln\left(\frac{S_{i+1}}{S_i}\right)^2 - \mathrm{VS}$$

where S_i is the stock price at t_i (with $t_0 = t$ and $t_N = T$). By market convention, the strike VS is fixed such that the price of the option quoted at t is zero. Therefore VS is

$$\mathrm{VS} = \mathbb{E}^{\mathbb{P}^T}[\frac{1}{N}\sum_{i=0}^{N-1} \ln\left(\frac{S_{i+1}}{S_i}\right)^2 |\mathcal{F}_t] \tag{3.5}$$

where we have used the forward measure $\mathbb{P}^T$ for convenience sake. Note that the discount factor P_{tT} cancels out.

Assumption 1: We assume that the stock price S_t can be described by an Itô diffusion process and that S_t follows the SDE in a risk-neutral measure $\mathbb{P}$

$$\frac{dS_t}{S_t} = r_t dt + \sigma_t . dW_t$$

We make no assumptions on r_t and σ_t, which therefore can be either deterministic or stochastic processes.

In the large N limit, VS, given by (3.5), converges to the integral

$$\mathrm{VS} = \frac{1}{T-t}\int_t^T \mathbb{E}^{\mathbb{P}^T}[\sigma_s^2|\mathcal{F}_t]ds$$

Then, thanks to a change of measure, the process S_t in the forward measure $\mathbb{P}^T$ is

$$\frac{dS_t}{S_t} = (r_t + \sigma_t.\sigma_t^P)dt + \sigma_t.dW_t^T \tag{3.6}$$

with σ_t^P the volatility of the bond P_{tT}. By applying Itô's lemma on $\ln(S_t)$, we have

$$\int_t^T \frac{\sigma_s^2}{2} ds = -\ln\left(\frac{S_T}{S_t}\right) + \int_t^T (r_s + \sigma_s.\sigma_s^P) ds + \int_t^T \sigma_s.dW_s^T \tag{3.7}$$

If we apply the pricing operator $\mathbb{E}^{\mathbb{P}^T}[\cdot|\mathcal{F}_t]$ on both sides of this equation, we deduce that the variance swap VS is related to a log-contract option with a payoff $\ln\left(\frac{S_T}{S_t}\right)$

$$\mathrm{VS} = -\frac{2}{T-t}\mathbb{E}^{\mathbb{P}^T}[\ln\left(\frac{S_T}{S_t}\right)|\mathcal{F}_t] + \frac{2}{T-t}\int_t^T \mathbb{E}^{\mathbb{P}^T}[(r_s + \sigma_s.\sigma_s^P)|\mathcal{F}_t]ds$$

where we have used that $\mathbb{E}^{\mathbb{P}^T}[\int_t^T \sigma_s dW_s^T|\mathcal{F}_t] = 0$ by Itô's isometry.

Assumption 2: Let us assume that the instantaneous rate is a deterministic constant process $r_s \equiv r(s)$ and therefore the volatility of the bond σ_t^P cancels. The value of VS simplifies to

$$\mathrm{VS} = -\frac{2}{T-t}\mathbb{E}^{\mathbb{P}^T}[\ln\left(\frac{S_T}{S_t}\right)|\mathcal{F}_t] + \frac{2}{T-t}\int_t^T r(s)ds$$

From (3.4), we obtain the static replication

$$\begin{aligned}\mathrm{VS} &= \frac{2P_{tT}^{-1}}{(T-t)}\left(\int_0^{S_t} dK\frac{\mathcal{P}^{\mathrm{BS}}(t,K,T)}{K^2} + \int_{S_t}^\infty dK\frac{\mathcal{C}^{\mathrm{BS}}(t,K,T)}{K^2}\right) \\ &+ \frac{2}{T-t}\left(\ln P_{tT}^{-1} + 1 - P_{tT}^{-1}\right)\end{aligned} \tag{3.8}$$

In this equation, we have used that $\ln P_{tT} = -\int_t^T r(s)ds$. $\square$

Example 3.3 Generalized Variance swap

The construction above can be generalized to any option which pays at the maturity date T

$$\frac{1}{N}\sum_{i=0}^{N-1} f(S_i)\left(\ln\left(\frac{S_{i+1}}{S_i}\right)\right)^2 - K$$

Here f is a measurable function. For

$$f(S) = 1\left(\frac{S}{S_t} \in [A,B]\right)$$

we have a *corridor variance swap* [71]. As for the variance swap, the strike K is fixed such that the value of the option quoted at t is zero. Thus K is under the forward measure $\mathbb{P}^T$ (for convenience sake) equal to

$$K = \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{E}^{\mathbb{P}^T} [f(S_i) \left(\ln \left(\frac{S_{i+1}}{S_i} \right) \right)^2 |\mathcal{F}_t]$$

In the large N limit, K becomes

$$K = \frac{1}{(T-t)} \int_t^T \mathbb{E}^{\mathbb{P}^T} [f(S_s)\sigma_s^2 |\mathcal{F}_t] ds \tag{3.9}$$

We want to specify the function $f(S_s)$ such that there still exists a static replication formula for the generalized variance swap. In this context, we note that using the Itô lemma, we have the relation

$$\begin{aligned} g(S_T) - g(S_t) = &\int_t^T S_s \partial_S g(S_s) \sigma_s . dW_s^T + \int_t^T S_s^2 \partial_S^2 g(S_s) \frac{\sigma_s^2}{2} ds \\ &+ \int_t^T S_s \partial_S g(S_s)(r_s + \sigma_s . \sigma_s^P) ds \end{aligned}$$

for a function $g \in C^2(\mathbb{R})$ and the Itô diffusion process S_t following (3.6). Moreover, taking the mean value under $\mathbb{P}^T$, we obtain

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^T}[g(S_T)|\mathcal{F}_t] - g(S_t) = &\int_t^T \mathbb{E}^{\mathbb{P}^T}[S_s^2 \partial_S^2 g(S_s) \frac{\sigma_s^2}{2} |\mathcal{F}_t] ds \\ &+ \int_t^T \mathbb{E}^{\mathbb{P}^T}[S_s \partial_S g(S_s)(r_s + \sigma_s . \sigma_s^P)|\mathcal{F}_t] ds \end{aligned} \tag{3.10}$$

Comparing (3.9) and (3.10), if we impose that the function $f(S_s)$ satisfies

$$\frac{S^2}{2} \partial_S^2 g(S) = f(S) \tag{3.11}$$

and that the interest rate is a deterministic process $r(s)$ (i.e., the bond's volatility cancels), then there is a static replication formula

$$K = \frac{1}{(T-t)} \left(\mathbb{E}^{\mathbb{P}^T}[g(S_T)|\mathcal{F}_t] - g(S_t) - \int_t^T r(s) \mathbb{E}^{\mathbb{P}^T}[S_s \partial_S g(S_s)|\mathcal{F}_t] ds \right)$$

The general solution of (3.11) can be written as a sum of a particular solution $g_p(S)$ and the homogeneous solution $aS + b$ with a and b two constants:

$$g(S) = g_p(S) + aS + b$$

We obtain

$$K = \frac{1}{(T-t)} \left(\mathbb{E}^{\mathbb{P}^T}[g_p(S_T)|\mathcal{F}_t] - g_p(S_t) - \int_t^T r(s) \mathbb{E}^{\mathbb{P}^T}[S_s \partial_S g_p(S_s)|\mathcal{F}_t] ds \right)$$

The term $\mathbb{E}^{\mathbb{P}^T}[g_p(S_T)|\mathcal{F}_t]$ (resp. $\mathbb{E}^{\mathbb{P}^T}[S_s\partial_S g_p(S_s)|\mathcal{F}_t]$) can be decomposed as a sum of call and put options with maturity T (resp. s) using (3.4). For $f(S)=1$, the relation (3.11) gives $g_p(S)=-2\ln(S)$ and we reproduce the variance swap replication formula (3.8). For a corridor variance swap with $f(S)=1\left(\frac{S}{S_t}\in[A,B]\right)$, we obtain

$$g_p(S)=1\left(\frac{S}{S_t}\in[A,B]\right)\ln\frac{S}{S_t}+1(S>BS_t)\left(\frac{S}{BS_t}+\ln B-1\right)$$
$$+1(S<AS_t)\left(\frac{S}{AS_t}+\ln A-1\right)$$

□

3.3 Forward starting options and dynamics of the implied volatility

From the fair value of European call-put options, the value of the implied volatility can be deduced. However, this surface evolves with time and describes a non-trivial dynamics. For European options, we have seen in the section on static replication that the implied volatility dynamics has no impact on their valuations as their fair values depend only on the initial implied volatility. However, path-dependent exotic options will depend strongly on implied volatility dynamics. It is therefore important to have at our disposal financial instruments that can inform on the dynamics followed by the implied volatility. Such exotic options are the forward-starting options and in practice the cliquet options. Using these instruments, we will define the forward implied volatility and the forward skew. As a foreword we give simple rules that have been used to estimate the evolution of the implied volatility.

3.3.1 Sticky rules

The observation that the implied volatility surface evolves with time has led practitioners to develop simple rules to estimate its evolution.

- Sticky delta: A commonly used rule is the "sticky moneyness" rule which stipulates that when viewed in relative coordinates $(m=\ln(\frac{K}{S_t}),\tau=T-t)$, the surface $\sigma_{\mathrm{BS}\ t}(\tau,m)$ remains constant from day to day:

$$\forall(\tau,m)\ \ \sigma_{\mathrm{BS}\ \ t+\Delta t}(\tau,m)=\sigma_{\mathrm{BS}\ \ t}(\tau,m)$$

- Sticky strike: Another well-known rule is the so-called "sticky rule" : in this case one assumes that the level of implied volatilities in absolute strikes does not change:

$$\forall(\tau, K) \;\; \sigma_{\mathrm{BS}\ t+\Delta t}(\tau, K) = \sigma_{\mathrm{BS}\ t}(\tau, K)$$

3.3.2 Forward-start options

A forward-start option on a single asset is a contract according to which the holder receives an option at T_1 with an expiry date $T_2 > T_1$. For a forward-start call option, the payoff at T_2 is

$$\max\left(\frac{S_{T_2}}{S_{T_1}} - K, 0\right)$$

Its fair price is given at time t by

$$\mathcal{C}(t, S_t) = P_{tT_1}\mathbb{E}^{\mathbb{P}^{T_1}}[S_{T_1}^{-1}\max\left(S_{T_2} - KS_{T1}, 0\right)|\mathcal{F}_t]$$

with $\mathbb{P}^{T_1}$ the forward measure. By using the tower property, we have

$$\begin{aligned}\mathcal{C}(t, S_t) &= P_{tT_1}\mathbb{E}^{\mathbb{P}^{T_1}}[S_{T_1}^{-1}\max\left(S_{T_2} - KS_{T_1}, 0\right)|\mathcal{F}_t] \\ &= P_{tT_1}\mathbb{E}^{\mathbb{P}^{T_1}}[S_{T_1}^{-1}\mathbb{E}^{\mathbb{P}^{T_1}}[\max\left(S_{T_2} - KS_{T_1}, 0\right)|\mathcal{F}_{T_1}]|\mathcal{F}_t]\end{aligned}$$

Assuming that S_t follows a Black-Scholes log-normal model with a volatility σ and that we have a deterministic rate, we obtain

$$\begin{aligned}\mathcal{C}(t, S_t) &= P_{tT_1}\mathbb{E}^{\mathbb{P}^{T_1}}[N(d_+) - KP_{T_1T_2}N(d_-)|\mathcal{F}_t] \\ &= P_{tT_1}N(d_+) - KP_{tT_2}N(d_-)\end{aligned} \tag{3.12}$$

with $d_\pm = -\frac{\ln(KP_{T_1T_2})}{\sigma_{\mathrm{BS}}\sqrt{T_2-T_1}} \pm \frac{\sigma_{\mathrm{BS}}\sqrt{T_2-T_1}}{2}$. From this Black-Scholes-like formula, we define the forward implied volatility:

DEFINITION 3.3 *The forward implied volatility is defined as the "wrong number" σ which when put in the formula (3.12) reproduces the fair value quoted on the market.*

We can also define *a forward skew* as the skew of the forward implied volatility. Usually the forward implied volatility is derived from more complicated forward-start options called cliquet options.

3.3.3 Cliquet options

A cliquet is a series of forward-start call options, with periodic settlements. The profit can be accumulated until final maturity, or paid out at each reset date. The payoff is the sum

$$\sum_{i=1}^{N}\max\left(\frac{S_{T_{i+1}}}{S_{T_i}} - K, 0\right)$$

In the next subsection, we present an example of a new derivative product which strongly depends on the dynamics of the (forward) implied volatility.

3.3.4 Napoleon options

Napoleon options are financial instruments that give the traders the opportunity to play with the forward volatility of a market. The main factors of the payoff of a Napoleon option are a fixed coupon and the worst return of an index over specified time periods.
Let $T_0 < T_1 < \cdots < T_N$ be several settlement period dates. The payoff will depend on the performance of a single stock or index $\{R_j \equiv \frac{S_{t_i^{j+1}}}{S_{t_i^j}} - 1\}_{t_i^j \in [T_i, T_{i+1}]}$ between each period $[T_i, T_{i+1}]$. It is given at each T_{i+1} by

$$\min\left(\text{cap}, \max\left(\text{coupon} + \min_j[R_j], \text{floor}\right)\right)$$

When the market volatility is small, the buyer of a Napoleon option will have a good chance to get a payoff around the fixed coupon. When the market volatility is large, the buyer will get at least the minimum return determined by the floor.
Napoleon options vary from one contract to another. Some Napoleon option contracts have multiple coupon calculation/payment periods and each coupon consists of performances of the underlying index on multiple time periods. Some have a single coupon payment but multiple performance calculation periods. Some have only one coupon payment and one performance calculation period also. Besides, caps and floors on performances and coupons are specified differently across contracts.
In the following section, we define the liquid instruments that are used to calibrate an interest-rate model. As usual, we will compute their fair values in a Black-Scholes-like model and this enables us to define an implied volatility.

3.4 Interest rate instruments

3.4.1 Bond

The first liquid interest-rate instrument is a bond which is a contract which pays 1 at a maturity date T. Its fair value at $t < T$ is noted P_{tT} and it is given in the risk-neutral measure $\mathbb{P}$ by

$$P_{tT} = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds} | \mathcal{F}_t]$$

Similarly, we can define a forward bond which is a contract which pays a bond, expiring at T_2, at maturity $T_1 < T_2$. Its value at $t < T_1$ is noted $P(t, T_1, T_2)$.

From a no-arbitrage argument, we have that

$$P(t, T_1, T_2) = \frac{P_{tT_2}}{P_{tT_1}} \tag{3.13}$$

A Libor $L(t, T_1, T_2)$ is defined as

$$P(t, T_1, T_2) = \frac{1}{1 + (T_2 - T_1)L(t, T_1, T_2)}$$

By using (3.13), we obtain

$$L(t, T_1, T_2) = \frac{1}{(T_2 - T_1)} \left(\frac{P_{tT_1}}{P_{tT_2}} - 1 \right)$$

3.4.2 Swap

A swap is a contract composed of two flux (we say two legs). One pays a Libor rate and the other one a fixed rate. At each instant T_i in $T_{\alpha+1}, ..., T_\beta$, the fixed legs pay ($\tau_i = T_i - T_{i-1}$)

$$\tau_i K$$

whereas the floating leg pays

$$\tau_i L(T_i, T_{i-1}, T_i)$$

The discounted payoff at time t for the payer of the fixed leg (receiving the floating leg) is

$$\sum_{i=\alpha+1}^{\beta} D_{tT_i} \tau_i \left(L(T_i, T_{i-1}, T_i) - K \right)$$

We remind the reader that D_{tT} is the discount factor as defined in (2.23). By using the forward measure $\mathbb{P}^{T_i}$ associated to the bond P_{tT_i}, we have that

$$\mathbb{E}^{\mathbb{P}}[D_{tT_i} \tau_i \left(L(T_i, T_{i-1}, T_i) - K \right) |\mathcal{F}_t] = P_{tT_i} \tau_i \mathbb{E}^{\mathbb{P}^{T^i}} \left[\left(L(T_i, T_{i-1}, T_i) - K \right) |\mathcal{F}_t \right]$$

Therefore, the price of the swap is

$$\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i} \mathbb{E}^{\mathbb{P}^{T_i}} \left[\left(L(T_i, T_{i-1}, T_i) - K \right) |\mathcal{F}_t \right] = \sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i} \left(L(t, T_{i-1}, T_i) - K \right)$$

where we have used that $L(t, T_{i-1}, T_i)$ is a martingale in the forward measure $\mathbb{P}^{T^i}$ as explained in 2.10.3.3. From the relation $L(t, T_{i-1}, T_i) = \frac{1}{\tau_i} \left(\frac{P_{tT_{i-1}}}{P_{tT_i}} - 1 \right)$, we obtain that the sum of the two legs is

$$-P_{tT_\beta} + P_{tT_\alpha} - K \sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}$$

By definition, the swap rate is the strike $K = s_{\alpha\beta,t}$ such that the value of the swap contract at time t is zero

$$s_{\alpha\beta,t} = \frac{P_{tT_\alpha} - P_{tT_\beta}}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}} \tag{3.14}$$

The natural next question is how one can reconstruct the yield curve between two dates t and T from the swap data (note that a Libor $L(t, T_{\alpha-1}, \alpha) = s_{\alpha-1\alpha,t}$ is an example of swap). Let us consider a set of m forward Libor rates $\{L_k\}$ that are associated to the tenor structure $\{T_k\}_k = 1, \cdots, n$. We want to reconstruct at any time the yield curve described by the $\{L_k \equiv L(t, T_{k-1}, T_k)\}$ Libor rates from the swaps characterized by the set of pairs

$$\mathcal{S} = \{\epsilon_j = (s(j), e(j)) \ , \ j = 1, \cdots m \ , \ 1 \leq s(j) < e(j) \leq m\}$$

In accordance with our previous notation, a swap $s_{\alpha\beta}$ has $s = \alpha+1$ and $e = \beta$. We have the following result

PROPOSITION 3.1 [135]
The yield curve can be built from the set $\mathcal{S}$ if and only if $n = m$ and $s(j) = j$. That is why, at each tenor date, we have to specify a unique swap.

3.4.3 Swaption

A swaption gives the right, but not the obligation, to enter into an interest rate swap at a pre-determined rate, noted K below, on an agreed future date. The maturity date for the swaption is noted T_α and coincides with the first tenor date of the swap contract. T_β is the expiry date for the swap. The value of the swap at T_α is

$$\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i} \left(L(T_i, T_{i-1}, T_i) - K \right)$$

Then the discounted value of the swaption at $t \leq T_\alpha$ is under the risk-neutral measure $\mathbb{P}$

$$\mathcal{C}_{\alpha\beta} = \mathbb{E}^{\mathbb{P}}[D_{tT_\alpha} \max \left(\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i} (L(T_i, T_{i-1}, T_i) - K), 0 \right)]$$

By using the definition (3.14) of the swap $s_{\alpha\beta,t}$, this can be written as

$$\mathcal{C}_{\alpha\beta} = \mathbb{E}^{\mathbb{P}}[D_{tT_\alpha} \sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i} \max \left(s_{\alpha\beta,T_\alpha} - K, 0 \right)]$$

We consider the measure $\mathbb{P}^{\alpha\beta}$ associated to the numéraire $\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}$. The Radon-Nikodym change of measure from $\mathbb{P}$ to $\mathbb{P}^{\alpha\beta}$ is

$$M_{T_\alpha} \equiv \frac{d\mathbb{P}^{\alpha\beta}}{d\mathbb{P}}|_{\mathcal{F}_{T_\alpha}} = \frac{\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i}}{D_{T_\alpha T_\alpha}}$$

By using (2.41), we obtain

$$\mathcal{C}_{\alpha\beta} = \sum_{i=\alpha+1}^{\beta} \tau_i \mathbb{E}^{\mathbb{P}^{\alpha\beta}}[D_{tT_\alpha} P_{T_\alpha T_i} \frac{M_t}{M_{T_\alpha}} \max\left(s_{\alpha\beta,T_\alpha} - K, 0\right)]$$

As

$$\frac{M_t}{M_{T_\alpha}} = \frac{\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i} D_{tT_\alpha}}$$

we get our final result

$$\mathcal{C}_{\alpha\beta} = \sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i} \mathbb{E}^{\mathbb{P}^{\alpha\beta}}[\max\left(s_{\alpha\beta,T_\alpha} - K, 0\right) |\mathcal{F}_t]$$

As the product of the swap rate $s_{\alpha\beta,t}$ with the numéraire $\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}$ is a difference of two bonds $P_{tT_\alpha} - P_{tT_\beta}$ and therefore a traded asset, the swap is a local martingale in the measure $\mathbb{P}^{\alpha\beta}$. Adopting a Black-Scholes-like model, we assume that the swap is log-normal (in the measure $\mathbb{P}^{\alpha\beta}$)

$$ds_{\alpha\beta,t} = \sigma_{\alpha\beta} s_{\alpha\beta,t} dW^{\alpha\beta} \tag{3.15}$$

with the constant volatility $\sigma_{\alpha\beta}$. By noting the similarity with the computation of the fair value of a European call option in the Black-Scholes model, we obtain the fair value of a swaption as

$$\mathcal{C}_{\alpha\beta} = \sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i} \mathcal{C}^{\mathrm{BS}}(K, T_\alpha, \sigma_{\alpha\beta} | s_{\alpha\beta,t}, t) \tag{3.16}$$

with $s_{\alpha\beta,t}$ the swap spot at t.
We can then define the *swaption implied volatility* as

DEFINITION 3.4 Swaption implied volatility *The swaption implied volatility is the volatility that when put in the formula (3.16) reproduces the market price of a swaption.*

The at-the-money swaption implied volatilities (i.e., $K = s_{\alpha\beta,0}$) are coded in a matrix, usually called *swaption matrix*. The columns represent the tenor of the swap (i.e., $T_\beta - T_\alpha$) and the rows represent the maturity of the swaption (i.e., T_α). Two particular lines are useful for calibration purposes (more details are given in chapter 8). The first one is the first column which corresponds to *caplets*. The second one is the anti-diagonal, called *co-terminal swaptions*. An extra dimension can be added with the strike K and we have a *swaption cube*.

3.4.4 Convexity adjustment and CMS option

Convexity correction arises when an interest rate is paid out at the "wrong time." As an example, we have a *Constant Maturity Swap* (CMS) which pays the swap rate $s_{\alpha\beta,T_\alpha}$ at the maturity date T_α. Therefore the fair value is in the forward measure $\mathbb{P}^{T_\alpha}$

$$\mathrm{CMS}_{\alpha\beta}(t) = P_{tT_\alpha} \mathbb{E}^{\mathbb{P}^{T_\alpha}}[s_{\alpha\beta,T_\alpha}|\mathcal{F}_t] \tag{3.17}$$

As discussed previously, the swap rate is not a local martingale in the forward measure $\mathbb{P}^{T_\alpha}$. This is the case under the swap measure $\mathbb{P}^{\alpha\beta}$ associated to the numéraire $C_{\alpha\beta}(t) = \sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}$. By using this natural measure $\mathbb{P}^{\alpha\beta}$, we can rewrite $\mathrm{CMS}_{\alpha\beta}(t)$ as ($\mathbb{E}^{\alpha\beta} \equiv \mathbb{E}^{\mathbb{P}^{\alpha\beta}}$)

$$\begin{aligned}\mathrm{CMS}_{\alpha\beta}(t) &= C_{\alpha\beta}(t)\mathbb{E}^{\alpha\beta}[\frac{1}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i}} s_{\alpha\beta,T_\alpha}|\mathcal{F}_t] \\ &= \underbrace{C_{\alpha\beta}(t)\mathbb{E}^{\alpha\beta}[\left(\frac{1}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i}} - 1\right) s_{\alpha\beta,T_\alpha}|\mathcal{F}_t]}_{\text{Convexity correction}} + C_{\alpha\beta}(t) s^0_{\alpha\beta}\end{aligned} \tag{3.18}$$

As $s_{\alpha\beta}$ is a local martingale under $\mathbb{P}^{\alpha\beta}$ (we assume that $s_{\alpha\beta,t}$ is also a martingale), the second term in (3.18) is given by the spot swap rate $s^0_{\alpha\beta} \equiv s_{\alpha\beta,t}$ times $C_{\alpha\beta}(t)$. The first term, usually small, is called the *convexity correction*. To compute this quantity, we assume that the numéraire ratio $R_{\alpha\beta} = \frac{P_{tT_\alpha}}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}}$ depends on the time t and the swap rate $s_{\alpha\beta,t}$

$$\frac{P_{tT_\alpha}}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}} - 1 \equiv G(t, s_{\alpha\beta,t}) \tag{3.19}$$

The convexity correction can be written as

$$C_{\alpha\beta}(t)\mathbb{E}^{\alpha\beta}[G(T_\alpha, s_{\alpha\beta,T_\alpha}) s_{\alpha\beta,T_\alpha}|\mathcal{F}_t]$$

The simple assumption (3.19) implies a strong constraint on the function $G(\cdot,\cdot)$. By definition $R_{\alpha\beta}$ is a local martingale under $\mathbb{P}^{\alpha\beta}$ and the function $G(t, s_{\alpha\beta})$ must be a local martingale under $\mathbb{P}^{\alpha\beta}$ meaning that its drift cancels.

As $s_{\alpha\beta}$ is a local martingale under $\mathbb{P}^{\alpha\beta}$, for example as given in (3.15), we have the PDE satisfied by G

$$\partial_t G(t, s_{\alpha\beta}) + \frac{1}{2}\sigma_{\alpha\beta}(s_{\alpha\beta})^2 \partial^2_{s_{\alpha\beta}} G(t, s_{\alpha\beta}) = 0 \tag{3.20}$$

with the initial condition

$$G(0, s^0_{\alpha\beta}) = \frac{P_{0T_\alpha}}{C_{\alpha\beta}(0)} - 1 \tag{3.21}$$

For example, assuming that the function G does not depend on the time, $G = G(s_{\alpha\beta})$, then the solution of (3.20) is a linear function of $s_{\alpha\beta}$

$$G(s_{\alpha\beta}) = a + bs_{\alpha\beta}$$

The initial condition (3.21) implies that $a + bs^0_{\alpha\beta} = \frac{P_{0T_\alpha}}{C_{\alpha\beta}(0)} - 1$.
Then, to compute the convexity correction, we use a static replication for the function $f(x) \equiv G(T_\alpha, x)x$ similar to (3.4)

$$\begin{aligned}
\mathbb{E}^{\alpha\beta}[f(s_{\alpha\beta})|\mathcal{F}_t] &= f(s^0_{\alpha\beta}) + f'(s^0_{\alpha\beta})\mathbb{E}^{\alpha\beta}[s_{\alpha\beta} - s^0_{\alpha\beta}|\mathcal{F}_t] \\
&\quad + \int_0^{s^0_{\alpha\beta}} f^{''}(x)\mathbb{E}^{\alpha\beta}[\max(x - s_{\alpha\beta}, 0)|\mathcal{F}_t]dx \\
&\quad + \int_{s^0_{\alpha\beta}}^{\infty} f^{''}(x)\mathbb{E}^{\alpha\beta}[\max(s_{\alpha\beta} - x, 0)|\mathcal{F}_t]dx
\end{aligned}$$

and finally the fair value of the CMS is

$$\begin{aligned}
\mathrm{CMS}_{\alpha\beta}(t) &= C_{\alpha\beta}(t)s^0_{\alpha\beta}G(T_\alpha, s^0_{\alpha\beta}) \\
&\quad + \int_0^{s^0_{\alpha\beta}} G''(T_\alpha, x)x\mathcal{P}_{\alpha\beta}(x)dx + \int_{s^0_{\alpha\beta}}^{\infty} G''(T_\alpha, x)x\mathcal{C}_{\alpha\beta}(x)dx \\
&\quad + 2\int_0^{s^0_{\alpha\beta}} G'(T_\alpha, x)\mathcal{P}_{\alpha\beta}(x)dx + 2\int_{s^0_{\alpha\beta}}^{\infty} G'(T_\alpha, x)\mathcal{C}_{\alpha\beta}(x)dx
\end{aligned}$$

where we have set $\mathcal{C}_{\alpha\beta}(x) \equiv C_{\alpha\beta}(t)\mathbb{E}^{\alpha\beta}[\max(s_{\alpha\beta,T_\alpha} - x, 0)|\mathcal{F}_t]$[2] the value of a call swaption with strike x. In the particular case when G is linear, we have

$$\mathrm{CMS}_{\alpha\beta}(t) = C_{\alpha\beta}(t)s^0_{\alpha\beta}(a + bs^0_{\alpha\beta}) + 2b\left(\int_0^{s^0_{\alpha\beta}} \mathcal{P}_{\alpha\beta}(x)dx + \int_{s^0_{\alpha\beta}}^{\infty} \mathcal{C}_{\alpha\beta}(x)dx\right) \tag{3.22}$$

[2] resp. $\mathcal{P}_{\alpha\beta}(x) \equiv C^0_{\alpha\beta}(t)\mathbb{E}^{\alpha\beta}[\max(x - s_{\alpha\beta,T_\alpha}, 0)|\mathcal{F}_t]$.

Therefore, if we want to match the CMS (3.22) and the initial condition (3.21), we should impose that the convexity adjustment function G is

$$G(s_{\alpha\beta}) = \frac{P_{tT_\alpha}}{C_{\alpha\beta}(t)} - 1 + \frac{\mathrm{CMS}_{\alpha\beta}(t) - s^0_{\alpha\beta}\left(P_{tT_\alpha} - C_{\alpha\beta}(t)\right)}{2(\int_0^{s^0_{\alpha\beta}} \mathcal{P}_{\alpha\beta}(x)dx + \int_{s^0_{\alpha\beta}}^{\infty} \mathcal{C}_{\alpha\beta}(x)dx)}(s_{\alpha\beta} - s^0_{\alpha\beta})$$

Modulo the fundamental assumption (3.19) on the dependence of the yield curve according to the unique swap rate $s_{\alpha\beta}$, the value that we have found for the CMS (3.22) is model-independent.
Before closing this section, we present an other convexity adjustment which has the inconvenience of being model dependent. We assume that G is a function of the swap rate $s_{\alpha\beta}$ only. The dynamics of $s_{\alpha\beta}$ follows a log-normal process (3.15) in $\mathbb{P}^{\alpha\beta}$. Then, the dynamics of $s_{\alpha\beta}$ in $\mathbb{P}^{T_\alpha}$ is

$$\frac{ds_{\alpha\beta}}{s_{\alpha\beta}} = \sigma^2_{\alpha\beta} s_{\alpha\beta} \frac{\partial_{s_{\alpha\beta}} G(s_{\alpha\beta})}{G(s_{\alpha\beta})} dt + \sigma_{\alpha\beta} dW^\alpha$$

By approximating the drift by its value at the spot swap rate $s^0_{\alpha\beta}$, the process $s_{\alpha\beta}$ becomes log-normal and the value of the CMS (3.17) is given by

$$\mathrm{CMS}_{\alpha\beta}(t) = P_{tT_\alpha} s^0_{\alpha\beta} e^{(T_\alpha - t)\sigma^2_{\alpha\beta} s^0_{\alpha\beta} \frac{\partial_{s_{\alpha\beta}} G(s^0_{\alpha\beta})}{G(s^0_{\alpha\beta})}}$$

3.5 Problems

Exercises 3.1 Volatility swap
A *Volatility swap* is a contract that pays at a maturity date T the realized volatility σ_{realized} (which is the square root of the realized variance) of a stock or index minus a strike K. By definition, the strike K is set such that the value of the volatility swap at t (i.e., today) is zero. Therefore, we have

$$K = \mathbb{E}^{\mathbb{P}}[\sqrt{V_{\text{realized}}}|\mathcal{F}_t]$$

with

$$V_{\text{realized}} \equiv \frac{1}{N}\sum_{i=0}^{N-1} \ln\left(\frac{S_{i+1}}{S_i}\right)^2$$

We recall that a *variance swap* K is

$$K = \mathbb{E}^{\mathbb{P}}[V_{\text{realized}}|\mathcal{F}_t]$$

Assuming a diffusion model for the stock price S_t, when we take the limit $N \to \infty$, the realized variance between t and T is

$$V_{\text{realized}} \equiv \frac{1}{T-t} \int_t^T \sigma_s^2 ds$$

Here σ_s is the instantaneous volatility defined by

$$dS_t = \sigma_t S_t dW_t$$

We assume zero interest rate. We have seen in this chapter that the value of a variance swap (i.e., K) is completely model-independent. Indeed, there exists a static replication where the value of the variance swap can be decomposed as an infinite sum of calls and puts which are priced on the market. In this problem, we propose to show that a similar model-independent solution exists for the volatility swap if we assume that the stock S_t and the volatility σ_t are independent processes. This problem is based on the article [72]. Below, we note $V_T = V_{\text{realized}}$.

1. As the processes σ_t and W_t are independent by assumption, prove that conditional on the filtration $\mathcal{F}_T^\sigma$ generated by the volatility $\{\sigma_t\}_{t \le T}$, $X_T = \ln(\frac{S_T}{S_t})$ is a normal process with mean $-\frac{1}{2}(T-t)V_T$ and variance $(T-t)V_T$.

2. As an application of the tower property, we have $\forall p \in \mathbb{C}$,

$$\mathbb{E}^{\mathbb{P}}[e^{pX_T}] = \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[e^{pX_T} | \mathcal{F}_T^\sigma]]$$

   By using question 1. and the equality above, prove that

$$\mathbb{E}^{\mathbb{P}}[e^{pX_T}] = \mathbb{E}^{\mathbb{P}}[e^{\lambda(p)(T-t)V_T}] \tag{3.23}$$

   with $\lambda(p) \equiv \frac{p^2}{2} - \frac{p}{2}$. Equivalently, we have $p(\lambda) = \frac{1}{2} \pm \sqrt{\frac{1}{4} + 2\lambda}$.

3. Let $f(t)$ be a real function, locally summable on $\mathbb{R}_+$. The Laplace transform of f, noted $\mathcal{L}(z)$, $z = x + iy \in \mathbb{C}$, is defined as

$$\mathcal{L}(z) = \int_0^\infty f(t) e^{-zt} dt$$

   If $\mathcal{L}(z)$ is defined for all $x \equiv \mathrm{Re}[z] > x_0$ and if $\mathcal{L}(z)$ is a summable function $\forall x > x_0$, then the inversion formula (called the Bromvitch formula) giving $f(t)$ is

$$1_{t \ge 0} f(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \mathcal{L}(\lambda) e^{\lambda x} d\lambda \tag{3.24}$$

   where $a > x_0$. By using the inversion formula (3.24), prove that the Laplace transform of the square root function is

$$\sqrt{t} = \frac{1}{2}\sqrt{\pi} \int_0^\infty \frac{1 - e^{-\lambda t}}{\lambda^{\frac{3}{2}}} d\lambda \tag{3.25}$$

4. By using (3.25) and (3.23), deduce that

$$\mathbb{E}^{\mathbb{P}}[\sqrt{(T-t)V_T}] = \frac{1}{2}\sqrt{\pi}\int_0^{\infty} \frac{1-\mathbb{E}^{\mathbb{P}}[e^{(\frac{1}{2}-\sqrt{\frac{1}{4}-2\lambda})X_T}|\mathcal{F}_t]}{\lambda^{\frac{3}{2}}} d\lambda$$

5. By using the static replication formula (3.4), decompose the European option $\mathbb{E}^{\mathbb{P}}[e^{(\frac{1}{2}-\sqrt{\frac{1}{4}-2\lambda})X_T}|\mathcal{F}_t]$ on an infinite sum of calls and puts.

6. Deduce the model-independent formula for the volatility swap.

Exercises 3.2 Ho-Lee equity hybrid model

The purpose of this exercise is to price exotic options depending on a single stock and characterized by a large maturity date. As a consequence, the assumption of deterministic rates should be relaxed. In this problem, we present a simple toy model which achieves this goal and we focus on calibration issues. It is assumed that the stock follows a Merton model

$$\frac{dS}{S} = r_t dt + \sigma(t) dW_t$$

where $\sigma(t)$ is a time-dependent volatility. The instantaneous short rate r_t is assumed to be driven by

$$dr_t = \theta(t)dt + \psi dZ_t$$

with a constant volatility ψ, a time-dependent drift $\theta(t)$ and dZ_t a Brownian motion correlated to dW_t: $dW_t dZ_t = \rho dt$. This is the short-rate Ho-Lee model [7]. The next questions deal with how to calibrate these model parameters $\sigma(\cdot)$, ψ and $\theta(\cdot)$.

1. Prove that the rate can be written as $r_t = \phi(t) + x_t$ with $\phi(t)$ a deterministic function that you will specify and x_t a process following

$$dx_t = \psi dZ_t$$

2. Prove that x_t is a Brownian process and compute its mean and its variance at the time t.

3. By using the results of the question above, prove that the value of a bond P_{tT} expiring at T and quoted at $t < T$ is

$$P_{tT} = A(t,T)e^{-B(t,T)x_t}$$

with $A(t,T)$ and $B(t,T)$ two functions that you will specify. A model such that the bond value can be written in this way is called *an affine model.*

4. Deduce that we can choose the function $\theta(t)$ in order to calibrate the initial yield curve P_{0T} , $\forall T$.

5. From the previous questions, deduce that the bond P_{tT} follows the SDE

$$\frac{dP_{tT}}{P_{tT}} = r_t dt + \sigma_P(t) dZ_t$$

with $\sigma_P(t) = -\psi(T-t)$.

6. Prove that the (equity) forward $f_t^T = \frac{S_t}{P_{tT}}$ in the forward measure $\mathbb{P}^T$ follows the SDE

$$\frac{df_t^T}{f_t^T} = \sigma(t) dW_t - \sigma_P(t) dZ_t$$

Why is it obvious that the forward is a local martingale?

7. As the forward obeys a log-normal process, prove that the implied volatility with maturity T is

$$\sigma_{\mathrm{BS},t}^2 (T-t) = \int_t^T \left(\sigma(s)^2 + \sigma_P(s)^2 - 2\rho\sigma(s)\sigma_P(s)\right) ds$$

8. Deduce the Merton volatility $\sigma(s)$ in order to calibrate the at-the-money market implied volatility.

Chapter 4

Differential Geometry and Heat Kernel Expansion

ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ[1]
—Plato

Abstract In this chapter, we present the key tool of this book: the heat kernel expansion on a Riemannian manifold. In the first section, in order to introduce this technique naturally, we remind the reader of the link between the multi-dimensional Kolmogorov equation and the value of a European option. In particular, an asymptotic implied volatility in the short-time limit will be obtained if we can find an asymptotic expansion for the multi-dimensional Kolmogorov equation. This is the purpose of the heat kernel expansion. Rewriting the Kolmogorov equation as a heat kernel equation on a Riemannian manifold endowed with an Abelian connection, we can apply Hadamard-DeWitt's theorem giving the short-time asymptotic solution to the Kolmogorov equation. An extension to the time-dependent heat kernel will also be presented as this case is particularly important in finance in order to include term structures. In the next chapters, we will present several applications of this technique, for example the calibration of local and stochastic volatility models.

4.1 Multi-dimensional Kolmogorov equation

In this part, we recall the link between the valuation of a multi-dimensional European option and the backward (and forward) Kolmogorov equation. For the sake of simplicity, we assume a zero interest rate.

[1] "Let no one inapt to geometry come in." Inscribed over the entrance to the Plato Academy in Athens.

4.1.1 Forward Kolmogorov equation

We assume that our market model depends on n Itô processes which can be traded assets or unobservable Markov processes such as a stochastic volatility. Let us denote the stochastic processes $x \equiv (x^i)_{i=1,\cdots,n}$ and $\alpha \equiv (\alpha^i)_{i=1,\cdots,n}$. These processes x^i satisfy the following SDEs in a risk-neutral measure $\mathbb{P}$

$$\begin{aligned} dx_t^i &= b^i(t, x_t)dt + \sigma^i(t, x_t)dW_i \\ dW_i dW_j &= \rho_{ij}(t)dt \end{aligned} \tag{4.1}$$

with the initial condition $x_0 = \alpha$.
Here $[\rho_{ij}]_{i,j=1,\cdots,n}$ is a *correlation matrix* (i.e., a symmetric non-degenerate matrix).

REMARK 4.1 In chapter 2, the SDEs were driven by independent Brownian motions. The SDEs (4.1) can be framed in this setting by applying a Cholesky decomposition: we write the correlation matrix as

$$\rho = LL^{\dagger}$$

or in components

$$\rho_{ij} = \sum_{k=1}^{n} L_{ik} L_{jk}$$

Then the correlated Brownians $\{W_i\}_{i=1,\cdots,n}$ can be decomposed over a basis of n independent Brownians $\{Z_i\}_{i=1,\cdots,n}$:

$$dW_i = \sum_{k=1}^{n} L_{ik} dZ_k$$

□

In order to ensure the existence and uniqueness of SDE (4.1), we impose the space-variable Lipschitz condition (2.17) and the growth condition (2.16). The no-arbitrage condition implies that the discounted traded assets are (local) martingales under this equivalent measure $\mathbb{P}$. For $\mathbb{P}$, the drifts b_i are consequently zero for the traded assets according to the theorem 2.4. Note that the measure $\mathbb{P}$ is not unique as the market is not necessarily complete according to the theorem 2.8. Finally, the fair value of a European option at t with payoff $f(x_T)$ at maturity T is given by the mean value of the payoff conditional on the filtration $\mathcal{F}_t$ generated by the processes $\{x^i_{s\leq t}\}$

$$\mathcal{C}(\alpha, t, T) = \mathbb{E}^{\mathbb{P}}[f(x_T)|\mathcal{F}_t] \tag{4.2}$$

No discount factors have been added as we assume a zero interest rate. By definition of the conditional mean value, the fair value $\mathcal{C}$ depends on the conditional probability density $p(T,x|t,\alpha)$ by

$$\mathcal{C}(\alpha,t,T) = \int \prod_{i=1}^{n} dx^i f(x) p(T,x|t,\alpha) \tag{4.3}$$

REMARK 4.2 It is not obvious that the SDE (4.1) admits a (smooth) conditional probability density. We will come back to this point when we discuss the Hörmander theorem in the last section. □

Independently, from Itô's formula, we have

$$f(x_T) = f(\alpha) + \int_t^T Df(x_s)ds + \int_t^T \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x^i}\sigma^i(t,x)dW_i \tag{4.4}$$

with D a second-order differential operator

$$D = b^i(t,x)\frac{\partial}{\partial x^i} + \frac{1}{2}\sigma^i(t,x)\sigma^j(t,x)\rho_{ij}(t)\frac{\partial^2}{\partial x^i \partial x^j} \tag{4.5}$$

In (4.5), we have used the Einstein summation convention, meaning that two identical indices are summed. For example, $b^i(t,x)\frac{\partial}{\partial x^i} = \sum_{i=1}^n b^i(t,x)\frac{\partial}{\partial x^i}$.

We will adopt this Einstein convention throughout this book.

Taking the conditional mean-value $\mathbb{E}^{\mathbb{P}}[\cdot|\mathcal{F}_t]$ on both sides of (4.4), we obtain

$$\mathcal{C}(\alpha,t,T) = f(\alpha) + \mathbb{E}^{\mathbb{P}}[\int_t^T Df(x_s)ds|\mathcal{F}_t] \tag{4.6}$$

where we have used that $\mathbb{E}^{\mathbb{P}}[\int_t^T \frac{\partial f(x)}{\partial x^i}\sigma^i(t,x)dW_i|\mathcal{F}_t] = 0$ by Itô isometry. Finally, differentiating (4.6) according to the variable T, we obtain

$$\begin{aligned}\frac{\partial \mathcal{C}(\alpha,t,T)}{\partial T} &= \mathbb{E}^{\mathbb{P}}[Df(x_T)|\mathcal{F}_t] \\ &\equiv \int \prod_{i=1}^{n} dx^i Df(x) p(T,x|t,\alpha)\end{aligned}$$

Integrating by parts and discarding the surface terms, we obtain

$$\frac{\partial \mathcal{C}(\alpha,t,T)}{\partial T} = \int \prod_{i=1}^{n} dx^i f(x) D^\dagger p(T,x|t,\alpha) \tag{4.7}$$

with

$$D^{\dagger} = -\frac{\partial}{\partial x^i} b^i(x,t) + \frac{1}{2}\rho_{ij}(t)\frac{\partial^2}{\partial x^i \partial x^j}\sigma^i(x,t)\sigma^j(x,t)$$

Moreover, differentiating (4.3) according to T, we have

$$\frac{\partial \mathcal{C}(\alpha,t,T)}{\partial T} = \int \prod_{i=1}^{n} dx^i f(x) \frac{\partial p(T,x|t,\alpha)}{\partial T} \tag{4.8}$$

Since $f(x)$ is an arbitrary function, we obtain by identifying (4.7) and (4.8) that the conditional probability density satisfies the *forward Kolmogorov equation (or Fokker-Planck equation)* given by

$$\frac{\partial p(T,x|t,\alpha)}{\partial T} = D^{\dagger} p(T,x|t,\alpha)$$

with the initial condition

$$\lim_{T \to t} p(T,x|t,\alpha) = \delta(x-\alpha)$$

The initial condition is to be understood in the weak sense, i.e.,

$$\lim_{T \to t^+} \int \prod_{i=1}^{n} dx^i p(T,x|t,\alpha) f(x) = f(\alpha)$$

for any compactly supported function f on $\mathbb{R}^n$.
Coming back to the surface terms in (4.7), we have the boundary conditions

$$\left(b^i - \frac{1}{2}\rho_{ij}\partial_i(\sigma_i\sigma_j)\right) f + \frac{1}{2}\rho_{ij}\sigma_i\sigma_j\partial_j f = 0 \ , \ \forall x \in \partial M$$

4.1.2 Backward Kolmogorov's equation

From the Feynman-Kac theorem, we have that the option fair value satisfies the PDE

$$-\partial_t \mathcal{C}(\alpha,t,T) = b^i(t,\alpha)\frac{\partial \mathcal{C}(\alpha,t,T)}{\partial \alpha^i} + \frac{1}{2}\rho_{ij}(t)\sigma^i(t,\alpha)\sigma^j(t,\alpha)\frac{\partial^2 \mathcal{C}(\alpha,t,T)}{\partial \alpha^i \partial \alpha^j}$$

From (4.3), we obtain that $p(T,x|t,\alpha)$ satisfies the backward Kolmogorov equation

$$\begin{aligned} -\frac{\partial p}{\partial t} &= b^i(t,\alpha)\frac{\partial p(T,x|t,\alpha)}{\partial \alpha^i} + \frac{1}{2}\rho_{ij}(t)\sigma^i(t,\alpha)\sigma^j(t,\alpha)\frac{\partial^2 p(T,x|t,\alpha)}{\partial \alpha^i \partial \alpha^j} \\ p(t=T,x|t,\alpha) &= \delta(\alpha - x) \end{aligned}$$

We assume in the following that b^i, σ^i and ρ_{ij} are time-independent (for an extension to the time-dependent case, see section (4.6)). Let us define

$\tau = T - t$. $p(T, x|t, \alpha) \equiv p(\tau, x|\alpha)$ only depends on the combination $T - t$ and not on t or T separately (i.e., solutions are time-homogeneous). Then $p(\tau, x|\alpha)$ satisfies the *backward Kolmogorov* PDE

$$\frac{\partial p(\tau, x|\alpha)}{\partial \tau} = Dp(\tau, x|\alpha) \tag{4.9}$$

with the initial condition

$$p(\tau = 0, x|\alpha) = \delta(x - \alpha)$$

and with D defined by (with $\partial_i = \frac{\partial}{\partial \alpha_i}$)

$$D = b^i(\alpha)\partial_i + \frac{1}{2}\rho_{ij}\sigma^i(\alpha)\sigma^j(\alpha)\partial^2_{ij}$$

The following section is dedicated to short-time asymptotic solutions for the multi-dimensional Kolmogorov equation (4.9).
Note that the coefficients in the Kolmogorov equation do not behave nicely under a change of variables. Indeed if we do a change of variables from $x = \{x^i\}_{i=1,\cdots,n}$ to $x^{i'} = f^{i'}(x)$ with f a $C^\infty(\mathbb{R}^n, \mathbb{R}^n)$-function, via Itô's formula one shows that the $x^{i'}$ variables satisfy the SDE

$$dx^{i'} = b^{i'}(x')dt + \sigma^{i'}(x')dW_{i'}$$

with

$$b^{i'}(x') = \frac{\partial f^{i'}}{\partial x^i} b^i(x) + \frac{1}{2}\rho_{ij}\sigma^i(x)\sigma^j(x)\frac{\partial^2 f^{i'}(x)}{\partial x^i \partial x^j} \tag{4.10}$$

$$\sigma^{i'}(x') = \frac{\partial f^{i'}(x)}{\partial x^i}\sigma^i(x) \tag{4.11}$$

The volatility $\sigma^i(x)$ transforms covariantly (4.11). This is not the case for the drift (4.10). The non-covariant term is due to the Itô additional term. We will see in the following how to transform the non-covariant Kolmogorov equation into an equivalent covariant equation, the *heat kernel* equation. For this purpose, we will introduce a metric and an Abelian connection, depending on the drift and the volatility terms, that transform covariantly under a change of variables.
In the next section, we introduce the reader to a few basic notions in differential geometry, useful to formulate a heat kernel equation on a Riemannian manifold and to state the short-time asymptotic solution to the Kolmogorov equation. Detailed proofs and complements can be found in [14] and [22].

4.2 Notions in differential geometry

> The beginner should not be discouraged if he finds that he does not have the prerequisite for reading the prerequisites.
> — P. Halmos

4.2.1 Manifold

A real n-dimensional *manifold* M is a space which looks like $\mathbb{R}^n$ around each point. More precisely, M is covered by open sets $\mathcal{U}_i$ (i.e., M is a topological space) which are homeomorphic to $\mathbb{R}^n$ meaning that there is a continuous application ϕ_i (and its inverse) from $\mathcal{U}_i$ to $\mathbb{R}^n$ for each i. $(\mathcal{U}_i, \phi_i)$ is called a *chart* and ϕ_i a map. Via a map $\phi_i : \mathcal{U}_i \to \mathbb{R}^n$, we can endow the open set $\mathcal{U}_i$ with a system of coordinates $\{x_k\}_{k=1,\cdots,n}$ by $x = \phi(p)$ with $p \in \mathcal{U}_i$.

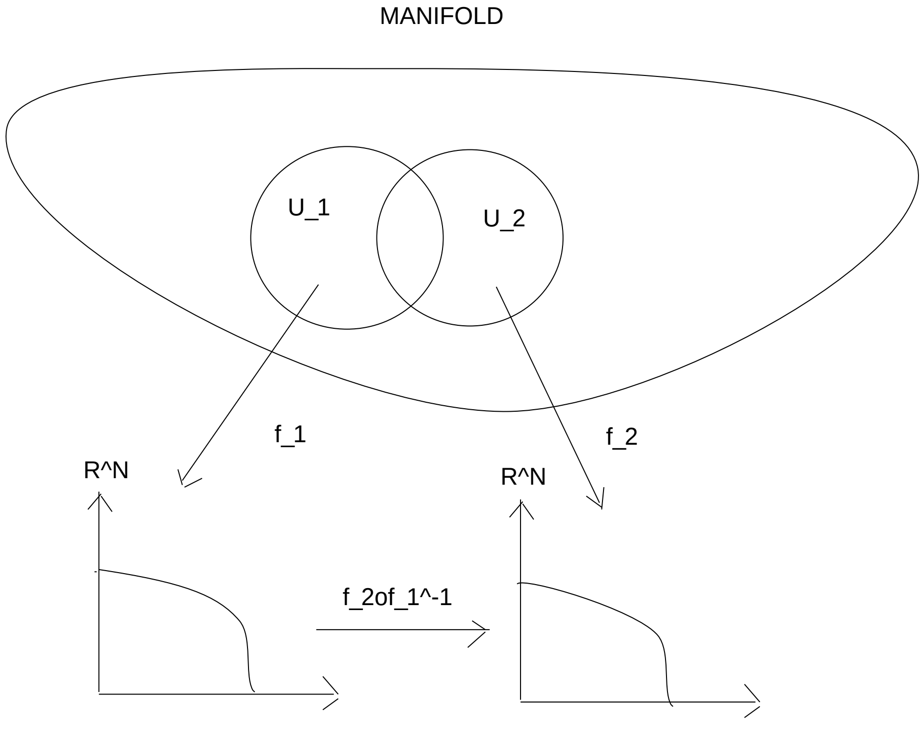

FIGURE 4.1: Manifold

The figure (4.1) shows that a point p belonging to the intersection of the two charts $\mathcal{U}_1$ and $\mathcal{U}_2$ corresponds to two systems of coordinates x_1 (resp. x_2) via the map ϕ_1 (resp. ϕ_2). We should give a rule indicating how to pass from a system of coordinates to the other. We impose that the applications $\phi_{i,j} \equiv \phi_i \circ \phi_j^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ are $C^\infty(\mathbb{R}^n)$. We say that we have a C^∞-manifold. In conclusion, we have a manifold structure if we define a chart and if the composition of two maps is an infinitely differentiable function. Let us see a simple example of a manifold: the 2-sphere.

Example 4.1 2-Sphere
The two-sphere

$$S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

can be covered with two patches: $\mathcal{U}_N$ and $\mathcal{U}_S$, defined respectively as S^2 minus the north pole and the south pole. The map ϕ_N (resp. ϕ_S) is obtained by a stereographic projection on $\mathcal{U}_N$ (resp. $\mathcal{U}_S$). This projection consists in taking the intersection of the equatorial plane with a line passing through the North (resp. South) pole and a point p on S^2 (see Fig. 4.2). The resulting maps are

$$\phi_N(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$
$$\phi_S(x,y,z) = \left(\frac{x}{1+z}, \frac{y}{1+z}\right)$$

The change of coordinates defined on $\mathcal{U}_N \cap \mathcal{U}_S$ is

$$\phi_N \circ \phi_S^{-1}(x,y) = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$$

and is C^∞. So S^2 is a C^∞*-manifold.* □

4.2.2 Maps between manifolds

Let us define $f : M \to N$ a map from an m-dimensional manifold M with a chart $\{\mathcal{U}_i, \phi_i\}_i$ to a n-dimensional manifold N with a chart $\{\mathcal{V}_i, \psi_i\}_i$. A point $p \in M$ is mapped to a point $f(p) \in N$. Let us take $\mathcal{U}_i$ an open set containing p with the map ϕ_i and $\mathcal{V}_i$ an open set containing $f(p)$ with the map ψ_i. By using the maps ϕ_i and ψ_i, f can be viewed as a map from $\mathbb{R}^m$ to $\mathbb{R}^n$

$$F \equiv \psi_i \circ f \circ \phi_i^{-1} : \mathbb{R}^m \to \mathbb{R}^n$$

If we set $\phi_i(p) = x$ and $\psi_i(f(p)) = y$, we can write this map as

$$y^j = f^j(x)$$

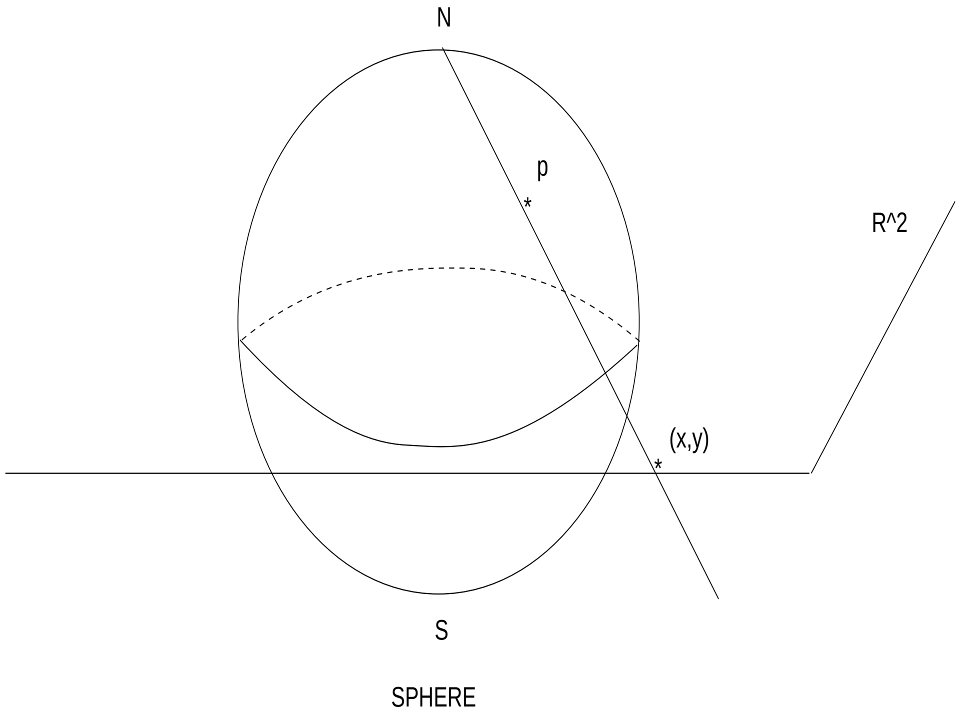

FIGURE 4.2: 2-sphere

We say that f is C^∞-differentiable if $f = \{F^j\}_{j=1,\cdots,n}$ is a C^∞-differentiable function from $\mathbb{R}^m$ to $\mathbb{R}^n$. Note that this definition is independent from the open sets $\mathcal{U}_i$ and $\mathcal{V}_i$ containing respectively p and $f(p)$. Indeed, let $\mathcal{U}'_i$ be another open set including p with a map ${\phi'}_i$. Therefore, we have the other representative of f:

$$\psi_i \circ f \circ {\phi'}_i^{-1} : \mathbb{R}^m \to \mathbb{R}^n$$

As $\phi_i \circ {\phi'}_i^{-1} : \mathbb{R}^m \to \mathbb{R}^m$ is a C^∞-differentiable function by definition of a smooth manifold M, therefore $\psi_i \circ f \circ {\phi'}_i^{-1}$ is if $\psi_i \circ f \circ \phi_i^{-1}$ is.
In the following, the space of C^∞-differentiable functions from M to $\mathbb{R}$ is noted $C^\infty(M)$.

4.2.3 Tangent space

On the space $C^\infty(M)$, we define a derivation:

DEFINITION 4.1 derivation *A derivation D at a point $x_0 \in M$ is a map $C^\infty(M) \to \mathbb{R}$ such that*

- *D is linear: $D(f + \lambda g) = D(f) + \lambda D(g)$ with $\lambda \in \mathbb{R}$.*
- *D satisfies the Leibnitz rule: $D(f.g) = D(f)g(x_0) + f(x_0)D(g)$.*

The set of derivations forms a vector space and also a $C^\infty(M)$-module as the product of a function $f \in C^\infty(M)$ with a derivation D is still a derivation. The vector space of all derivations at a point x_0 is denoted by $T_{x_0}M$ and called the *tangent space* to the manifold M at a point x_0.
Let $x_1, \cdots x_n$ be the local coordinates on the chart $\mathcal{U}$ containing the point x_0. The partial derivatives $(\frac{\partial}{\partial x^i})_{i=1,\cdots,n}$ evaluated at x_0 are independent and belong to $T_{x_0}M$. One can show that they form a basis for $T_{x_0}M$:

THEOREM 4.1
On a n-dimensional manifold M, $T_{x_0}M$ is a n-dimensional vector space generated by $\left(\frac{\partial}{\partial x^i}\right)_{i=1,\cdots,n}$

Therefore an element X of $T_{x_0}M$, called a *vector field*, can be uniquely decomposed as

$$X = X^i \frac{\partial}{\partial x^i}$$

X^i are the components of X in the basis $\{\frac{\partial}{\partial x^i}\}_{i=1,\cdots,n}$.
Let us consider another chart $\mathcal{U}'$ with local coordinates $x'_1, \cdots x'_n$ containing the point x_0. By using these new coordinates, the vector field can be decomposed as $X = X^{i'} \frac{\partial}{\partial x^{i'}}$. As we have

$$X = X^{i'} \frac{\partial x^i}{\partial x^{i'}} \frac{\partial}{\partial x^i} = X^i \frac{\partial}{\partial x^i}$$

we find that the components of a vector field written in two different charts transform covariantly as

$$X^i = X^{i'} \frac{\partial x^i}{\partial x^{i'}} \tag{4.12}$$

4.2.4 Metric

A metric $g_{ij}(x)$ written with the local coordinates $x \equiv \{x^i\}_{i=1,\cdots,n}$ (corresponding to a particular chart $\mathcal{U}$) is a symmetric non-degenerate and C^∞-differentiable tensor. It allows us to measure the distance between infinitesimally nearby points p with coordinates x^i and $p+dp$ with coordinates $x^i + dx^i$ by

$$ds^2 = g_{ij} dx^i dx^j \tag{4.13}$$

As $p + dp$ is assumed to be infinitesimally close to the point p, it belongs to the same chart and has a common system of coordinates with p.
If a point p (and $p + dp$) belongs to two different charts $\mathcal{U}$ and $\mathcal{U}'$ then the distance can be computed using two different systems of coordinates x^i and $x^{i'} = f(x)$ with $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^n)$. However, the result of the measure between p and $p + dp$ should be the same if the two points are considered to be on $\mathcal{U}$ or $\mathcal{U}'$, meaning that

$$g_{ij}(x)dx^i dx^j = g_{i'j'}(x')dx^{i'}dx^{j'} \tag{4.14}$$

As $dx^{i'} = \frac{\partial x^{i'}}{\partial x^i} dx^i$, we deduce that under a change of coordinates, the metric is not invariant but on the contrary changes in a contravariant way by

$$g_{ij}(x) = g_{i'j'}(x') \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^{j'}}{\partial x^j} \tag{4.15}$$

A manifold endowed with a metric is called a *Riemannian manifold.*
From this metric, we can measure the length of a curve $\mathcal{C}$. The basic idea is to divide this curve into infinitesimal pieces whose square of the length is given by (4.13). The rule (4.15) enables to glue the pieces belonging to different charts (i.e., different systems of coordinates). Let $\mathcal{C} : [0, 1] \to M$ be a C^1-differentiable or piecewise C^1-differentiable curve (parameterized by $x^i(t)$) joining the two points $x(0) = x$ and $x(1) = y$. Its length $l(\mathcal{C})$ is defined by

$$l(\mathcal{C}) = \int_0^1 \sqrt{g_{ij}(x(t)) \frac{dx^i(t)}{dt} \frac{dx^j(t)}{dt}} dt \tag{4.16}$$

4.2.5 Cotangent space

We can associate to a vector space E its dual E^* defined as the vector space of linear forms on E. Assuming that E is a finite n-dimensional vector space generated by the basis $\{e_i\}_{i=1,\cdots,n}$, an element $x \in E$ can be decomposed uniquely as

$$x = x^i e_i$$

For a linear form l on E, we have by linearity that

$$l(x) = x^i l(e_i)$$

Therefore, the map l is completely fixed if we know the value $l(e_i)$. E^* is a n-dimensional vector space with a (dual) basis $\{e^j\}_{j=1,\cdots,n}$ defined by

$$e^j(e_i) = \delta_i^j$$

with δ_i^j the Kronecker symbol. l can therefore be decomposed over this basis as

$$l = l_i e^i$$

with $l(e_i) = l_i$.
Let M be a smooth manifold and x_0 a point of M. The dual vector space to $T_{x_0}M$ is called the *cotangent vector space* and noted $T^*_{x_0}M$. An element of $T^*_{x_0}M$ is a linear form ω which when applied to a vector field X in $T_{x_0}M$ gives a real number. Elements of $T^*_{x_0}M$ are called *one-form*. A particular one-form (called *exact one-form*) is defined by

$$df(X) = X(f) \ , \ X \in T_{x_0}M$$

with $f \in C^\infty(M)$.
If $x^1, \cdots, x^n$ are local coordinates of x_0 and $(\frac{\partial}{\partial x^i})_{i=1,\cdots,n}$ a basis of $T_{x_0}M$, then we write $(dx^i)_{i=1,\cdots,n}$ as the dual basis of $T^*_{x_0}M$ as

$$dx^j\left(\frac{\partial}{\partial x^i}\right) = \frac{\partial x^j}{\partial x^i} = \delta_i^j$$

By using this canonical basis, a one-form ω can be uniquely decomposed as

$$\omega = \omega_i dx^i$$

For an exact one-form df, we have

$$df = \frac{\partial f}{\partial x^i} dx^i$$

Note that if we use another system of coordinates $x^{1'}, \cdots, x^{n'}$ of x_0, we have

$$\omega = \omega_{i'} dx^{i'} = \omega_{i'} \frac{\partial x^{i'}}{\partial x^i} dx^i$$

Therefore, the components of a one-form transform contravariantly under a change of coordinates as

$$\omega_i = \omega_{i'} \frac{\partial x^{i'}}{\partial x^i} \tag{4.17}$$

4.2.6 Tensors

In this section, we see how to map vectors to one-forms using the metric which defines an isomorphism between the tangent and cotangent vector spaces. Then, we will define general (r, p)-tensors where $(1, 0)$ is a vector and $(0, 1)$ is a one-form.

The data of a non-degenerate bilinear form $g(\cdot,\cdot)$ on a vector space E is equivalent to the data of an isomorphism between E and its dual E^*, $\mathbf{g}: E \to E^*$, according to the relation

$$g(X, Y) = \mathbf{g}(X)(Y)$$

By using a basis $\{e_i\}$ of E and its dual basis $\{e^i\}$ of E^*, the above-mentioned relation can be written as

$$e_i = g_{ij} e^j$$

with $g_{ij} = g(e_i, e_j)$. Similarly, through multiplying the equation above by the inverse of the matrix $[g_{ij}]$, noted g^{ij}, we obtain

$$e^i = g^{ij} e_j$$

Similar expressions can be obtained on the components of an element of E or E^*:

$$\omega = \omega_i e^i = \omega_i g^{ij} e_j \equiv \omega^i e_i$$

and therefore

$$\omega^i = g^{ij} \omega_j$$
$$\omega_i = g_{ij} \omega^j$$

Applying this isomorphism between the tangent space $T_{x_0}M$ and the cotangent space $T^*_{x_0}M$ using the metric, we can map the components of a vector field X^i to the components of a one-form X_i by

$$X^i = g^{ij} X_j$$
$$X_i = g_{ij} X^j$$

The metric is therefore a machine to lower and raise the indices of vectors and one-forms.
A *tensor* of type (r, p) is a multi-linear object which maps r elements of $T^*_{x_0}M$ and p elements of $T_{x_0}M$ to a real number. The set of tensors of type (r, p) forms a vector space noted $T^{(r,p)}_{x_0}M$. An element T of $T^{(r,p)}_{x_0}M$ is written in the basis $\{\frac{\partial}{\partial x^i}\}_i$ and $\{dx^i\}_i$ as

$$T = T^{i_1\cdots i_r}_{j_1\cdots j_p} \frac{\partial}{\partial x_{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x_{i_r}} dx^{j_1} \otimes \cdots \otimes dx^{j_p}$$

By using another system of coordinates $x^{i'}$, the components of T transform as

$$T^{i'_1\cdots i'_r}_{j'_1\cdots j'_p} = T^{i_1\cdots i_r}_{j_1\cdots j_p} \frac{\partial x^{i'_1}}{\partial x_{i_1}} \cdots \frac{\partial x^{i'_r}}{\partial x_{i_r}} \frac{\partial x^{j_1}}{\partial x^{j'_1}} \cdots \frac{\partial x^{j_r}}{\partial x^{j'_r}} \tag{4.18}$$

A special $(0,p)$-tensor is a *p-differential form* which is a totally antisymmetric $(0,p)$-tensor. Let us define the wedge product $\wedge$ of p one-forms by the totally antisymmetric tensor product

$$dx^{i_1} \wedge \cdots dx^{i_p} = \sum_{P \in S_p} \mathrm{sgn}(P) dx^{i_{P(1)}} \otimes \cdots \otimes dx^{i_{P(p)}} \tag{4.19}$$

where P is an element of S_p, the symmetric group of order p and $\mathrm{sgn}(P) = +1$ (resp. $\mathrm{sgn}(P) = -1$) for even (resp. odd) permutations.

Example 4.2

$$dx^i \wedge dx^j = dx^i \otimes dx^j - dx^j \otimes dx^i$$

$\square$

The elements (4.19) form a basis of the vector space of r-forms, $\Omega^r(M)$, and an element $\omega \in \Omega^r(M)$ is decomposed as

$$\omega = \frac{1}{n!} \omega_{i_1 \cdots i_n} dx^{i_1} \wedge \cdots \wedge dx^{i_n}$$

We can then define the *exterior derivative* $d : \Omega^r(M) \to \Omega^{r+1}(M)$ acting on r-forms as

$$d\omega = \frac{1}{n!} \partial_k \omega_{i_1 \cdots i_n} dx^k \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_n}$$

One easily checks that d is a linear operator and $d^2 = 0$.

Example 4.3
The exterior derivative d on a 1-form $\mathcal{A} = \mathcal{A}_i dx^i$ gives

$$\begin{aligned} d\mathcal{A} &= \partial_j \mathcal{A}_i dx^j \wedge dx^i \\ &= (\partial_j \mathcal{A}_i - \partial_i \mathcal{A}_j) \, dx^j \otimes dx^i \end{aligned}$$

$\square$

Tensors, metrics, forms, vector fields have been locally defined on a neighborhood of x_0 as elements of $T^{(r,p)}_{x_0} M$ and we have explained how these objects are glued together (4.12), (4.17), (4.18) on the intersection of two patches. We can unify these definitions with the notion of *vector bundles*. The tensor fields will then be seen as *sections* of particular vector bundles.

4.2.7 Vector bundles

The main idea of a vector bundle (more generally of differential geometry) is to define a vector space over each chart of a manifold M. To a point x_0 belonging to two different patches, we can associate two different vector spaces and we need to specify how to glue them.

A *real vector bundle* $\mathcal{E}$ of rank m is defined as follows: one starts with an open covering of M, $\{\mathcal{U}_\alpha\}_{\alpha \in A}$, and for each $(\alpha, \beta) \in A$, a smooth transition function

$$g_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \rightarrow \mathrm{GL}(m, \mathbb{R})$$

where $\mathrm{GL}(m, \mathbb{R})$ is the space of invertible real matrix of dimension m. We impose that the function $g_{\alpha\beta}$ satisfies

$$g_{\alpha\alpha} = 1_m$$

and the *co-cycle* condition on the triple intersection $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \cap \mathcal{U}_\gamma$

$$g_{\alpha\beta} g_{\beta\gamma} g_{\gamma\alpha} = 1_m \ \ on \ \ \mathcal{U}_\alpha \cap \mathcal{U}_\beta \cap \mathcal{U}_\gamma \tag{4.20}$$

with 1_m being the identity (linear) map on $\mathbb{R}^m$.

Let us denote $\tilde{E}$ the set of all triples $(\alpha, p, v) \in A \times M \times \mathbb{R}^m$ such that $p \in \mathcal{U}_\alpha$. Let us define an equivalence relation $\sim$ on $\tilde{E}$ by

$$\begin{aligned} (\alpha, p, v) \sim (\beta, q, w) \ \Leftrightarrow \ & p = q \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta \\ & v = g_{\alpha\beta} w \end{aligned} \tag{4.21}$$

Let us denote the equivalence classes of (α, p, v) by $[\alpha, p, v]$ and the set of equivalence classes by E and define a projection map

$$\pi : E \rightarrow M \ , \ \pi([\alpha, p, v]) = p$$

Let us define $\tilde{\mathcal{U}}_\alpha = \pi^{-1}(\mathcal{U}_\alpha)$ and a bijection by

$$\psi_\alpha = \tilde{\mathcal{U}}_\alpha \rightarrow \mathcal{U}_\alpha \times \mathbb{R}^m \ , \ \psi_\alpha([\alpha, p, v]) = (p, v)$$

There is a unique manifold structure on E which makes ψ_α into a diffeomorphism.

A real vector bundle of rank m is a pair (E, π) constructed as above and a collection of transition functions $g_{\alpha\beta}$ which satisfy the co-cycle condition (4.20). The *fiber* of E over $p \in M$ is the m-dimensional vector space $E_p = \pi^{-1}(p)$. When $m = 1$, the real vector bundle is called a *line bundle* (see Fig. 4.3).

A *section* σ of E is defined by its local representatives σ on each $\mathcal{U}_\alpha$:

$$\sigma|_{\mathcal{U}_\alpha} \equiv \sigma_\alpha : \mathcal{U}_\alpha \rightarrow \mathbb{R}^m$$

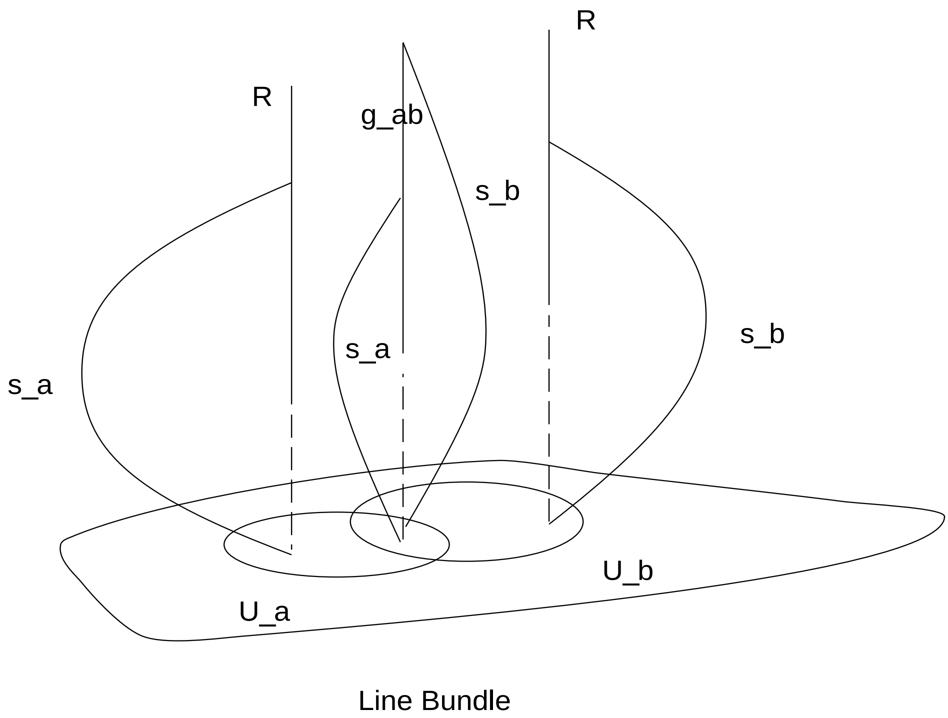

FIGURE 4.3: Line bundle

and they are related to each other by the formula

$$\sigma_\alpha = g_{\alpha\beta}\sigma_\beta \tag{4.22}$$

on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$. The smooth sections on E are denoted $\Gamma(E)$.
In the following, we give classical examples of real vector bundles.

Example 4.4 Trivial vector bundle
The simplest real bundle I of rank m which can be constructed on a manifold M is the trivial vector bundle given by the transition function

$$g_{\alpha\beta} = 1_m$$

The co-cycle condition (4.20) is trivially satisfied. One can show that I is isomorphic to the product $M \times \mathbb{R}^m$. ∎

Example 4.5 (Co)-Tangent vector bundle
The tangent space TM to a manifold M is a vector bundle of rank n. The

transition functions are given by

$$g_{\alpha'\alpha} : \mathcal{U}_{\alpha'} \cap \mathcal{U}_{\alpha} \to \mathrm{GL}(n, \mathbb{R})$$
$$g_{\alpha'\alpha}(x) = \frac{\partial x^{\alpha'}}{\partial x^{\alpha}}$$

A vector field is then a section of TM .
The cotangent vector bundle (of rank n) T^*M is the dual vector bundle of TM with a transition function given by

$$g^*_{\alpha'\alpha} : \mathcal{U}_{\alpha'} \cap \mathcal{U}_{\alpha} \to \mathrm{GL}(n, \mathbb{R})$$
$$g^*_{\alpha'\alpha} \equiv g^{-1\dagger}_{\alpha'\alpha} = \frac{\partial x^{\alpha}}{\partial x^{\alpha'}}$$

A one-form is a section of T^*M.
The tensor vector bundle of type (r, s) , $\otimes_{p=1}^r TM \otimes_{q=1}^s T^*M$, is then defined as the n-rank real vector bundle with transitions $\otimes_{p=1}^r g_{\alpha'\alpha} \otimes_{q=1}^s g^*_{\alpha'\alpha}$. A tensor field is a section of the tensor vector bundle and transforms as (4.18) conformingly to (4.22). ∎

4.2.8 Connection on a vector bundle

DEFINITION 4.2 Connection 1 *A connection d_A on a vector bundle E is a map*

$$d_A : \Gamma(E) \to \Gamma(T^*M \otimes E)$$

which satisfies the following axiom:

$$d_A(f\sigma + \tau) = df \otimes \sigma + f d_A \sigma + d_A \tau \tag{4.23}$$

with $f \in C^\infty(M)$, $\sigma \in \Gamma(E)$ and $\tau \in \Gamma(E)$.
Firstly, we will characterize a connection on a trivial bundle $E = M \times \mathbb{R}^m$. Then using that a vector bundle is locally trivial, we will generalize to a general vector bundle.
Let us consider a trivial bundle $E = M \times \mathbb{R}^m$ defined by the transition functions $g_{\alpha\beta} = \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{R})$, $g_{\alpha\beta} = 1_m$. A section is then represented by a global vector-valued map $\sigma : M \to \mathbb{R}^m$ which can be decomposed on the constant sections e_i:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} , \; e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \cdots , \; e_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

$$\sigma(x) = \sigma^i(x)e_i$$

By definition (4.23), we have

$$d_A\sigma = d\sigma^i \otimes e_i + \sigma^i d_A e_i$$

As $d_A e_i$ should be an element of $\Gamma(T^*M \otimes E)$, there exists a rank m matrix $A_i^j \in T^*M$ of 1-form, $A_i^j = A_{ki}^j dx^k$, such that

$$d_A e_i = A_i^j \otimes e_j$$

Therefore

$$d_A\sigma = (d\sigma^i + \sigma^j A_j^i) \otimes e_i$$

and the components of $d_A\sigma \equiv (d_A\sigma)^i e_i$ are

$$\begin{aligned}(d_A\sigma)^i &= d\sigma^i + \sigma^j A_j^i \\ &= (\partial_k\sigma^i + A_{kj}^i\sigma^j)dx^k\end{aligned}$$

We note the components of d_A, $d_A\sigma \equiv \nabla_k\sigma^i dx^k \otimes e_i$, and

$$\nabla_k\sigma^i = \partial_k\sigma^i + A_{kj}^i\sigma^j \tag{4.24}$$

Similarly, on a non-trivial bundle, a connection can be characterized as follows: A section $\sigma \in \Gamma(E)$ possesses a local representative σ_α which is a vector-valued map on $\mathcal{U}_\alpha$. So, proceeding as above, we can write

$$(d_A\sigma)_\alpha = d\sigma_\alpha + A_\alpha\sigma_\alpha$$

with A_α a rank m matrix of 1-form on $\mathcal{U}_\alpha$. Now, we need to specify how to glue together these local connections $(d_A\sigma)_\alpha$ and $(d_A\sigma)_\beta$ defined on two different patches $\mathcal{U}_\alpha$ and $\mathcal{U}_\beta$. For a section defined on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$, we have

$$\begin{aligned}(d_A\sigma)_\alpha &= d\sigma_\alpha + A_\alpha\sigma_\alpha \\ (d_A\sigma)_\beta &= d\sigma_\beta + A_\beta\sigma_\beta\end{aligned}$$

By using the transition function $g_{\alpha\beta}$, we impose that

$$\begin{aligned}\sigma_\beta &= g_{\beta\alpha}\sigma_\alpha \\ (d_A\sigma)_\beta &= g_{\beta\alpha}(d_A\sigma)_\alpha\end{aligned}$$

It gives

$$A_\beta = g_{\alpha\beta}^{-1}dg_{\alpha\beta} + g_{\alpha\beta}^{-1}A_\alpha g_{\alpha\beta} \tag{4.25}$$

DEFINITION 4.3 Connection 2 *A connection on a real vector bundle E of rank m defined by a covering $\{\mathcal{U}_\alpha : \alpha \in A\}$ and transition functions $g_{\alpha\beta}$: $\alpha, \beta \in A$ is a collection of first-order differential operators $\{d + A_\alpha : \alpha \in A\}$ where d is the exterior derivative on $\mathbb{R}^m$-valued functions and A_α is a rank m matrix of one-forms on $\mathcal{U}_\alpha$ which transform according to (4.25). When E is a line bundle, i.e., $m = 1$, A_α is called an Abelian connection.*

Applying to the tangent vector bundle E, a connection d_A is called a *covariant derivative* and A^i_{jk} is noted usually Γ^i_{jk}. The covariant derivative of the components of a vector field $X = X^i \partial_i$ is according to (4.24)

$$\nabla_i X^j = \partial_i X^j + \Gamma^j_{ik} X^k \tag{4.26}$$

The action of a covariant derivative on a vector field can be extended to any (r, s)-tensor fields by the rules:

1. If T is a tensor field of type (r, s), then ∇T is a tensor field of type $(r, s + 1)$.
2. ∇ is linear and commutes with contractions.
3. For any tensor fields T_1 and T_2, we have the Leibnitz rules
$$\nabla(T_1 \otimes T_2) = \nabla(T_1) \otimes T_2 + T_1 \otimes \nabla(T_2)$$
4. $\nabla f = df$ for any function $f \in C^\infty(M)$.

If we contract a one-form ω with a vector field X, $< \omega, X >$ is a function on M: $\omega_j X^j$. By using the rules 3 and 4, we have

$$\begin{aligned} \nabla_i(\omega_j X^j) &= \partial_i(\omega_j X^j) \\ &= \nabla_i(\omega_j) X^j + \omega_j \nabla_i(X^j) \end{aligned}$$

Taking $X = \partial_j$, we obtain

$$\nabla_i \omega_j = \partial_i \omega_j - \Gamma^k_{ij} \omega_k$$

Note the sign $-$ in front of the connection instead of the sign $+$ in (4.26). It is easy to generalize these results to a (q, p)-tensor field T

$$\begin{aligned} \nabla_\nu T^{j_1 \cdots j_p}_{i_1 \cdots i_q} &= \partial_\nu T^{j_1 \cdots j_p}_{i_1 \cdots i_q} + \Gamma^{j_1}_{\nu k} T^{k \cdots j_p}_{i_1 \cdots i_q} + \cdots \Gamma^{j_p}_{\nu k} T^{j_1 \cdots k}_{i_1 \cdots i_q} \\ &\quad - \Gamma^k_{\nu i_1} T^{j_1 \cdots j_p}_{k \cdots i_q} - \cdots - \Gamma^k_{\nu i_q} T^{j_1 \cdots j_p}_{i_1 \cdots k} \end{aligned} \tag{4.27}$$

Example 4.6 Levi-Cevita connection
Endowing M with a Riemannian metric g, the *Levi-Cevita connection* Γ is the unique symmetric connection on TM which preserves the metric

$$\nabla_k g_{ij} = 0$$

This equation is equivalent to (using (4.27))

$$\partial_k g_{ij} - \Gamma^p_{ki} g_{pj} - \Gamma^p_{kj} g_{pi} = 0 \tag{4.28}$$

Cyclic permutations of (k, i, j) yield

$$\partial_j g_{ki} - \Gamma^p_{jk} g_{pi} - \Gamma^p_{ik} g_{pj} = 0 \tag{4.29}$$
$$\partial_i g_{jk} - \Gamma^p_{ij} g_{pk} - \Gamma^p_{ji} g_{pk} = 0 \tag{4.30}$$

The combination of $-(4.28)+(4.29)+(4.30)$ yields

$$\begin{aligned}&-\partial_k g_{ij} + \partial_j g_{ki} + \partial_i g_{jk}\\&+(\Gamma^p_{ki} - \Gamma^p_{ik}) g_{pj} + (\Gamma^p_{kj} - \Gamma^p_{jk}) g_{pi} - (\Gamma^p_{ij} + \Gamma^p_{ji}) g_{pk} = 0\end{aligned}$$

If we impose that the connection is symmetric $\Gamma^p_{ij} = \Gamma^p_{ji}$ (the antisymmetric part is called the *torsion*), we obtain

$$-\partial_k g_{ij} + \partial_j g_{ki} + \partial_i g_{jk} - 2\Gamma^p_{ij} g_{pk} = 0$$

Solving for Γ^p_{ij}, the connection is uniquely given by

$$\Gamma^p_{ij} = \frac{1}{2} g^{pk} \left(-\partial_k g_{ij} + \partial_j g_{ki} + \partial_i g_{jk}\right) \tag{4.31}$$

Γ^p_{ij} are called the *Christoffel symbol.* ∎

4.2.9 Parallel gauge transport

Pullback bundle

If (E, π) is a vector bundle over M defined by a covering $\{\mathcal{U}_\alpha : \alpha \in A\}$ and transition functions $g_{\alpha\beta} : \alpha, \beta \in A$ and $F : N \to M$ a smooth map between two manifolds N and M, the *pullback bundle* (F^*E, π^*) is the vector bundle over N defined by the open covering $\{F^{-1}(\mathcal{U}_\alpha) : \alpha \in A\}$ and the transition functions $\{\tilde{g}_{\alpha\beta} = g_{\alpha\beta} \circ F : \alpha, \beta \in A\}$. It is easy to see that $\tilde{g}_{\alpha\beta}$ satisfies the co-cycle condition (4.20).
If we define a connection d_A on the vector bundle E, we can induce (we say *pullback*) a connection $d_{\pi^* A}$ on the pullback bundle $\pi^* E$ in the following way: Locally on $\mathcal{U}_\alpha$, d_A is represented by

$$d_A|_{\mathcal{U}_\alpha} = d + A_\alpha$$

with $A_\alpha = A_{\alpha k} dx^k$. On the open covering $F^{-1}(\mathcal{U}_\alpha)$, the pullback connection $d_{\pi^* A}$ is defined as

$$d_{\pi^* A}|_{F^{-1}(\mathcal{U}_\alpha)} = d + \pi^* A_\alpha$$

with $\pi^* A_\alpha = A_{\alpha k} \frac{\partial F^k}{\partial x^\alpha} dx^\alpha$.

Applying this construction to a curve $\mathcal{C} = [0,1] \to M$ and a connection d_A on a vector bundle (E, π) on M, we induce a pullback connection on $\mathcal{C}$, $d_{\mathcal{C}^*A}$. Now, we consider the sections σ on $\mathcal{C}^*E$ which are preserved by the connection $d_{\mathcal{C}^*A}$ meaning that

$$d_{\mathcal{C}^*A}\sigma = 0 \tag{4.32}$$

Locally on $\mathcal{C}^{-1}(\mathcal{U}_\alpha)$, this condition reads

$$d\sigma_\alpha + (\mathcal{C}^*A_\alpha)\sigma_\alpha = 0$$

Written in local coordinates, we have the first-order differential equation

$$\frac{d\sigma_\alpha(t)}{dt} + (\mathcal{C}^*A_{dt})\sigma_\alpha(t) = 0 \tag{4.33}$$

It follows from the theory of ordinary differential equations (ODEs) that given an element $\sigma_0 \in (\mathcal{C}^*E)_0$, there is a unique solution of (4.33) which satisfies the initial condition $\sigma(0) = \sigma_0$. We can define the isomorphism

$$\mathcal{P} : (\mathcal{C}^*E)_0 \to (\mathcal{C}^*E)_1$$

by setting $\mathcal{P}(\sigma_0) = \sigma(1)$ where σ is the unique solution to (4.33) which satisfies $\sigma(0) = \sigma_0$. $\mathcal{P}$ is called the *parallel gauge transport*. When E is a line bundle, $\mathcal{C}^*A_{dt}$ is no more a matrix but a real number. In this case, $\mathcal{P}$ is given by

$$\mathcal{P} = e^{-\int_0^1 \mathcal{C}^*A_{dt}(u)du}$$

4.2.10 Geodesics

We apply the previous construction to the parallel transport of a vector, tangent to a curve $\mathcal{C}$, along this curve. The ODE (4.33) will be called a *geodesic equation* and the curve $\mathcal{C}$, a *geodesic curve*.
Given a curve on a Riemannian manifold M, we may define the parallel (gauge) transport of a vector field tangent to a curve $\mathcal{C}$ using the Levi-Cevita connection:
Let V be a vector field tangent to a curve $\mathcal{C} = [0,1] \to M$. For simplicity, we assume that the curve is covered by a single chart $(\mathcal{U}, \phi)$ with coordinates $\{x_i\}_{i=1,\cdots,n}$. Therefore, the vector V tangent to the curve $\mathcal{C}$ can be written as

$$V|_{\mathcal{C}} = \frac{dx^i(\mathcal{C}(t))}{dt}\frac{\partial}{\partial x^i}$$

The (tangent) vector V is said to be parallel transport along its curve $\mathcal{C}(t)$ if V satisfies the condition (4.32) (with ∇ the Levi-Cevita connection)

$$\nabla_{\mathcal{C}^*V}V = 0$$

This condition is written in terms of components as

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt} = 0 \tag{4.34}$$

Indeed, we obtain this ODE from (4.33) by observing that $\sigma(t) \equiv \{\frac{dx^i(t)}{dt}\}_i$ and $\mathcal{C}^*\Gamma_{dt} = \{\Gamma^i_{jk}\frac{dx^j}{dt}\}^i_k$. This ODE is called the *geodesic equation.* We recall that the Christoffel coefficients Γ^i_{jk} depend on the metric and its first derivatives by (see example 4.6)

$$\Gamma^p_{ij} = \frac{1}{2}g^{pk}\left(-\partial_k g_{ij} + \partial_j g_{ki} + \partial_i g_{jk}\right) \tag{4.35}$$

The curve $\mathcal{C}$ satisfying this equation is called a *geodesic curve.*
Geodesics are, in some sense, the straightest possible curves in a Riemannian manifold. To understand this point, we try to minimize the length (4.16) of a curve $\mathcal{C}$ joining two points $x(0) = x$ and $x(1) = y$:

$$\min_{\mathcal{C}} \int_0^1 \sqrt{g_{ij}(x(t))\frac{dx^i(t)}{dt}\frac{dx^j(t)}{dt}}dt \tag{4.36}$$

By using the Euler-Lagrange equation giving the critical point of this functional, we obtain the geodesics equation (4.34).
As the geodesics are the main mathematical objects which appear in the heat kernel expansion technique, we give below examples of metrics on two-dimensional manifolds and solve explicitly the geodesic equations. These metrics will re-emerge naturally when we will discuss implied volatility asymptotics of stochastic volatility models.

Example 4.7 Hyperbolic surface
Let us consider the metric on the hyperbolic surface which is the complex upper half-plane $\mathbb{H}^2 = \{x \in \mathbb{R}, y > 0\}$

$$ds^2 = \frac{dx^2 + dy^2}{y^2} \tag{4.37}$$

We will give more information on this geometry when we will look at the SABR model in chapter 6. For the moment, we focus on the derivation of the geodesics curves and the geodesic distance which are the two main ingredients in the heat kernel expansion. There are only three non-zero Christoffel symbols[2](4.35) ($x_1 = x$, $x_2 = y$)

$$\Gamma^1_{12} = \Gamma^2_{22} = -\Gamma^2_{11} = -\frac{1}{y}$$

[2]Note that this computation can be easily done using Mathematica with the package *tensor* http://home.earthlink.net/ djmp/TensorialPage.html.

So the geodesic equations (4.34) reduce to

$$\ddot{x} - \frac{2}{y}\dot{x}\dot{y} = 0 \tag{4.38}$$

$$\ddot{y} + \frac{1}{y}\dot{x}^2 - \frac{1}{y}\dot{y}^2 = 0 \tag{4.39}$$

The dot means derivatives according to the parameter s which we have chosen to parameterize the geodesic curve. First note that the geodesic for x (4.38) can be simplified to read

$$\frac{d}{ds}\left(\ln(\dot{x}) - \ln(y^2)\right) = 0$$

or

$$\dot{x} = Cy^2 \tag{4.40}$$

for some unspecified constant C.
If we choose $C = 0$, we have x constant and y is a vertical line from (4.39):

$$\begin{aligned} y(s) &= y_0 e^{cs} \\ x(s) &= x_0 \end{aligned} \tag{4.41}$$

with c an integration constant. We assume below that $C \neq 0$.
The geodesic equation (4.39) for y is equivalent to

$$\frac{d}{ds}\left(\frac{\dot{x}^2 + \dot{y}^2}{y^2}\right) = 0$$

that we integrate to

$$\dot{x}^2 + \dot{y}^2 = y^2$$

where we have chosen an undetermined constant to 1. In fact, this equation is the definition of the parameter s (4.37). When we substitute in the known value (4.40) of $\frac{dx}{ds}$, we find that[3]

$$\frac{dy}{ds} = y\sqrt{1 - C^2y^2}$$

Next, we use the chain rule to find

$$\frac{\frac{dx}{ds}}{\frac{dy}{ds}} = \frac{dx}{dy} = \frac{Cy}{\sqrt{1 - C^2y^2}}$$

[3]We have chosen the sign + in front of the square root.

and

$$(x - c) = \int^{y}_{.} \frac{Cy'dy'}{\sqrt{1 - C^2 y^2}}$$
$$= -\sqrt{\frac{1}{C^2} - y^2}$$

where c is another undetermined constant. Upon squaring both sides, we find

$$y^2 + (x - c)^2 = \frac{1}{C^2} \tag{4.42}$$

indicating that the geodesic for $C \neq 0$ is a semi-circle with radius $\frac{1}{C}$ centered at $(c, 0)$.
In conclusion, geodesics are semi-circles centered at $(c, 0)$ on the x-axis and of radius $\frac{1}{C}$ (4.42) and vertical lines (4.41).
As the geodesic curve has been parameterized by the affine coordinate s representing the length of the geodesic curve, the square of the geodesic distance d between two points (x_1, y_1) and (x_2, y_2) joining by a geodesic curve is

$$\begin{aligned} d^2 &\equiv (s_2 - s_1)^2 \\ &= \int_{y_1}^{y_2} \frac{dy'}{y' \sqrt{1 - C^2 y^2}} \\ &= \ln \frac{y_2}{y_1} - \ln \left(\frac{1 + \sqrt{1 - C^2 y_2^2}}{1 + \sqrt{1 - C^2 y_1^2}} \right) \end{aligned}$$

C is fixed by the constraint

$$x_2 - x_1 = -\sqrt{\frac{1}{C^2} - y_2^2} + \sqrt{\frac{1}{C^2} - y_1^2}$$

which can be solved explicitly. Then after some simple algebraic transformations, we obtain the final expression for the geodesic distance

$$d = \cosh^{-1} \left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2 y_2 y_1} \right) \tag{4.43}$$

□

Example 4.8 Metric on a Riemann surface with one Killing vector
Let us consider a general two-dimensional Riemannian manifold (M, g). In a local coordinate (u, v), the metric takes the form

$$ds^2 = g_{uu} du^2 + 2 g_{uv} du dv + g_{vv} dv^2 \tag{4.44}$$

Note g the determinant of the metric, $g = g_{uu}g_{vv} - g_{uv}^2$ and rewrite (4.44) as

$$ds^2 = (\sqrt{g_{uu}}du + \frac{g_{uv} + \imath\sqrt{g}}{\sqrt{g_{uu}}}dv)(\sqrt{g_{uu}}du + \frac{g_{uv} - \imath\sqrt{g}}{\sqrt{g_{uu}}}dv)$$

According to the theory of differential equations, there exists an integrating factor $\lambda(u,v) = \lambda_1(u,v) + \imath\lambda_2(u,v)$ such that

$$\lambda(\sqrt{g_{uu}}du + \frac{g_{uv} + \imath\sqrt{g}}{\sqrt{g_{uu}}}dv) = dx + \imath dy$$
$$\lambda^*(\sqrt{g_{uu}}du + \frac{g_{uv} - \imath\sqrt{g}}{\sqrt{g_{uu}}}dv) = dx - \imath dy$$

Then in the new coordinates (x,y), we have by setting $|\lambda|^{-2} = e^{2\phi(x,y)}$

$$ds^2 = e^{2\phi(x,y)}(dx^2 + dy^2) \tag{4.45}$$

The coordinates (x,y) are called the *isothermal coordinates.*
We consider in the following the most general metric on a Riemann surface with one *Killing vector* meaning that $\phi(x,y)$ only depends on x: $\phi(x,y) \equiv \frac{1}{2}\ln(F(x))$. So, we have

$$ds^2 = F(y)(dx^2 + dy^2) \tag{4.46}$$

The Christoffel symbols are

$$\Gamma^1_{12} = \Gamma^2_{22} = -\Gamma^2_{11} = \frac{F'(y)}{2F(y)}$$

and the geodesic equation reads

$$\ddot{x} + \frac{F'(y)}{F(y)}\dot{x}\dot{y} = 0 \tag{4.47}$$
$$\ddot{y} - \frac{F'(y)}{2F(y)}\dot{x}^2 + \frac{F'(y)}{2F(y)}\dot{y}^2 = 0 \tag{4.48}$$

The dot $\cdot$ (resp. prime) means derivative according to the parameter s (resp. y). The first equation (4.47) is

$$\dot{x} = CF(y)^{-1} \tag{4.49}$$

for some constant C. The geodesic equation (4.48) for y is equivalent to

$$\frac{d}{ds}\left(F(y)(\dot{x}^2 + \dot{y}^2)\right) = 0$$

that we integrate to

$$\dot{x}^2 + \dot{y}^2 = \frac{1}{F(y)} \tag{4.50}$$

where we have chosen the undetermined constant to 1. In fact, this equation is the definition of the parameter s (4.46). From (4.49), the equation (4.50) simplifies to

$$\frac{dy}{ds} = \frac{\sqrt{F(y) - C^2}}{F(y)}$$

and from (4.49), we have

$$\frac{dx}{dy} = \frac{C}{\sqrt{F(y) - C^2}}$$

As usual, the geodesic curve has been parameterized by s and therefore the geodesic distance between the points (x_1, y_1) (reached at s_1) and (x_2, y_2) (reached at s_2) is

$$\begin{aligned} d &\equiv |s_2 - s_1| \\ &= | \int_{y_1}^{y_2} \frac{F(y')dy'}{\sqrt{F(y') - C^2}} | \end{aligned}$$

with the constant $C = C(x_1, y_1, x_2, y_2)$ determined by the equation

$$x_2 - x_1 = \int_{y_1}^{y_2} \frac{C}{\sqrt{F(y') - C^2}} dy'$$

$\square$

Cut-locus

Every geodesic can be extended in both directions indefinitely and every pairs of points can be connected by a distance-minimizing geodesic.[4] For each unit vector[5] $e \in T_oM$, there is a unique geodesic $C_e : [0, \infty) \to M$ such that $\dot{C}_e(0) = e$. The *exponential map* $\exp : T_oM \to M$ is

$$\exp(te) \equiv C_e(t) \tag{4.51}$$

For small t, the geodesic $C_e([0, t])$ is the *unique* distance-minimizing geodesic between its endpoints. Let $t(e)$ be the largest t such that the geodesic $C_e([0, t(e)])$ is the unique distance-minimizing from $C_e(0)$ to $C_e(t(e))$. Let us define

$$\tilde{C}_0 = \{t(e)e \ : \ e \in T_oM \ , \ |e| = 1\}$$

[4] Assuming that the Riemannian manifold is complete. This is the Hopf-Rinow theorem. Note that this result is not applicable for Lorentzian manifolds and this is why we must live with black holes!

[5] $||e||^2 = g_{ij}(o)e^i e^j = 1$.

Then the *cut-locus* is the set

$$C_o = \exp \tilde{C}_0$$

The set within the cut-locus is the star-shaped domain

$$\tilde{E}_o = \{t(e)e \; : \; e \in T_o M \; , \;\; 0 \leq t < t(e) \; , \; |e| = 1\}$$

On M the set within the cut-locus is $E_o = \exp \tilde{E}_o$.
The cut-locus can be characterized with the following proposition

PROPOSITION 4.1
A point $x \in C_o$ if and only

- *there exists a non-minimizing geodesic from o to x.*
- *or/and there exists two minimizing geodesics from o to x.*

The *injectivity radius* at a point o is defined as the largest strictly positive number r such that every geodesic γ starting from o and of length $l(\gamma) \leq r$ is minimizing

$$\mathrm{inj}(o) = \min_{e \in T_o M \; , \; |e|=1} t(e)$$

We have the following basic results [8]

THEOREM 4.2

1. *The map* $\exp : \tilde{E}_0 \to E_0$ *is a diffeomorphism.*
2. *The cut-locus C_o is a closed subset of measure zero.*
3. *If $x \in C_y$, then $y \in C_x$.*
4. *E_0 and C_0 are disjoint sets and $M = E_0 \cup C_0$.*

Example 4.9 Cut-locus of a 2-sphere
On S^2, the cut-locus C_o of a point o is its antipodal point: $C_o = \{-o\}$. For example, the cut-locus of the North pole N is the South pole S. Indeed, there exists an infinity of minimizing geodesics between N and S. □

4.2.11 Curvature of a connection

The curvature of a connection d_A on a rank m-vector bundle E is defined as the map

$$d_A^2 \equiv d_A \circ d_A : \Gamma(E) \to \Omega^2(M) \otimes \Gamma(E)$$

with $\Omega^2(M)$ the space of 2-forms on M. It is easy to verify that d_A^2 is a well defined tensor as

$$\begin{aligned} d_A^2(f\sigma + \tau) &= d_A(df \otimes \sigma + f d_A \sigma + d_A \tau) \\ &= d^2 f\sigma - df d_A \sigma + df d_A \sigma + f d_A^2 \sigma + d_A^2 \tau \\ &= f d_A^2 \sigma + d_A^2 \tau \end{aligned}$$

where we have used that $d^2 = 0$. Locally on $\mathcal{U}_\alpha$, d_A^2 is represented by a rank m-matrix 2-form

$$\Omega_\alpha = \Omega_{\alpha ij} dx^i \wedge dx^j$$

and

$$d_A^2 \sigma_\alpha = \Omega_\alpha \sigma_\alpha$$

As locally $d_A = d + A_\alpha$, we have

$$\begin{aligned} d_A^2 &= (d + A_\alpha)(d + A_\alpha) \\ &= d^2 + (dA_\alpha) + A_\alpha d - A_\alpha d + A_\alpha \wedge A_\alpha \end{aligned}$$

and finally

$$\Omega_\alpha = dA_\alpha + A_\alpha \wedge A_\alpha \tag{4.52}$$

In local coordinates, the matrix-valued 2-form Ω_α is

$$[\Omega_\alpha] = \Omega^i_{jkl} dx^k \wedge dx^l$$

and the connection is $A_\alpha = A^i_{kj} dx^k$. From (4.52), we obtain

$$\Omega^i_{jkl} = -\partial_l A^i_{kj} + \partial_k A^i_{lj} + A^i_{kr} A^r_{lj} - A^i_{lr} A^r_{kj} \tag{4.53}$$

As d_A^2 is a tensor, we have

$$\begin{aligned} d_A^2 \sigma_\beta &= d_A^2(g_{\beta\alpha}\sigma_\alpha) \\ &= g_{\beta\alpha} d_A^2 \sigma_\alpha \\ &= g_{\beta\alpha} \Omega_\alpha \sigma_\alpha = \Omega_\beta \sigma_\beta \end{aligned}$$

Therefore on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$, the curvatures Ω_α and Ω_β glue together as

$$\Omega_\beta = g_{\beta\alpha} \Omega_\alpha g_{\beta\alpha}^{-1}$$

Example 4.10 Riemann tensor, Ricci tensor and scalar curvature
Let us take the Levi-Cevita connection on TM with Γ^i_{kj} the Christoffel symbols. The components (4.53) of the curvature Ω^i_{jkl} are preferably noted R^i_{jkl} and called the *Riemann tensor*

$$R^i_{jkl} = -\partial_l \Gamma^i_{kj} + \partial_k \Gamma^i_{lj} + \Gamma^i_{kr}\Gamma^r_{lj} - \Gamma^i_{lr}\Gamma^r_{kj}$$

From this tensor, two other important tensors can be defined by contraction. The first one is the *Ricci tensor* obtained by contracting the upper indices and the lower second indices in the Riemann tensor

$$R_{jl} \equiv R^i_{jil}$$

Then the *scalar curvature* is obtained by taking the trace of the Ricci tensor

$$R \equiv g^{ij} R_{ij}$$

□

4.2.12 Integration on a Riemannian manifold

On a n-dimensional Riemannian M, we can build a *partition of unity*:

DEFINITION 4.4 Partition of unity *Given an atlas $(\mathcal{U}_\alpha, \phi_\alpha)$ on a n-dimensional manifold M, a set of C^∞ functions ϵ_α is called a partition of unity if*

1. $0 \leq \epsilon_\alpha \leq 1$.
2. *the support of ϵ_α, i.e., the closure of the set $\{p \in M : \epsilon_\alpha(p) \neq 0\}$, is contained in the corresponding $\mathcal{U}_\alpha$.*
3. $\sum_\alpha \epsilon_\alpha = 1$ *for all* $p \in M$.

By using this partition of unity, we define the integral of a measurable function f on M as

$$\int_M f d\mu(g) \equiv \sum_\alpha \int_{\phi_\alpha(\mathcal{U}_\alpha)} (\epsilon_\alpha \sqrt{g} f)(\phi_\alpha^{-1}(x)) d^n x$$

g is the determinant of the metric $[g_{ij}]$ and $d^n x \equiv \prod_{i=1}^n dx^i$ is the Lebesgue measure on $\mathbb{R}^n$. In order to have a consistent definition, the integral above should be independent of the partition of unity and of the atlas. The independence with respect to the atlas comes from the fact that the measure $\sqrt{g}(x) d^n x$ is invariant under an arbitrary change of coordinates:

PROOF Indeed the metric changes as $g_{ij} = g_{i'j'} \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^{j'}}{\partial x^j}$ and therefore $g = \det(g_{ij})$ changes as $\sqrt{g} = \det\left(\frac{\partial x^{i'}}{\partial x^i}\right)\sqrt{g'}$. Moreover, the element $\prod_{i=1}^n dx^i$ changes as $\prod_{i=1}^n dx^i = [\det\left(\frac{\partial x^{i'}}{\partial x^i}\right)]^{-1} \prod_{i'=1}^n dx^{i'}$ and we deduce the result. □

DEFINITION 4.5 Density bundle *We call $|\Lambda|_x$ the density bundle on M whose sections can be locally written as $f(x)\sqrt{g}$ with $f(x)$ a smooth function on M.*

Now, everything is in place to define the heat kernel on a Riemannian manifold and formulate the Minakshisundaram-Pleijel-De Witt-Gilkey theorem giving an expansion of the heat kernel in the short-time limit and representing an asymptotic solution for the conditional probability density to the backward Kolmogorov equation.

4.3 Heat kernel on a Riemannian manifold

In this section, the PDE (4.9) will be interpreted as a heat kernel on a general smooth n-dimensional Riemannian manifold M endowed with a metric g_{ij} (here we have that $i, j = 1 \cdots n$) and an Abelian connection $\mathcal{A}$.

REMARK 4.3 Note that the coordinates $\{\alpha_i\}_i$ (resp. $\{x_i\}_i$) will be noted $\{x_i\}_i$ (resp. $\{y_i\}_i$) below in order to be consistent with our previous (geometric) notation. □

The inverse of the metric g^{ij} is defined by

$$g^{ij}(x) = \frac{1}{2}\rho_{ij}\sigma_i(x)\sigma_j(x)$$

Note that in this relation although two indices are repeated, there is no implicit summation over i and j as the result is a symmetric tensor dependent precisely on these two indices. The metric (ρ^{ij} inverse of ρ_{ij}, i.e., $\rho^{ij}\rho_{jk} = \delta^i_k$) is

$$g_{ij}(x) = 2\frac{\rho^{ij}}{\sigma_i(x)\sigma_j(x)} \tag{4.54}$$

The differential operator

$$D = b^i(x)\partial_i + g^{ij}(x)\partial_{ij}$$

which appears in (4.9) is a second-order *elliptic operator* of Laplace type.

REMARK 4.4 Elliptic operator We recall to the reader the definition of an elliptic operator. We need to introduce the *symbol* of D:
The symbol of D is given by

$$\sigma(x,k) = b^i(x)k_i + g^{ij}(x)k_ik_j$$

It corresponds to the Fourier transform of D when x is fixed. Its *leading symbol* is defined as

$$\sigma_2(x,k) = g^{ij}(x)k_ik_j$$

DEFINITION 4.6 Second-order elliptic operator *A second-order operator is elliptic if for any open set $\Omega \subset M$, its corresponding leading symbol $\sigma_2(x,k)$ is always non-zero for non-zero k (i.e., non-degenerate quadratic form).*

From the definition, we see that D is elliptic if and only if g_{ij} is a metric. $\square$

We can then show that there is a unique connection ∇ on $\mathcal{L}$, a line bundle over M, and a unique smooth section $Q(x)$ of $End(\mathcal{L}) \simeq \mathcal{L} \otimes \mathcal{L}^*$ such that

$$\begin{aligned} D &= g^{ij}\nabla_i\nabla_j + Q \\ &= g^{-\frac{1}{2}}\left(\partial_i + \mathcal{A}_i\right) g^{\frac{1}{2}} g^{ij}\left(\partial_j + \mathcal{A}_j\right) + Q \end{aligned} \tag{4.55}$$

Here $g = \det[g_{ij}]$ and $\mathcal{A}_i$ are the components of an Abelian connection. Then, the backward Kolmogorov equation (4.9) can be written in the covariant way

$$\frac{\partial p(\tau, x|y)}{\partial \tau} = Dp(\tau, x|y) \tag{4.56}$$

If we take $\mathcal{A}_i = 0$, $Q = 0$ then D becomes the Laplace-Beltrami operator (or Laplacian)

$$\Delta = g^{-\frac{1}{2}}\partial_i\left(g^{\frac{1}{2}}g^{ij}\partial_j\right) \tag{4.57}$$

For this configuration, (4.56) will be called the *Laplacian heat kernel* equation. As expected, Δ is invariant under a change of coordinates:

PROOF From $\{x^i\}_{i=1,\cdots,n}$ to $\{x^{i'}\}_{i'=1,\cdots,n}$, the metric g_{ij} transforms as

$$g_{i'j'}(x') = \partial_{i'}x^i\partial_{j'}x^j g_{ij}(x)$$

Then by plugging the expression for $g_{i'j'}(x')$ into (4.57), we obtain

$$\Delta(g') = \Delta(g)$$

□

We may express the connection $\mathcal{A}^i$ and Q as a function of the drift b_i and the metric g_{ij} by identifying in (4.55) the terms ∂_i and ∂_{ij} with those in (4.9). We find

$$\mathcal{A}^i = \frac{1}{2}\left(b^i - g^{-\frac{1}{2}}\partial_j\left(g^{\frac{1}{2}}g^{ij}\right)\right) \tag{4.58}$$

$$Q = g^{ij}\left(\mathcal{A}_i\mathcal{A}_j - b_j\mathcal{A}_i - \partial_j\mathcal{A}_i\right) \tag{4.59}$$

Note that the Latin indices i,$j \cdots$ can be lowered or raised using the metric g_{ij} or its inverse g^{ij} as explained in 4.2.6. For example $\mathcal{A}_i = g_{ij}\mathcal{A}^j$ and $b_i = g_{ij}\lfloor^j$. The components $\mathcal{A}_i$ define locally a one-form $\mathcal{A} = \mathcal{A}_i dx^i$. We deduce that under a change of coordinates $x^{i'}(x^j)$, $\mathcal{A}_i$ undergoes the vector transformation $\mathcal{A}_{i'}\partial_i x^{i'} = \mathcal{A}_i$. Note that the components b_i don't transform as a vector (see 4.10). This results from the fact that the SDE (4.1) has been derived using the Itô calculus and not the Stratonovich one (see section 4.4). The equation (4.58) can be re-written using the Christoffel symbol (4.35)

$$\mathcal{A}^i = \frac{1}{2}\left(b^i - g^{pq}\Gamma^i_{pq}\right)$$

To summarize, a heat kernel equation on a Riemannian manifold M is constructed from the following three pieces of geometric data:

1. A metric g on M, which determines the second-order piece.
2. A connection $\mathcal{A}$ on a line bundle $\mathcal{L}$, which determines the first-order piece.
3. A section Q of the bundle $End(\mathcal{L}) \simeq \mathcal{L} \otimes \mathcal{L}^*$, which determines the zeroth-order piece.

The fundamental solution of the heat kernel equation (4.56), called a *heat kernel*, is defined as follows

DEFINITION 4.7 *A heat kernel for a heat kernel equation (4.56) is a continuous section $p(\tau, x|y)$ of the bundle*[6] $(\mathcal{L} \boxtimes \mathcal{L}^* \otimes |\Lambda|_y)$ *over* $\mathbb{R}_+ \times M \times M$, *satisfying the following properties:*

1. *$p(\tau, x|y)$ is C^1 with respect to τ, that is, $\partial_\tau p(\tau, x|y)$ is continuous in (τ, x, y)*

[6] Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two vector bundles on M and pr_1 (resp. pr_2) be the projections from $M \times M$ onto the first (resp. second) factor. We denote the external product $\mathcal{E}_1 \boxtimes \mathcal{E}_2$ the (pullback) vector bundle $pr_1^*\mathcal{E}_1 \otimes pr_2^*\mathcal{E}_2$ over $M \times M$.

2. *$p(\tau, x|y)$ is C^2 with respect to x, that is the partial derivatives $\frac{\partial^2 p(\tau,x|y)}{\partial x_i \partial x_j}$ are continuous in (τ, x, y) for any coordinate system x.*

3. *$p(\tau, x|y)$ satisfies the boundary condition at $\tau = 0$:*

$$\lim_{\tau \to 0} p(\tau, x|y) = \delta(x-y)$$

This boundary condition means that for any smooth section s of $\mathcal{L}$, then

$$\lim_{\tau \to 0} \int_M p(\tau, x|y) s(y) d^n y = s(x)$$

In order to be considered as a probability density for a backward Kolmogorov equation, the heat kernel $p(\tau, x|y)$ should be a positive section and normalized by

$$\int_M p(\tau, x|y) d^n y = 1 \tag{4.60}$$

Note that it may happen that

$$\int_M p(\tau, x|y) d^n y < 1$$

Probabilistically it means that the Itô diffusion process x_t following (4.1) associated to the backward Kolmogorov equation may not run for all time and may go off the manifold in a finite amount of time. In the next chapters about local and stochastic volatility models, we discuss the conditions on the metric and the Abelian connection under which there is no explosion of the process x_t in a finite amount of time. If the heat kernel satisfies (4.60), the pair (M, D) is called *stochastically complete.*
Let us see now the simplest example of a heat kernel equation, that is the heat kernel on $\mathbb{R}^n$ for which there is an analytical solution.

Example 4.11 Heat kernel equation on $\mathbb{R}^n$
On $\mathbb{R}^n$ with the flat metric $g_{ij} = \delta_{ij}$, with a zero-Abelian connection and a zero section $Q = 0$, the heat kernel equation (4.56) reduces to

$$\frac{\partial p(\tau, x|y)}{\partial \tau} = \partial_i^2 p(\tau, x|y)$$

It is easy to see that the heat kernel is

$$p(\tau, x|y) = \frac{1}{(4\pi\tau)^{\frac{n}{2}}} e^{-\frac{|x-y|^2}{4\tau}} \tag{4.61}$$

where $|x-y|^2 = \sum_{i=1}^n (x_i - y_i)^2$ is the Euclidean distance which coincides with the geodesic distance. We can verify the identity (4.60) and $(\mathbb{R}^n, \Delta)$ is stochastically complete. $\square$

4.4 Abelian connection and Stratonovich's calculus

In our presentation, one of the motivations for introducing an Abelian connection was that $\mathcal{A}$ is a one-form and therefore transforms covariantly under a change of variables. Another possibility to achieve this goal is to use the Stratonovich calculus.
Let M, N be two continuous semi-martingales. The Stratonovich product $\diamond$ is defined as follows

DEFINITION 4.8 Stratonovich product

$$\int_0^T M_s \diamond dN_s = \int_0^T M_s dN_s + \frac{1}{2} < M, N >_t$$

with $< M, N >_t$ *the quadratic variation (see definition 2.13). This writes formally*

$$M_t \diamond dN_t = M_t dN_t + \frac{1}{2} d < M, N >_t \tag{4.62}$$

By using the definition (4.62), an Itô diffusion SDE

$$\begin{aligned} dx^i &= b^i(x)dt + \sigma^i(x)dW_i \\ dW_i dW_j &= \rho_{ij}(t)dt \end{aligned}$$

can be transformed into the Stratonovich diffusion SDE

$$dx^i = \mathcal{A}_s^i(x)dt + \sigma^i(x) \diamond dW_i \tag{4.63}$$

where the drift $\mathcal{A}_s^i(x)$ in (4.63) is

$$\mathcal{A}_s^i(x) = b^i(x) - \frac{1}{2}\rho_{ik}(t)\sigma^k(x)\partial_k \sigma^i(x)$$

From the transformation of the drift (4.10) and the volatility (4.11) under a change of variables $x^{i'} = f^{i'}(x)$, we obtain that the drift $\mathcal{A}_i^s$ changes covariantly as a vector field, i.e.,

$$\mathcal{A}_s^{i'} = \partial_i f^{i'} \mathcal{A}_s^i$$

In the Stratonovich calculus, the complicated drift in Itô's lemma cancels and we have

$$d\mathcal{C}(t,x) = \partial_t \mathcal{C}(t,x)dt + \partial_{x^i}\mathcal{C}(t,x)dx^i$$

where dx^i is given by (4.63).

PROOF From Itô's lemma, we have

$$d\mathcal{C}(t,x) = \partial_t \mathcal{C}(t,x)dt + \frac{1}{2}\rho_{ij}\sigma^i\sigma^j\partial_{ij}\mathcal{C}(t,x) + \sigma^i\partial_i\mathcal{C}(t,x)dW_i$$

As

$$\sigma^i\partial_i\mathcal{C}(t,x)dW_i = \sigma^i\partial_i\mathcal{C}(t,x)\diamond dW_i - \frac{1}{2}\rho_{ki}\sigma^k\partial_k\left(\sigma^i\partial_i\mathcal{C}(t,x)\right)$$

we get our final result. ∎

Example 4.12 GBM process
As an example, the GBM process $\frac{df_t}{f_t} = \sigma dW_t$ becomes the Stratonovich diffusion

$$\frac{df_t}{f_t} = -\frac{1}{2}\sigma^2 dt + \sigma \diamond dW_t$$

∎

Note that there is a difference between $\mathcal{A}_s^i(x)$ and the Abelian connection $\mathcal{A}^i(x)$ which is given by

$$\mathcal{A}^i = \frac{1}{2}\left(b^i + \frac{1}{2}\rho_{ij}\sigma^i\sigma^j\frac{\partial_j\sigma^p}{\sigma^p} - \frac{1}{2}\sigma^i\partial_i\sigma^i + \frac{1}{2}\left(\sigma^i\right)^2\frac{\partial_i\sigma^p}{\sigma^p}\right)$$

The main motivation for using $\mathcal{A}^i$ instead of $\mathcal{A}_s^i$ is that we can apply gauge transformations on the heat kernel equation.

4.5 Gauge transformation

In order to reduce the complexity of the heat kernel equation (4.56), two transformations can be used. The first one is the group of diffeomorphism $\mathrm{Diff}(M)$ which acts on the metric and the Abelian connection by

$$(f^*g)_{ij} = g_{pk}\partial_i f^p(x)\partial_j f^k(x)$$
$$(f^*\mathcal{A})_i = \mathcal{A}_p\partial_i f^p(x)\ ,\ f \in \mathrm{Diff}(M)$$

The second one is a *gauge transformation*. Before introducing a geometrical definition, let us give a simple definition without any formalism.
If we define a new section $p' \in (\mathcal{L} \boxtimes \mathcal{L}^* \otimes |\Lambda|_y)$ by

$$p'(\tau,x|y) = e^{\chi(\tau,x)-\chi(0,y)}p(\tau,x|y) \tag{4.64}$$

with $\chi(\tau, x)$ a smooth function on $\mathbb{R}_+ \times M$, then it is easy to check that $p'(\tau, x|y)$ satisfies the same heat kernel equation as $p(\tau, x|y)$ (4.56) but with a new Abelian connection $\mathcal{A}' = \mathcal{A}'_i dx^i$ and a new section Q'

$$\begin{aligned} \mathcal{A}'_i &\equiv \mathcal{A}_i - \partial_i \chi \\ Q' &\equiv Q + \partial_\tau \chi \end{aligned} \tag{4.65}$$

The transformation (4.64) is called a *gauge transformation.* The reader should be warned that the transformation (4.65) only applies to the connection $\mathcal{A}_i$ with lower indices. Moreover, the constant phase $e^{\chi(0,y)}$ has been added in (4.64) so that $p(\tau, x|y)$ and $p'(\tau, x|y)$ satisfy the same boundary condition at $\tau = 0$.

This transformation is particularly useful in the sense that if the one-form $\mathcal{A}$ is exact (meaning that there exists a smooth function χ such that $\mathcal{A} = d\chi$), then the new connection $\mathcal{A}'$ (4.65) vanishes. The complexity of the heat kernel equation is consequently reduced as the Abelian connection has disappeared and the operator D has become a Laplace-Beltrami operator (4.57).

In order to understand mathematically the origin of the gauge transformation, let us consider the group of automorphisms $\mathrm{Aut}(\mathcal{L})$ of the line bundle $\mathcal{L}$. We have seen previously that the line bundle $\mathcal{L}$ can be defined as the set $\tilde{\mathcal{L}}$ of all triples $(\alpha, p, v) \in A \times M \times \mathbb{R}_+$ modulo the equivalence relation $\sim$ (4.21). An automorphism of $\mathcal{L}$ can be seen as a map on $\tilde{\mathcal{L}}$ which commutes with the projection π of $\mathcal{L}$. Therefore an element $g \in \mathrm{Aut}(\mathcal{L})$ maps the triple (α, p, v) to $(\alpha, p, g.v)$ with $g.v \in \mathbb{R}_+$, i.e., it leaves the fiber unchanged. This element g can be seen locally as a map from $\mathcal{U}_\alpha$ to $\mathbb{R}_+$. Moreover it induces an action on the Abelian connection by (4.25) which reduces to (4.65) for

$$g = e^{-\chi(\tau,x)}$$

Interpretation of a gauge transformation as a change of measure

In this section, we show with a simple example that a gauge transformation can be interpreted as a change of measure according to the Girsanov theorem. For the sake of simplicity, we assume that we have a one-dimensional Itô process X_t given by

$$dX_t = dW_t$$

with W_t a one-dimensional Brownian motion. We have the initial condition $X_t = x$. The option fair value $\mathcal{C}(t, x) = \mathbb{E}^{\mathbb{P}}[f(X_T)|\mathcal{F}_t]$ satisfies the PDE

$$\partial_t \mathcal{C}(t, x) + \frac{1}{2} \partial_x^2 \mathcal{C}(t, x) = 0$$

with the terminal condition $\mathcal{C}(T,x) = f(X_T)$. We apply a time-independent gauge transformation on the function (section) $\mathcal{C}(t,x)$:

$$\mathcal{C}'(t,x) = e^{\Lambda(x)}\mathcal{C}(t,x)$$

$\mathcal{C}'(t,x)$ satisfies the PDE

$$\partial_t \mathcal{C}'(t,x) + \frac{1}{2}\partial_x^2 \mathcal{C}'(t,x) - \partial_x \Lambda \partial_x \mathcal{C}'(t,x) + \frac{1}{2}\left((\partial_x \Lambda)^2 - \partial_x^2 \Lambda\right)\mathcal{C}'(t,x) = 0$$

with the boundary condition $\mathcal{C}'(T,x) = e^{\Lambda(x_T)} f(X_T)$. By using the Feynman-Kac theorem, we obtain the identity

$$\mathbb{E}^{\mathbb{P}}[f(X_T)|\mathcal{F}_t] = e^{-\Lambda(x)}\mathbb{E}^{\mathbb{P}}[e^{\Lambda(X'_T)} f(X'_T) e^{\frac{1}{2}\int_t^T \left((\partial_x \Lambda(X'_s))^2 - \partial_x^2 \Lambda(X'_s)\right) ds}|\mathcal{F}_t]$$

with $dX'_t = dW_t - \partial_x \Lambda(X'_t)dt$. By using that

$$\begin{aligned}
\Lambda(X'_T) &= \Lambda(x) + \int_t^T \partial_x \Lambda(X'_s)\left(dW_s - \partial_x \Lambda(X'_s)ds\right) + \frac{1}{2}\int_t^T \partial_x^2 \Lambda(X'_s)ds \\
&= \Lambda(x) + \int_t^T \partial_x \Lambda(X'_s)dW_s + \frac{1}{2}\int_t^T \partial_x^2 \Lambda(X'_s)ds - \int_t^T (\partial_x \Lambda(X'_s))^2 ds
\end{aligned}$$

the identity above becomes

$$\mathbb{E}^{\mathbb{P}}[f(X_T)|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[f(X'_T) e^{\int_t^T \partial_x \Lambda(X'_s)dW_s - \frac{1}{2}\int_t^T (\partial_x \Lambda(X'_s))^2 ds}|\mathcal{F}_t] \tag{4.66}$$

This relation can be obtained according to a change of measure from $\mathbb{P}$ to the new measure $\mathbb{P}'$ defined by the Radon-Nikodym derivative

$$\frac{d\mathbb{P}'}{d\mathbb{P}}\,|_{\mathcal{F}_T} = e^{\int_0^T \partial_x \Lambda(X'_t)dW_t - \frac{1}{2}\int_0^T \left(\partial_x \Lambda(X'_t)\right)^2 dt}$$

and (4.66) can be written as

$$\mathbb{E}^{\mathbb{P}}[f(X_T)|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}'}[f(X'_T)|\mathcal{F}_t]$$

4.6 Heat kernel expansion

The asymptotic resolution of a heat kernel (4.56) in the short time is an important problem in theoretical physics and in mathematics. In physics, it corresponds to the solution of an Euclidean Schrödinger equation on a fixed space-time background [13] and in mathematics, the heat kernel, corresponding to the determination of the spectrum of the Laplacian, can give topological information (e.g., the Atiyah-Singer index theorem) [19]. The following theorem proved by DeWitt-Gilkey-Pleijel-Minakshisundaram (in short the DeWitt theorem) gives the complete asymptotic solution for a heat kernel on a Riemannian manifold. Beforehand, let us introduce the cut-off function ϕ:

DEFINITION 4.9 Cut-off *If $\phi : \mathbb{R}_+ \to [0,1]$ is a smooth function such that*

$$\phi(s) = 1 \text{ if } s < \frac{\epsilon^2}{4}$$
$$\phi(s) = 0 \text{ if } s > \frac{\epsilon^2}{4}$$

with $\epsilon \in \mathbb{R}_+^*$, we will call ϕ a cut-off function.

THEOREM 4.3 Heat kernel expansion [59]
Let M be a Riemannian n-dimensional manifold and $p(\tau, x|y)$ be the heat kernel of the heat kernel equation (4.56). Then there exists smooth sections $a_n(x,y) \in \Gamma(M \times M, \mathcal{L} \boxtimes \mathcal{L}^*)$ such that for every $N > \frac{n}{2}$, $p^N(\tau, x|y)$ defined by the formula

$$p^N(\tau, x|y) = \phi(d(x,y)^2) \frac{\sqrt{g(y)}}{(4\pi\tau)^{\frac{n}{2}}} \sqrt{\Delta(x,y)} \mathcal{P}(y,x) e^{-\frac{\sigma(x,y)}{2\tau}} \sum_{n=0}^{N} a_n(x,y)\tau^n \tag{4.67}$$

is asymptotic to $p(\tau, x|y)$:

$$\| \partial_\tau^k \left(p(\tau, x|y) - p^N(\tau, x|y) \right) \|_l = O\left(t^{N - \frac{n}{2} - \frac{l}{2} - k} \right)$$

with $\| \phi \|_l = \sup_{k \leq l} \sup_{x \in \mathbb{R}^n} \mid \frac{d^k}{dx^k} \phi(x) \mid$.

- ϕ is a cut-off function where ϵ is chosen smaller than the injectivity radius of the manifold M.
- $\sigma(x,y)$ is the Synge world function equal to one half of the square of geodesic distance $d(x,y)$ between x and y for the metric g.
- $\Delta(x,y)$ is the so-called Van Vleck-Morette determinant

$$\Delta(x,y) = g(x)^{-\frac{1}{2}} \det \left(-\frac{\partial^2 \sigma(x,y)}{\partial x \partial y} \right) g(y)^{-\frac{1}{2}} \tag{4.68}$$

  with $g(x) = \det[g_{ij}(x,x)]$.
- $\mathcal{P}(y,x) \in \Gamma(M \times M, \mathcal{L} \boxtimes \mathcal{L}^*)$ is the parallel transport of the Abelian connection along the geodesic curve $\mathcal{C}$ from the point y to x

$$\mathcal{P}(y,x) = e^{-\int_{\mathcal{C}(y,x)} \mathcal{A}_i dx^i} \tag{4.69}$$

- The functions $a_i(x,y)$, called the *heat kernel coefficients*, are smooth sections $\Gamma(M \times M, \mathcal{L} \boxtimes \mathcal{L}^*)$. The first coefficient is simple

$$a_0(x,y) = 1 \ , \ \forall\ (x,y) \in M \times M$$

The other coefficients are more complex. However, when evaluated on the diagonal $x = y$, they depend on geometric invariants such as the scalar curvature R. As for the non-diagonal coefficients, they can be computed as a Taylor series when x is in a neighborhood of y.

The first diagonal coefficients are fairly easy to compute by hand. Recently $a_n(x,x)$ has been computed up to the order $n = 8$. The formulas become exponentially more complicated as n increases. For example, the a_6 formula has 46 terms. The first diagonal coefficients are given below [51]

$$a_1(x,x) = P(x) \equiv \frac{1}{6}R + Q(x) \tag{4.70}$$

$$\begin{aligned} a_2(x,x) &= \frac{1}{180}\left(R_{ijkl}R^{ijkl} - R_{ij}R^{ij}\right) + \frac{1}{2}P^2 \\ &+ \frac{1}{12}\mathcal{F}_{ij}\mathcal{F}^{ij} + \frac{1}{6}\Delta Q + \frac{1}{30}\Delta R \end{aligned} \tag{4.71}$$

with R_{ijkl} the Riemann tensor, R_{ij} the Ricci tensor, R the scalar curvature and $\mathcal{F}_{ij} = \partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i$ the curvature associated to the Abelian connection $\mathcal{A}$.

REMARK 4.5 When p is truncated at the N^{th} order, we call the solution an asymptotic solution of order N. □

Let us, before giving a sketch of the proof of the theorem (4.3), explain how to use the heat kernel expansion (4.67) with a simple example, namely the GBM process.

Example 4.13 GBM process
The SDE is $df_t = \sigma f_t dW$ with the initial condition $f_0 = f$. By using the definition(4.54), we obtain the following one-dimensional metric

$$g_{ff} = \frac{2}{\sigma^2 f^2}$$

When written with the coordinate $s = \sqrt{2}\frac{\ln(f)}{\sigma}$, the metric is flat $g_{ss} = 1$ and all the heat kernel coefficients depending on the Riemann tensor vanish. The geodesic distance between two points s_0 and s is given by the classical Euclidean distance

$$d(s_0, s) = |s - s_0|$$

and the Synge function is

$$\sigma(s_0, s) = \frac{1}{2}(s - s_0)^2$$

In the old coordinate $[f]$, $\sigma(f_0, f)$ is equal to

$$\sigma(f_0, f) = \frac{1}{\sigma^2} \ln\left(\frac{f}{f_0}\right)^2$$

Furthermore, the connection $\mathcal{A}$ (4.58) and the section Q (4.59) are given by

$$\mathcal{A} = -\frac{1}{2f} df$$
$$Q = -\frac{\sigma^2}{8}$$

Therefore the parallel transport $\mathcal{P}(f_0, f)$ is given by

$$\mathcal{P}(f_0, f) = e^{\frac{1}{2}\ln\left(\frac{f}{f_0}\right)}$$

By plugging these expressions into (4.67), we obtain the following first-order asymptotic solution for the log-normal process

$$p_1(\tau, f|f_0) = \frac{1}{f(2\pi\sigma_0^2\tau)^{\frac{1}{2}}} e^{-\frac{\ln(\frac{f}{f_0})^2}{2\sigma_0^2\tau} - \frac{1}{2}\ln\left(\frac{f}{f_0}\right)} \left(1 - \frac{\sigma^2\tau}{8}\right) \tag{4.72}$$

REMARK 4.6 Note that in the last equation, we have replaced the notation f (resp. f_0) by f (resp. f_0) according to our convention (remark 4.3). □

At the second-order, the second heat kernel coefficient is given by

$$a_2 = \frac{1}{2}Q^2 = \frac{\sigma^4\tau^2}{128}$$

and the asymptotic solution is

$$p_2(\tau, f|f_0) = \frac{1}{f(2\pi\sigma^2\tau)^{\frac{1}{2}}} e^{-\frac{\ln\left(\frac{f}{f_0}\right)^2}{2\sigma^2\tau} - \frac{1}{2}\ln\left(\frac{f}{f_0}\right)} \left(1 - \frac{\sigma^2\tau}{8} + \frac{\sigma^4\tau^2}{128}\right) \tag{4.73}$$

We end this example with the use of gauge transformations to solve the heat kernel:
We note that $\mathcal{A}$ and Q are exact, meaning that

$$\mathcal{A} = d\Lambda$$
$$Q = -\partial_\tau \Lambda$$

with $\Lambda = -\frac{1}{2}\ln(f) + \frac{\sigma^2\tau}{8}$. Modulo a gauge transformation

$$p' = e^{-\frac{1}{2}\ln\left(\frac{f}{f_0}\right)+\frac{\sigma^2\tau}{8}} p \frac{\sigma f_0}{\sqrt{2}}$$

p' satisfies the (Laplacian) heat kernel on $\mathbb{R}$

$$\partial_s^2 p' = \partial_\tau p'$$

The constant factor $\frac{\sigma f_0}{\sqrt{2}}$ has been added in order to have the initial condition $\lim_{\tau\to 0} p'(\tau, s) = \delta(s - s_0)$. The solution is the normal distribution (4.61). So the exact solution for $p = p' e^{\frac{1}{2}\ln(\frac{f}{f_0}) - \frac{\sigma^2\tau}{8}}$ is given by (see remark 4.6)

$$p(\tau, f|f_0) = \frac{1}{f(2\pi\sigma^2\tau)^{\frac{1}{2}}} e^{-\frac{(\ln(\frac{f}{f_0})+\frac{\sigma^2\tau}{2})^2}{2\sigma^2\tau}} \tag{4.74}$$

To be thorough, we conclude this section with a brief sketch of the derivation of the heat kernel expansion.

PROOF We start with the Schwinger-DeWitt ansatz

$$p(t, x|y) = \frac{\sqrt{g(y)}}{(4\pi t)^{\frac{n}{2}}} \Delta(x, y)^{\frac{1}{2}} \mathcal{P}(y, x) e^{-\frac{\sigma(x,y)}{2t}} \Omega(t, x|y) \tag{4.75}$$

By plugging (4.75) into the heat kernel equation (4.56), we derive a PDE satisfied by the function $\Omega(t, x|y)$

$$\partial_t \Omega = \left(-\frac{1}{t}\sigma^i \nabla_i + \mathcal{P}^{-1}\Delta^{-\frac{1}{2}}(D + Q)\Delta^{\frac{1}{2}}\mathcal{P}\right)\Omega \tag{4.76}$$

with $\nabla_i = \partial_i + \mathcal{A}_i$ and $\sigma_i = \nabla_i \sigma$, $\sigma^i = g^{ij}\sigma_j$. The regular boundary condition is $\Omega(t, x|x) = 1$. We solve this equation by writing the function Ω as a formal series in t:

$$\Omega(t, x|y) = \sum_{n=0}^{\infty} a_n(x, y) t^n \tag{4.77}$$

By plugging this series (4.77) into (4.76) and identifying the coefficients in t^n, we obtain an infinite system of coupled ODEs:

$$a_0 = 1$$
$$(1 + \frac{1}{k}\sigma^i\nabla_i)a_k = \mathcal{P}^{-1}\Delta^{-\frac{1}{2}}(D + Q)\Delta^{\frac{1}{2}}\mathcal{P}\frac{a_{k-1}}{k} \quad k \neq 0$$

The calculation of the heat kernel coefficients in the general case of arbitrary background offers a complex technical problem. The Schwinger-DeWitt method is quite simple but is not effective at higher orders. By means of it, only the two lowest order terms were calculated. For other advanced methods see [51]. ∎

From the DeWitt theorem, we see that the most difficult part in obtaining the asymptotic expression for the conditional probability is to derive the geodesics distance.

Heat kernel on a time-dependent Riemannian manifold

In most financial models, the drift and volatility terms can explicitly depend on time. In this case, we obtain a time-dependent metric, Abelian connection and section Q. It is therefore useful to generalize the heat kernel expansion from the previous section to the case of a time-dependent metric defined by

$$g_{ij}(t,x) = 2\frac{\rho^{ij}(t)}{\sigma^i(t,x)\sigma^j(t,x)} \tag{4.78}$$

This is the purpose of this section.
We consider the time-dependent heat kernel equation

$$\frac{\partial}{\partial t}p(t,x|y) = \mathcal{D}p(t,x|y) \tag{4.79}$$

where the differential operator $\mathcal{D}$ is a time-dependent family of operators of Laplace type given by

$$\mathcal{D} = b^i(t,x)\partial_i + g^{ij}(t,x)\partial_{ij} \tag{4.80}$$

Let ; (resp. $,t$) denote the multiple covariant differentiation according to the Levi-Cevita connection (resp. to the time t). We expand $\mathcal{D}$ in a Taylor series expansion in t to write $\mathcal{D}$ invariantly in the form

$$\mathcal{D}u = Du + \sum_{r>0} t^r \left(\mathcal{G}_r^{ij}u_{;ij} + \mathcal{F}_r^i u_{;i} + Q_r\right)$$

with the operator D depending on the connection $\mathcal{A}_i$ and the smooth section Q given by (4.55) (with $g_{ij} \equiv g_{ij}(t=0)$ and $b^i \equiv b^i(t=0)$). The tensor $\mathcal{G}_1^{ij}$ is given by

$$\begin{aligned}\mathcal{G}_1^{ij}(x) &= g^{ij}_{,t}(0,x) \\ &= \frac{\rho_{ij,t}(0)}{2}\sigma^i(0,x)\sigma^j(0,x) + \rho_{ij}(0)\sigma^i_{,t}(0,x)\sigma^j(0,x)\end{aligned} \tag{4.81}$$

The asymptotic resolution of the heat kernel (4.79) in the short time limit in a time-dependent background is an important problem in quantum cosmology. When the spacetime varies slowly, the time-dependent metric describing the cosmological evolution can be expanded in a Taylor series with respect to t. In this situation, the index r is related to the adiabatic order [4]. The subsequent theorem obtained in [92, 93] gives the complete asymptotic solution for the time-dependent heat kernel on a Riemannian manifold.

THEOREM 4.4
Let M be a Riemannian n-dimensional manifold and $p(t, x|y)$ be the heat kernel of the time-dependent heat kernel equation (4.79). Then there exists smooth sections $a_n(x, y) \in \Gamma(M \times M, \mathcal{L} \boxtimes \mathcal{L}^*)$ such that for every $N > \frac{n}{2}$, $p^N(t, x|y)$ defined by the formula

$$p^N(t, x|y) = \phi(d(x, y)^2) \frac{\sqrt{g(y)}}{(4\pi t)^{\frac{n}{2}}} \sqrt{\Delta(x, y)} \mathcal{P}(y, x) e^{-\frac{\sigma(x,y)}{2t}} \sum_{n=0}^{N} a_n(x, y) t^n \tag{4.82}$$

is asymptotic to $p(x, y, t)$:

$$\| \partial_\tau^k \left(p(t, x|y) - p^N(t, x|y) \right) \|_l = O\left(t^{N - \frac{n}{2} - \frac{l}{2} - k} \right)$$

with $\| \phi \|_l = \sup_{k \le l} \sup_{x \in \mathbb{R}^n} |\frac{d^k}{dx^k} \phi(x)|$.
σ, Δ and $\mathcal{P}$ are computed with the metric $g_{ij}(t = 0)$ and the connection $\mathcal{A}(t = 0)$.

The diagonal heat kernel coefficients $a_i(x, x)$ depend on geometric invariants such as the scalar curvature R. The coefficients a_n have been computed up to the fourth-order $(a_0(x, y) = 1)$. The first coefficient is given by

$$a_1(x, x) = \frac{1}{6} R + Q - \frac{1}{4} \mathcal{G}_{1,ii}$$

where $\mathcal{G}_{1,ii} \equiv g_{ij}(t = 0) \mathcal{G}_1^{ij}$.

4.7 Hypo-elliptic operator and Hörmander's theorem

4.7.1 Hypo-elliptic operator

As seen in the previous sections, the heat kernel expansion can be applied if D is an elliptic operator. However, in some situations such as the pricing of path-dependent options, D is no longer elliptic.

Example 4.14 Asian options
For the pricing of an Asian option in the case of a Black-Scholes model, the market model is given by

$$dS_t = \sigma S_t dW_t \qquad (4.83)$$
$$dA_t = S_t dt$$

where S_t is the asset price and $A_t = \int_0^t S_s ds$. Here we have taken a zero interest rate for the sake of simplicity. The second-order differential operator D is

$$D = \frac{1}{2}\sigma^2 S^2 \partial_S^2 + S\partial_A$$

Its leading symbol $\sigma_2(S, A, k_1, k_2) = \frac{1}{2}\sigma^2 S^2 k_1^2$ is a degenerate quadratic form and D is not elliptic. This situation precisely appears when our market model involves path-dependent variables which are not driven by a Brownian motion.
$\square$

The heat kernel expansion can be extended to an *hypo-elliptic* operator [56].

DEFINITION 4.10 Hypo-ellipticity *A differential operator D is hypo-elliptic if and only if the condition Du is C^∞ in an open set $\Omega \subset M$ implies that u is C^∞ in Ω.*

In particular it can be proved that an elliptic operator is an hypo-elliptic operator.
The definition above is difficult to use except when we have an explicit solution of the equation $Du = f$. There exists a sufficient (not necessary) condition to prove that D is an hypo-elliptic operator as given by the Hörmander theorem.

4.7.2 Hörmander's theorem

We consider a market model (2.20) written using the Stratonovich calculus:

$$dx_t^i = \mathcal{A}_s^i(x_t)dt + \sum_{j=1}^m \sigma_j^i(x_t) \diamond dW_t^j \ , \ i = 1 \cdots n$$

We assume that the coefficients of the SDE above are infinitely differentiable with bounded derivatives of all orders and do not depend on time. We define the $m+1$ n-dimensional vector fields

$$V_0 = \mathcal{A}_s^i \partial_i$$
$$V_j = \sigma_j^i \partial_i \ , \ j = 1, \cdots m$$

The operator D can be put in the Hörmander form

$$D = \frac{1}{2}\left(V_0 + \sum_{j=1}^{m} V_j V_j\right)$$

To formulate the Hörmander theorem, we consider the Lie bracket (see exercise 4.3) between two vector fields $X = X^i\partial_i$ and $Y = Y^i\partial_i$:

$$[X, Y] = (X^i\partial_i Y^j - Y^i\partial_i X^j)\partial_j$$

THEOREM 4.5 Hörmander [33]
If the vector space spanned by the vector fields

$$\begin{aligned} &V_1, \cdots V_m \ , \ [V_i, V_j] \ , \ 0 \le i, j \le m \\ &[V_i, [V_j, V_k]] \ , \ 0 \le i, j, k \le m \cdots \end{aligned} \tag{4.84}$$

at a point x_0 is $\mathbb{R}^n$, then D is an hypo-elliptic operator. Furthermore, for any $t > 0$, the process x_t has an infinitely differentiable density.

Note that if D is an elliptic operator then the Hörmander condition (4.84) holds.

Example 4.15 Asian options
From (4.83), we have the two vector fields

$$\begin{aligned} V_0 &= -\frac{\sigma^2}{2} S\partial_S + S\partial_A \\ V_1 &= \sigma S\partial_S \end{aligned}$$

As $[V_0, V_1] = -\sigma S\partial_A$, the Hörmander condition (4.84) holds for $S \neq 0$. Therefore D is an hypo-elliptic operator and (S_t, A_t) has an infinitely differentiable density. □

Example 4.16 Heat kernel on Heisenberg group
We consider the following SDE which corresponds to a Brownian motion on $\mathbb{R}^2$ and its Lévy area (see Appendix B)

$$\begin{aligned} dx &= \diamond dW^X \\ dy &= \diamond dW^Y \\ dt &= 2\left(y \diamond dW^X - x \diamond dW^Y\right) \end{aligned}$$

The vector fields associated to this SDE are

$$\begin{aligned} X &= \partial_x + 2y\partial_t \\ Y &= \partial_y - 2x\partial_t \end{aligned}$$

D is hypo-elliptic in the Hörmander sense as $(X, Y, T \equiv \partial_t)$ generates the *Heisenberg Lie algebra*:[7]

$$
\begin{aligned}
[X, Y] &\equiv -4T \\
[X, T] &= 0 \; , \; [Y, T] = 0
\end{aligned}
$$

As a consequence, the Markov process admits a smooth density with respect to the Lebesgue measure on $\mathbb{R}^3$. The (hypo-elliptic) heat kernel reads

$$\partial_\tau p(\tau, Z, T|0) = Dp(\tau, Z, T|0)$$

with $D = \frac{1}{2}\left(X^2 + Y^2\right)$ and $Z \equiv x + \imath y$. This heat kernel is solvable [90] and gives the joint probability law for $W^X, W^Y, \int \left(W^Y dW^X - W^Y dW^Y\right)$:

$$p(\tau, Z, T|0) = \frac{1}{(2\pi\tau)^2} \int_{-\infty}^{\infty} dk \frac{2k}{\sinh 2k} e^{\frac{ikT}{\tau}} e^{-\frac{|Z|^2}{2\tau k} \frac{2}{\tanh 2k}} dk$$

□

4.8 Problems

Exercises 4.1 Conformal metrics on a Riemann surface
Let us consider the metric on a Riemann surface

$$ds^2 = F(y)(dx^2 + dy^2)$$

1. Compute the Christoffel symbols Γ^k_{ij}.
2. Prove that the geodesic equations are given by the ODEs (4.47) and (4.48).
3. Compute the Riemann curvature R^p_{ijk}.
4. Compute the Ricci curvature R_{ij}.
5. Compute the scalar curvature R.

[7] In Quantum Mechanics, this reads $[X, P] = \imath\hbar$.

Exercises 4.2 Christoffel symbols
Prove that under a change of coordinates from $\{x^i\}$ to $\{x^{i'}\}$, the Christoffel symbols do not transform in a covariant way but into

$$\Gamma^{k'}_{i'j'} = \underbrace{\frac{\partial x^{k'}}{\partial x^p}\frac{\partial^2 x^p}{\partial x^{i'}\partial x^{j'}}}_{\text{Non-covariant term}} + \Gamma^p_{qr}\frac{\partial x^q}{\partial x^{i'}}\frac{\partial x^r}{\partial x^{j'}}\frac{\partial x^{k'}}{\partial x^p}$$

Exercises 4.3 Lie algebra
A vector space $\mathcal{G}$ is called a *Lie algebra* if there exists a product (called Lie bracket) $[\cdot,\cdot] : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ with the following properties

- Antisymmetry: $[x, y] = -[y, x]\ , \forall x, y \in \mathcal{G}$
- Jacobi identity: $[x, [y, z]] + [z, [x, y]] + [y, [z, x]] = 0\ , \forall x, y, z \in \mathcal{G}$

1. Prove that the space of real antisymmetric n-dimensional matrix

$$\mathcal{U}(n) = \{A \in \mathrm{GL}(n, \mathbb{R})\ ,\ A^\dagger = -A\}$$

is a (finite) Lie algebra with the bracket

$$[A, B] = AB - BA$$

2. Prove that the space of vector fields of TM is an (infinite) Lie algebra with the Lie bracket

$$[X, Y] = (X^i \partial_i Y^j - Y^i \partial_i X^j)\partial_j$$

where we have set $X = X^i \partial_i$ and $Y = Y^i \partial_i$.

Exercises 4.4 Asian option and Heisenberg algebra
An Asian call option on a single stock with strike K is defined by the following payoff at the maturity date T

$$\max\left(\frac{1}{T}\int_0^T f_u du - K, 0\right)$$

with f_u the stock forward at time u (we take a zero interest rate for the sake of simplicity). We explain in this exercise a link with the *Heisenberg Lie algebra.* We note $A_t = \int_0^t f_u du$. The SDEs generated by the processes (f_t, A_t) in the Black-Scholes model are

$$\frac{df_t}{f_t} = \sigma dW_t$$
$$dA_t = f_t dt$$

1. Convert the Itô diffusion in a Stratonovich diffusion

$$\begin{aligned}\frac{df_t}{f_t} &= -\frac{1}{2}\sigma^2 dt + \sigma \diamond dW_t \\ dA_t &= f_t dt\end{aligned}$$

2. Write the diffusion operator in the Hörmander form.

3. Convert these SDEs into ODEs by replacing the Brownian W_t by the path $\omega_\pm = \pm t$.

$$\begin{aligned}\frac{df_t}{f_t} &= (-\frac{1}{2}\sigma^2 \pm \sigma) dt \\ dA_t &= f_t dt\end{aligned}$$

4. These flows (see B.6 in appendix B) are generated by the following vector fields

$$V_\pm = f\partial_A + (-\frac{1}{2}\sigma^2 \pm \sigma) f \partial_f$$

Prove that the vector fields V_+, V_-, $V_{+-} \equiv [V_+, V_-]$ satisfy the following Lie algebra (see exercise 4.3), called Heisenberg algebra

$$[V_{+-}, V_+] = 0 \;,\; [V_{+-}, V_-] = 0$$

Exercises 4.5 Mehler's formula

In this exercise, we give an example of a one-dimensional heat kernel equation (HKE) which admits an analytical solution. We consider the following HKE

$$\partial_t \bar{p}(t,x|x_0) = (\partial_x^2 - \frac{r^2 x^2}{16} - f)\bar{p}(t,x|x_0) \tag{4.85}$$

with r and f two constants.

1. By scaling the variables x and t and doing a time-dependent gauge transform, show that the HKE (4.85) can be reduced to

$$\partial_t p(t,x|x_0) = (\partial_x^2 - x^2) p(t,x|x_0)$$

This PDE corresponds to a HKE with a zero Abelian connection $\mathcal{A}$ and a quadratic potential Q (called harmonic potential). By observing that the operator D is quadratic in differentiation and multiplication, we seek a solution which is a Gaussian function of x and x_0. In addition, the solution must clearly be symmetric in x and x_0 since the operator D is self-adjoint. We therefore try the following ansatz

$$p(t,x|x_0) = \exp\left(a(t)\frac{x^2}{2} + b(t) x x_0 + a(t) x_0^2 + c(t)\right)$$

2. Prove that the coefficients $a(t), b(t)$ and $c(t)$ satisfy the following ODEs

$$\frac{\dot{a}(t)}{2} = a(t)^2 - 1 = b(t)^2$$
$$\dot{c}(t) = a(t)$$

3. Prove that the solutions are given by

$$a(t) = -\coth(2t + C)$$
$$b(t) = \csc(2t + C)$$
$$c(t) = -\frac{1}{2}\ln\sinh(2t + C) + D$$

with C and D two integration constants.

4. By using the initial condition $p(0, x|x_0) = \delta(x - x_0)$, show that the values of the integration constants are

$$C = 0$$
$$D = \ln(2\pi)^{-\frac{1}{2}}$$

5. Finally, prove that the solution of (4.85) is

$$\bar{p}(t, x|x_0) = \sqrt{\frac{t\frac{r}{2}}{4\pi t\sinh(t\frac{r}{2})}} \exp\left(-\frac{r}{8t}\left(\coth(t\frac{r}{2})(x^2 + x_0^2) - 2\csc(t\frac{r}{2})xx_0\right) - tf\right)$$

Chapter 5

Local Volatility Models and Geometry of Real Curves

Abstract The existence of an implied volatility indicates that the Black-Scholes assumption that assets have a constant volatility should be relaxed. The simplest extension of the log-normal Black-Scholes model is to assume that assets still follow a one-dimensional Itô diffusion process but with a volatility function $\sigma_{\text{loc}}(t, f)$ depending on the underlying forward f and the time t. As shown by Dupire [81], prices of European call-put options determine the diffusion term $\sigma_{\text{loc}}(t, f)$ uniquely.
In this chapter, we show how to apply the theorem (4.4) to find an asymptotic solution to the backward Kolmogorov (Black-Scholes) equation and derive an asymptotic implied volatility in the context of local volatility models (LVM). Before moving on to the general case in the second section, we consider in the first section a specific separable local volatility function: $\sigma_{\text{loc}}(t, f) = A(t)C(f)$.
Throughout this chapter, deterministic interest rates are assumed for the sake of simplicity.

5.1 Separable local volatility model

Let us consider an asset with a price S_t whose forward is $f_t = \frac{S_t}{P_{tT}}$. As the product of the forward with the bond P_{tT} is a traded asset, the forward f_t should be a local martingale under the forward measure $\mathbb{P}^T$. It follows a driftless process and we assume the following dynamics

$$df_t = A(t)C(f_t)dW_t \tag{5.1}$$

$C : \mathbb{R}_+ \to \mathbb{R}_+$ and $A : \mathbb{R}_+ \to \mathbb{R}_+$ are continuous deterministic functions. We also assume that $C(0) = 0$ which ensures that the forward cannot go below zero.
After a change of local time

$$t'(t) = \int_0^t A^2(s)ds$$

TABLE 5.1: Example of separable LV models satisfying $C(0) = 0$.

LV Model	$C(f)$
Black-Scholes	f
Quadratic	$f(af+b)$
CEV	$f^{\beta}\ ,\ \beta > 0$
LCEV	$f\min(f^{\beta-1}, \epsilon^{\beta-1})\ ,\ \epsilon > 0$
Exponential	$f(1+ae^{-bf})\ ,\ a,b > 0$

we obtain the simpler time-homogeneous SDE[1]

$$df_{t'} = C(f_{t'})dW_{t'} \tag{5.2}$$

Below, we write t instead of t'.
In table (5.1), we list examples of commonly used separable local volatility models. For most of these models (except Black-Scholes), the function $C(f)$ does not satisfy the global Lipschitz condition (2.17). Therefore we cannot conclude that the SDE (5.2) admits a unique strong solution.
This point will be discussed below: it is shown that these models admit a weak solution up to an explosion time.

5.1.1 Weak solution

Let $T \in (0, \infty)$ be a fixed time horizon and D an open connected subset of $\mathbb{R}^n$. We consider the n-dimensional SDE

$$dx_t = b(t, x_t)dt + \sum_{j=1}^{m} \sigma_j(t, x_t)dW_t^j \tag{5.3}$$

for continuous functions $b : [0,T] \times D \to \mathbb{R}^n$, $\sigma_j : [0,T] \times D \to \mathbb{R}^n$, $j = 1, \cdots, m$ with an m-dimensional Brownian motion $\{W_t^j\}_{j=1,\cdots,m}$.

THEOREM 5.1 Kunita
If the coefficients b and σ_j, $j = 1, \cdots, m$ on $[0,T] \times D$ are locally Lipschitz-continuous in x, uniformly in t, i.e.,

$$\| G(t,x) - G(t,y) \| \leq K_n \| x - y \| \ , \ \forall \ \| x \|, \| y \| \leq n \ , \ \forall\, t \in [0,T]$$

with $G \in \{b, \sigma_1, \cdots, \sigma_m\}$, then the SDE (5.3) has a unique weak solution up to an explosion time.

The local Lipschitz condition is satisfied for all models listed in table 5.1 (except for the Constant Elasticity of Variance (CEV) model $0 < \beta < 1$). From

[1]Use the formal rule $< dW_{t'}, dW_{t'} >= dt' = A(t)^2 < dW_t, dW_t >= A(t)^2 dt$.

this theorem, we conclude that these models admit a unique weak solution up to an explosion time τ that will be computed in the section 5.1.2.

REMARK 5.1 Note that the linear growth condition

$$\exists K > 0 \text{ such that } ||G(t,x)||^2 \leq K\left(1 + ||x||^2\right) \ , \forall\, x > 0$$

ensures that the process does not explode (i.e., $\tau = \infty$ a.s.). ∎

CEV model

As the local Lipschitz condition is not satisfied for the *CEV model* for $0 < \beta < 1$ at $f = 0$, the theorem (5.1) can not be used. As the CEV process will often appear in the remainder, here is a list of a few useful results about the CEV model below [80], [44].

LEMMA 5.1

$$df_t = f_t^{\beta} dW_t \ , \ \beta > 0 \tag{5.4}$$

1. *All solutions to (5.4) are non-explosive.*
2. *For $\beta \geq \frac{1}{2}$, the SDE has a unique solution.*
3. *For $0 < \beta < 1$, $f = 0$ is an attainable boundary.*
4. *For $\beta \geq 1$, $f = 0$ is an unattainable boundary.*
5. *For $0 < \beta < \frac{1}{2}$, the SDE (5.4) does not have a unique solution unless a separate boundary condition is specified for the boundary behavior at $f = 0$.*
6. *For $\frac{1}{2} \leq \beta \leq 1$, the SDE (5.4) has an unique solution with an absorbing condition at $f = 0$.*
7. *If $f_{t=0} > 0$, the solution is positive $\forall t > 0$.*

According to (5), if $0 < \beta < \frac{1}{2}$, the behavior at $f = 0$ is not unique and requires us to choose between the two possible boundary conditions: absorbing or reflecting. If we require f_t to be a martingale even when starting at $f = 0$, we should have an absorbing condition for $f = 0$ [80]. Indeed a reflecting boundary condition leads to an arbitrage opportunity: if there is a positive probability of reaching zero, one just has to wait for that event to happen and buy at zero cost the forward which will have a strictly positive value an instant later due to the reflecting boundary.

For $0 < \beta < 1$, the origin is an attainable boundary, not such a nice feature from a modeling point of view. To overcome this drawback, we introduce the *LCEV model* [44]

$$C(f) = f \min\left(f^{\beta-1}, \epsilon^{\beta-1}\right) \ , \ \epsilon > 0 \tag{5.5}$$

ϵ is a small number for $\beta < 1$ and a large number when $\beta > 1$.

5.1.2 Non-explosion and martingality

Being a driftless process (5.2), f_t is a local martingale as imposed by the no-arbitrage condition. Whilst strictly local martingale discounted stock prices do not involve strict arbitrage opportunities, they can create numerical problems, for example, causing the inability to price derivatives using the pricing formula (2.31). The Monte-Carlo pricing of derivatives is therefore inaccurate. Moreover, the *put-call parity* is no longer valid as the forward is not priced conveniently [29], [114]:

PROPOSITION 5.1
If f_t is a strict local martingale then the put-call parity is no longer valid:

$$\mathbb{E}^{\mathbb{P}^T}[\max(f_T - K, 0)|\mathcal{F}_t] - \mathbb{E}^{\mathbb{P}^T}[\max(K - f_T, 0)|\mathcal{F}_t] \neq f_0 - K$$

PROOF The payoff of a call option at the maturity date T with strike K can be decomposed as

$$\max(f_T - K, 0) = \max(K - f_T, 0) + f_T - K$$

The first term $\max(K - f_T, 0)$ is the payoff of a European put option. As well as being a bounded local martingale, it is also a true martingale [27]. Consequently, $\max(f_t - K, 0)$ is a martingale if and only if f_t is a martingale. In this case, we obtain the put-call parity

$$\mathcal{C} = \mathcal{P} + f_0 - K$$

with $\mathcal{C}$ (resp. $\mathcal{P}$) the fair value of a call (resp. put) option. □

In order to preserve the put-call parity and the ability to use the pricing formula (2.31), it is required that the forward is not only a local martingale but also a *true* martingale. Let us see under which conditions a positive local martingale can be a martingale.
We have

$$\mathbb{E}^{\mathbb{P}^T}[f_T|\mathcal{F}_t] = f_t \mathbb{E}^{\mathbb{P}^T}[e^{-\frac{1}{2}\int_t^T a_s^2 ds + \int_t^T a_s dW_s}|\mathcal{F}_t] \tag{5.6}$$

with $a_t = \frac{C(f_t)}{f_t}$ the volatility of the forward. Under $\mathbb{P}^T$, a_t satisfies the SDE

$$da_t = \frac{C(f)^2}{2} \frac{d^2 \left(\frac{C(f)}{f}\right)}{df^2} dt + C(f) \frac{d\left(\frac{C(f)}{f}\right)}{df} dW_t \tag{5.7}$$

The exponential term in (5.6) is the Radon-Nikodym derivative

$$M_t \equiv \frac{d\mathbb{P}^f}{d\mathbb{P}^T} |_{\mathcal{F}_t} = e^{-\frac{1}{2}\int_0^t a_s^2 ds + \int_0^t a_s dW_s}$$

corresponding to the change of measure from the forward measure $\mathbb{P}^T$ to the spot measure $\mathbb{P}^f$ associated to the forward itself. Therefore, assuming that M_t is a well-defined martingale, when doing the change of measure from $\mathbb{P}^T$ to $\mathbb{P}^f$, we obtain

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^T}[f_T|\mathcal{F}_t] &= f_t \mathbb{E}^{\mathbb{P}^T}[\frac{M_T}{M_t}|\mathcal{F}_t] \\ &= f_t \mathbb{E}^{\mathbb{P}^f}[1|\mathcal{F}_t] = f_t \end{aligned}$$

and f_t is a martingale. In the section on Girsanov's theorem 2.7, we have encountered sufficient conditions (2.39) and (2.40) to ensure that M_t is not only a local martingale but also a true martingale. The following theorem gives the necessary and sufficient condition [140, 107, 47].

THEOREM 5.2
Let us suppose that a_t is the unique non-exploding (at infinity) strong (or weak) solution of the SDE under $\mathbb{P}^T$

$$da_t = b(a_t)dt + \sigma(a_t)dW_t$$

where $b(\cdot)$ and $\sigma(\cdot)$ are continuous function and $\sigma(\cdot)^2 > 0$. Then, the exponential local martingale M_t defined by

$$M_t = e^{-\frac{1}{2}\int_0^t a_s^2 ds + \int_0^t a_s dZ_s}$$

is a martingale if and only if there is a non-exploding weak solution of the SDE

$$da_t = (b(a_t) + \rho a_t \sigma(a_t))\, dt + \sigma(a_t) dB_t \tag{5.8}$$

where $dZ_t dB_t = \rho dt (= dW_t dZ_t)$.

PROOF
$\Rightarrow$: If M_t is a martingale, we define the probability measure $\mathbb{P}^f$ by the Radon-Nikodym derivative: $\frac{d\mathbb{P}^f}{d\mathbb{P}^T} = M_t$. From the change of measure from $\mathbb{P}^T$ to $\mathbb{P}^f$,

a_t satisfies the SDE (5.8). a_t does not explode under $\mathbb{P}^f$ as $\mathbb{P}^f$ is equivalent to $\mathbb{P}^T$. Besides, a_t does not explode under $\mathbb{P}^T$ by definition.
$\Leftarrow$: Let us denote $\tau_n = \inf\{t \in \mathbb{R}_+ \ : \ a_t \geq n\}$ and $\tau_\infty = \lim_{n\to\infty} \tau_n$ the explosion time for a_t. As a_t does not explode under $\mathbb{P}^T$, we have $1_{\tau_\infty > T} = 1$ $\mathbb{P}^T$-a.s. and

$$\mathbb{E}^{\mathbb{P}^T}[f_T] = \mathbb{E}^{\mathbb{P}^T}[f_T 1_{\tau_\infty > T}] = \mathbb{E}^{\mathbb{P}^T}[f_T \lim_{n\to\infty} 1_{\tau_n > T}] \tag{5.9}$$

By observing that

$$0 \leq f_T 1_{\tau_n > T} \leq f_T 1_{\tau_{n+1} > T} \ , \ n = 1, 2, \ldots$$

Thanks to the monotone convergence theorem [26], we conclude that we can permute the mean-value and the lim sign in (5.9)

$$\begin{aligned}\mathbb{E}^{\mathbb{P}^T}[f_T] &= \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}^T}[f_T 1_{\tau_n > T}] \\ &= f_0 \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}^T}[M_T 1_{\tau_n > T}]\end{aligned}$$

Let us define $M_t^n = M_{t\wedge\tau_n}$.[2] We have

$$\mathbb{E}^{\mathbb{P}^T}[M_T 1_{\tau_n > T}] = \mathbb{E}^{\mathbb{P}^T}[M_T^n 1_{\tau_n > T}]$$

As the coefficients in M_t^n are bounded, M_t^n is a martingale (the Novikov condition is satisfied). Therefore, if we define the measure $\mathbb{P}_n^f$ with the Radon-Nikodym derivative $\frac{d\mathbb{P}_n^f}{d\mathbb{P}^T} = M_t^n$ then

$$\mathbb{E}^{\mathbb{P}^T}[M_T^n 1_{\tau_n > T}] = \mathbb{E}^{\mathbb{P}_n^f}[1_{\tau_n > T}]$$

and

$$\mathbb{E}^{\mathbb{P}^T}[f_T] = f_0 \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}_n^f}[1_{\tau_n > T}]$$

As the process a_t satisfies the same SDE under $\mathbb{P}_n^f$ and $\mathbb{P}^f$ up to an explosion time τ_∞, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}^T}[f_T] &= f_0 \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}^f}[1_{\tau_n > T}] \\ &= f_0 \mathbb{E}^{\mathbb{P}^f}[1_{\tau_\infty > T}]\end{aligned}$$

The monotone convergence theorem was used in the last equality. As a_t does not explode under $\mathbb{P}^f$, we deduce that $\mathbb{E}^{\mathbb{P}^T}[f_T] = f_0$ and f_T is a martingale under $\mathbb{P}^T$. ∎

[2] $t_1 \wedge t_2 = \min(t_1, t_2)$.

According to the theorem above, f_t is a martingale if and only if the volatility process a_t does not explode in the measure $\mathbb{P}^T$ (5.7) and $\mathbb{P}^f$ (5.10): In the spot measure $\mathbb{P}^f$, we have

$$da_t = C(f)^2 \left(\frac{1}{2} \frac{d^2 \left(\frac{C(f)}{f} \right)}{df^2} + \frac{1}{f} \frac{d \left(\frac{C(f)}{f} \right)}{df} \right) dt + C(f) \frac{d \left(\frac{C(f)}{f} \right)}{df} dW_t^f \tag{5.10}$$

The SDEs (5.7) and (5.10) correspond to one-dimensional Itô diffusion processes for which there exists a necessary and sufficient criteria to test if there are non-exploding solutions (Theorem VI. 3.2 in [24]):

THEOREM 5.3 Feller non-explosion test

Let $I = (c_1, c_2)$ with $-\infty \leq c_1 < c_2 \leq \infty$. Let $\sigma(\cdot)$ and $b(\cdot)$ be continuous functions on I such that $\sigma(x) > 0 \; \forall x \in I$. Define for fixed $c \in I$

$$s(x) = \int_c^x e^{-2\int_c^y \frac{b(z)}{\sigma(z)^2} dz} dy$$
$$l(x) = \int_c^x s'(y) dy \int_c^y \frac{1}{\sigma(z)^2 s(z)'} dz$$

Let $\mathbb{P}_x$ be the diffusion measure of an Itô process X_t starting from x where the infinitesimal generator is

$$L = \frac{1}{2}\sigma(x)^2 \frac{d^2}{dx^2} + b(x) \frac{d}{dx}$$

Below, we refer to

$$e = \inf\{t \geq 0 \; : \; X_t \notin (c_1, c_2)\}$$

as the exit time from I.

- $\mathbb{P}_x[e = \infty] = 1$ *if and only if* $l(c_1) = l(c_2) = \infty$.
- $\mathbb{P}_x[e < \infty] < 1$ *if and only if one of the following three conditions is satisfied*
  1. $l(c_1) < \infty$ *and* $l(c_2) < \infty$.
  2. $l(c_1) < \infty$ *and* $s(c_2) = \infty$.
  3. $l(c_2) < \infty$ *and* $s(c_1) = -\infty$.

Applying the criteria to the processes (5.7) and (5.10), we obtain

$$l_{\mathbb{P}^T}(a) = \int_{\phi^{-1}(c)}^{\phi^{-1}(a)} df \int_{\phi^{-1}(c)}^{f} \frac{dx}{C(x)^2} \tag{5.11}$$

$$l_{\mathbb{P}^f}(a) = \int_{\phi^{-1}(c)}^{\phi^{-1}(a)} \frac{df}{f^2} \int_{\phi^{-1}(c)}^{f} dx \frac{x^2}{C(x)^2} \tag{5.12}$$

TABLE 5.2: Feller criteria for the CEV model.

	$0 < \beta < \frac{1}{2}$	$\frac{1}{2} \leq \beta < 1$	$\beta > 1$
$l_{\mathbb{P}^T}(f=0)$	$\neq \infty$	$\neq \infty$	∞
$l_{\mathbb{P}^T}(f=\infty)$	∞	∞	∞
$l_{\mathbb{P}^f}(f=0)$	∞	∞	$\neq \infty$
$l_{\mathbb{P}^f}(f=\infty)$	∞	∞	$\neq \infty$

where $\phi^{-1}(\cdot)$ is the inverse of the monotone function $\phi(f) = \frac{C(f)}{f}$.

Example 5.1 CEV model
Applying the formula above (5.11,5.12) to the CEV model for which $C(x) = x^\beta$, $\phi^{-1}(a) = a^{\frac{1}{\beta-1}}$, $\beta \neq 1$, we obtain

- For all $\beta \in \mathbb{R}/\{1, \frac{1}{2}\}$

$$l_{\mathbb{P}^T}(a) = \frac{1}{(1-2\beta)}[\frac{f^{2-2\beta}}{2-2\beta} - \phi^{-1}(c)^{1-2\beta} f]_{\phi^{-1}(c)}^{\phi^{-1}(a)}$$
$$l_{\mathbb{P}^f}(a) = \frac{1}{(3-2\beta)}[\frac{f^{2-2\beta}}{2-2\beta} + \frac{\phi^{-1}(c)^{3-2\beta}}{f}]_{\phi^{-1}(c)}^{\phi^{-1}(a)}$$

- For $\beta = \frac{1}{2}$

$$l_{\mathbb{P}^T}(a) = [-f\left(1 + \ln\phi^{-1}(c)\right) + f\ln f]_{\phi^{-1}(c)}^{\phi^{-1}(a)}$$
$$l_{\mathbb{P}^f}(a) = \frac{1}{2}[\frac{\left(\phi^{-1}(c)\right)^2}{f} + f]_{\phi^{-1}(c)}^{\phi^{-1}(a)}$$

From table 5.2, we deduce the properties (1), (3) and (4) in the lemma 5.1 and that the CEV model defines a martingale if and only if $\beta \leq 1$.

□

Example 5.2 Quadratic volatility model
In a *quadratic volatility model*, the forward is driven under $\mathbb{P}^T$ by

$$df_t = f_t(af_t + b)dW_t \ , \ a \neq 0 \ b \neq 0 \tag{5.13}$$

These models have been studied in detail by [145]. Applying the Feller criteria (in particular formula (5.11) and (5.12)), we obtain that

$$l_{\mathbb{P}^T}(f=0) < \infty \ , \ l_{\mathbb{P}^T}(f=\infty) = \infty$$
$$l_{\mathbb{P}^f}(f=0) = \infty \ , \ l_{\mathbb{P}^f}(f=\infty) < \infty$$

We conclude that the SDE (5.13) under $\mathbb{P}^T$ does not explode to infinity contrary to the intuition but converges to $f = 0$ $\mathbb{P}^T$-a.s. Moreover, f_t is *not* a true martingale but only a local martingale. □

5.1.3 Real curve

In our geometrical framework, the model (5.2) corresponds to a (one-dimensional) real curve endowed with the time-independent metric

$$g_{ff} = \frac{2}{C(f)^2}$$

For the new coordinate $u(f) = \sqrt{2}\int_{f_0}^{f} \frac{dx}{C(x)}$, the metric is flat: $g_{uu} = 1$. The distance is given by the classical Euclidean distance

$$d(u,u')^2 = |u-u'|^2 = 2\left(\int_{f'}^{f} \frac{dx}{C(x)}\right)^2 \tag{5.14}$$

The connection $\mathcal{A}$ (4.58) and the section Q (4.59) are given by

$$\begin{aligned}\mathcal{A} &= -\frac{1}{2} d\ln C(f)\\ Q(f) &= \frac{C(f)^2}{4}\left(\left(\frac{C''(f)}{C(f)}\right) - \frac{1}{2}\left(\frac{C'(f)}{C(f)}\right)^2\right)\end{aligned} \tag{5.15}$$

As $\mathcal{A}$ is an exact form, the parallel transport between two points f' and f is given by

$$\mathcal{P}(f',f) = e^{-\int_{f'}^{f}\mathcal{A}} = \sqrt{\frac{C(f)}{C(f')}} \tag{5.16}$$

Moreover, the first (4.70) and second heat kernel coefficients (4.71) (in the limit $f' = f$) are given by

$$a_1(f,f) = Q(f) \tag{5.17}$$

$$a_2(f,f) = \frac{1}{2}\left(Q(f)^2 + \frac{\Delta Q(f)}{3}\right) \tag{5.18}$$

with the Laplacian

$$\Delta = \partial_u^2 = \frac{C(f)}{2}\partial_f(C(f)\partial_f)$$

In reality, to compute the conditional probability density $p(t,f|f_0)$ using the heat kernel expansion, we need the non-diagonal coefficients $a_1(f,f_0)$ and $a_2(f,f_0)$. From the diagonal heat kernel coefficients (5.17) and (5.18), we use the following approximation for the non-diagonal terms that we justify in section 5.2.2.1

$$a_1(f,f_0) = a_1(f_{\mathrm{av}}, f_{\mathrm{av}}) \tag{5.19}$$

$$a_2(f,f_0) = a_2(f_{\mathrm{av}}, f_{\mathrm{av}}) \tag{5.20}$$

with $f_{\rm av} = \frac{f_0+f}{2}$. By plugging the expressions (5.14), (5.15), (5.16), (5.19), (5.20) in the heat kernel expansion (4.67), we obtain the probability density for the old coordinate f at the second-order in time (see remark 4.6)

$$p(t', f|f_0) = \frac{1}{C(f)\sqrt{2\pi t'}}\sqrt{\frac{C(f_0)}{C(f)}}e^{-\frac{\sigma(f,f_0)}{2t'}} \left(1 + Q(f_{\rm av})t' + \frac{1}{2}\left(Q^2(f_{\rm av}) + \frac{\Delta Q(f_{\rm av})}{3}\right)t'^2\right) \tag{5.21}$$

with

$$\sigma(f, f_0) = \left(\int_f^{f_0} \frac{dx}{C(x)}\right)^2$$

and

$$t' = \int_0^t A(s)^2 ds$$

Whilst this result has already been obtained in [99] using a more direct approach, our method produces a quicker answer.
Specializing our asymptotic solution (5.21) to the CEV model, we obtain

Example 5.3 CEV model
For the CEV model ($C(f) = f^\beta$, $A(t) = \sigma$), the asymptotic solution at the second-order to the backward Kolmogorov equation is from (5.21)

$$p(t, f|f_0) = \frac{f^{-\beta}}{\sqrt{2\pi t'}}e^{\frac{-\left(\int_f^{f_0} df' f'^{-\beta}\right)^2}{2t'}}\left(\frac{f_0}{f}\right)^{\frac{\beta}{2}} \left(1 + \frac{1}{8}f_{\rm av}^{2\beta-2}\beta(\beta-2)t' + \frac{\beta(\beta-2)(3\beta-2)(3\beta-4)}{128}f_{\rm av}^{4\beta-4}(t')^2\right) \tag{5.22}$$

with $t' = \sigma^2 t$. We can check that this asymptotic solution for $\beta = 1$ reproduces the asymptotic solution previously given in the case of the GBM process (4.73).
In the case of the CEV model, an exact solution (with absorbing condition at $f = 0$) exists for the Kolmogorov equation [76] that we derive in chapter 9 using a spectral decomposition

$$p(t, f|f_0) = \frac{f^{\frac{1}{2}-2\beta}}{(1-\beta)t'}\sqrt{f_0}e^{-\frac{f^{2(1-\beta)}+f_0^{2(1-\beta)}}{2(1-\beta)^2 t'}}\mathrm{I}_{\frac{1}{2(1-\beta)}}\left(\frac{(ff_0)^{1-\beta}}{(1-\beta)^2 t'}\right) \tag{5.23}$$

where $\mathrm{I}_\nu(z)$ is the modified Bessel function of the first kind given by

$$I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma(\nu+k+1)}$$

$\Gamma(\cdot)$ is the gamma function.

This exact solution will be used to test the validity of our asymptotic solution at the second-order. In the following graphs (Fig. 5.1, Fig. 5.2), the asymptotic conditional probability (5.22) has been plotted against the exact solution (5.23) for $\beta = 0.33$ and $\beta = 0.6$. There is a good match between the exact and approximate solutions.

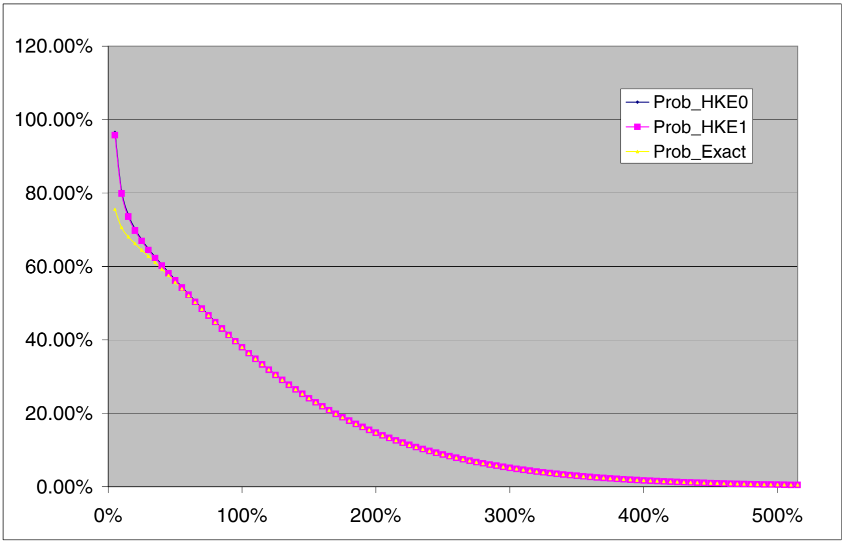

FIGURE 5.1: Comparison of the asymptotic solution at the first-order (resp. second-order) against the exact solution (5.23). $f_0 = 1$, $\sigma = 0.3$, $\beta = 0.33$, $\tau = 10$ years.

$\square$

In the next paragraph, we focus on the general case where we assume that the volatility function $\sigma_{\text{loc}}(t, f)$ is a general function of the forward and the time. This model is called a local volatility model (LVM). What is particular with this LVM is that the market is complete and the model can be automatically calibrated to the initial implied volatility as coded in the Dupire formula.

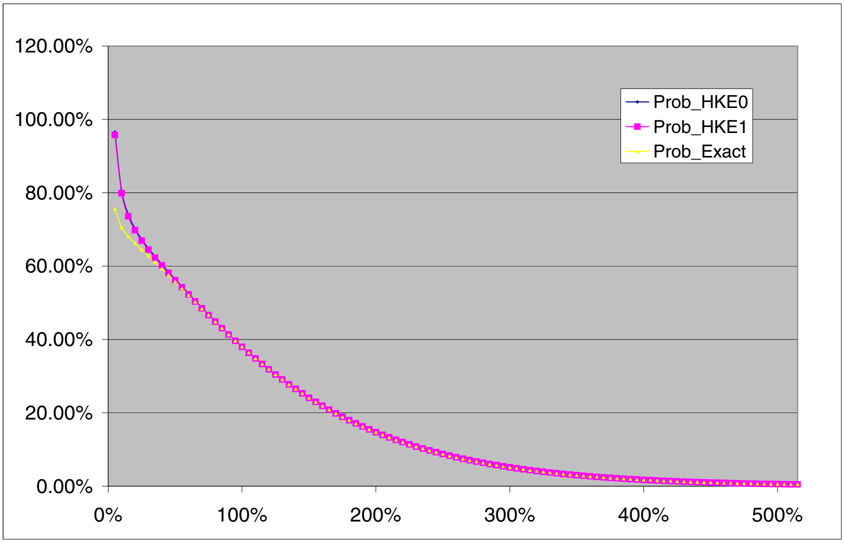

FIGURE 5.2: Comparison of the asymptotic solution at the first-order (resp. second-order) against the exact solution (5.23). $f_0 = 1$, $\sigma = 0.3$, $\beta = 0.6$, $\tau = 10$ years.

5.2 Local volatility model

5.2.1 Dupire's formula

We derive the Dupire local volatility formula [81]. We assume that the asset price S_t follows the following SDE in the risk-neutral measure $\mathbb{P}$

$$dS_t = r_t S_t dt + S_t \sigma(t, S_t) dW_t$$

with $\sigma(\cdot, \cdot)$ a continuous function of time t and S_t satisfying the uniform ellipticity condition: there exists a $\lambda \in \mathbb{R}^+$ such that

$$\lambda^{-1} \leq \sigma(t, S) \leq \lambda$$

As a result f_t admits a non-explosive strong solution and f_t is a true martingale.

We try to choose the local volatility (LV) $\sigma(t, S_t)$ such that when we price European call options $C(K, T|S_0)$ for any strike price K and any maturity T, we match the market prices $C^{\text{mkt}}(K, T|S_0)$.
By applying Itô-Tanaka's formula [27] (which is a generalization of the Itô formula for non-smooth functions) on the payoff $\max(S_t - K, 0)$, we obtain

$$\begin{aligned} d\max(S_t - K, 0) &= 1(S_t - K)\left(r_t S_t dt + S_t \sigma(t, S_t) dW_t\right) \\ &\quad + \frac{1}{2} K^2 \sigma(t, K)^2 \delta(S_t - K) dt \end{aligned} \tag{5.24}$$

with $\delta(\cdot)$ the Dirac function. Then, by taking the mean value operator $\mathbb{E}^{\mathbb{P}}[\cdot|\mathcal{F}_0]$ on both sides of equation (5.24), we have

$$\begin{aligned} d\mathbb{E}^{\mathbb{P}}[\max(S_t - K, 0)|\mathcal{F}_0] &= r_t \mathbb{E}^{\mathbb{P}}[1(S_t - K) S_t|\mathcal{F}_0] dt \\ &\quad + \frac{1}{2} K^2 \sigma^2(t, K) \mathbb{E}^{\mathbb{P}}[\delta(S_t - K)|\mathcal{F}_0] dt \end{aligned}$$

As the fair value of a call option with a maturity t and a strike K is

$$\mathcal{C}(t, K|S_0) = P_{0t} \mathbb{E}^{\mathbb{P}}[\max(S_t - K, 0)|\mathcal{F}_0]$$

we obtain

$$\begin{aligned} d\mathcal{C}(K, t|S_0) &= r_t dt(-\mathcal{C}(t, K|S_0) + P_{0t} \mathbb{E}^{\mathbb{P}}[1(S_t - K) S_t|\mathcal{F}_0]) \\ &\quad + \frac{1}{2} P_{0t} K^2 \sigma^2(t, K) \mathbb{E}^{\mathbb{P}}[\delta(S_t - K)|\mathcal{F}_0] dt \end{aligned}$$

Using the following identities

$$\begin{aligned} \frac{\partial^2 \mathcal{C}(t, K|S_0)}{\partial K^2} &= P_{0t} \mathbb{E}^{\mathbb{P}}[\delta(S_t - K)|\mathcal{F}_0] \\ \mathbb{E}^{\mathbb{P}}[1(S_t - K) S_t|\mathcal{F}_0] &= \mathbb{E}^{\mathbb{P}}[\max(S_t - K, 0)|\mathcal{F}_t] + K \mathbb{E}^{\mathbb{P}}[1(S_t - K)|\mathcal{F}_0] \\ \frac{\partial \mathcal{C}(t, K|S_0)}{\partial K} &= -P_{0t} \mathbb{E}^{\mathbb{P}}[1(S_t - K)|\mathcal{F}_0] \end{aligned}$$

we finally obtain

$$d\mathcal{C}(t, K|S_0) = \left(-r_t K \frac{\partial \mathcal{C}(t, K|S_0)}{\partial K} + \frac{1}{2} K^2 \sigma^2(t, K) \frac{\partial^2 \mathcal{C}(K, t|S_0)}{\partial K^2}\right) dt$$

Inverting this formula, we obtain the Dupire LV

DEFINITION 5.1 Local volatility *The Dupire LV, denoted $\sigma_{\text{loc}}(t, S)$, is defined as*

$$\sigma_{\text{loc}}(t, K)^2 \equiv 2 \frac{\frac{\partial \mathcal{C}(t,K|S_0)}{\partial t} + r_t K \frac{\partial \mathcal{C}(t,K|S_0)}{\partial K}}{K^2 \frac{\partial^2 \mathcal{C}(t,K|S_0)}{\partial^2 K}} \tag{5.25}$$

We have shown that European call options $C(t,K|S_0)$ for every strike $K \in [0,\infty)$ and maturity $t \in [0,\infty)$ are automatically calibrated (therefore so is the initial implied volatility) if the instantaneous volatility $\sigma(t,S)$ is equal to the Dupire LV $\sigma_{\mathrm{loc}}(t,S)$ defined by (5.25). By definition of the implied volatility, we have

$$C^{\mathrm{mkt}}(T,K|S,t) = C^{\mathrm{BS}}(K,T,\sigma_{\mathrm{BS}t}(K,T)|S,t)$$

By substituting the expression (3.1) in (5.25), we obtain a relation between the implied volatility and the Dupire LV. A straightforward computation gives

$$\sigma_{\mathrm{loc}}(T,y)^2 = \frac{\frac{\partial \omega(T,y)}{\partial T}}{1 - \frac{y}{\omega(T,y)}\frac{\partial \omega(T,y)}{\partial y} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{\omega(T,y)} + \frac{y^2}{\omega(T,y)^2}\right)\left(\frac{\partial \omega(T,y)}{\partial y}\right)^2 + \frac{1}{2}\frac{\partial^2 \omega}{\partial y^2}} \tag{5.26}$$

with $\omega(T,y) \equiv \sigma_{\mathrm{BS}}(K,T|S,t)^2(T-t)$ and $y = \ln\left(\frac{K}{f_0}\right)$.
In Fig. (5.3), we have plotted the market implied volatility (one year maturity) for the index SP500 versus the Dupire local volatility.

Example 5.4 Merton model
When we have an implied volatility with no skew, i.e., $\frac{\partial \omega}{\partial y} = 0$, the expression (5.26) reduces to

$$\sigma_{\mathrm{loc}}(T)^2 = \frac{\partial \omega(T)}{\partial T}$$

Inverting this equation, we deduce that the initial implied volatility (with no skew) is perfectly calibrated with a Merton model characterized by a time-dependent volatility $\sigma_{\mathrm{loc}}(\cdot)$ such that

$$\sigma^2_{\mathrm{BS}}(T|t) = \frac{1}{T-t}\int_t^T \sigma_{\mathrm{loc}}(s)^2 ds$$

$\square$

In the following, we obtain an asymptotic relation in the short-time limit between the LV function and the implied volatility using the heat kernel expansion on $\mathbb{R}$ with a time-dependent metric. This map will be particularly useful when discussing stochastic volatility models in chapter 6.

5.2.2 Local volatility and asymptotic implied volatility

Let us assume that the forward f_t follows a LVM in the forward measure $\mathbb{P}^T$

$$df = C(t,f)dW_t \; ; \; f(t=0) = f_0 \tag{5.27}$$

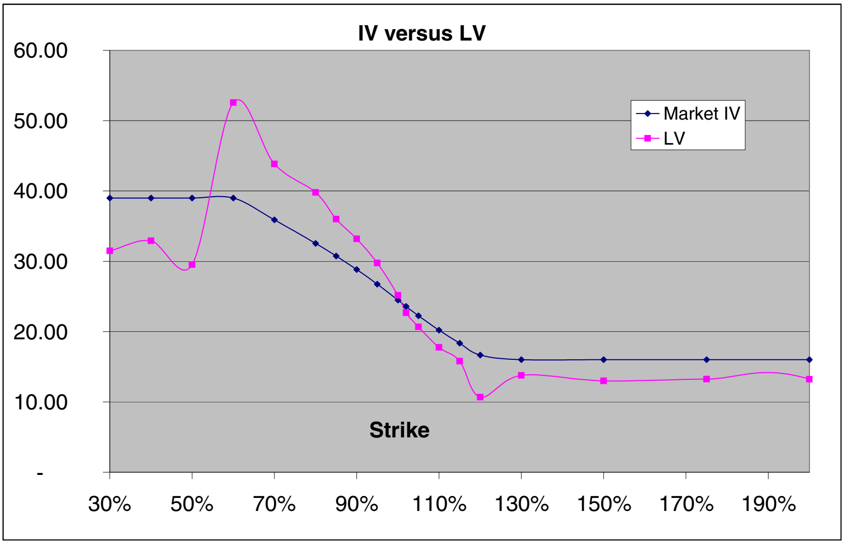

FIGURE 5.3: Market implied volatility (SP500, 3-March-2008) versus Dupire local volatility (multiplied by $\times 100$). $T = 1$ year. Note that the local skew is twice the implied volatility skew (see remark 5.2).

Repeating what was done in the previous section, we apply the Itô-Tanaka lemma on the payoff $\max(f_t - K, 0)$ and obtain

$$d\max(f_t - K, 0) = 1(f_t - K)df_t + \frac{1}{2}C(t,K)^2\delta(f_t - K)dt \qquad (5.28)$$

Taking the mean value operator $\mathbb{E}^{\mathbb{P}^T}[\cdot|\mathcal{F}_0]$ on both sides of this equation (5.28) and integrating over the time from $t = 0$ to T, we obtain that the fair value of a European call option (with a maturity date T and a strike K)

$$\frac{\mathcal{C}(T,K|f_0)}{P_{0T}} = \mathbb{E}^{\mathbb{P}^T}[\max(f_T - K, 0)|\mathcal{F}_0]$$

is given by

$$\mathcal{C}(T,K|f_0) = P_{0T}\left(\max(f_0 - K, 0) + \frac{1}{2}\int_0^T C(t,K)^2 p(t,K|f_0)dt\right) \qquad (5.29)$$

$p(t, K|f_0)$ is the conditional probability associated to the process (5.27) in the forward measure $\mathbb{P}^T$ which satisfies the forward Kolmogorov equation

$$\begin{aligned}\partial_t p(t, f|f_0) &= \frac{1}{2}\partial_f^2 \left(C(t,f)^2 p(t,f|f_0)\right) \\ &= \frac{1}{2}C^2\partial_f^2 p(t,f|f_0) + 2C\partial_f C\partial_f p(t,f|f_0) \\ &+ \left(C\partial_f^2 C + (\partial_f C)^2\right) p(t,f|f_0)\end{aligned}$$

with the terminal condition $\lim_{t\to 0} p(t, f|f_0) = \delta(f - f_0)$.
In our geometrical framework, the LV model corresponds to a (one-dimensional) real curve endowed with the time-dependent metric

$$g_{ff}(t) = \frac{2}{C(t,f)^2}$$

For the new coordinate $u(f) = \sqrt{2}\int_{f_0}^{f} \frac{dx}{C(x)}$ (with $C(f) \equiv C(0, f)$), the metric at $t = 0$ is flat

$$g_{uu} = 1$$

The distance between two points f' and f is then given by the classical Euclidean distance

$$d(f, f_0) = |u(f) - u(f_0)| = \sqrt{2}|\int^{f_0} f\frac{dx}{C(x)}| \tag{5.30}$$

The connection $\mathcal{A}$ (4.58) and the function Q (4.59) are given by

$$\mathcal{A} = \frac{3}{2}d\ln C(f) \tag{5.31}$$

$$Q = \frac{C(f)^2}{4}\left(\left(\frac{C''(f)}{C(f)}\right) - \frac{1}{2}\left(\frac{C'(f)}{C(f)}\right)^2\right) \tag{5.32}$$

As the Abelian connection is an exact form, the parallel transport between the point f' and f is obtained by direct integration

$$\mathcal{P}(f_0, f) = \sqrt{\frac{C(f_0)}{C(f)}\frac{1}{C(f)}}$$

Furthermore, the first heat kernel diagonal coefficient $a_1(f, f)$ is given by

$$a_1(f,f) = Q(f) - \frac{1}{4}\mathcal{G}(f)$$

with the coefficient $\mathcal{G}(f)$ (4.81) equal to

$$\mathcal{G}(f) = 2\partial_t \ln C(0, f) \tag{5.33}$$

By using the same approximation as for the separable LVM, the non-diagonal first heat kernel coefficient is approximated by

$$a_1(K,f) = Q(f_{\rm av}) - \frac{1}{4}\mathcal{G}(f_{\rm av}) \tag{5.34}$$

with $f_{\rm av} = \frac{K+f}{2}$. From the heat kernel expansion on a time-dependent manifold (4.82), the first-order conditional probability at time t is then

$$p(\tau,K|0,f) = \frac{1}{C(K)\sqrt{2\pi\tau}}\sqrt{\frac{C(f)}{C(K)}}e^{\frac{-\left(\int_K^f \frac{dx}{C(x)}\right)^2}{2\tau}}\left(1+\left(Q(f_{\rm av}) - \frac{\mathcal{G}\left(f_{\rm av}\right)}{4}\right)\tau\right)$$

By plugging this expression in (5.29) and using that

$$C(t,f)^2 = C(f)^2\left(1+t\mathcal{G}(f)\right) + o(t^2)$$

we obtain

$$\frac{\mathcal{C}(T,K|f_0)}{P_{0T}} = \max(f_0-K,0) + \tag{5.35}$$
$$\frac{\sqrt{C(K)C(f_0)}}{2}\int_0^T d\tau\frac{1}{\sqrt{2\pi\tau}}e^{\frac{-\left(\int_K^{f_0}\frac{dx}{C(x)}\right)^2}{2\tau}}\left(1+\left(Q(f_{\rm av})+\frac{3\mathcal{G}(f_{\rm av})}{4}\right)\tau\right)$$

The integration over τ can be performed and we obtain

PROPOSITION 5.2

The value of a European call option $\mathcal{C}(T,K|f_0)$, at the first-order in the maturity T, is for the local volatility model (5.27),

$$P_{0T}^{-1}\mathcal{C}(T,K|f_0) = \max(f_0-K,0) + \frac{\sqrt{C(K)C(f_0)T}}{2\sqrt{2\pi}} \tag{5.36}$$
$$\left(H_1(\omega) + \left(Q(f_{\rm av}) + \frac{3\mathcal{G}(f_{\rm av})}{4}\right)TH_2(\omega)\right)$$

with

$$H_1(\omega) = 2\left(e^{-\omega^2} + \sqrt{\pi\omega^2}(N(\sqrt{2}|\omega|)-1)\right)$$
$$H_2(\omega) = \frac{2}{3}\left(e^{-\omega^2}(1-2\omega^2) - 2|\omega|^3\sqrt{\pi}(N(\sqrt{2}|\omega|)-1)\right)$$
$$\omega = \int_{f_0}^K \frac{df'}{\sqrt{2T}C(f')}$$

with $Q(f)$ and $\mathcal{G}(f)$ given respectively by (5.32) and (5.33).

In the case of a constant volatility $C(t,f) = \sigma f$, the formula above reduces to

Example 5.5 Black-Scholes model

$$P_{0T}^{-1}\mathcal{C}(T,K|f_0) = \max(f_0-K,0) + \frac{\sqrt{Kf_0\sigma^2 T}}{2\sqrt{2\pi}}\left(H_1(\bar{\omega}) - \frac{\sigma^2 T}{8}H_2(\bar{\omega})\right) \tag{5.37}$$

with $\bar{\omega} = \frac{\ln(\frac{K}{f_0})}{\sqrt{2T}\sigma}$. □

By identifying the formula (5.36) with the same formula obtained with an implied volatility $\sigma = \sigma_{\mathrm{BS}}(K,T)$ (5.37), we deduce that the implied volatility $\sigma_{\mathrm{BS}}(K,T)$ at the first-order satisfies the non-linear equation

$$\sigma_{\mathrm{BS}}(K,T) = \frac{\sqrt{C(K)C(f_0)}}{\sqrt{Kf_0}}\frac{H_1(\omega)}{H_1(\bar{\omega})} \tag{5.38}$$
$$\left(1 + \left(Q(f_{\mathrm{av}}) + \frac{3\mathcal{G}(f_{\mathrm{av}})}{4}\right)T\frac{H_2(\omega)}{H_1(\omega)} + \frac{\sigma_{\mathrm{BS}}^2(K,T)T}{8}\frac{H_2(\bar{\omega})}{H_1(\bar{\omega})}\right)$$

with $\bar{\omega} = \frac{\ln(\frac{K}{f_0})}{\sqrt{2T}\sigma_{\mathrm{BS}}(K,T)}$. At the zero-order (i.e., independent of the maturity T), we obtain $\omega = \bar{\omega}$, i.e.,

$$\lim_{T\to 0}\sigma_{\mathrm{BS}}(T,K) = \frac{\ln\left(\frac{K}{f_0}\right)}{\int_{f_0}^{K}\frac{df'}{C(f')}} \tag{5.39}$$

The formula (5.39) has already been found in [59], [46] and we will call it the *BBF* relation in the following. Then using the recurrence equation (5.38), we obtain at the first-order

$$\sigma_{\mathrm{BS}}(K,T) = \frac{\ln(\frac{K}{f_0})}{\int_{f_0}^{K}\frac{df'}{C(f')}}\left(1 + \frac{T}{3}\left(\frac{1}{8}\left(\frac{\ln(\frac{K}{f_0})}{\int_{f_0}^{K}\frac{df'}{C(f')}}\right)^2 + Q(f_{\mathrm{av}}) + \frac{3\mathcal{G}(f_{\mathrm{av}})}{4}\right)\right) \tag{5.40}$$

This expression can be approximated by

$$\sigma_{\mathrm{BS}}(K,T) \simeq \frac{\ln(\frac{K}{f_0})}{\int_{f_0}^{K}\frac{df'}{C(f')}}(1 + \frac{C^2(f_{\mathrm{av}})T}{24}\left(2\frac{C''(f_{\mathrm{av}})}{C(f_{\mathrm{av}})} - \left(\frac{C'(f_{\mathrm{av}})}{C(f_{\mathrm{av}})}\right)^2 + \frac{1}{f_{\mathrm{av}}^2}\right)$$
$$+ T\frac{\partial_t C(0,f_{\mathrm{av}})}{2C(f_{\mathrm{av}})})$$

In the case of a time-homogeneous LV function $C(t,f) = C(f)$, we reproduce the asymptotic implied volatility obtained by [99].

REMARK 5.2 Local skew versus implied volatility skew From the BBF formula (5.39), we obtain that the *local skew* $\mathcal{S}_{\rm loc} \equiv f_0 \partial_f \sigma(f_0)$ ($\sigma(f) = \frac{C(f)}{f}$) is twice the implied volatility skew: $\mathcal{S}_{\rm loc} = 2\mathcal{S}$ as observed in Fig. 5.3.
□

Example 5.6 CEV model
For the CEV model, $C(f) = f^\beta$, $A(t) = \sigma$, the first-order asymptotic implied volatility is from (5.40)

$$\sigma_{\rm BS}(K,T) = \sqrt{\frac{T'}{T}} \frac{(1-\beta)\ln\frac{K}{f_0}}{K^{1-\beta} - f_0^{1-\beta}} \left(1 + \frac{(\beta-1)^2 T'}{24} f_{\rm av}^{2\beta-2}\right) \tag{5.41}$$

with $T' = \sigma^2 T$. In the case of the CEV model, an analytical formula for the fair value of a call option exists [76]: For $0 < \beta < 1$, we have

$$\frac{\mathcal{C}(T,K|f_0)}{P_{0T}} = f_0 \left(1 - \chi(a, b+2, c)^2\right) - K\chi(c,b,a) \tag{5.42}$$

and for $\beta > 1$

$$\frac{\mathcal{C}(T,K|f_0)}{P_{0T}} = f_0 \left(1 - \chi(c, -b, a)^2\right) - K\chi(a, 2-b, a)$$

The variables a, b, c are given by

$$a = \frac{K^{2(1-\beta)}}{(1-\beta)^2\sigma^2 T}\ , \ b = \frac{1}{1-\beta}\ , \ c = \frac{f_0^{2(1-\beta)}}{(1-\beta)^2\sigma^2 T}$$

and χ is the non-central Chi-squared distribution.
This exact solution is used to test the validity of our asymptotic implied volatility. In Fig. 5.4 and Fig. 5.5, the asymptotic implied volatility (5.40) is plotted against the exact solution (5.42). We obtain a good match between the exact and approximate solutions.

□

The formula (5.40) will be useful for deriving an implied volatility for a stochastic volatility model. More precisely, we will compute the LV associated to the stochastic volatility model (the square of the LV function is the mean value of the square of the stochastic volatility conditional to the spot) and finally we will use the relation (5.40) in order to obtain the corresponding implied volatility.
In the next section, we present an alternative method to derive an implied volatility from a LV. This second derivation will justify the approximation (5.34) that we have used to compute the non-diagonal heat kernel coefficients from the diagonal coefficients. A similar computation can be found in [46].

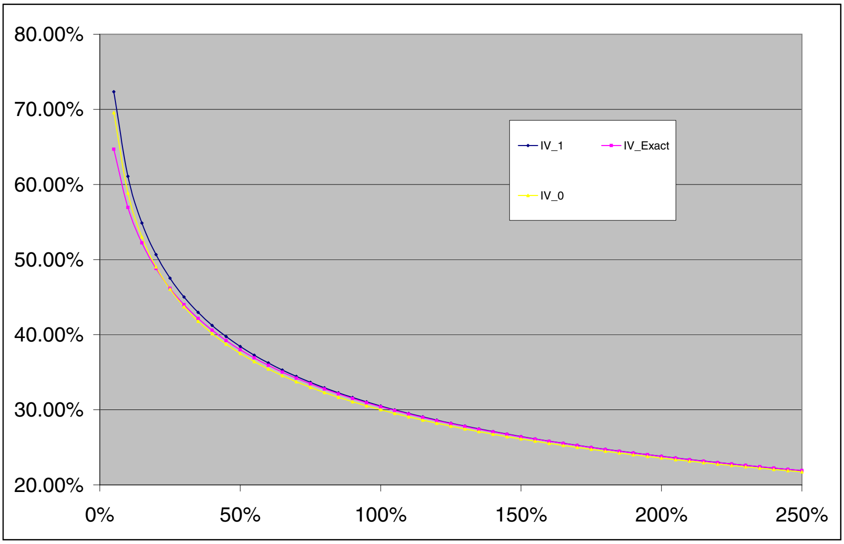

FIGURE 5.4: Comparison of the asymptotic implied volatility (5.41) at the zero-order (resp. first-order) against the exact solution (5.42). $f_0 = 1$, $\sigma = 0.3$, $\tau = 10$ years, $\beta = 0.33$.

5.2.2.1 Alternative method

We seek a small T expansion of the implied volatility of the form

$$\sigma_{\mathrm{BS}}(y, T) = \sum_{i=0}^{\infty} T^i \sigma_i(y) \tag{5.43}$$

with the moneyness $y \equiv \ln\left(\frac{K}{f_0}\right)$. Note that (5.43) only admits integer powers of T. Similarly, we do a Taylor expansion of the Dupire LV function around $T = 0$

$$\sigma_{\mathrm{loc}}(T, y) = \sum_{i=0}^{\infty} T^i \sigma_{i,\mathrm{loc}}(y) \tag{5.44}$$

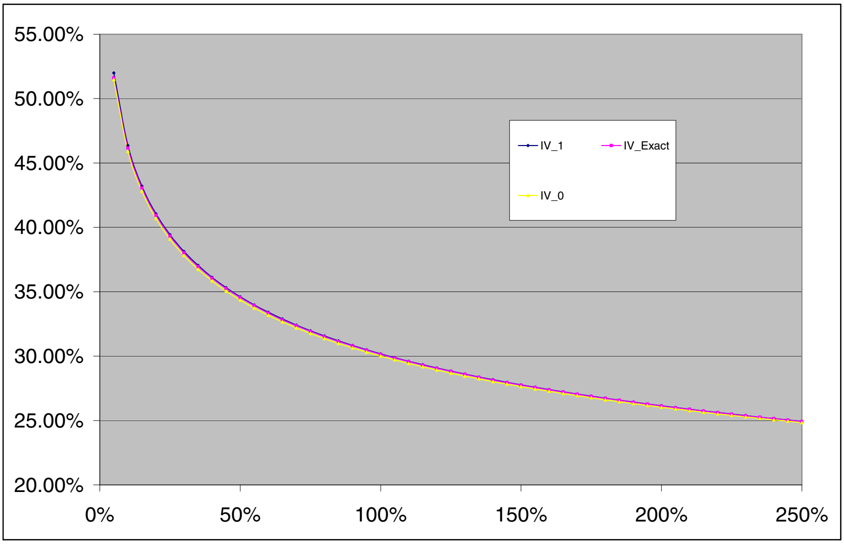

FIGURE 5.5: Comparison of the asymptotic implied volatility (5.41) at the zero-order (resp. first-order) against the exact solution (5.42). $f_0 = 1$, $\sigma = 0.3\%$, $\tau = 10$ years, $\beta = 0.6$.

We recall the Dupire formula (5.26) giving a map between the LV function and the implied volatility

$$\sigma_{\text{loc}}(T, y)^2 = \frac{\sigma_{\text{BS}}^2 + 2\sigma_{\text{BS}} T \partial_T \sigma_{\text{BS}}}{\left(1 - \frac{y}{\sigma_{\text{BS}}} \partial_y \sigma_{\text{BS}}\right)^2 - \frac{\sigma_{\text{BS}}^2 T^2}{4} \left(\partial_y \sigma_{\text{BS}}\right)^2 + T \sigma_{\text{BS}} \partial_y^2 \sigma_{\text{BS}}}$$

Substituting (5.43) and (5.44) into the equation above and matching term of order T^0 gives

$$\sigma_{0,\text{loc}}(y)^2 = \frac{\sigma_0(y)^2}{\left(1 - y \partial_y \ln \sigma_0(y)\right)^2}$$

Taking the square-root of the equation above and rearranging leads to the first-order ODE

$$\frac{1}{\sigma_{0,\text{loc}}(y)} = \partial_y \left(\frac{y}{\sigma_0(y)}\right) \tag{5.45}$$

Solving (5.45), subject to the boundary condition that the limit must be finite for y goes to zero, leads to

$$\sigma_0(y) = \frac{y}{\int_0^y \frac{dy'}{\sigma_{0,\mathrm{loc}}(y')}}$$

In the K coordinate, we have

$$\sigma_0(K) = \frac{\ln(\frac{K}{f_0})}{\int_{f_0}^K \frac{df'}{f'\sigma_{0,\mathrm{loc}}(f')}}$$

which is nothing but the BBF relation (5.39).
Next, matching the term T^1 and rearranging, we obtain the first-order ODE for the function $\phi(y) \equiv \frac{\sigma_1(y)}{\sigma_0(y)}$

$$y\partial_y\phi(y) + 2\frac{\sigma_0(y)}{\sigma_{0,\mathrm{loc}}(y)}\phi(y) = \frac{\sigma_{0,\mathrm{loc}}(y)}{2}\partial_y^2\sigma_0(y) + \frac{\sigma_0(y)\sigma_{1,\mathrm{loc}}(y)}{\sigma_{0,\mathrm{loc}}(y)^2}$$

Solving this first-order ODE, subject to the boundary condition that the limit must be finite for y goes to zero, leads to

$$\begin{aligned}\phi(y) = -\frac{1}{2}\left(\frac{\sigma_0(y)}{y}\right)^2 \\ \left(\ln\left(\frac{\sigma_0(y)^2}{\sigma_{0,\mathrm{loc}}(y)\sigma_{0,\mathrm{loc}}(0)}\right) - \int_0^y dy' \frac{\sigma_{1,\mathrm{loc}}(y')}{\sigma_{0,\mathrm{loc}}(y')}\partial_{y'}\left(\frac{y'}{\sigma_0(y')}\right)^2\right)\end{aligned}$$

Finally, in the forward coordinate, we obtain that the implied volatility at the first-order is given by

$$\begin{aligned}\sigma_{\mathrm{BS}}(K,T) = \sigma_0(K)\left(1 - \frac{T}{2\left(\int_{f_0}^K \frac{df'}{f'\sigma_{0,\mathrm{loc}}(f')}\right)^2}\right. \\ \left.\left(\ln\left(\frac{\sigma_0(K)^2}{\sigma_{0,\mathrm{loc}}(K)\sigma_{0,\mathrm{loc}}(f_0)}\right) - \int_{f_0}^K \frac{\sigma_{1,\mathrm{loc}}(f')}{\sigma_{0,\mathrm{loc}}(f')}\partial_{f'}\left(\frac{\ln\frac{f'}{f_0}}{\sigma_0(f')}\right)^2 df'\right)\right)\end{aligned}$$

Approximating the last integral by ($f_{\mathrm{av}} = \frac{f_0+K}{2}$)

$$\begin{aligned}\int_{f_0}^K \frac{\sigma_{1,\mathrm{loc}}(f')}{\sigma_{0,\mathrm{loc}}(f')}\partial_{f'}\left(\frac{\ln\frac{f'}{f_0}}{\sigma_0(f')}\right)^2 df' &\simeq \frac{\sigma_{1,\mathrm{loc}}(f_{\mathrm{av}})}{\sigma_{0,\mathrm{loc}}(f_{\mathrm{av}})}\int_{f_0}^K \partial_{f'}\left(\frac{\ln\frac{f'}{f_0}}{\sigma_0(f')}\right)^2 df' \\ &= \frac{\sigma_{1,\mathrm{loc}}(f_{\mathrm{av}})}{\sigma_{0,\mathrm{loc}}(f_{\mathrm{av}})}\left(\frac{\ln\frac{K}{f_0}}{\sigma_0(K)}\right)^2\end{aligned}$$

we obtain

$$\sigma_{\mathrm{BS}}(K,T) = \sigma_0(K) \left(1 - \frac{T}{2\left(\int_{f_0}^{K} \frac{df'}{f'\sigma_{0,\mathrm{loc}}(f')}\right)^2} \ln\left(\frac{\sigma_0(K)^2}{\sigma_{0,\mathrm{loc}}(K)\sigma_{0,\mathrm{loc}}(f_0)}\right) + \frac{T}{2}\frac{\sigma_{1,\mathrm{loc}}(f_{\mathrm{av}})}{\sigma_{0,\mathrm{loc}}(f_{\mathrm{av}})}\right)$$

As

$$\frac{1}{2\left(\int_{f_0}^{K} \frac{df'}{f'\sigma_{0,\mathrm{loc}}(f')}\right)^2} \ln\left(\frac{\sigma_0(K)^2}{\sigma_{0,\mathrm{loc}}(K)\sigma_{0,\mathrm{loc}}(f_0)}\right) \simeq \frac{C^2(f_{\mathrm{av}})}{24}\left(2\frac{C''(f_{\mathrm{av}})}{C(f_{\mathrm{av}})} - \left(\frac{C'(f_{\mathrm{av}})}{C(f_{\mathrm{av}})}\right)^2 + \frac{1}{f_{\mathrm{av}}^2}\right)$$

with $C(f) \equiv f\sigma_{0,\mathrm{loc}}(f)$, we reproduce our previous result (5.40) and therefore justify the approximations (5.19), (5.20) and (5.34).
In the next paragraph, following [18], we derive another expression for the implied volatility in terms of the LV function. This expression is simpler to use than (5.40) in order to study the large-time behavior of the implied volatility.

5.3 Implied volatility from local volatility

The PDEs satisfied by a European call option in the Black-Scholes and LV models are

$$\frac{\partial \mathcal{C}_{\mathrm{BS}}(t,f)}{\partial t} + \frac{1}{2}\sigma_{\mathrm{BS}}^2 f^2 \frac{\partial^2 \mathcal{C}_{\mathrm{BS}}(t,f)}{\partial f^2} = 0$$

$$\frac{\partial \mathcal{C}(t,f)}{\partial t} + \frac{1}{2}\sigma_{\mathrm{loc}}(t,f)^2 f^2 \frac{\partial^2 \mathcal{C}(t,f)}{\partial f^2} = 0$$

Setting $\Phi(t,f) \equiv \mathcal{C}(t,f) - \mathcal{C}_{\mathrm{BS}}(t,f)$ and

$$\Gamma^{\mathrm{BS}}(t,f) \equiv \frac{\partial^2 \mathcal{C}_{\mathrm{BS}}(t,f)}{\partial f^2}$$

the so-called Gamma of a European call option in the Black-Scholes model, we obtain that Φ satisfies

$$\frac{\partial \Phi(t,f)}{\partial t} + \frac{1}{2}\sigma_{\mathrm{loc}}(t,f)^2 f^2 \frac{\partial^2 \Phi(t,f)}{\partial f^2} + \frac{1}{2} f^2 \Gamma^{\mathrm{BS}}(t,f)\left(\sigma_{\mathrm{loc}}(t,f)^2 - \sigma_{\mathrm{BS}}^2\right) = 0 \tag{5.46}$$

subject to the terminal condition $\phi(T,f) = 0$ as the two options have the same payoff $\max(f_T - K, 0)$ at the maturity T. By using the Feynman-Kac theorem (with source), the solution at $t = 0$ of (5.46) can be represented in the integral form by

$$\Phi(0, f_0) = \frac{1}{2}\int_0^T dt \int_0^\infty df f^2 \left(\sigma_{\mathrm{loc}}(t,f)^2 - \sigma_{\mathrm{BS}}^2\right) \Gamma^{\mathrm{BS}}(t,f) p(t,f|f_0)$$

By definition of the implied volatility, we have

$$\Phi(0, f_0) = 0$$

and we obtain the following exact relation between the implied volatility and the LV function

$$\sigma_{\mathrm{BS}}^2(K,T) = \frac{\int_0^T \int_0^\infty f^2 \sigma_{\mathrm{loc}}(t,f)^2 \Gamma_{\mathrm{BS}}(t,f) p(t,f|f_0) dt df}{\int_0^T \int_0^\infty f^2 \Gamma_{\mathrm{BS}}(t,f) p(t,f|f_0) dt df} \tag{5.47}$$

By using this expression (5.47), we obtain the following approximation between the local and the implied volatilities

PROPOSITION 5.3

Assuming that in the forward measure $\mathbb{P}^T$, the forward satisfies the driftless process $df_t = f_t\left(\sigma + \epsilon(t, f_t)\right) dW_t$, then the implied volatility can be approximated at the first-order in ϵ by

$$\sigma_{\mathrm{BS}}(K,T)^2 \approx \sigma^2 + 2\sigma \int_0^1 dt \int_{-\infty}^{\infty} \frac{du}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \epsilon\left(tT, f_0 e^{t \ln \frac{K}{f_0} + \sqrt{t(1-t)}\sigma u}\right)$$

PROOF At the zero-order in ϵ, the conditional probability for the forward is a log-normal distribution

$$p(t,f|f_0) = \frac{1}{f\sqrt{2\pi t}\sigma} e^{-\frac{\left(\ln \frac{f}{f_0} + \frac{\sigma^2 t}{2}\right)^2}{2\sigma^2 t}}$$

The exact Black-Scholes Gamma is

$$\Gamma_{\mathrm{BS}}(t,f) = \frac{1}{f\sqrt{2\pi(T-t)}\sigma_{\mathrm{BS}}(K,T)} e^{-\frac{\left(\ln \frac{f}{K} + \frac{\sigma_{\mathrm{BS}}(K,T)^2(T-t)}{2}\right)^2}{2\sigma_{\mathrm{BS}}(K,T)^2(T-t)}}$$

By using that at the zero-order $\sigma_{\mathrm{BS}}(K,T) = \sigma$, the product of the Gamma and the conditional probability is

$$f^2 \Gamma_{\mathrm{BS}}(t,f) p(t,f|f_0) = \frac{1}{\sqrt{2\pi(T-t)}\sigma} e^{-\frac{\left(\ln \frac{f}{K} + \frac{\sigma^2(T-t)}{2}\right)^2}{2\sigma^2(T-t)}} \frac{1}{\sqrt{2\pi t}\sigma} e^{-\frac{\left(\ln \frac{f}{f_0} + \frac{\sigma^2 t}{2}\right)^2}{2\sigma^2 t}} \tag{5.48}$$

By plugging (5.48) into (5.47), we obtain at the first-order in ϵ

$$\sigma_{\rm BS}(K,T)^2 = \sigma^2 + 2\sigma \frac{\int_0^T dt \int_0^\infty df \frac{1}{\sqrt{t(T-t)}} \epsilon(t,f) e^{-\frac{\left(\ln\frac{f}{K}+\frac{\sigma^2(T-t)}{2}\right)^2}{2\sigma^2(T-t)}} e^{-\frac{\left(\ln\frac{f}{f_0}+\frac{\sigma^2 t}{2}\right)^2}{2\sigma^2 t}}}{\int_0^T dt \int_0^\infty df \frac{1}{\sqrt{t(T-t)}} e^{-\frac{\left(\ln\frac{f}{K}+\frac{\sigma^2(T-t)}{2}\right)^2}{2\sigma^2(T-t)}} e^{-\frac{\left(\ln\frac{f}{f_0}+\frac{\sigma^2 t}{2}\right)^2}{2\sigma^2 t}}}$$

Doing a change of variable $y = \ln\frac{f}{f_0}$, we get

$$\sigma_{\rm BS}(K,T)^2 = \sigma^2 + 2\sigma \frac{\int_0^T dt \int_{-\infty}^\infty dy \frac{1}{\sqrt{t(T-t)}} \epsilon(t,f_0 e^y) e^{-\left(\frac{1}{T-t}+\frac{1}{t}\right)\frac{\left(y-\frac{t}{T}\ln\frac{K}{f_0}\right)^2}{2\sigma^2}}}{\int_0^T dt \int_{-\infty}^\infty dy \frac{1}{\sqrt{t(T-t)}} e^{-\left(\frac{1}{T-t}+\frac{1}{t}\right)\frac{\left(y-\frac{t}{T}\ln\frac{K}{f_0}\right)^2}{2\sigma^2}}}$$

With a new change of variable from y to $u = \frac{1}{\sigma}\sqrt{\left(\frac{1}{T-t}+\frac{1}{t}\right)}\left(y - \frac{t}{T}\ln\frac{K}{f_0}\right)$, we get our result. ∎

Example 5.7 Skew averaging
We consider the following local volatility

$$\sigma_{\rm loc}(t,f) = \sigma + \lambda(t)\ln\frac{f}{f_0} \tag{5.49}$$

By using the proposition 5.3, we get the implied volatility at the first-order in λ

$$\sigma_{\rm BS}(K,T)^2 = \sigma^2 + \frac{2\sigma}{T^2}\ln\frac{K}{f_0}\int_0^T t\lambda(t)dt \tag{5.50}$$

We try to recover the same implied volatility using a time-independent local volatility

$$\sigma_{\rm loc}(t,f) = \sigma + \bar{\lambda}\ln\frac{f}{f_0} \tag{5.51}$$

$\bar{\lambda}$ can be interpreted as twice the skew $\mathcal{S}$: From the BBF formula (5.39), we have at the zero-order in the maturity T

$$\sigma_{\rm BS}(f,T) = \sigma + \bar{\lambda}\ln\frac{f_{\rm av}}{f_0}$$

and

$$\mathcal{S} \equiv K\partial_K \sigma_{\mathrm{BS}}(K,T)|_{K=f_0} = \frac{\bar{\lambda}}{2}$$

From (5.50), the implied volatilities associated to the LV models (5.49) and (5.51) are equivalent at the first-order in λ if and only if

$$\bar{\lambda} = \frac{2}{T^2}\int_0^T t\lambda(t)dt$$

This is called a *skew averaging* which was previously obtained in [136] using a different but close enough approach. $\square$

Exercises 5.1 Generalized skew averaging
Let us consider the LV model

$$df = f\left(\sigma + \sum_{i=1}^{n} \lambda_i(t)\left(\ln \frac{f}{f_0}\right)^i\right) dW_t$$

By using the proposition 5.3, prove that this model produces an equivalent implied at the first-order in λ_i that the time-homogeneous LV model defined by

$$df = f\left(\sigma + \sum_{i=1}^{n} \bar{\lambda}_i\left(\ln \frac{f}{f_0}\right)^i\right) dW_t$$

Specify the parameters $\bar{\lambda}_i$.

Chapter 6

Stochastic Volatility Models and Geometry of Complex Curves

The shortest path between two truths in the real domain passes through the complex domain.
— J. Hadamard

Abstract In this chapter we pursue the application of the heat kernel expansion for stochastic volatility models (SVM). In our geometric framework, a SVM corresponds to a complex curve, also called Riemann surfaces. By using the classification of conformal metrics on a Riemann surface, we show that SVMs fall into two classes. In particular, the SABR model corresponds to the Poincaré hyperbolic surface.
We derive the first-order asymptotics for implied volatility for any time-homogeneous SVM. This general formula, particularly useful for calibration purposes, reproduces and improves the well-known asymptotic implied volatility in the case of the SABR model. This expression only depends on the geometric objects (metric, connection) characterizing a specific SVM. We apply this formula to the SABR model with a mean-reverting drift and the Heston model.
Finally, in order to show the strength of our geometrical framework, we give an exact solution to the Kolmogorov equation for the normal (resp. log-normal) SABR model. The solutions are connected to the Laplacian heat kernel on the two-dimensional (resp. three-dimensional) hyperbolic surface.
Deterministic interest rates will be assumed throughout this chapter.

6.1 Stochastic volatility models and Riemann surfaces

6.1.1 Stochastic volatility models

As previously discussed, a LVM can automatically be calibrated to the initial implied volatility via the Dupire formula. However, the dynamics produced by the LVM is not very realistic. For example, the LVM process is not time-homogeneous and therefore the dynamics is not invariant by time translation.

An alternative to produce a better dynamics for the implied volatility is to introduce time-homogeneous SVMs.

A SVM is defined as a set of two correlated SDEs: one for the forward f_t and one for the instantaneous stochastic volatility a_t. As the latter variable is not directly observable on the market (this is not a traded asset), the market model is incomplete in opposition to the case of LVMs. As a consequence, the risk-neutral measure $\mathbb{P}$ is not unique here according to the theorem 2.8.

We define a general (one-factor) SVM in the forward measure $\mathbb{P}^T$ by the process

$$df_t = a_t C(f_t) dW_t \tag{6.1}$$

$$da_t = b(a_t)dt + \sigma(a_t)dZ_t \tag{6.2}$$

$$dW_t dZ_t = \rho dt$$

with Z_t and W_t two correlated Brownian motions and with the initial values $f_{t=0} = f_0$ and $a_{t=0} = \alpha$. σ is called by the practitioners the *volatility of volatility* (or in short vol of vol).

As usual W_t can be decomposed over a basis of uncorrelated Brownian motions Z_t, $Z_t^\perp$:

$$W_t = \rho Z_t + \sqrt{1-\rho^2} Z_t^\perp$$

Note that as a_t is not a traded asset, a_t is not a local martingale and we have a priori a drift term $b(a)$. We assume that $b(\cdot)$ and $\sigma(\cdot)$ are only functions of the volatility process a_t. We could assume that $b(\cdot)$ and $\sigma(\cdot)$ depend on the forward f_t as well but we do not pursue this route as examples we look at do not exhibit this dependence.

Below is a list of commonly used SVMs (see Table 6.1).

Non-explosion and martingality

As in section 5.1.2, we preserve the put-call parity and ensure the ability to price derivatives using the pricing formula (2.31) if f_t defines not only a local martingale but also a martingale. According to the theorem 5.2, f_t is a martingale if and only if the instantaneous (log-normal) volatility process $\xi_t = \frac{C(f_t)}{f_t} a_t$ does not explode under the measure $\mathbb{P}^f$ associated to the forward as numéraire.

Under $\mathbb{P}^T$, we have

$$\begin{aligned} d\xi &= aC(f)\partial_f \left(\frac{C(f)}{f}\right) dW + \frac{C(f)}{f}\sigma(a)dZ \\ &+ \left(\frac{C(f)}{f} b(a) + \partial_f \left(\frac{C(f)}{f}\right) C(f) a\rho\sigma(a) + \frac{a^3}{2} C(f)^2 \partial_f^2 \left(\frac{C(f)}{f}\right)\right) dt \end{aligned} \tag{6.3}$$

TABLE 6.1: Example of SVMs.

Name	SDE
Stein-Stein	$\frac{df_t}{f_t} = a_t dW_t$ $da_t = \lambda(a_t - \bar{a})dt + \zeta dZ_t \ , \ dW_t dZ_t = 0$
Geometric	$\frac{df_t}{f_t} = a_t dW_t$ $da_t = \lambda a_t dt + \zeta a_t \alpha dZ_t \ , \ dW_t dZ_t = \rho dt$
3/2-model	$\frac{df_t}{f_t} = a_t dW_t$ $da_t^2 = \lambda(a_t^2 - \bar{v} a_t^4)dt + \zeta a_t^3 \alpha dZ_t \ , \ dW_t dZ_t = \rho dt$
SABR	$\frac{df_t}{f_t} = a_t f_t^{\beta-1} dW_t$ $da_t = \nu a_t dZ_2 \ , \ dW_t dZ_t = \rho dt$
Scott-Chesney	$\frac{df_t}{f_t} = e^y f_t dW_t$ $dy = \lambda\left(e^y - \bar{y}\right)dt + \zeta dZ_t \ , \ dW_t dZ_t = 0$
Heston	$\frac{df_t}{f_t} = a_t dW_t$ $da_t^2 = \lambda(\bar{v} - a_t^2)dt + \zeta a_t dZ_t \ , \ dW_t dZ_t = \rho dt$

and under $\mathbb{P}^f$

$$\begin{aligned} d\xi &= aC(f)\partial_f \frac{C(f)}{f} dW^f + \frac{C(f)}{f}\sigma(a)dZ^f + \\ &\left(\frac{C(f)}{f}b(a) + \partial_f\left(\frac{C(f)}{f}\right)C(f)a\rho\sigma(a) + \frac{a^3}{2}C(f)^2\partial_f^2\left(\frac{C(f)}{f}\right)\right. \\ &\left. +\xi\left(aC(f)\partial_f\left(\frac{C(f)}{f}\right) + \rho\frac{C(f)}{f}\sigma(a)\right)\right)dt \end{aligned} \tag{6.4}$$

For the special case when $C(f) = f$ ($\xi = a$), the above-mentioned SDEs reduce to

$$da_t = b(a_t)dt + \sigma(a_t)dZ_t \ \text{ under } \mathbb{P}^T \tag{6.5}$$

$$da_t = \left(b(a_t) + \rho a_t \sigma(a_t)\right)dt + \sigma(a_t)dZ_t^f \ \text{ under } \mathbb{P}^f \tag{6.6}$$

In this case the Feller non-explosion criterion 5.3 can be used to check if the processes (6.5,6.6) do not explode. We apply this test to the log-normal SABR and Heston models.

Example 6.1 Log-normal SABR model
The log-normal SABR model [99] is defined by the following SDEs under $\mathbb{P}^T$

$$\begin{aligned} df_t &= a_t f_t dW_t \\ da_t &= \nu a_t dZ_t \end{aligned}$$

Using (6.6), the process $\xi_t = a_t$ follows under $\mathbb{P}^f$

$$da_t = \rho\nu a_t^2 dt + \nu a_t dZ_t^f \tag{6.7}$$

Therefore, we have

$$l_{\mathbb{P}^T}(a) = -\frac{1}{\nu^2}\left(\ln\frac{a}{c} - \frac{a}{c} + 1\right) \tag{6.8}$$

$$l_{\mathbb{P}^f}(a) = \frac{1}{\nu^2}\int_{\bar{\rho}c}^{\bar{\rho}a} e^{-v}dv \int_{\bar{\rho}c}^{v} u^{-2}e^{u}du \tag{6.9}$$

with $\bar{\rho} = \frac{2\rho}{\nu} \leq 0$. We deduce that a_t does not explode under $\mathbb{P}^T$ and $a = 0$ is an unattainable boundary. Moreover a_t does not explode under $\mathbb{P}^f$ if and only if $\bar{\rho} \leq 0$. In conclusion, the log-normal SABR model defines a martingale if and only if $\bar{\rho} \leq 0$. This condition is satisfied when we calibrate this model to the market implied volatility ($\rho < 0$ and $\nu > 0$). It corresponds to a negative skew. □

Example 6.2 Heston model
The Heston model [106] is defined by the following SDEs under $\mathbb{P}^T$

$$df_t = a_t f_t dW_t$$
$$da_t^2 = \lambda\left(\bar{v} - a_t^2\right) + \zeta a_t dZ_t$$

By using (6.6), the process $v_t \equiv a_t^2$ follows under $\mathbb{P}^f$

$$dv_t = ((-\lambda + \rho\zeta)v_t + \lambda\bar{v}) + \zeta\sqrt{v_t}dZ_t^f \tag{6.10}$$

Therefore, we have

$$l_{\mathbb{P}^T}(v) = \frac{1}{\zeta^2}\int_c^v e^{\frac{-2\lambda y}{\zeta^2}} y^{\frac{2\lambda\bar{v}}{\zeta^2}} dy \int_c^y e^{\frac{2\lambda z}{\zeta^2}} z^{-\frac{2\lambda\bar{v}}{\zeta^2}-1}dz$$
$$l_{\mathbb{P}^f}(v) = \frac{1}{\zeta^2}\int_c^v e^{\frac{2(-\lambda+\rho\zeta)y}{\zeta^2}} y^{\frac{2\lambda\bar{v}}{\zeta^2}} dy \int_c^y e^{-\frac{2(-\lambda+\rho\zeta)z}{\zeta^2}} z^{-\frac{2\lambda\bar{v}}{\zeta^2}-1}dz$$

As $l_{\mathbb{P}^T}(\infty) = \infty$, we deduce that a_t does not explode under $\mathbb{P}^T$. $a = 0$ is an unattainable boundary if and only if $2\lambda\bar{v} > \zeta^2$ as in this case $l_{\mathbb{P}^T}(0) = \infty$. Moreover a_t does not explode under $\mathbb{P}^f$ and f_t is a true positive martingale. □

In the general case (i.e., $C(f) \neq f$), it is difficult to check that the two-dimensional processes (6.3,6.4) do not explode as the Feller criterion is not applicable anymore. In this case, one can use the Hasminskii non-explosion test [38] which gives a sufficient criterion for a multi-dimensional diffusion process:

$$dx_t^i = b^i(t, x_t)dt + \sum_{j=1}^{m} \sigma_j^i(t, x_t)dW_t^j \ , \ i = 1, \cdots, n \tag{6.11}$$

with W_t^j being uncorrelated Brownian motions.

THEOREM 6.1 Hasminskii non-explosion test
Suppose that for each T, there exists an $r > 0$ and continuous functions $A_T : [r, \infty) \to (0, \infty)$ and $B_T : (r, \infty) \to (0, \infty)$ such that for $\rho > (2r)^{\frac{1}{2}}$, $0 \leq t \leq T$ and $\sqrt{\sum_{i=1}^{n}(x^i)^2} = \rho$,

$$A_T\left(\frac{\rho^2}{2}\right) \geq \sum_{i,j=1}^{n}\sum_{k=1}^{m} x^i \sigma_k^i \sigma_k^j x^j$$

$$\sum_{i,j=1}^{n}\sum_{k=1}^{m} x^i \sigma_k^i \sigma_k^j x^j B_T\left(\frac{\rho^2}{2}\right) \geq \sum_{i=1}^{n}\sum_{k=1}^{m} \sigma_k^i \sigma_k^i + 2\sum_{i=1}^{n} x^i b^i(t, x)$$

and

$$\int_r^{\infty} C_T(\rho)^{-1} d\rho \int_r^{\rho} C_T(\sigma) A_T(\sigma)^{-1} d\sigma = \infty$$

where

$$C_T(\rho) = e^{\int_r^{\rho} B_T(\sigma) d\sigma}$$

Then the process x_t (6.11) does not explode.

In our geometrical framework, a SVM corresponds to a two-dimensional Riemannian manifold, called also a Riemann surface. Surprisingly, there is a classification of Riemann surfaces, corresponding to a classification of SVMs.

6.1.2 Riemann surfaces

On a Riemann surface we can show that a metric can always be locally written in a neighborhood of a point (using the so-called isothermal coordinates)

$$g_{ij}(x) = e^{\phi(x)} \delta_{ij} \ , \ i, j = 1, 2 \tag{6.12}$$

and therefore a metric is locally conformally flat. The coordinates $x = \{x_i\}$ are called the isothermal coordinates (see example 4.8). Furthermore, two metrics on a Riemann surface, g_{ij} and h_{ij} (in local coordinates), are called conformally equivalent if there exists a (globally defined) smooth function $\phi(x)$ such that

$$g_{ij}(x) = e^{\phi(x)} h_{ij}(x)$$

The following theorem follows from the observations above:

THEOREM 6.2 Uniformization
Every metric on a simply connected Riemann surface[1] is conformally equivalent to a metric of constant scalar curvature R:

[1] The non-simply connected Riemann surfaces can also be classified by taking the double cover.

1. *$R = +1$: the Riemann sphere S^2.*

2. *$R = 0$: the complex plane $\mathbb{C}$.*

3. *$R = -1$: the upper half plane $\mathbb{H}^2 = \{z \in \mathbb{C} \mid \Im(z) > 0\}$, called also the Poincaré hyperbolic surface.*

By the uniformization theorem, all surfaces fall into one of these three types. We conclude that there are a priori three types of SVMs (modulo the conformal equivalence). If we discard the two-sphere S^2 corresponding to a compact manifold, we are left with two classes: the flat manifold $\mathbb{C}$ and the upper half-plane $\mathbb{H}^2$. The SVMs corresponding to $\mathbb{H}^2$ and $\mathbb{C}$ envelop these universality classes and thus are generic frameworks allowing all possible behaviors of implied volatility.
The metric associated to a SVM defined by (6.1,6.2) is (using (4.54))

$$\begin{aligned} ds^2 &= g_{ij} dx^i dx^j \\ &= \frac{2}{a^2(1-\rho^2)} \left(\frac{df^2}{C(f)^2} - 2\rho \frac{adf da}{C(f)\sigma(a)} + \frac{a^2 da^2}{\sigma(a)^2} \right) \end{aligned}$$

Let us introduce the variable $q(f) = \int_{f_0}^{f} \frac{df'}{C(f')}$ and $\xi(a) = \int_{\cdot}^{a} \frac{u}{\sigma(u)} du$. We have

$$ds^2 = \frac{2}{a^2(1-\rho^2)} \left(dq^2 - 2\rho dq d\xi + d\xi^2 \right)$$

Completing the square with the new coordinates

$$x = q(f) - \rho \xi(a) \tag{6.13}$$

$$y = (1-\rho^2)^{\frac{1}{2}} \xi(a) \tag{6.14}$$

the metric becomes in the coordinates $[x, y]$ (i.e., isothermal coordinates)

$$ds^2 = e^{\phi(y)} (dx^2 + dy^2) \tag{6.15}$$

with the conformal factor

$$F(y) \equiv e^{\phi(y)} = \frac{2}{a(y)^2(1-\rho^2)} \tag{6.16}$$

Note that this metric exhibits a Killing vector ∂_x as the conformal factor does not depend on the coordinate x. In the example 4.8, we have computed explicitly the geodesic distance for such metrics. We reproduce the result here:
The geodesic distance d between two points (x_1, y_1) and (x_2, y_2) is

$$d = \int_{y_1}^{y_2} \frac{F(y')dy'}{\sqrt{F(y') - C^2}} \tag{6.17}$$

TABLE 6.2: Example of metrics for SVMs. $\sim$ means modulo a multiplicative constant factor.

Name	Conformal factor	Scalar curvature
Geometric	$F(y) \sim y^{-2}$	$R = -1$
3/2-model	$F(y) \sim e^{\frac{-2y}{\sqrt{1-\rho^2}}}$	$R = 0$
SABR	$F(y) \sim y^{-2}$	$R = -1$
Heston	$F(y) \sim y^{-1}$	$R = -2a^{-2} < 0$

with the constant $C = C(x_1, y_1, x_2, y_2)$ determined by the equation

$$x_2 - x_1 = \int_{y_1}^{y_2} \frac{C}{\sqrt{F(y') - C^2}} dy' \tag{6.18}$$

Using the results from exercise 4.1, the Ricci and scalar curvatures are given by

$$\begin{aligned} R_{ij} &= -\frac{F(y)F''(y) - (F'(y))^2}{2F(y)^2}\delta_{ij} \\ R &= -\frac{F(y)F''(y) - (F'(y))^2}{F(y)^3} \end{aligned}$$

Plugging the expression (6.16) for F in the equations above, we obtain

$$\begin{aligned} R_{ij} &= \frac{\sigma(a)^2}{(1-\rho^2)a^3}\left(\frac{\sigma'(a)}{\sigma(a)} - \frac{2}{a}\right)\delta_{ij} \\ R &= \frac{\sigma(a)^2}{a}\left(\frac{\sigma'(a)}{\sigma(a)} - \frac{2}{a}\right) \end{aligned} \tag{6.19}$$

In particular, for a volatility of volatility given by $\sigma(a) = a^p$, we have

$$\begin{aligned} R_{ij} &= \frac{1}{(1-\rho^2)} a^{2p-4}(p-2)\delta_{ij} \\ R &= a^{2p-2}(p-2) \end{aligned} \tag{6.20}$$

and the conformal factor is

$$\begin{aligned} F(y) &= \frac{2(1-\rho^2)^{\frac{p-1}{2-p}} y^{\frac{2}{p-2}}}{(2-p)^{\frac{2}{2-p}}} \ , \ \forall p \neq 2 \\ &= \frac{2}{(1-\rho^2)} e^{\frac{-2y}{\sqrt{1-\rho^2}}} \ , \ p = 2 \end{aligned}$$

Table 6.2 shows the conformal factors and the curvatures associated to the SVMs listed in table 6.1.

The metric associated to the SABR model is the metric on $\mathbb{H}^2$ and the 3/2-model corresponds to the metric on $\mathbb{C} \approx \mathbb{R}^2$. The scalar curvature is always non-negative and therefore we do not have a SVM exhibiting the metric on S^2. A SVM is therefore connected to a non-compact Riemann surface.

REMARK 6.1 Heston model The most commonly used SVM, the Heston model, has a negative scalar curvature diverging at $a = 0$. This is a true singularity[2] and the manifold is not complete.
The heat kernel expansion explained in chapter 4 is no more applicable as long as the volatility process can reach the singularity $a = 0$. As seen in example 6.2, we see that $a = 0$ is not reached if and only if

$$2\lambda\bar{v} > \zeta^2$$

□

Cut-locus and Cartan-Hadamard manifold

In our discussion on the heat kernel, we have seen that the short-time behavior of the fundamental solution $p(t, x|y)$ to the heat kernel is valid when x and y are not on each other cut-locus. Surprising, for commonly used SVMs, the underlying manifold is a *Cartan-Hadamard manifold* for which the cut-locus is empty.

DEFINITION 6.1 Cartan-Hadamard manifold *Let M be a Riemann surface. If the scalar curvature[3] is non-positive then M is called a Cartan-Hadamard manifold. For such a manifold, the cut-locus is empty [22].*

It means that the map exp as defined in (4.51) realized a diffeomorphism of $\mathbb{R}^n$ to M. Two points on M can then be joined by a unique minimizing geodesic. As shown by the expression for the scalar curvature (6.19), a SVM is a Cartan-Hadamard manifold (assuming that there is no singularity at $a = 0$) if the volatility of volatility satisfies the following inequality

$$a\frac{\sigma'(a)}{\sigma(a)} \leq 2$$

In the next section, we explain the link between LV and SV models.

[2] The Heston model behaves as an (Euclidean) black hole.

[3] For a general n-dimensional manifold, the scalar curvature is replaced by the sectional curvature.

6.1.3 Associated local volatility model

For a generality purpose, we assume that the forward follows the driftless process (in the forward measure $\mathbb{P}^T$)

$$df_t = f_t \sigma_t dW_t \tag{6.21}$$

where σ_t is a $\mathcal{F}_t$ adapted process that can depend on time t, on the forward f_t and other Markov processes noted generally a_t. This model can be viewed as a general SVM. In our case, we have assumed that $\sigma_t = a_t \frac{C(f_t)}{f_t}$.
Applying the Itô-Tanaka formula on the payoff $\max(f_t - K, 0)$, we obtain

$$d \max(f_t - K, 0) = 1(f_t - K) f_t \sigma_t dW_t + \frac{1}{2} f_t^2 \sigma_t^2 \delta(f_t - K) dt$$

Then, taking the mean value operator $\mathbb{E}^{\mathbb{P}^T}[\cdot|\mathcal{F}_0]$ on both sides of this equation, we have

$$d\mathbb{E}^{\mathbb{P}^T}[\max(f_t - K, 0)|\mathcal{F}_0] = \frac{K^2}{2} \mathbb{E}^{\mathbb{P}^T}[\sigma_t^2 \delta(f_t - K)|\mathcal{F}_0] dt \tag{6.22}$$

We recall that for the Dupire LVM, we have found in the previous chapter

$$d\mathbb{E}^{\mathbb{P}^T}[\max(f_t - K, 0)|\mathcal{F}_0] = \frac{K^2 \sigma_{\text{loc}}^2(t, K)}{2} \mathbb{E}^{\mathbb{P}^T}[\delta(f_t - K)|\mathcal{F}_0] dt \tag{6.23}$$

Comparing (6.22) and (6.23), we obtain that the square of the Dupire local volatility function is equal to the mean of the square of the stochastic volatility when the forward is fixed to the strike

$$\begin{aligned} \sigma_{\text{loc}}^2(t, K) &= \frac{\mathbb{E}^{\mathbb{P}^T}[\sigma_t^2 \delta(f_t - K)|\mathcal{F}_0]}{\mathbb{E}^{\mathbb{P}^T}[\delta(f_t - K)|\mathcal{F}_0]} \\ &\equiv \mathbb{E}^{\mathbb{P}}[\sigma_t^2 | f_t = K] \end{aligned} \tag{6.24}$$

This derivation obtained by Dupire in [81] can be put on a rigorous ground and we obtain

THEOREM 6.3 Gyöngy [98]
Let $X_t = \{X_t^i\}_{i=1,\cdots n}$ be an n-dimensional Itô process

$$dX_t^i = b_t^i dt + \sum_{i=1}^m \sigma_{t,j}^i dW_t^j$$

where b_t^i and $\sigma_t \equiv [\sigma_j^i]$ are bounded $\mathcal{F}_t$-adapted processes. We assume that σ_t satisfies the uniform ellipticity condition

$$\exists C \in \mathbb{R}_+^* \text{ such as } \sigma_t \sigma_t^\dagger \geq C1$$

with $\dagger$ *the transpose and* 1 *the* $m \times m$ *identity matrix.*
Then there exists bounded measurable functions $\Sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathrm{M}_{m,n}(\mathbb{R})$ *and* $B : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ *defined* $\forall (t,x) \in \mathbb{R}_+ \times \mathbb{R}^n$ *by*

$$\Sigma(t,x)\Sigma(t,x)^\dagger = \mathbb{E}[\sigma_t \sigma_t^\dagger | x_t = x]$$
$$B(t,x) = \mathbb{E}[b_t | x_t = x]$$

such that the following SDE

$$dx_t^i = B^i(t,x_t)dt + \sum_{j=1}^m \Sigma(t,x_t)^i_j dW^j \ , \ x_0 = X_0$$

has a weak solution with the same one-dimensional marginals as X_t*.*

REMARK 6.2 For most commonly used SVMs, the boundness and ellipticity conditions are not satisfied. This technical complication will be overlooked in the following. □

The formula (6.24) involving the Dirac function means

$$\mathbb{E}^{\mathbb{P}}[\sigma_t^2 | f_t = K] \equiv \lim_{\epsilon \to 0^+} \frac{\mathbb{E}^{\mathbb{P}}[\sigma_t^2 1(f_t - K \in (-\epsilon, \epsilon))]}{\mathbb{E}^{\mathbb{P}}[1(f_t - K \in (-\epsilon, \epsilon))]}$$

with $1(\cdot)$ the Heaviside function. Assuming that $\sigma_t = \sigma(t, f_t, a_t)$ with a_t a Itô process, then by definition of the conditional mean value, the expression above becomes

$$\sigma_{\text{loc}}(t,K)^2 = \frac{\int_0^\infty \sigma(t,K,a)^2 p(t,K,a|f_0,\alpha) da}{\int_0^\infty p(t,K,a|f_0,\alpha) da}$$

where $p(t,K,a|f_0,\alpha)$ is the conditional probability associated to the process (6.21). Therefore we have shown that a SVM can be calibrated exactly to the initial implied volatility if the stochastic volatility satisfies (6.24). Equivalently, a LV model and a SV model, which satisfies the relation (6.24), have the same marginals.
For example, let us assume that $\sigma_t = a_t \frac{C(t,f_t)}{f_t}$ with a_t a Itô process and $C(t,f_t)$ a function depending on the forward and the time. The implied volatility is matched if $C(t,f_t)$ satisfies

$$K^2 \sigma_{\text{loc}}(t,K)^2 = C(t,K)^2 \mathbb{E}^{\mathbb{P}^T}[a_t^2 | f_t = K]$$

This model, although automatically calibrated to the IV with

$$C(t,K)^2 = \frac{K^2 \sigma_{\text{loc}}(t,K)^2}{\mathbb{E}^{\mathbb{P}^T}[a_t^2 | f_t = K]}$$

without specifying the dynamics for the process a_t, is however not time-homogeneous. Consequently we impose that a general SVM is defined by the processes (6.1) and (6.2).
In the next paragraph, using our relation between a LVM and a SVM, we obtain an asymptotic implied volatility for a general SVM defined by (6.1) and (6.2). Before explaining the derivation, we state the result.

6.1.4 First-order asymptotics of implied volatility

The general asymptotic implied volatility at the first-order for any time-homogeneous SVM, depending implicitly on the metric g_{ij} (4.54) and the connection $\mathcal{A}_i$ (4.58) on our Riemann surface, is given by

$$\sigma_{\mathrm{BS}}(K,T) = \frac{\ln\frac{K}{f_0}}{\int_{f_0}^{K} \frac{df'}{\sqrt{2g^{ff}(a_{\min})}}} \Bigg(1 \tag{6.25}$$
$$+\frac{g^{ff}(a_{\min})T}{12}\left(-\frac{3}{4}\left(\frac{\partial_f g^{ff}(a_{\min})}{g^{ff}(a_{\min})}\right)^2 + \frac{\partial_f^2 g^{ff}(a_{\min})}{g^{ff}(a_{\min})} + \frac{1}{f_{\mathrm{av}^2}}\right)$$
$$+\frac{g^{ff'}(a_{\min})T}{2g^{ff}(a_{\min})\phi''(a_{\min})}\left(\ln(\Delta g \mathcal{P}^2)'(a_{\min}) - \frac{\phi'''(a_{\min})}{\phi''(a_{\min})} + \frac{g^{ff''}(a_{\min})}{g^{ff'}(a_{\min})}\right)\Bigg)$$

with $f_{\mathrm{av}} \equiv \frac{f_0+K}{2}$ and with $a_{\min}$ the volatility a which minimizes the geodesic distance $d(a, f_{\mathrm{av}}|\alpha, f_0)$ on the Riemann surface ($\phi = d^2$). Δ is the Van Vleck-Morette determinant (4.68), g is the determinant of the metric and $\mathcal{P}$ is the parallel gauge transport (4.69). The prime symbol $'$ indicates a derivative according to a. The formula (6.25) is particularly useful as we can use it to rapidly calibrate any SVM. In section 6.3 (resp. 6.5), we apply it to the λ-SABR model (resp. Heston model). In order to use the formula, the only computation needed is the calculation of the geodesic distance. This was achieved for our class of SVMs in equation (6.17).

PROOF A time-homogeneous SVM is coded by a metric g on a Riemann surface, an Abelian connection $\mathcal{A}$ and a section Q. This is the necessary data to apply the heat kernel expansion and deduce from it the asymptotic formula for the probability density at the first-order. We obtain ($x = (f_0, \alpha)$, $y = (f, a)$)

$$p(\tau, x|y) = \frac{\sqrt{g(y)}}{(4\pi\tau)}\sqrt{\Delta(y,x)}\mathcal{P}(y,x)e^{-\frac{d^2(x,y)}{4\tau}}\left(1 + a_1(y,x)\tau\right) \tag{6.26}$$

The computation of an asymptotic expression for the implied volatility involves two steps. The first step consists in computing the LV function $\sigma(T, f)$ associated to the SVM. In the second step, we deduce the implied volatility from the LV function.

As seen previously (6.24), the mean value of the square of the SV when the forward is fixed is the LV given by

$$\begin{aligned}\sigma(T,f)^2 &= C(f)^2\mathbb{E}^{\mathbb{P}^T}\left[a_T^2|f_T=f\right]\\ &= f^2\sigma_{\mathrm{loc}}(T,f)^2\end{aligned}$$

By definition of the mean value, we have

$$\sigma(T,f)^2 = \frac{2\int_0^\infty g^{ff}(f,a)p(T,f_0,\alpha|f,a)da}{\int_0^\infty p(T,f_0,\alpha|f,a)da} \tag{6.27}$$

$p(T,x|y)$ is the conditional probability given in the short-time limit at the first-order by (6.26). We set $\phi(x,y)=d(x,y)^2$. By plugging our asymptotic expression for the conditional probability (6.26) in (6.27), we obtain

$$\sigma(T,f)^2 = \frac{\int_0^\infty f(T,a)e^{\epsilon\phi(a)}da}{\int_0^\infty h(T,a)e^{\epsilon\phi(a)}da} \tag{6.28}$$

with $\phi(a)=d^2(x,y)$, $h(T,a)=\sqrt{g}\sqrt{\Delta(x,y)}\mathcal{P}(y,x)(1+a_1(y,x)T)$, $f(T,a)=2h(T,a)g^{ff}$ and $\epsilon=-\frac{1}{4T}$. Using the saddle-point method (see appendix A), we find an asymptotic expression for the LV:
At the zero-order, σ^2 is given by

$$\sigma(0,f)^2 = 2g^{ff}(a_{\min})$$

with $a_{\min}$ the stochastic volatility which minimizes the geodesic distance on our Riemann surface:

$$a_{\min} \equiv a \,|\, \min_a \phi(a) \tag{6.29}$$

At the first-order, we find the following expression for the numerator in (6.28) (see Appendix A for a sketch of the proof)

$$\begin{aligned}\int_0^\infty f(T,a)e^{\epsilon\phi(a)}da = &\sqrt{\frac{2\pi}{-\epsilon\phi''(a_{\min})}}f(T,a_{\min})e^{\epsilon\phi(a_{\min})}\,(1\\ &+\frac{1}{\epsilon}\left(-\frac{f''(0,a_{\min})}{2f(0,a_{\min})\phi''(a_{\min})}+\frac{\phi^{(4)}(a_{\min})}{8\phi''(a_{\min})^2}\right.\\ &\left.+\frac{f'(0,a_{\min})\phi'''(a_{\min})}{2f(0,a_{\min})\phi''(a_{\min})^2}-\frac{5(\phi'''(a_{\min}))^2}{24(\phi''(a_{\min}))^3}\right))\end{aligned}$$

The prime symbol $'$ indicates a derivative according to a. Computing the denominator in (6.28) in a similar way, we obtain the first-order correction to the LV

$$\begin{aligned}\sigma(T,f)^2 = &\,2g^{ff}(a_{\min})\,(1\\ &+\frac{1}{\epsilon}\left(-\frac{1}{2\phi''(a_{\min})}\left(\frac{f''(a_{\min})}{f(a_{\min})}-\frac{h''(a_{\min})}{h(a_{\min})}\right)\right.\\ &\left.+\frac{\phi'''(a_{\min})}{2\phi''(a_{\min})^2}\left(\frac{f'(a_{\min})}{f(a_{\min})}-\frac{h'(a_{\min})}{h(a_{\min})}\right)\right))\end{aligned}$$

Plugging the expression for f and g, we finally obtain

$$\sigma(T,f) = \sqrt{2g^{ff}(a_{\min})}\,(1+$$
$$\frac{T}{\phi''(a_{\min})}\left(\frac{g^{ff'}(a_{\min})}{g^{ff}(a_{\min})}\left(\ln(\Delta g\mathcal{P}^2)'(a_{\min}) - \frac{\phi'''(a_{\min})}{\phi''(a_{\min})}\right) + \frac{g^{ff''}(a_{\min})}{g^{ff}(a_{\min})}\right)\Bigg)$$

This expression depends solely on the metric and the connection $\mathcal{A}$ on our Riemann surface and not on the first heat kernel coefficient $a_1(y,x)$ the contribution of which has disappeared when we have taken the ratio of the numerator and denominator.
The final step is to use the (asymptotic) relation between a LV function $\sigma(t,f)$ and the implied volatility (5.40) that we have obtained in the previous chapter using the heat kernel expansion on a time-dependent one-dimensional real line. This gives (6.25). □

REMARK 6.3 Boundary conditions Note that in our proof, we have disregarded the boundary conditions. In the heat kernel expansion, these boundary conditions only affect the heat kernel coefficients [92]. As a result, our formula (6.25) does not depend on some specific boundary conditions. □

Computation of the saddle-point $a_{\min}$

By definition, the saddle-point $a_{\min}$ minimizes the geodesic distance when the forward and the strike are fixed: $\frac{\partial d(a_{\min})}{\partial a} = 0$. From the exact expression (6.17) for the geodesic distance, we obtain

$$\frac{F(y_{\min})}{\sqrt{F(y_{\min}) - C(y_{\min})^2}} + \int_{y_0}^{y_{\min}} \frac{F(y)}{(F(y) - C(y_{\min}))^{\frac{3}{2}}} C(y_{\min})\partial_y C(y_{\min}) = 0 \tag{6.30}$$

where $y_{\min} = \sqrt{1-\rho^2}\zeta(a_{\min})$. Differentiating the equation (6.18) with respect to a, we get

$$\frac{C(y_{\min})}{\sqrt{F(y_{\min}) - C(y_{\min})^2}} + \partial_y C(y_{\min})\left(\int_{y_0}^{y_{\min}} \frac{F(y)dy}{(F(y) - C(y_{\min}))^{\frac{3}{2}}}\right) = -\frac{\rho}{\sqrt{1-\rho^2}}$$

Using (6.30), the equation above simplifies and reduces to

$$C(y_{\min})^2 = F(y_{\min})\left(1-\rho^2\right)$$

and therefore from (6.18), $a_{\min}$ satisfies the following equation

$$q(K) - \rho\left(\zeta(a_{\min}) - \zeta(\alpha)\right) = \pm\frac{a_{\min}^2}{\sqrt{1-\rho^2}} \int_{\frac{\sqrt{1-\rho^2}\alpha}{a_{\min}}}^{\sqrt{1-\rho^2}} \frac{u^2}{\sigma\left(\frac{a_{\min}}{\sqrt{1-\rho^2}}u\right)\sqrt{1-u^2}} du \tag{6.31}$$

6.2 Put-Call duality

Let us have a look at the general duality that the implied volatility associated to a SVM should satisfy. It turns out that this relation allows to constrain the asymptotic series (in time) for the implied volatility.
In the forward measure $\mathbb{P}^T$, a European call option, with strike K and maturity T, is given by

$$\mathcal{C}(T, K|f_0) = P_{0T}\mathbb{E}^{\mathbb{P}^T}[\max(f_T - K, 0)]$$

which can be rewritten as

$$\mathcal{C}(T, K|f_0) = P_{0T}K\mathbb{E}^{\mathbb{P}^T}[f_T \max(\frac{1}{K} - \frac{1}{f_T}, 0)] \tag{6.32}$$

Using the measure $\mathbb{P}^f$ associated to the martingale f_t (6.1), the expression above becomes

$$\mathcal{C}(T, K|f_0) = P_{0T}f_0K\mathbb{E}^{\mathbb{P}^f}[\max(\frac{1}{K} - \frac{1}{f_T}, 0)] \tag{6.33}$$

Setting $X_t = \frac{1}{f_t}$ and applying Itô's lemma, we have that X_t satisfies in the measure $\mathbb{P}^T$

$$\frac{dX_t}{X_t} = a_t^2\left(\frac{C(f_t)}{f_t}\right)^2 dt - a_t\frac{C(f_t)}{f_t}dW_t$$

Thanks to a change of measure, X_t satisfies the SDE in the measure $\mathbb{P}^f$

$$\begin{aligned} dX_t &= -a_tX_t^2C(\frac{1}{X_t})dW_t \\ da_t &= \left(b(a_t) + X_tC(\frac{1}{X_t})\rho a_t\sigma(a_t)\right)dt + \sigma(a_t)dZ_t \end{aligned}$$

It does not come as a surprise that X_t is a martingale: As the product of X_t with f_t is 1, X_t is driftless in the measure associated to the numéraire f_t (i.e., $\mathbb{P}^f$).
The forementioned SDEs for X_t and a_t define a SVM, dual to the SVM defined by (6.1, 6.2). It will be called the SV model II, and (6.1,6.2) the SV model I. From the relation (6.33), we obtain

$$\mathcal{C}^{\mathrm{I}}(T, K|f_0) = f_0K\mathcal{P}^{\mathrm{II}}(T, \frac{1}{K}|\frac{1}{f_0}) \tag{6.34}$$

with $\mathcal{C}(T, K|f_0)$ (resp. $\mathcal{P}(T, K|f_0)$) the fair value of a call (resp. put) option. From the Black-Scholes formula (3.1), we deduce that the implied volatilities for the SV models I and II are the same when the strike K and the forward f_0 are inverted

$$\sigma_{\mathrm{BS}}^{\mathrm{I}}(T, K|f_0) = \sigma_{\mathrm{BS}}^{\mathrm{II}}(T, \frac{1}{K}|\frac{1}{f_0}) \tag{6.35}$$

REMARK 6.4 Zero-correlation case When the correlation is zero and $C(f) = f$ (i.e., conditional on a, we have a log-normal process for the forward), the processes f_t and X_t satisfy the same SDE (except that dW_t is changed into $-dW_t$). Moreover, in this case, the implied volatility is a function of the moneyness $y = \ln(\frac{K}{f_0})$ as shown in 3.2. Therefore, the relation (6.35) reduces to

$$\sigma_{\mathrm{BS}}(T, y) = \sigma_{\mathrm{BS}}(T, -y) \tag{6.36}$$

meaning that the implied volatility is a symmetric function in y. $\square$

REMARK 6.5 Put-Call duality for LVMs For a LVM ($a = 1$), the put-call duality becomes

$$\sigma_{\mathrm{BS}}^{\mathrm{I}}(T, K|f_0) = \sigma_{\mathrm{BS}}^{\mathrm{II}}(T, \frac{1}{K}|\frac{1}{f_0})$$

where I (resp. II) is associated to the LV model $df_t = C(f_t)dW_t$ (resp. $df_t = f_t^2 C(\frac{1}{f_t})dW_t$). As a check of our asymptotic relation between a local volatility and an implied volatility, we can show that the relation (5.40) is preserved under the put-call duality. In particular this is the case for the BBF relation (5.39) and the potential Q. More generally, the implied volatility asymptotics series depend on quantities invariant under the (projective) transformation

$$C(f) \to f^2 C(\frac{1}{f})$$
$$f_0, K \to \frac{1}{f_0}, \frac{1}{K}$$

The invariant projective functions have been classified in [2] and are

$$I_2 = 2C''C - C'^2\,,\, I_3 = 2C'''C^2\,, \cdots, I_n = CI'_{n-1}$$

I_2 is precisely the potential Q. $\square$

REMARK 6.6 Put-Call symmetry for LVMs We assume that the local volatility satisfies the relation

$$f^2 C(\frac{1}{f}) = C(f) \tag{6.37}$$

In this case, models I and II are identical and we obtain the so-called *put-call symmetry*

$$\mathcal{C}^{\mathrm{I}}(T,K|f_0) = f_0 K \mathcal{P}^{\mathrm{I}}(T,\frac{1}{K}|\frac{1}{f_0}) \tag{6.38}$$

As an application, we can derive a closed form formula for a *down-and-out call* option, $DO(T,K,B|f_0)$, in this model. We recall that a down-and-out call pays a call option with strike K at maturity T if the barrier level B is not reached by the asset during the interval $[0,T]$. Here, we take $f_0 > B$ and $K > B$. We have

$$DO(T,K,B|f_0) = \mathcal{C}^{\mathrm{I}}(T,K|f_0) - \frac{f_0}{B}\mathcal{C}^{\mathrm{I}}(T,K|\frac{B^2}{f_0})$$

The reader can check as an exercise (see the similar exercise 2.8) that

$$DO(t,f) \equiv \mathcal{C}^{\mathrm{I}}(T,K|t,f) - \frac{f}{B}\mathcal{C}^{\mathrm{I}}(T,K|t,\frac{B^2}{f})$$

is indeed a solution of the Black-Scholes PDE:

$$\partial_t DO + \frac{1}{2}C(f)^2\partial_f DO = 0 \ , \ \forall f > B$$

subject to $C(\cdot)$ satisfying (6.37). Moreover the boundary conditions

$$DO(T,f) = \max(K-f,0)1_{f>B}$$

and $DO(t,B) = 0$ are satisfied: Indeed, if the asset does not reach the barrier B then the second call $\mathcal{C}^{\mathrm{I}}(T,K|\frac{B^2}{f_0})$ is worthless. Furthermore, if the asset reaches the barrier, the two calls cancel.
Finally, using (6.38), we obtain a static replication (i.e., model-independent modulo the relation (6.37)) for the value of a down-and-out call

$$DO(T,K,B|f_0) = \mathcal{C}^{\mathrm{I}}(T,K|f_0) - \frac{K}{B}\mathcal{P}^{\mathrm{I}}(T,\frac{B^2}{K}|f_0)$$

□

6.3 λ-SABR model and hyperbolic geometry

We introduce the λ-SABR SVM. This model corresponds to the hyperbolic surface $\mathbb{H}^2$. By applying the formula (6.25), we obtain an asymptotic implied volatility.

6.3.1 λ-SABR model

The volatility a_t is not a tradable asset. Consequently, a_t can have a drift in the forward measure $\mathbb{P}^T$. A popular choice is to make the volatility process mean-reverting. We investigate the λ-SABR model defined by the following SDE (this model has also been considered in [60] with $C(f) = f$)

$$\begin{aligned} df_t &= a_t C(f_t) dW_t \\ da_t &= \lambda(a_t - \bar{\lambda})dt + \nu a_t dZ_t \\ C(f) &= f^\beta \ , \ a_0 = \alpha \ , f_{t=0} = f_0 \end{aligned}$$

where W_t and Z_t are two Brownian processes with correlation $\rho \in (-1, 1)$. The model depends on 6 (constant) parameters: α, β, λ, $\bar{\lambda}$, ν and ρ. For $\lambda = 0$, the λ-SABR model degenerates into the SABR model introduced in [100].

Conditional to the volatility a_t, f_t is a CEV process. From lemma 5.1, we have that for $\beta < 1$, f_t can reach 0 with a positive probability. In order to preserve the martingality condition ($\beta \leq 1/2$) we assume an absorbing boundary condition at 0 as usual.

In the following section, we present our asymptotic smile for the λ-SABR model and postpone the derivation to the next section.

6.3.2 Asymptotic implied volatility for the λ-SABR

The asymptotic implied volatility (with strike f, maturity date T and forward f_0) at the first-order associated to the stochastic λ-SABR model is

$$\sigma_{\mathrm{BS}}(f, T) = \frac{\ln \frac{f}{f_0}}{\mathrm{vol}(f)} \left(1 + \sigma_1 \left(\frac{f + f_0}{2} \right) T \right) \tag{6.39}$$

with

$$\begin{aligned} \sigma_1(f) = &\frac{(C(f) a_{\min}(f))^2}{24} \left(\frac{1}{f^2} + \frac{2\partial_{ff} \left(C(f) a_{\min}(f) \right)}{C(f) a_{\min}(f)} - \left(\frac{\partial_f (C(f) a_{\min}(f))}{C(f) a_{\min}(f)} \right)^2 \right) \\ &+ \frac{\alpha \nu^2 \ln(\mathcal{P})'(f)(1 - \rho^2) \sinh(d(f))}{2d(f)} \end{aligned}$$

with

$$
\begin{aligned}
\mathrm{vol}(f) &= \frac{1}{\nu}\ln\left(\frac{q\nu+\alpha\rho+\sqrt{\alpha^2+q^2\nu^2+2q\alpha\nu\rho}}{\alpha(1+\rho)}\right)\\
a_{\min}(f) &= \sqrt{\alpha^2+2\alpha\nu\rho q+\nu^2q^2}\\
d(f) &= \cosh^{-1}\left(\frac{-q\nu\rho-\alpha\rho^2+a_{\min}(f)}{\alpha(1-\rho^2)}\right)\\
q &= \frac{f^{(1-\beta)}-f_0^{(1-\beta)}}{(1-\beta)}\ (\beta\neq 1);\ ,\ q=\ln\frac{f}{f_0}\ (\beta=1)
\end{aligned}
$$

Besides, we have

$$
\begin{aligned}
\ln\left(\frac{\mathcal{P}}{\mathcal{P}^{\mathrm{SABR}}}\right)'(f) &= \frac{\lambda}{\nu^2}\left(\frac{\bar{\lambda}(\alpha-a_{\min})}{a_{\min}\alpha}+\ln\frac{a_{\min}}{\alpha}\right. \\
&\left.-\frac{\rho}{\sqrt{1-\rho^2}}\left(G_0(\theta_2,A_0)\theta_2'-G_0(\theta_1,A_0)\theta_1'+A_0'\left(G_1(\theta_2)-G_1(\theta_1)\right)\right)\right)
\end{aligned}
\tag{6.40}
$$

with

$$
\begin{aligned}
\ln\left(\mathcal{P}^{\mathrm{SABR}}\right)'(f) &= -\frac{\beta\rho}{2\sqrt{1-\rho^2}(1-\beta)}\left(F_0(\theta_2,A,B)\theta_2'\right.\\
&\left.-F_0(\theta_1,A,B)\theta_1'-A'\left(F_1(\theta_2,A,B)-F_1(\theta_1,A,B)\right)\right)
\end{aligned}
$$

and with

$$
\begin{aligned}
G_1(x) &= \left(\ln\left(\tan\frac{x}{2}\right)\right)\\
G_0(x,a) &= 1+\frac{a}{\sin x}\\
A_0 &= -\frac{\bar{\lambda}\sqrt{1-\rho^2}}{a_{\min}}\ ,\ A_0'=\frac{\bar{\lambda}\sqrt{1-\rho^2}(\alpha\rho+\nu q)}{a_{\min}^2(\nu q+\rho(\alpha-a_{\min}))}\\
F_0(x,a,b) &= \frac{\cos(x)}{a+\cos(x)+b\sin(x)}\\
F_1(x,a,b) &= \int^x\frac{\cos(\theta)}{\left(a+\cos(\theta)+b\sin(\theta)\right)^2}d\theta\\
\tan\theta_2 &= -\frac{\alpha\sqrt{1-\rho^2}}{\alpha\rho+\nu q}\ ,\ \theta_2'=\frac{\sqrt{1-\rho^2}}{\nu q+\rho(\alpha-a_{\min})}\\
\tan\theta_1 &= -\frac{\sqrt{1-\rho^2}}{\rho}\ ,\ \theta_2'=\theta_1'\\
A &= \frac{f^{1-\beta}\nu}{(1-\beta)a_{\min}}\ ,\ A'=\frac{f\nu(\alpha\rho+\nu q)+f^\beta(\beta-1)a_{\min}^2}{(\beta-1)f^\beta a_{\min}^2\left(\nu q+\rho\left(\alpha-a_{\min}\right)\right)}\\
B &= \frac{\rho}{\sqrt{1-\rho^2}}
\end{aligned}
$$

Note that although $F_1(x, a, b)$ can be integrated analytically, we do not reproduce this lengthy expression here.

6.3.3 Derivation

In order to use our general formula for the implied volatility (6.25), we compute the geodesic distance and the connection associated to the λ-SABR model in the next subsection. We show that the λ-SABR metric is diffeomorphic equivalent to the metric on $\mathbb{H}^2$, the so-called hyperbolic Poincaré plane.

Hyperbolic Poincaré plane

Let us introduce the variable $q = \int_{f_0}^{f} \frac{df'}{C(f')}$. As shown in paragraph (6.1.2), by introducing the new coordinates

$$x = \nu q - \rho a \tag{6.41}$$

$$y = (1 - \rho^2)^{\frac{1}{2}} a \tag{6.42}$$

the metric is the standard hyperbolic metric on the Poincaré half-plane $\mathbb{H}^2$ in the coordinates $[x, y]$

$$ds^2 = \frac{2}{\nu^2} \frac{dx^2 + dy^2}{y^2} \tag{6.43}$$

The unusual factor $\frac{2}{\nu^2}$ in front of the metric (6.43) can be eliminated by scaling the time $\tau' = \frac{\nu^2}{2}\tau$ in the heat kernel (4.56) (and Q becomes $\frac{2}{\nu^2}Q$). This is what we will use in the following.
As the connection between the λ-SABR model and $\mathbb{H}^2$ is quite intriguing, we investigate several useful properties of the hyperbolic space (for example the geodesics).

Isometry $\mathrm{PSL}(2, \mathbb{R})$

By introducing the complex variable $z = x + \imath y$, the metric becomes

$$ds^2 = \frac{dz d\bar{z}}{\Im(z)^2}$$

In this coordinate system, it can be shown that $\mathrm{PSL}(2, \mathbb{R})$[4] is an isometry, meaning that the distance is preserved. The action of $\mathrm{PSL}(2, \mathbb{R})$ on z is *transitive* and given by

$$z' = \frac{az + b}{cz + d}$$

[4] $\mathrm{PSL}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/\mathbb{Z}_2$ with $\mathrm{SL}(2, \mathbb{R})$ the group of two by two real matrices with determinant one. $\mathbb{Z}_2$ acts on $A \in \mathrm{SL}(2, \mathbb{R})$ by $\mathbb{Z}_2 A = -A$.

By transitive we mean that given two complex points z and z', there exists a unique element $A \in \mathrm{PSL}(2,\mathbb{R})$ such that $z' = A.z$.
Note that when the real coefficients (a,b,c,d) are replaced by $(-a,-b,-c,-d)$, the action is unchanged and this is why we have taken the quotient of the Lie group $\mathrm{SL}(2,\mathbb{R})$ by the discrete group $\mathbb{Z}_2$ to get a transitive action.

Models and coordinates for the hyperbolic plane.

There are four models (i.e., geometric representations) commonly used for hyperbolic geometry: the Klein model, the Poincaré disc model, the upper half-plane model and the Minkowski model. We explain below the relation between the last three ones.
Let us define a *Moebius transformation* T as an element of $\mathrm{PSL}(2,\mathbb{R})$ which is uniquely given by its values at 3 points: $T(0) = 1$, $T(i) = 0$ and $T(\infty) = -1$. If $\Im(z) > 0$ then $|T(z)| < 1$ so T maps the *upper half-plane* on the *Poincaré disk*

$$\mathcal{D} = \{z = x + \imath y \in \mathbb{C} \mid \; |z| \leq 1\}$$

Then if we define $x_0 = \frac{1+|z|^2}{1-|z|^2}$, $x_1 = \frac{2\Re(z)}{1-|z|^2}$, $x_3 = \frac{2\Im(z)}{1-|z|^2}$, we obtain that $\mathcal{D}$ is mapped to the *Minkowski pseudo-sphere*

$$-x_0^2 + x_1^2 + x_2^2 = -1$$

On this space, we have the Lorentzian metric

$$ds^2 = -dx_0^2 + dx_1^2 + dx_3^2$$

We can then deduce the induced metric. On the upper-half plane, this gives (6.43) (without the scale factor $\frac{2}{\nu^2}$).

Geodesics

In the upper half-plane, the geodesics correspond to vertical lines and to semi-circles centered on the horizon $\Im(z) = 0$, and in the Poincaré disk $\mathcal{D}$ the geodesics are circles orthogonal to $\mathcal{D}$ (Fig. 6.1).
In the example 4.7, we have obtained that the geodesic distance (invariant under $\mathrm{PSL}(2,\mathbb{R})$) between the complex points $z = x + \imath y$, $z' = x' + \imath y'$ on $\mathbb{H}^2$ is given by

$$d(z,z') = \cosh^{-1}\left(1 + \frac{|z-z'|^2}{2yy'}\right) \tag{6.44}$$

This specific expression for the geodesic distance (6.44) allows us to find that the saddle-point $a_{\min}$ which minimizes the geodesic distance when the forward is fixed to the strike has the following expression

$$a_{\min}(f) = \sqrt{\alpha^2 + 2\alpha\nu\rho q + \nu^2 q^2} \tag{6.45}$$

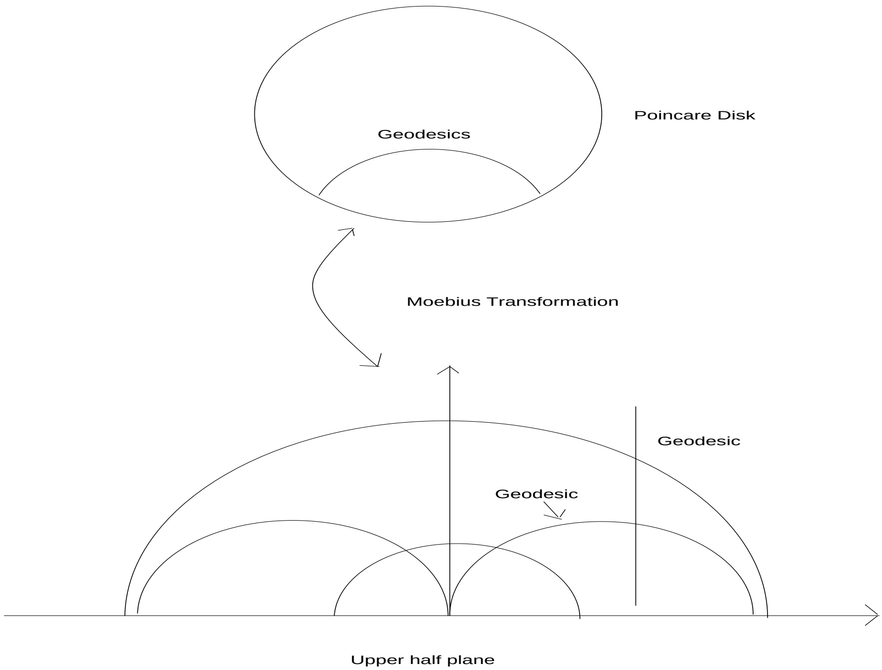

FIGURE 6.1: Poincaré disk $\mathcal{D}$ and upper half-plane $\mathbb{H}^2$ with some geodesics. In the upper half-plane, the geodesics correspond to vertical lines and to semi-circles centered on the horizon $\Im(z) = 0$ and in $\mathcal{D}$ the geodesics are circles orthogonal to $\mathcal{D}$.

Finally, by using this explicit expression for the geodesic distance, the Van Vleck-Morette determinant is

$$\Delta(z, z') = \frac{d(z, z')}{\sinh(d(z, z'))}$$

As Δ only depends on the geodesic distance d and d is minimized for $a = a_{\min}$, we have

$$(\ln \Delta)'(a_{\min}) = 0$$

Furthermore, we have the following identities (using Mathematica)

$$d(a_{\min}) = \cosh^{-1}\left(\frac{-q\nu\rho - \alpha\rho^2 + a_{\min}}{\alpha(1-\rho^2)}\right) \tag{6.46}$$

$$a_{\min}\phi''(a_{\min}) = \frac{2d(a_{\min})}{\alpha(1-\rho^2)\sinh(d(a_{\min}))} \tag{6.47}$$

$$\frac{\phi'''(a_{\min})}{\phi''(a_{\min})} = -\frac{3}{a_{\min}} \tag{6.48}$$

$$(\ln g)'(a_{\min}) = -\frac{4}{a_{\min}} \tag{6.49}$$

This formula (6.46) has already been obtained independently in [60].

Connection for the λ-SABR

In the coordinates $[a, f]$, the connection $\mathcal{A}$ is

$$\mathcal{A} = \frac{1}{2(1-\rho^2)}\left(\frac{2\lambda\bar{\lambda}\rho - 2\lambda\rho a - \nu a^2 C'(f)}{\nu C(f) a^2}df + \frac{-2\lambda\bar{\lambda} + 2\lambda a + \nu\rho a^2 C'(f)}{\nu^2 a^2}da\right) \tag{6.50}$$

In the new coordinates $[x, y]$, the connection above is given by

$$\mathcal{A} = \mathcal{A}^{\text{SABR}} + \frac{\lambda\left(\bar{\lambda}\sqrt{1-\rho^2} - y\right)}{\nu^2 y^2\sqrt{1-\rho^2}}\left(\rho dx - \sqrt{1-\rho^2}dy\right)$$

with $\mathcal{A}^{\text{SABR}}$ the connection for the SABR model given by

$$\mathcal{A}^{\text{SABR}} = -\frac{d\ln C(f)}{2(1-\rho^2)} + \frac{\rho\beta}{2\sqrt{1-\rho^2}(1-\beta)}\frac{dy}{\left(x + \frac{\rho}{\sqrt{1-\rho^2}}y + \frac{\nu f_0^{1-\beta}}{(1-\beta)}\right)} \quad \text{for} \quad \beta \neq 1$$

$$\mathcal{A}^{\text{SABR}} = -\frac{dx}{2(1-\rho^2)\nu} \quad \text{for} \quad \beta = 1$$

The pullback of the connection on a geodesic curve $\mathcal{C}$ satisfying $(x - x_0(a,f))^2 + y^2 = R^2(a,f)$ is given by $(\beta \neq 1)$[5]

$$i^*\mathcal{A} = i^*\mathcal{A}^{\text{SABR}} + \frac{\lambda(-\bar{\lambda}\sqrt{1-\rho^2} + y)}{\nu^2 y^2}\left(\frac{\rho y}{\sqrt{(R^2(a,f) - y^2)(1-\rho^2)}} + 1\right)dy$$

and with

$$i^*\mathcal{A}^{\text{SABR}} = \frac{\rho\beta}{2\sqrt{1-\rho^2}(1-\beta)}\frac{dy}{(\hat{x_0}(a,f) + \sqrt{R^2(a,f) - y^2} + \frac{\rho}{\sqrt{1-\rho^2}}y)} - \frac{d\ln C(f)}{2(1-\rho^2)}$$

[5]The case $\beta = 1$ will be treated in the next section.

with $i : \mathcal{C} \to \mathbb{H}^2$ the embedding of the geodesic $\mathcal{C}$ on the Poincaré plane and $\hat{x}_0 = x_0 + \frac{\nu f_0^{1-\beta}}{(1-\beta)}$. We have used that $i^* dx = -\frac{y dy}{\sqrt{R^2 - y^2}}$. Note that the two constants x_0 and R are determined by using the fact that the two points $z_1 = -\rho\alpha + i\sqrt{1-\rho^2}\alpha$ and $z_2 = \nu \int_{f_0}^{f} \frac{df'}{C(f')} - \rho a + i\sqrt{1-\rho^2}a$ pass through the geodesic curves. The algebraic equations giving R and x_0 can be exactly solved:

$$x_0(a,f) = \frac{x_1^2 - x_2^2 + y_1^2 - y_2^2}{2(x_1 - x_2)}$$

$$R(a,f)^2 = y_1^2 + \frac{\left((x_1 - x_2)^2 - y_1^2 + y_2^2\right)^2}{4(x_1 - x_2)^2}$$

Using the polar coordinates $x - x_0 = R\cos(\theta)$, $y = R\sin(\theta)$, we obtain that the parallel gauge transport is

$$\begin{aligned}\ln(\mathcal{P})(a) = &\frac{\lambda}{\nu^2}\left(\int_{y_1}^{y_2} \frac{\bar{\lambda}\sqrt{1-\rho^2} - y}{y^2} dy - \frac{\rho}{\sqrt{1-\rho^2}} \int_{\theta_1}^{\theta_2} \left(1 + \frac{A_0}{\sin(\theta)}\right) d\theta\right) \\ &+ \ln(\mathcal{P}^{\mathrm{SABR}})(a)\end{aligned}$$

with

$$\begin{aligned}\ln \mathcal{P}^{\mathrm{SABR}}(z,z') = &-\frac{\beta\rho}{2\sqrt{1-\rho^2}(1-\beta)} \int_{\theta_1}^{\theta_2} \frac{\cos(\theta) d\theta}{(\cos(\theta) + \frac{\hat{x}_0}{R} + B\sin(\theta))} \\ &+ \frac{1}{2(1-\rho^2)} \ln \frac{C(f)}{C(f_0)}\end{aligned}$$

with $\theta_i(a,f) = \arctan\left(\frac{y_i}{x_i - x_0}\right)$, $i = 1, 2$, $B = \frac{\rho}{\sqrt{1-\rho^2}}$ and $A_0 = -\frac{\bar{\lambda}\sqrt{1-\rho^2}}{R}$. The two integration bounds θ_1 and θ_2 explicitly depend on a and the coefficient $A \equiv \frac{\hat{x}_0}{R}$. Doing the integration over θ, we obtain (6.40). Note that according to remark 4.3, we have interchanged f_0 (resp. α) with f (resp. a).

By plugging all these results (6.40,6.46,6.47,6.48, 6.49) in (6.25), we obtain our final expression for the asymptotic smile (6.39) at the first-order associated to the stochastic λ-SABR model.

Below, we have plotted the probability density for the λ-SABR (see Fig. 6.2). Note that the density remains positive.

REMARK 6.7 SABR original formula We can now see how the Hagan-al asymptotic smile [100] formula can be obtained in the case $\lambda = 0$ and show that our formula gives a better approximation.

First we approximate $a_{\min}$ (6.57) by the following expression

$$a_{\min} \simeq \alpha + q\rho\nu$$

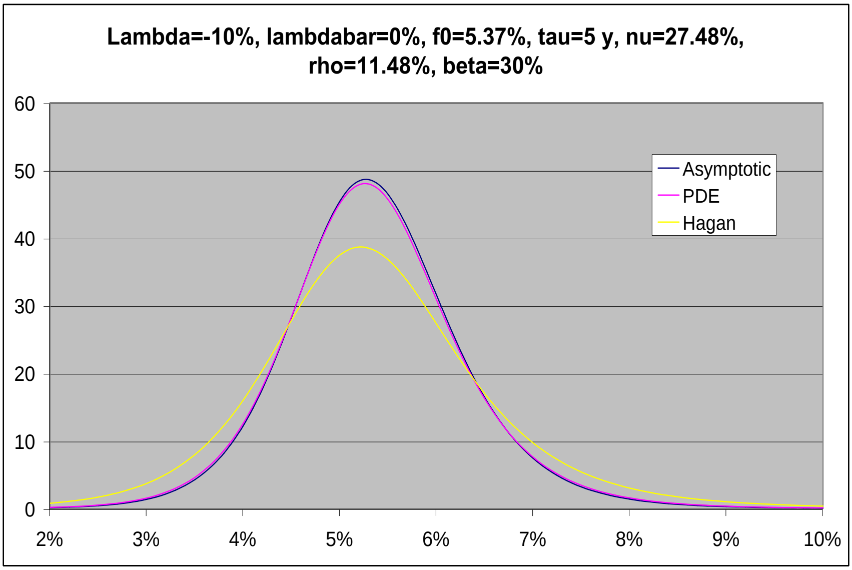

FIGURE 6.2: Probability density $p(K,T|f_0) = \frac{\partial^2 \mathcal{C}(T,K)}{\partial^2 K}$. Asymptotic solution vs numerical solution (PDE solver). The Hagan-al formula has been plotted to see the impact of the mean-reverting term. Here f_0 is a swap spot and α has been fixed such that the Black volatility $\alpha f_0^{\beta-1} = 30\%$.

In the same way, we have

$$\frac{\sinh(d(a_{\min}))}{d(a_{\min})} \simeq 1$$

Furthermore, for $\lambda = 0$, the connection (6.50) reduces to

$$\mathcal{A}^{\text{SABR}} = \frac{1}{2(1-\rho^2)} \left(-d\ln(C(f)) + \frac{\rho}{\nu} \partial_f C da \right)$$

Therefore, the parallel gauge transport is obtained by integrating this one-form along a geodesic $\mathcal{C}$

$$\mathcal{P}^{\text{SABR}} = e^{\frac{1}{2(1-\rho^2)} \left(-\ln \frac{C(f)}{C(f_0)} + \int_{\mathcal{C}} \frac{\rho}{\nu} \partial_f C da \right)}$$

The component $\mathcal{A}_f$ of the connection is an exact form and therefore has easily been integrated. The result doesn't depend on the geodesic curve but only on its endpoints. However, this is not the case for the component $\mathcal{A}_a$. By approximating ($f_{\rm av} = \frac{f_0+f}{2}$) $\partial'_f C(f') \simeq \partial_f C(f_{\rm av})$, the component $\mathcal{A}_a$ becomes an exact form and can therefore be integrated by

$$\int_{\mathcal{C}} \frac{\rho}{\nu} \partial_f C da \simeq \frac{\rho}{\nu} \partial_f C(f_{\rm av})(a-\alpha)$$

Finally, by plugging these approximations into our formula (6.25), we reproduce the Hagan-al original formula [100]

$$\sigma_{\rm BS}(K,T) = \frac{\ln \frac{K}{f_0}}{\mathrm{vol}(K)} \left(1 + \sigma_1 (\frac{K+f_0}{2}) T \right) \tag{6.51}$$

with

$$\sigma_1(f) = \frac{(\alpha C(f))^2}{24} \left(\frac{1}{f^2} + \frac{2\partial_{ff}C(f)}{C(f)} - \left(\frac{\partial_f C(f)}{C(f)} \right)^2 \right) + \frac{\alpha \nu \partial_f C(f) \rho}{4} + \frac{2-3\rho^2}{24} \nu^2$$

Therefore, the Hagan-al formula corresponds to the approximation of the Abelian connection by an exact form. The latter can be integrated outside the parametrization of the geodesic curve.

□

In the following graphs (see Figs. 6.3, 6.4, 6.5) we have compared the accuracy of our formula (6.39) (denoted $\mathbb{H}^2$ formula) against the Hagan-al formula (6.51) using a Monte-Carlo pricer. As the vol of vol ν is large (i.e., $\nu = 100\%$), we are outside the domain of validity of the perturbation expansion when the maturity τ is greater than two years. In this case, the dimensionless parameter $\frac{\nu^2 \tau}{2}$ is greater than 1. We remark that our asymptotic formula performs better than the Hagan-al expansion. In this context, we have observed experimentally that in our asymptotic expansion (6.39), $C(f) a_{\min}$ should be replaced by $C(f)\,(\alpha + q\rho\nu)$ if $q > 0$. This is what we have used.

REMARK 6.8 $\mathbb{H}^2$-model In the previous section, we have seen that the λ-SABR model corresponds to the geometry of $\mathbb{H}^2$. This space is particularly nice in the sense that the geodesic distance and the geodesic curves are known analytically. A similar result holds if we assume that $C(f)$ is a general function ($C(f) = f^\beta$ for λ-SABR). In the following, we will try to fix this arbitrary function in order to fit the *short-term smile*. In this case, we can use our general asymptotic smile formula at the zero-order: The short-term smile will

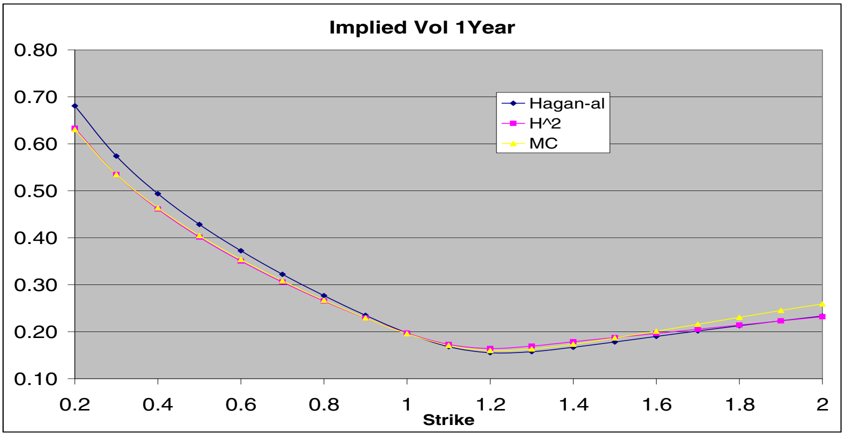

FIGURE 6.3: Implied volatility for the SABR model $\tau = 1$Y. $\alpha = 0.2$, $\rho = -0.7$, $\frac{\nu^2}{2}\tau = 0.5$ and $\beta = 1$.

be automatically calibrated by construction if the short-term local volatility $\sigma_{\mathrm{loc}}(f)$ is

$$\sigma_{\mathrm{loc}}(f) = C(f)a_{\min}(f) \tag{6.52}$$

with

$$a_{\min}(f)^2 = \alpha^2 + 2\rho\alpha\nu q + \nu^2 q^2$$
$$q = \int_{f_0}^{f} \frac{df'}{C(f')}$$

By *short-term*, we mean a maturity date less than 1 year in practice. Solving (6.52) according to q, we obtain

$$\nu q = -\rho\alpha\nu + \sqrt{\alpha^2(-1+\rho^2) + \frac{\sigma_{\mathrm{loc}}(f)^2}{C(f)^2}}$$

and if we derive under f, we have ($\psi(f) \equiv \frac{\sigma_{\mathrm{loc}}(f)}{C(f)}$)

$$\frac{d\psi}{\sqrt{\psi^2 - \alpha^2(1-\rho^2)}} = \frac{\nu}{\sigma_{\mathrm{loc}}(f)} df$$

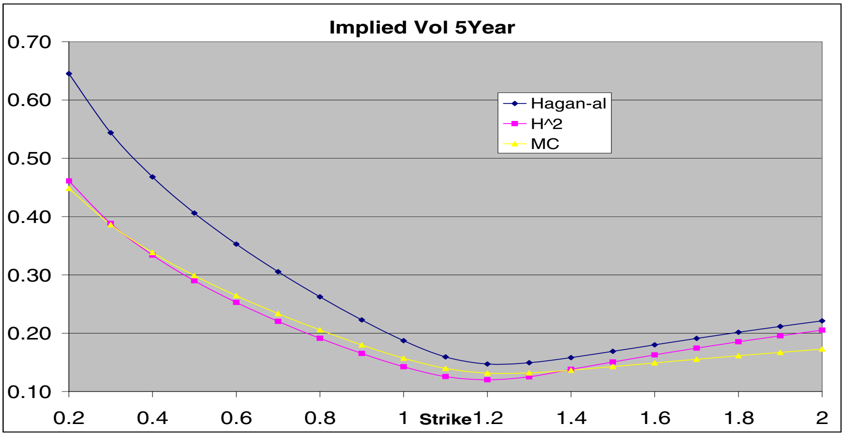

FIGURE 6.4: Implied volatility for the SABR model $\tau = 5$Y. $\alpha = 0.2$, $\rho = -0.7$, $\frac{\nu^2}{2}\tau = 2.5$ and $\beta = 1$.

Solving this ODE, we obtain that $C(f)$ is fixed to

$$C(f) = \frac{\sigma_{\text{loc}}(f)}{\alpha\sqrt{1-\rho^2}\cosh\left(\nu\int_{f_0}^{f}\frac{df'}{\sigma_{\text{loc}}(f')}\right)}$$

If $\nu = 0$, we have $C(f) = \sigma_{\text{loc}}(f)$ and we reproduce the Dupire formula. Using the BBF formula (5.39), we have

$$C(f) = \frac{f\sigma_{\text{BS}}(f)\left(1 - f\ln\left(\frac{f}{f_0}\right)\frac{\sigma'_{\text{BS}}(f)}{\sigma_{\text{BS}}(f)}\right)}{\alpha\sqrt{1-\rho^2}\cosh\left(\nu\frac{\ln\left(\frac{f}{f_0}\right)}{\sigma_{\text{BS}}(f)}\right)}$$

The short term smile is automatically calibrated when using this function in the case of the λ-SABR model. $\square$

In the following, we will show how to find an exact solution to the SABR model with $\beta = 0$ and $\beta = 1$ for the conditional probability. These exact solutions will allow us to test the validity of our asymptotic solution. A (complete) treatment of solvable stochastic volatility models will be done in chapter 9.

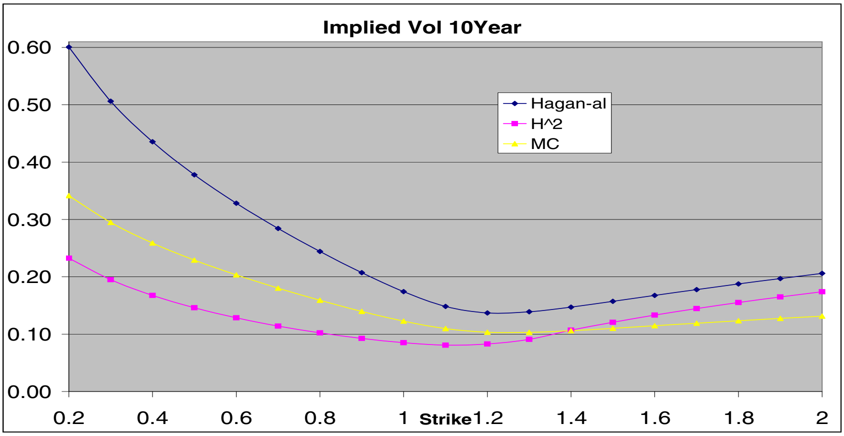

FIGURE 6.5: Implied volatility for the SABR model $\tau = 10$Y. $\alpha = 0.2$, $\rho = -0.7$, $\frac{\nu^2}{2}\tau = 5$ and $\beta = 1$.

6.4 Analytical solution for the normal and log-normal SABR model

6.4.1 Normal SABR model and Laplacian on $\mathbb{H}^2$

For the SABR model, the connection $\mathcal{A}$ and the function Q are given by

$$\mathcal{A} = \frac{1}{2\left(1-\rho^2\right)}\left(-d\ln(C) + \frac{\rho}{\nu}\partial_f C da\right) \tag{6.53}$$

$$Q = \frac{a^2}{4}\left(C\partial_f^2 C - \frac{\left(\partial_f C\right)^2}{2\left(1-\rho^2\right)}\right) \tag{6.54}$$

For $\beta = 0$, the function Q and the potential $\mathcal{A}$ vanish. p satisfies a heat kernel equation where the differential operator D reduces to the Laplace-Beltrami operator on $\mathbb{H}^2$:

$$\frac{\partial p}{\partial \tau'} = \Delta_{\mathbb{H}^2} p \tag{6.55}$$

$$\equiv y^2(\partial_x^2 + \partial_y^2) p \tag{6.56}$$

with $\tau' = \frac{\nu^2\tau}{2}$ and the coordinates $[x, y]$ defined by (6.41,6.42) (with $q = f - f_0$). Therefore solving the Kolmogorov equation for the normal SABR model (i.e., $\beta = 0$) is equivalent to solving this (Laplacian) heat kernel on $\mathbb{H}^2$. Surprisingly, there is an analytical solution for the heat kernel on $\mathbb{H}^2$ (6.55) found by McKean [116]. It is connected to the *Selberg trace formula* [21]. The *exact conditional probability* density p depends on the hyperbolic distance $d(z, z')$ and is given by

$$p(\tau', d) = \frac{\sqrt{2}e^{-\frac{\tau'}{4}}}{(4\pi\tau')^{\frac{3}{2}}} \int_{d(z,z')}^{\infty} \frac{be^{-\frac{b^2}{4\tau'}}}{(\cosh b - \cosh d(z, z'))^{\frac{1}{2}}} db$$

The conditional probability in the old coordinates $[a, f]$ is

$$p(\tau', f, a) df da = \frac{\nu}{\sqrt{1-\rho^2}} \frac{df da}{a^2} \frac{\sqrt{2}e^{-\frac{\tau'}{4}}}{(4\pi\tau')^{\frac{3}{2}}} \int_{d(z,z')}^{\infty} \frac{be^{-\frac{b^2}{4\tau'}}}{(\cosh b - \cosh d(z, z'))^{\frac{1}{2}}} db$$

where the term $\frac{\nu}{\sqrt{1-\rho^2 a^2}}$ corresponds to the invariant measure $\sqrt{g}$ on $\mathbb{H}^2$. We have compared this exact solution (Fig. 6.6) with a numerical solution (PDE) of the normal SABR model and found an agreement. A similar result was obtained independently in [101] for the conditional probability.
The value of a European option is

$$\mathcal{C}(T, K|f_0) = \max(f_0 - K, 0) + \frac{1}{2} \int_0^T d\tau \int_0^{\infty} da \; a^2 p(\tau, x_1 = K, a|\alpha)$$

This expression can be obtained by applying the Itô-Tanaka lemma on the payoff $\max(f_t - K, 0)$. In order to integrate over a, we use a small trick: we interchange the order of integration over b and a. The half space $b \geq d$ with a arbitrary is then mapped to the half-strip $a_{\min} \leq a \leq a_{\max}$ and $b \geq l_{\min}$ where[6]

$$\frac{(a_{\max} - a_{\min})^2}{4(1-\rho^2)} = -(K - f_0)^2\nu^2 - \alpha\left(\alpha + 2\nu\rho(K - f_0) + \alpha\rho^2\right)$$
$$+ \alpha \cosh b \left(2\rho\left((K - f_0)\nu + \alpha\rho\right) + \alpha(1-\rho^2)\cosh b\right) \tag{6.57}$$
$$\frac{a_{\min} + a_{\max}}{2} = \rho\left(\nu(K - f_0) + \alpha\rho\right) + \alpha(1-\rho^2)\cosh b \tag{6.58}$$
$$\cosh l_{\min} = \frac{-\rho\left((K - f_0)\nu + \alpha\rho\right) + \sqrt{\alpha^2 + 2\alpha\nu\rho(K - f_0) + \nu^2(K - f_0)^2}}{\alpha(1-\rho^2)} \tag{6.59}$$

[6] All these algebraic computations have been done with Mathematica.

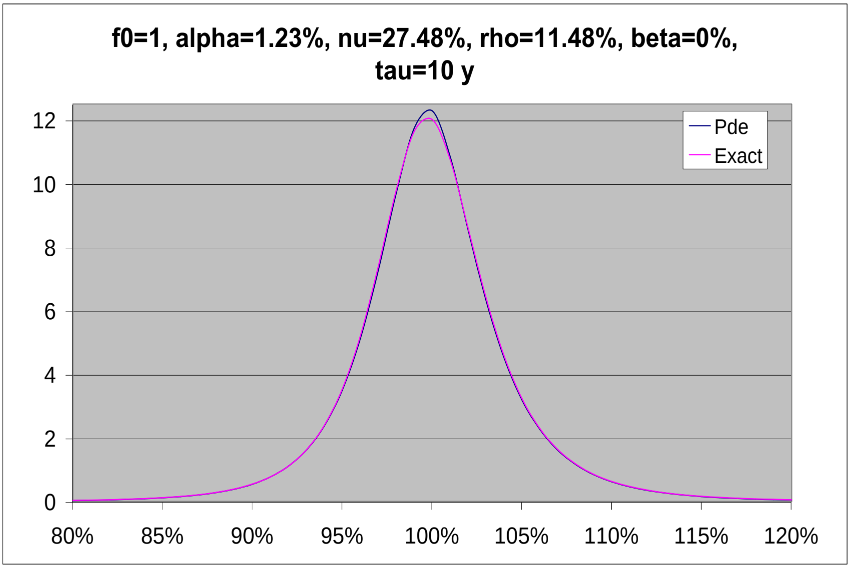

FIGURE 6.6: Exact conditional probability for the normal SABR model versus numerical PDE.

Performing the integration according to a leads to an *exact solution* for a European call option for the normal SABR model

$$\mathcal{C}(T,K|f_0) = \frac{\sqrt{2}}{\nu\sqrt{1-\rho^2}} \int_0^{\frac{\nu^2 T}{2}} dt \frac{e^{-\frac{t}{4}}}{(4\pi t)^{\frac{3}{2}}} \int_{l_{\min}}^{\infty} \frac{b(a_{\max}-a_{\min})e^{-\frac{b^2}{4t}}}{\sqrt{\cosh b - \cosh l_{\min}}} db \\ + \max(f_0 - K, 0)$$

with $a_{\min}$, $a_{\max}$ and $l_{\min}$ defined by (6.57,6.58,6.59).

6.4.2 Log-normal SABR model and Laplacian on $\mathbb{H}^3$

A similar computation can be carried out for the log-normal SABR model (i.e., $\beta = 1$). By using (6.53), the Abelian connection reduces to

$$\mathcal{A} = \frac{1}{2(1-\rho^2)} \left(-d\ln f + \frac{\rho}{\nu} da \right)$$

The potential $\mathcal{A}$ is exact, meaning there exists a smooth function Λ such that $\mathcal{A} = d\Lambda$ with

$$\Lambda(f,a) = \frac{1}{2(1-\rho^2)}\left(-\ln(f) + \frac{\rho}{\nu}a\right)$$

Furthermore, using (6.54), we have

$$Q = -\frac{a^2}{8(1-\rho^2)} = -\frac{y^2}{8(1-\rho^2)^2}$$

Applying an Abelian gauge transformation (4.64)

$$p' = e^{\Lambda(f,a)-\Lambda(f_0,\alpha)}p$$

we find that p' satisfies the following equation

$$y^2\left(\partial_x^2 + \partial_y^2 - \frac{1}{4\nu^2(1-\rho^2)^2}\right)p' = \partial_{\tau'}p' \tag{6.60}$$

How do we solve this equation? It turns out that the solution corresponds in some fancy way to the solution of the (Laplacian) heat kernel on the three dimensional hyperbolic space $\mathbb{H}^3$.

Geometry $\mathbb{H}^3$

This space can be represented as the upper-half space

$$\mathbb{H}^3 = \{x = (x_1, x_2, x_3) | x_3 > 0\}$$

In these coordinates, the metric takes the following form

$$ds^2 = \frac{(dx_1^2 + dx_2^2 + dx_3^2)}{x_3^2}$$

and the geodesic distance between two points x and x' in $\mathbb{H}^3$ is given by[7]

$$\cosh(d(x,x')) = 1 + \frac{|x-x'|^2}{2x_3x'_3}$$

As in $\mathbb{H}^2$, the geodesics are straight vertical lines or semi-circles orthogonal to the boundary of the upper-half space. An interesting property, useful to solve the heat kernel, is that the group of isometries of $\mathbb{H}^3$ is $\mathrm{PSL}(2,\mathbb{C})$.[8] If we represent a point $p \in \mathbb{H}^3$ as a quaternion[9] whose fourth components equal

[7] $|\cdot|$ is the Euclidean distance in $\mathbb{R}^3$.

[8] $\mathrm{PSL}(2,\mathbb{C})$ is identical to $\mathrm{PSL}(2,\mathbb{R})$, except that the real field is replaced by the complex field.

[9] The quaternionic field is generated by the unit element $\mathbf{1}$ and the basis $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ which satisfy the multiplication table $\mathbf{i}.\mathbf{j} = \mathbf{k}$ and the other cyclic products.

zero, then the action of an element $g \in \mathrm{PSL}(2,\mathbb{C})$ on $\mathbb{H}^3$ can be described by the formula

$$p' = g.p = \frac{ap+b}{cp+d} \tag{6.61}$$

with $p = x_1\mathbf{1} + x_2\mathbf{i} + x_3\mathbf{j}$.
The Laplacian on $\mathbb{H}^3$ in the coordinates $[x_1, x_2, x_3]$ is given by

$$\Delta_{\mathbb{H}^3} = x_3^2(\partial_{x_1}^2 + \partial_{x_2}^2 + \partial_{x_3}^2)$$

and the (Laplacian) heat kernel equation is

$$\partial_{\tau'} p' = \Delta_{\mathbb{H}^3} p'$$

Analytical probability density

The exact solution for the conditional probability density $p'(\tau', x|x')$, depending on the geodesic distance $d(x, x')$, is [95]

$$p'(\tau', x|x') = \frac{1}{(4\pi\tau')^{\frac{3}{2}}} \frac{d(x,x')}{\sinh(d(x,x'))} e^{-\tau' - \frac{d(x,x')^2}{4\tau'}}$$

Let us apply a Fourier transformation on p along the coordinate x_1 (or equivalently x_2)

$$\hat{p}(\tau', k, x_2, x_3|x') = \int_{-\infty}^{\infty} \frac{dx_1}{\sqrt{2\pi}} e^{-ikx_1} p'(\tau', x|x')$$

Then $\hat{p}'$ satisfies the following PDE

$$\partial_{\tau'}\hat{p}' = x_3^2(-k^2 + \partial_{x_2}^2 + \partial_{x_3}^2)\hat{p}' \tag{6.62}$$

By comparing (6.62) with (6.60), we deduce that the exact solution for the conditional probability (see remark 4.3) for the SABR model with $\beta = 1$ is (with $x \equiv x_2 = \nu \ln \frac{f_0}{f} - \rho\alpha$, $y \equiv x_3 = \sqrt{1-\rho^2}\alpha$, $k \equiv \frac{1}{2\nu(1-\rho^2)}$, $x_1' = 0$, $x_2' = -\rho a$, $x_3' = \sqrt{1-\rho^2}a$)

$$p'(x, y, x', y', \tau') = e^{\frac{1}{2(1-\rho^2)}(\ln(\frac{f_0}{f}) + \frac{\rho}{\nu}(a-\alpha))}$$
$$\int_{-\infty}^{\infty} \frac{dx_1}{\sqrt{2\pi}} \frac{1}{(4\pi\tau')^{\frac{3}{2}}} \frac{d(x,x')}{\sinh(d(x,x'))} e^{-\tau' - \frac{d(x,x')^2}{4\tau'}} e^{-\frac{ix_1}{2\nu(1-\rho^2)}}$$

A previous solution for the SABR model with $\beta = 1$ was obtained by [29], although only in terms of Gauss hypergeometric series.

6.5 Heston model: a toy black hole

In this section, we look at the Heston model, a typical example of SVM frequently used by practitioners. The dynamics of the forward and the instantaneous volatility is

$$\begin{aligned} \frac{df_t}{f_t} &= a_t dW_t \\ da_t^2 &= \lambda(\bar{v} - a_t^2)dt + \zeta a_t dZ_t \ , \ dW_t dZ_t = \rho dt \end{aligned}$$

or equivalently

$$da_t = \left(-\frac{\lambda}{2}\left(1 - \frac{\bar{v}}{a_t^2}\right) - \frac{\zeta}{8a_t^2}\right) a_t dt + \frac{\zeta}{2} dZ_t$$

The model has five parameters: $a_{t=0} \equiv \alpha$, λ, $\bar{v}$, ζ and ρ. The time $\tau = \frac{1}{\lambda}$ is a cutoff which separates the short from the long maturities. In [61], the main properties of the Heston model are examined carefully. In particular, this model admits an analytical characteristic function. As a consequence the fair value of a call option can be written as the one-dimensional Fourier transform of the characteristic function [106]. This feature is one of the main reasons why the Heston model has drawn the attention of the practitioners despite many disadvantages such as its delicate numerical simulation [43].

6.5.1 Analytical call option

The PDE satisfied by a call option $\mathcal{C}$ is

$$\partial_\tau \mathcal{C} = \frac{1}{2} v \partial_y \mathcal{C} + \frac{\zeta^2}{2} v \partial_v^2 \mathcal{C} + \zeta v \rho \partial_{vy} \mathcal{C} - \lambda(v - \bar{v}) \partial_v \mathcal{C} - \frac{1}{2} v \partial_y \mathcal{C}$$

using the variables $\tau = T - t$, $v = a^2$ and the moneyness $y = \ln \frac{f}{f_0}$. By analogy with the Black-Scholes formula, we guess a solution of the form

$$\mathcal{C}(\tau, y, v) = K\left(e^y P_1(\tau, y, v) - P_0(\tau, y, v)\right) \tag{6.63}$$

with P_0 and P_1 two-independent solutions. Plugging (6.63) into the pricing PDE, we obtain the PDEs

$$\begin{aligned} \partial_\tau P_i = \frac{1}{2} v \partial_y^2 P_i - \left(\frac{1}{2} - j\right) v \partial_y P_i + \frac{1}{2} \zeta^2 v \partial_v^2 P_i + \rho \zeta v \partial_{xv} P_i \\ + (a - b_j v) \partial_v P_j \end{aligned}$$

with $a = \lambda \bar{v}$ and $b_j = \lambda - j\rho\zeta \ , \ j = 0, 1$. In order to satisfy the terminal condition, these PDEs are subject to the terminal condition $P_j(0, y, v) = 1(y)$.

By applying a Fourier transform with respect to the variable y

$$\hat{P}_j(\tau, k, v) = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-\imath k y} P_j(\tau, y, v)$$

we have[10]

$$\partial_\tau \hat{P}_j = v\left(\alpha_j \hat{P}_j - \beta_j \partial_v \hat{P}_j + \gamma \partial_v^2 \hat{P}_j\right) + a \partial_v \hat{P}_j \tag{6.64}$$

with

$$\begin{aligned}
\alpha &= -\frac{k^2}{2} - \imath\frac{k}{2} + \imath j u \\
\beta &= \lambda - \rho\zeta j - \rho\zeta \imath k \\
\gamma &= \frac{\zeta^2}{2}
\end{aligned}$$

As an ansatz for the solution, we guess

$$\hat{P}_j = A_j(\tau, k) e^{B_j(\tau, k) v}$$

By plugging this guess into (6.64), we obtain that A_j and B_j satisfy two Riccati ODEs that can be solved. Finally, we obtain

$$P_j(\tau, y, v) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} du \Re\left(\frac{e^{C_j(\tau,v)\bar{v} + D_j(\tau,u)v + iuy}}{iu}\right)$$

with

$$\begin{aligned}
D(\tau, u) &= r_- \frac{1 - e^{-d\tau}}{1 - g e^{-d\tau}} \\
C(\tau, u) &= \lambda\left(r_- \tau - \frac{2}{\zeta^2}\ln\left(\frac{1 - g e^{-d\tau}}{1 - g}\right)\right)
\end{aligned}$$

where we define

$$\begin{aligned}
r_\pm &\equiv \frac{\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\gamma} \equiv \frac{\beta \pm d}{\zeta^2} \\
g &\equiv \frac{r_-}{r_+}
\end{aligned}$$

The integration over u can be achieved with a numerical scheme. Note that this integration can be tricky as the integrand is typically of oscillatory nature giving rise to numerical instability. A careful treatment of this problem is achieved in [112].

[10] Here we follow closely the notations in [18].

6.5.2 Asymptotic implied volatility

By specifying our general results in section 6.1.2, we obtain that the Heston model corresponds to the metric

$$ds^2 = \frac{2}{\sqrt{1-\rho^2}\zeta} \frac{dx^2 + dy^2}{y}$$

with the coordinates

$$x = \ln \frac{f}{f_0} - \rho \frac{a^2}{\zeta}$$
$$y = \sqrt{1-\rho^2} \frac{a^2}{\zeta}$$

Note that as explained in remark 6.1, the metric is smooth if the volatility can not reach $a = 0$. This is satisfied if and only if

$$2\lambda\bar{v} > \zeta^2 \tag{6.65}$$

Unfortunately, this condition is not usually satisfied when the Heston model is calibrated to market implied volatility surfaces. Although the use of the heat kernel expansion can appear quite problematic in this case, we conjecture that our general first-order asymptotic implied volatility (6.25) is still valid when the condition (6.65) is not satisfied. For an other (probabilistic) approach validating this computation, the reader can consult [83].

Geodesic distance

The geodesic distance d (see (6.17)) between two points $(x_1 = -\rho\frac{\alpha^2}{\zeta}, y_1 = \sqrt{1-\rho^2}\frac{\alpha^2}{\zeta})$ and (x_2, y_2) is

$$d = \frac{2\delta}{\zeta C} [\arcsin(\sqrt{u})]_{C^2 \frac{y_1 \zeta}{\delta}}^{C^2 \frac{y_2 \zeta}{\delta}} \tag{6.66}$$

where $\delta = \frac{2}{\sqrt{1-\rho^2}}$. The constant $C = C(x_1, y_1, x_2, y_2) > 0$ is determined by the equation

$$\mathrm{sign}((x_2 - x_1)(y_2 - y_1))(x_2 - x_1) = \frac{\delta}{\zeta C^2} [-\sqrt{u(1-u)} + \arcsin(\sqrt{u})]_{\frac{C^2 \zeta y_1}{\delta}}^{\frac{C^2 \zeta y_2}{\delta}}$$

By doing the change of variables $\frac{C^2 \zeta y_i}{\delta} = \sin^2 \theta_i$, $\theta_{i=1,2} \in [0, \pi)$, we get

$$d = 2\sqrt{\frac{\delta}{\zeta} \frac{\sqrt{y_i}}{\sin \theta_i}} \mid \theta_2 - \theta_1 \mid$$
$$\mathrm{sign}((x_2 - x_1)(y_2 - y_1)) \frac{x_2 - x_1}{y_2} \sin^2 \theta_2 = [-\sin\theta |\cos\theta| + \theta]_{\theta_1}^{\theta_2}$$

with $\frac{\sin\theta_2}{\sin\theta_1} = \sqrt{\frac{y_2}{y_1}}$.

REMARK 6.9 Computation of $a_{\min}$
From (6.31), the saddle-point $a_{\min}$ satisfies the equation

$$k - \frac{\rho}{\zeta}\left(a_{\min}^2 - \alpha^2\right) = \pm\frac{a_{\min}^2}{\sqrt{1-\rho^2}\zeta}[\arcsin(u) - u\sqrt{1-u^2}]_{\frac{\sqrt{1-\rho^2}\alpha}{a_{\min}}}^{\sqrt{1-\rho^2}}$$

with $k = \ln\frac{K}{f_0}$. We set $\sin\theta \equiv \sqrt{1-\rho^2}\frac{\alpha}{a_{\min}}$, $\theta \in (0,\pi)$. This gives

$$\frac{\sin^2\theta(k)}{\sqrt{1-\rho^2}}\left(\frac{k\zeta}{\alpha^2} + \rho\right) \pm (\theta(k) - \sin\theta(k)|\cos\theta(k)|) = \pm\phi$$

with $\phi = \arccos\rho$. The zero-order implied volatility (5.39) is then

$$\lim_{T\to 0}\sigma_{\mathrm{BS}}(T,K) = \frac{\alpha\sqrt{1-\rho^2}k}{\int_0^k dy\sin\theta(y)}$$

with $k = \ln\frac{K}{f_0}$. ∎

Parallel transport

Throughout this paragraph, let us note $i : \mathcal{C} \to \Sigma$, the immersion of the geodesic curve $\mathcal{C}$ on the non-compact Riemann surface $\Sigma \simeq \mathbb{R}\times\mathbb{R}_+^*$.
From the definition (4.58), the Abelian connection is given by

$$\mathcal{A} = -\frac{d\ln f}{(1-\rho^2)}\left(\frac{1}{2} - \frac{\rho\lambda}{\zeta}\left(1 - \frac{\bar{v}}{a^2}\right)\right) + \frac{ada}{(1-\rho^2)\zeta}\left(\rho - \frac{2\lambda}{\zeta}\left(1 - \frac{\bar{v}}{a^2}\right)\right)$$

In the new coordinates $[x,y]$, we obtain

$$\mathcal{A} = -\frac{dx}{(1-\rho^2)}\left(\frac{1}{2} - \frac{\rho\lambda}{\zeta}\left(1 - \sqrt{1-\rho^2}\frac{\bar{v}}{\zeta y}\right)\right) - \frac{\lambda dy}{\zeta\sqrt{1-\rho^2}}\left(1 - \sqrt{1-\rho^2}\frac{\bar{v}}{\zeta y}\right)$$

The pullback of the Abelian connection on the geodesic curve $\mathcal{C}$ is

$$i^*\mathcal{A} = -\frac{1}{(1-\rho^2)}\sqrt{\frac{y}{\frac{\delta}{C^2\zeta} - y}}dy\left(\frac{1}{2} - \frac{\rho\lambda}{\zeta}\left(1 - \sqrt{1-\rho^2}\frac{\bar{v}}{\zeta y}\right)\right)$$
$$-\frac{\lambda dy}{\zeta\sqrt{1-\rho^2}}\left(1 - \sqrt{1-\rho^2}\frac{\bar{v}}{\zeta y}\right)$$

where we have used that $i^*dx = \sqrt{\frac{y}{\frac{\delta}{C^2\zeta} - y}}dy$. We deduce that the parallel gauge transport is

$$\ln\mathcal{P}(x,y) = \left(\frac{1}{2} - \frac{\rho\lambda}{\zeta}\right)\frac{x_2 - x_1}{1-\rho^2} + \frac{\lambda}{\zeta}\frac{y_2 - y_1}{\sqrt{1-\rho^2}} - \frac{\bar{v}\lambda}{\zeta^2}\ln\frac{y_2}{y_1}$$
$$+ \frac{2\rho\lambda\bar{v}}{\zeta^2\sqrt{1-\rho^2}}[\arctan\sqrt{\frac{y}{\frac{\delta}{C^2\zeta} - y}}]_{y_1}^{y_2}$$

Furthermore, we have

$$-\ln g'(a_{\min}) = -\frac{2}{a_{\min}}$$

From the expression for the geodesic distance and the parallel gauge transport, we can deduce a first-order asymptotics for the implied volatility. The Van Vleck-Morette determinant is computed numerically.

6.6 Problems

Exercises 6.1 Mixing solution and Hull-White decomposition
For general SVMs, there is no closed-form formula for European call options. To circumvent this difficulty, we can use asymptotic methods as described in this chapter. However, for long maturity date, such methods are no longer applicable and one needs to rely on Monte-Carlo (MC) simulation. When the forward conditional to the instantaneous volatility is a log-normal Itô process, the MC simulation can be considerably simplified: This is called the *Mixing solution* [29].
Let us consider the following SVM defined by

$$\begin{aligned} df_t &= a_t f_t \left(\rho dZ_t + \sqrt{1-\rho^2} dW_t\right) \\ da &= b(a_t)dt + \sigma(a_t)dZ_t \end{aligned}$$

with W_t and Z_t two independent Brownian motions and the initial conditions $f_{t=0} = f_0$ and $a_0 = \alpha$.

1. Using Itô's formula, prove that

$$d\ln f_t = -\frac{1}{2}a_t^2 dt + a_t \left(\rho dZ_t + \sqrt{1-\rho^2} dW_t\right)$$

2. Conditional to the path of the second Brownian Z_t (with filtration $\mathcal{F}^Z$), f_t is a log-normal process with mean m_t and variance V_t. Prove that

$$\begin{aligned} \tilde{f}_0 &\equiv \mathbb{E}[f_t|\mathcal{F}^Z] = f_0 e^{-\frac{1}{2}\rho^2 \int_0^t a_s^2 ds + \rho \int_0^t a_s dZ_s} \\ V_t &= (1-\rho^2) \int_0^t a_s^2 ds \end{aligned}$$

3. Deduce that the fair value $\mathcal{C}$ at time t of a European call option with strike K and maturity date $T \geq t$ is

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}}[\mathcal{C}^{\mathrm{BS}}(K, T, \sqrt{(1-\rho^2)\frac{1}{T-t}\int_t^T a_s^2 ds} | \tilde{f}_0, t)|\mathcal{F}_t] \tag{6.67}$$

with $\mathcal{C}^{\mathrm{BS}}$ the Black-Scholes formula as given by (3.1). This formula is called the mixing solution. For $\rho = 0$, we obtain the Hull-White decomposition [109]:

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}}[\mathcal{C}^{\mathrm{BS}}(K, T, \sqrt{\frac{1}{T-t}\int_t^T a_s^2 ds} | f_0, t)|\mathcal{F}_t] \tag{6.68}$$

From (6.67) and (6.68), we see that the MC pricing of a call option only requires the simulation of the volatility a_t. The forward f_t has been integrated out.

Exercises 6.2 Variance swap

Compute the variance swap for the Heston and SABR model:

$$\mathrm{VS}_T = \frac{1}{T}\int_0^T \mathbb{E}^{\mathbb{P}}[a_s^2] ds$$

Chapter 7

Multi-Asset European Option and Flat Geometry

Abstract A standard method to price a multi-asset European option incorporating an implied volatility is to use a local volatility Monte-Carlo computation. Although straightforward, this method is quite time-consuming, particularly when the number of assets is large and we evaluate the Greeks. Applying our geometrical framework to this multi-dimensional problem, we explain how to obtain accurate approximations of multi-asset European options.

We use the heat kernel expansion to obtain an asymptotic solution to the Kolmogorov equation for a n-dimensional local volatility model. The resulting manifold is the flat Euclidean space $\mathbb{R}^n$. We present two applications. The first application we look at is the derivation of an asymptotic implied volatility for a basket option. In particular, we try to reconstruct the basket implied volatility from the implied volatility of each asset. In the second application, we obtain accurate approximation for Collateralized Commodity Obligations (CCO), which are recent commodity derivatives that mimic the Collateralized Debt Obligations (CDO).

7.1 Local volatility models and flat geometry

In the forward measure $\mathbb{P}^T$, each forward f_t^i $(i = 1, \cdots, n)$ is a local martingale and we assume that they follow a local volatility model

$$df_t^i = C^i(t, f_t^i) dW_i \; ; \; dW_i dW_j = \rho_{ij} dt \tag{7.1}$$

with a deterministic rate and with the initial condition $f_{t=0}^i = f_0^i$. The metric (4.78) at $t = 0$ underlying this model is

$$ds^2 = 2\rho^{ij} \frac{df^i}{C^i(f^i)} \frac{df^j}{C^j(f^i)} \tag{7.2}$$

where we have set $C^i(f^i) \equiv C^i(0, f^i)$. By using the Cholesky decomposition, we write the inverse of the correlation matrix as

$$\rho^{-1} = L^\dagger L$$

or in components

$$\rho^{ij} = L_{ki} L_{kj}$$

By convention ρ^{ij} denotes the components of the inverse of the correlation matrix. $[L]_{ik}$ is a $n \times n$-matrix. Similarly the correlation ρ can be written as

$$\rho = L^{-1} \left(L^{-1}\right)^\dagger$$

or in components

$$\rho_{ij} = L^{ik} L^{jk}$$

Here L^{ij} are the components of the inverse of the Cholesky matrix L.
If we introduce the new coordinates

$$u^i(f) = L_{ij} \int_{f_0^j}^{f^j} \frac{dx^j}{C^j(x^j)} \tag{7.3}$$

we obtain that the metric (7.2) (at $t = 0$) is flat (the factor 2 is introduced for a convenience purpose)

$$ds^2 = 2du^i du^i$$

The geodesic distance between the two points $f_0 \equiv \{f_0^i\}_i$ and $f \equiv \{f^i\}_i$ is then given by the Euclidean distance

$$d(u)^2 = 2u.u \equiv 2\sum_{i=1}^n (u^i)^2 \tag{7.4}$$

After some algebraic manipulations, the connection $\mathcal{A}$ (4.58) which appears in the time-dependent heat kernel expansion (4.79) is given by

$$\mathcal{A} = -\frac{1}{2}\rho^{ij} \frac{\partial_j C^j(f^j)}{C^i(f^i)} df^i \tag{7.5}$$

In the new coordinates u, the connection (7.5) is given by

$$\mathcal{A} = -\frac{1}{2} L_{kj} L_{pj} \frac{\partial \ln(C^j(u))}{\partial u_p} du^k \tag{7.6}$$

Note that unlike the one-dimensional case, the connection (7.6) is not an exact form and to obtain the parallel gauge transport we need to pullback this form on a geodesic curve. The geodesic curve joining together the spot forward f^0 (i.e., $u = 0$) and the forward f (i.e., u) is a straight line (that we parameterize by $\lambda \in [0, 1]$) on the flat manifold $\mathbb{R}^n$

$$u^k(\lambda) = u^k \lambda$$

By expanding at the first-order $\frac{\partial \ln(C^j(u))}{\partial u_p}$ around $u = 0$, we deduce that the connection pullback to this curve is approximated by

$$\mathcal{A} \simeq -\frac{1}{2} L_{kj} \left(\partial_j C^j(f_0^j) + \lambda C^j(f_0^j) \partial_j^2 C^j(f_0^j) L^{jp} u^p \right) u^k d\lambda$$

From this expression, we derive the parallel gauge transport from the point f^0 to f

$$\mathcal{P}(f_0, f) \approx e^{\frac{1}{2} u^\dagger C u + u^\dagger B} \tag{7.7}$$

with the n-dimensional matrix $[C]_{kp}$ and the vector $[B]_k$ defined by

$$C_{kp} = \frac{1}{4} C^j(f_0^j) \partial_j^2 C^j(f_0^j)(L_{kj} L^{jp} + L_{pj} L^{jk}) \tag{7.8}$$

$$B_k = \frac{1}{2} L_{kj} \partial_j C^j(f_0^j) \tag{7.9}$$

By plugging the expression for the geodesic distance (7.4) and the parallel gauge transport (7.7) in the time-dependent heat kernel expansion (4.82), we obtain the conditional probability density $p(t, u|0)$ at the first-order

$$p(t, u|0) = \frac{e^{\frac{-u^\dagger (\mathbb{I} - tC) u}{2t} - u^\dagger B}}{(2\pi t)^{\frac{n}{2}}} \left(1 + a_1(u, 0)t + o\left(t^2\right)\right) \tag{7.10}$$

We will use this asymptotic solution to price a basket option and a *Collateralized Commodity Obligation* (CCO). Note the minus sign in front of B as f_0 is written as f in our geometric notation (see remark 4.3). As for the first heat kernel coefficient $a_1(u, 0)$ between the point u and $u = 0$, it is not written explicitly as it does not take any part in the calculation.

7.2 Basket option

A basket option is a European option the payoff of which is linked to a portfolio or "basket" of assets. The basket can be any weighted sum of assets as long as all the weights are positive. Basket options are usually cash settled. A classical example is a call option on the France CAC 40 stock index.

Basket options are also popular for hedging foreign exchange risk. A corporation with multiple currency exposures can hedge the combined exposure less expensively by purchasing a basket option than by purchasing options in each currency individually.
The payoff of a basket option, based on an index of n assets, with strike K and maturity T, is given by

$$\max\left(\sum_{i=1}^{n}\omega_i S_T^i - K, 0\right)$$

with S_T^i the stock price at maturity and ω_i the weight of the asset i. By convention, the weights ω_i are normalized by $\sum_{i=1}^{n}\omega_i = 1$. Let us denote the following dimensionless parameters

$$\begin{aligned}\hat{\omega}_i &= \frac{\omega_i S_0^i}{\sum_{i=1}^{n}\omega_i S_0^i}; \quad \hat{K} = \frac{K}{\sum_{i=1}^{n}\omega_i S_0^i} \\ \hat{f}_t^i &= \frac{S_t^i}{S_0^i P_{tT}}\ ;\ \hat{A}_t = \sum_{i=1}^{n}\hat{\omega}_i\hat{f}_t^i \end{aligned} \tag{7.11}$$

with S_0^i the spot price of asset i, S_t^i the price at time t and P_{tT} the value of the bond quoted at t and expiring at T. In the forward measure $\mathbb{P}^T$, the fair value of a basket call option equals

$$\mathcal{C} = P_{0T}\mathbb{E}^{\mathbb{P}^T}[\max\left(\sum_{i=1}^{n}\omega_i S_T^i - K, 0\right)|\mathcal{F}_0] \tag{7.12}$$

$$= \sum_{i=1}^{n}\omega_i S_0^i P_{0T}\mathbb{E}^{\mathbb{P}^T}[\max\left(\hat{A}_T - \hat{K}, 0\right)|\mathcal{F}_0] \tag{7.13}$$

This representation (7.13) is introduced because it appears to give a better approximation of a basket option than formula (7.12).
Basket options are usually priced by treating the basket value $\hat{A}_T$ as a single asset satisfying a log-normal process with a constant volatility σ_{BS}. Therefore, the fair value (7.13) becomes equivalent to the fair value of a European call option on a single asset following a log-normal process. This leads to a Black-Scholes formula

$$\mathcal{C} = \sum_{i=1}^{n}\omega_i S_0^i P_{0T}\mathcal{C}^{\mathrm{BS}}(K, T, \sigma_{\mathrm{BS}}|\hat{A}_0, 0) \tag{7.14}$$

From this formula, we can define the *basket implied volatility* as the volatility σ_{BS} which when put in the formula above reproduces the market price. We show in the following how to derive an asymptotic implied volatility for a basket in the context of a multi-dimensional local volatility model (7.1). We will check this result against a Monte-pricer and the classical second-moment matching approximation that we review in the next section. Note that even if each asset has a constant volatility (i.e., no skew), the resulting basket implied volatility is skewed by the fact that a weighted sum of log-normal random variables is not log-normal. More generally, we try to understand how to construct the basket implied volatility from the implied volatility of each of its constituents. This problem was explored for the first time in [50].

7.2.1 Basket local volatility

We assume that each forward $f_t^i = \frac{S_t^i}{P_{tT}}$ follows a local volatility model (7.1). The process $\hat{f}_i$ (see Eq. (7.11)) satisfies the SDE under $\mathbb{P}^T$

$$d\hat{f}_t^i = \hat{C}^i(t, \hat{f}_t^i)dW_i \; ; \; dW_i dW_j = \rho_{ij} dt \tag{7.15}$$

with $\hat{C}^i(t, \hat{f}_t^i) = \frac{C^i(t, S_0^i \hat{f}_t^i)}{S_0^i}$. Being the sum of all assets, the index $\hat{A}_t$ is a traded asset and the process $\hat{A}_t$ is a local martingale under the forward measure $\mathbb{P}^T$. From the SDEs (7.15), the process $\hat{A}_t$ satisfies the SDE under $\mathbb{P}^T$

$$d\hat{A}_t = \sum_{i=1}^n \hat{\omega}_i \hat{C}^i(t, \hat{f}_t^i) dW_i \tag{7.16}$$

As discussed in section (6.1.3) in chapter 6 the process (7.16) has the same marginals as the local volatility model

$$d\hat{A}_t = C(t, \hat{A}_t) dW_t \tag{7.17}$$

where the square of the Dupire local volatility function $C(t, \hat{A})$ is equal to the square of the volatility in (7.16) when the index $\hat{A}$ is fixed to $\sum_{i=1}^N \hat{\omega}_i \hat{f}_t^i$ at time t

$$C(t, \hat{A})^2 = \sum_{i,j=1}^n \rho_{ij} \hat{\omega}_i \hat{\omega}_j \mathbb{E}^{\mathbb{P}^T} [\hat{C}^i(t, \hat{f}_t^i) \hat{C}^j(t, \hat{f}_t^j) | \hat{A} = \sum_{i=1}^n \hat{\omega}_i \hat{f}_t^i] \tag{7.18}$$

For a completeness purpose, we give another proof.

PROOF Applying Itô-Tanaka's lemma on the payoff $\max(\hat{A}_t - K, 0)$, we obtain

$$d\max(\hat{A}_t - K, 0) = \frac{1}{2}\delta(\hat{A}_t - K) \sum_{i,j=1}^n \rho_{ij} \hat{\omega}_i \hat{\omega}_j \hat{C}^i(t, \hat{f}_t^i) \hat{C}^j(t, \hat{f}_t^j) dt$$
$$+ 1(\hat{A}_t - K) d\hat{A}_t$$

Taking the operator $\mathbb{E}^{\mathbb{P}^t}[\cdot | \mathcal{F}_0]$ on both sides of this equation, we have

$$d\mathbb{E}^{\mathbb{P}^t}[\max(\hat{A}_t - K, 0) | \mathcal{F}_0] = \frac{1}{2} \sum_{i,j=1}^n \hat{\omega}_i \hat{\omega}_j \rho_{ij} \mathbb{E}^{\mathbb{P}^t} [\hat{C}^i \hat{C}^j \delta(\hat{A}_t - K)] dt \tag{7.19}$$

By definition of Dupire local volatility function, we have

$$d\mathbb{E}^{\mathbb{P}^t}[\max(\hat{A}_t - K, 0)|\mathcal{F}_0] = \frac{1}{2}C(t,K)^2\mathbb{E}^{\mathbb{P}^t}[\delta(\hat{A}_t - K)]dt \tag{7.20}$$

Identifying the two equations (7.19) and (7.20), we obtain the expected relation (7.18). ∎

By definition of the mean value in (7.18), the basket local volatility is

$$C(t,\hat{A})^2 = \sum_{i,j=1}^{n} \rho_{ij}\hat{\omega}_i\hat{\omega}_j \frac{\int_{\mathbb{B}} \hat{C}^i(t,\hat{f}^i)\hat{C}^j(t,\hat{f}^j)p(t,\hat{f}|\hat{f}_0)\prod_k d\hat{f}_k}{\int_{\mathbb{B}} p(t,\hat{f}|\hat{f}_0)\prod_k d\hat{f}_k} \tag{7.21}$$

with $p(t,\hat{f}|\hat{f}_0)$ the conditional probability associated to the SDEs (7.15) at t and $\mathbb{B}$ the hyperplane defined by the linear equation $\sum_{i=1}^n \hat{\omega}_i\hat{f}^i = \hat{A}$. As usual the computation of an asymptotic implied volatility will be done in two steps. First of all, from the definition (7.21), we calculate a local volatility at the first-order in time using our first-order asymptotic conditional probability density (7.10). Then, as a second step, we use our map between an asymptotic implied volatility and an asymptotic local volatility (5.40).

PROPOSITION 7.1
The basket local volatility is approximated at the first-order in time by

$$C(t,\hat{A})^2 = \sum_{i,j=1}^{n} \rho_{ij}\hat{\omega}_i\hat{\omega}_j\hat{C}^i(t,\hat{f}_i^*)\hat{C}^j(t,\hat{f}_j^*)\left(1+\frac{t}{2}\mathrm{Tr}(M^{ij}D_0)\right)$$

with

$$[M]^{ij}_{tp} = \partial_i\left(\hat{C}_i\partial_i\hat{C}_i\right)L^{it}L^{ip} + 2\partial_i\hat{C}_i\partial_j\hat{C}_jL^{ip}L^{jt} + \partial_j\left(\hat{C}_j\partial_j\hat{C}_j\right)L^{jt}L^{jp} \;|_{\hat{f}_i=\hat{f}_i^0} \tag{7.22}$$

$$D_t = (1-tC)^{-1} - \frac{(1-tC)^{-1}\bar{\omega}\bar{\omega}^\dagger(1-tC)^{-1}}{\left(\bar{\omega}^\dagger(1-tC)^{-1}\bar{\omega}\right)} \tag{7.23}$$

$$[\bar{\omega}]_{i=1,\cdots,n} = \sum_{j=1}^{n}\hat{\omega}_j\hat{f}_0^j\sigma^j_{\mathrm{BS}}L^{ji}$$

$$\frac{\hat{f}_i^*}{\hat{f}_i^0} = e^{\sigma^i_{\mathrm{BS}}L^{ip}(u^*)^p}$$

$$u^* = (1-tC)^{-1}\left(-tB + \bar{\omega}\frac{\left(\hat{A}-\hat{A}_0+t\bar{\omega}^\dagger(1-tC)^{-1}B\right)}{\left(\bar{\omega}^\dagger(1-tC)^{-1}\bar{\omega}\right)}\right) \tag{7.24}$$

$[C]$ and $[B]$ are given by (7.8), (7.9). σ^i_{BS} denotes the ATM implied volatility for asset i and $\hat{C}_i \equiv \hat{C}_i(0,\hat{f}_i)$.

Note that $D_0 \equiv \lim_{t \to 0} D_t = 1 - \frac{\bar{\omega}\bar{\omega}^\dagger}{\bar{\omega}^\dagger \bar{\omega}}$.

PROOF From (7.21), the local volatility function is

$$C(t, \hat{A})^2 = \sum_{i,j=1}^{n} \rho_{ij} \hat{\omega}_i \hat{\omega}_j \frac{\int_{\mathbb{B}} \hat{C}^i(t,u) \hat{C}^j(t,u) p(t,u|0) du}{\int_{\mathbb{B}} p(t,u|0) du}$$

where $du \equiv \prod_{i=1}^{n} du_i$. By plugging the first-order solution to the backward Kolmogorov equation as given in (7.10) in the equation above, we have

$$C(t, \hat{A})^2 = \sum_{i,j=1}^{n} \rho_{ij} \hat{\omega}_i \hat{\omega}_j \frac{\int_{\mathbb{B}} \hat{C}^i(t,u) \hat{C}^j(t,u) e^{-\frac{1}{2t} u^\dagger (1 - tC) u - u^\dagger B} du}{\int_{\mathbb{B}} e^{-\frac{1}{2t} u^\dagger (1 - tC) u - u^\dagger B} du}$$

Note that at this order, the first heat kernel coefficient does not contribute. The new coordinates u for which the metric at $t = 0$ is flat is

$$u^i = \sum_{j=1}^{n} L_{ij} \int_{\hat{f}_0^j}^{\hat{f}^j} \frac{dx}{\hat{C}^j(0,x)}$$

By using the BBF formula (5.39), we have

$$\ln \left(\frac{\hat{f}^j}{\hat{f}_0^j} \right) = \sum_{i=1}^{n} \sigma_{\mathrm{BS}}^j(0, \hat{f}^j) L^{ji} u^i$$

By assuming that $\ln \left(\frac{\hat{f}^j}{\hat{f}_0^j} \right) \approx \frac{\hat{f}^i}{\hat{f}_0^i} - 1$, we obtain the relation

$$\begin{aligned}
\hat{A} - \hat{A}_0 &= \sum_{j=1}^{n} \hat{\omega}_j \hat{f}_0^j \left(\frac{\hat{f}^j}{\hat{f}_0^j} - 1 \right) \approx \sum_{j=1}^{n} \hat{\omega}_j \hat{f}_0^j \ln \left(\frac{\hat{f}^j}{\hat{f}_0^j} \right) \\
&= \sum_{i,j=1}^{n} \hat{\omega}_j \hat{f}_0^j \sigma_{\mathrm{BS}}^j(0, \hat{f}^j) L^{ji} u^i \approx \sum_{i,j=1}^{n} \hat{\omega}_j \hat{f}_0^j \sigma_{\mathrm{BS}}^j(0, \hat{f}_0^j) L^{ji} u^i
\end{aligned}$$

Setting $\beta_j = \hat{\omega}_j \hat{f}_0^j \sigma_{\mathrm{BS}}^j(0, \hat{f}_0^j)$ and $\bar{\omega}_i = \sum_{j=1}^{n} \beta_j L^{ji}$, the hyperplane $\mathbb{B}$ is then represented by the following linear form (in the u coordinates)

$$\mathbb{B} : u^\dagger \bar{\omega} = \hat{A} - \hat{A}_0$$

The Dirac function over the submanifold $\mathbb{B}$ can be replaced by its Fourier transform

$$\delta \left(u^\dagger \bar{\omega} - \hat{A} + \hat{A}_0 \right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\imath q \left(u^\dagger \bar{\omega} - \hat{A} + \hat{A}_0 \right)} dq$$

We obtain

$$C(t,\hat{A})^2 = \sum_{i,j=1}^{n} \rho_{ij}\hat{\omega}_i\hat{\omega}_j \frac{\int_{\mathbb{R}^{n+1}} \hat{C}^i(t,u)\hat{C}^j(t,u) e^{-\frac{1}{2t}u^\dagger(1-tC)u - u^\dagger B + \imath q\left(u^\dagger\bar{\omega} - \hat{A} + \hat{A}_0\right)} du dq}{\int_{\mathbb{R}^{n+1}} e^{-\frac{1}{2t}u^\dagger(1-tC)u - u^\dagger B + \imath q\left(u^\dagger\bar{\omega} - \hat{A} + \hat{A}_0\right)} du dq}$$

Let us introduce the partition function

$$Z[j] \equiv \int_{\mathbb{R}^{n+1}} e^{-\frac{1}{2t}u^\dagger(1-tC)u - u^\dagger B + \imath q\left(u^\dagger\bar{\omega} - \hat{A} + \hat{A}_0\right) + j^\dagger(u-u^*)} du dq$$

where u^* is a n-dimensional vector that we specify below. By using the Gaussian integration formula

$$\int_{\mathbb{R}^n} dX e^{-\frac{1}{2}X^\dagger AX + j^\dagger X} = \left(\frac{(2\pi)^{\frac{n}{2}}}{\sqrt{\det A}}\right) e^{\frac{1}{2}j^\dagger A^{-1}j}$$

we obtain

$$\chi[j] \equiv \frac{Z[j]}{Z[0]} = e^{\frac{t}{2}j^\dagger D_t j + j^\dagger(T-u^*)}$$

with D_t a $n \times n$-dimensional matrix and T a n-dimensional vector given respectively by (7.23) and (7.24). We choose $u^* = T$. This gives

$$\chi[j] = e^{\frac{t}{2}j^\dagger D_t j}$$

By expanding $\Phi(u) = \hat{C}^i(t,u)\hat{C}^j(t,u)$ at the second-order in u around $u = u^*$, we have

$$\Phi(u) = \Phi(u^*) + \nabla_u\Phi(u^*)(u-u^*) + \frac{1}{2}(u-u^*)^\dagger\nabla_u^2\Phi(u^*)(u-u^*)$$

Finally, we obtain

$$\begin{aligned} C(t,\hat{A})^2 &= \sum_{i,j=1}^{n} \rho_{ij}\hat{\omega}_i\hat{\omega}_j \\ &\quad \left(\Phi(u^*) + \nabla_u\Phi(u^*)^\dagger \frac{\partial\chi[j]}{\partial j}(j=0) + \frac{1}{2}\mathrm{Tr}[\nabla_u^2\Phi(u^*)\frac{\partial^2\chi[j]}{\partial j^2}(j=0)]\right) \\ &= \sum_{i,j=1}^{n} \rho_{ij}\hat{\omega}_i\hat{\omega}_j \hat{C}^i(t,u^*)\hat{C}^j(t,u^*)\left(1 + \frac{t}{2}\mathrm{Tr}[\frac{\nabla_u^2\hat{C}^i\hat{C}^j}{\hat{C}^i\hat{C}^j}D_t]\right) \end{aligned}$$

An explicit computation shows that $\frac{\nabla_u^2\hat{C}^i\hat{C}^j}{\hat{C}^i\hat{C}^j}$ is given by (7.22). At the first-order, D_t can be replaced by D_0.

□

REMARK 7.1 Correlation smile By analogy with the smile, the correlation smile is defined as the correlation (matrix) $\rho_{ij}(K,T)$ that must be plugged in the Black-Scholes formula (7.14) in order to reproduce the market price of basket option with a strike K and a maturity T. We assume that each asset is calibrated to its implied volatility surface. By analogy with the local volatility, we can match the correlation smile by introducing a local correlation matrix [121] which depends on the time t and the level of the basket $\hat{A}_t$ by

$$\rho_{ij}(t,\hat{A}_t) = \bar{\rho}(t,\hat{A}_t)\rho_{ij}^{\mathrm{hist}} + \left(1 - \bar{\rho}(t,\hat{A}_t)\right)$$

$\rho_{ij}^{\mathrm{hist}}$ is the historical correlation matrix and $\bar{\rho}(t,\hat{A}_t) \in [0,1]$. The matrix ρ_{ij} is still definite positive if $\rho_{ij}^{\mathrm{hist}}$ is so. From our previous computation, a good guess for $\bar{\rho}(t,\hat{A}_t)$ is

$$\bar{\rho}(t,\hat{A}_t) = \frac{C_{\mathrm{mkt}}(t,\hat{A})^2 - \sum_{i,j=1}^n \hat{\omega}_i\hat{\omega}_j \hat{C}^i_{\mathrm{mkt}}(t,f_i^*)\hat{C}^j_{\mathrm{mkt}}(t,f_j^*)}{\sum_{i,j=1}^n \left(\rho_{ij}^{\mathrm{hist}} - 1\right)\hat{\omega}_i\hat{\omega}_j \hat{C}^i_{\mathrm{mkt}}(t,f_i^*)\hat{C}^j_{\mathrm{mkt}}(t,f_j^*)}$$

where C_{mkt} (resp. $\hat{C}^i_{\mathrm{mkt}}$) is the local volatility derived from the basket (asset) implied volatility via the Dupire formula. Note that as the LV $C(t,\hat{A})$ is computed by conditioning on $\hat{A}$, our asymptotic LV computed using a constant correlation $\rho_{ij}^{\mathrm{hist}}$ is still valid. □

In the next section, we present a classical approximation for the value of a basket option in the case when each asset has a constant volatility. We will check this approximation against our asymptotic implied volatility.

7.2.2 Second moment matching approximation

Let us assume that each forward $\hat{f}_t^i$ follows a log-normal process (with a volatility σ_i) and denote the geometric average by

$$\hat{G}_t = \prod_{i=1}^n \left(\hat{f}_t^i\right)^{\hat{\omega}_i}$$

By using this new variable, we rewrite the fair value (7.13) as

$$\mathcal{C} = P_{0T}\sum_{i=1}^n \hat{\omega}_i S_0^i \mathbb{E}^{\mathbb{P}^T}[\max(\hat{G}_T - (-\hat{A}_T + \hat{G}_T) - \hat{K}, 0)|\mathcal{F}_0]$$

We approximate the expression above by

$$\mathcal{C} \approx P_{0T}\sum_{i=1}^n \hat{\omega}_i S_0^i \mathbb{E}^{\mathbb{P}^T}[\max(\hat{G}_T - \widetilde{K}, 0)|\mathcal{F}_0]$$

where we have introduced the modified strike

$$\widetilde{K} \equiv \hat{K} + \mathbb{E}^{\mathbb{P}^T}[\hat{G}_T - \hat{A}_T|\mathcal{F}_0]$$

From the Jensen inequality which states that

$$\mathbb{E}[\Phi(S_T)|\mathcal{F}_0] \geq \Phi\left(\mathbb{E}[S_T|\mathcal{F}_0]\right)$$

for a convex function Φ, we deduce that our approximation gives a lower bound.
$\hat{G}_t$ is a log-normal process and $\hat{G}_T$ is given by

$$\hat{G}_T = P_{0T}^{-1} e^{W_T - \frac{c_2 T}{2}}$$

with $W_T = \sum_{i=1}^n \hat{\omega}_i \sigma_i W_T^i$ and $c_2 = \sum_{i=1}^n \hat{\omega}_i \sigma_i^2$. W_T follows a normal law with a zero mean and a variance $v^2 = \sum_{i,j=1}^n \rho_{ij} \hat{\omega}_i \hat{\omega}_j \sigma_i \sigma_j$. Therefore, the modified strike $\widetilde{K}$ is

$$\widetilde{K} = \hat{K} + P_{0T}^{-1}\left(e^{\frac{(v^2 - c_2)T}{2}} - 1\right) \tag{7.25}$$

Finally, we obtain the lower bound

PROPOSITION 7.2
The approximate value of the price $\mathcal{C}$ of a basket call option with strike price K and expiry date T equals

$$\mathcal{C} = P_{0T} \sum_{i=1}^n \omega_i S_i^0 \mathcal{C}_{\mathrm{BS}}(\tilde{K}, T, \sigma|\hat{G}_0)$$

with $\sigma^2 = \sum_{i,j=1}^k \rho_{ij} \hat{\omega}_i \hat{\omega}_j \sigma_{\mathrm{BS}}^i \sigma_{\mathrm{BS}}^j$ and $\hat{G}_0 = P_{0T}^{-1}$.

By construction, for $k = 1$, the formula above reduces to the Black-Scholes formula. In the graphs 7.1, 7.2 and 7.3, we compare our asymptotic implied volatility against the second moment-matching formula and an exact Monte-Carlo pricing.

7.3 Collaterized Commodity Obligation

The payoff of a Collaterized Commodity Obligation (CCO) is given by

$$\min(\max(\theta_T, K_{\min}), K_{\max})$$

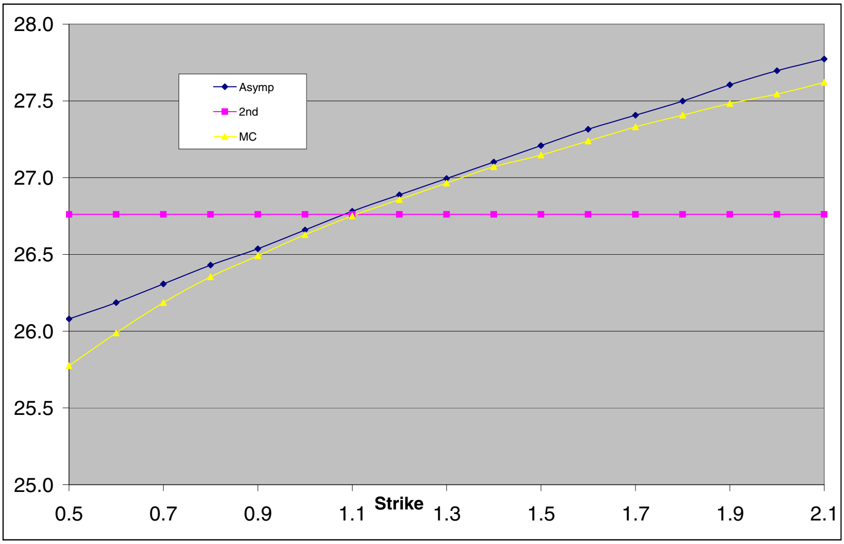

FIGURE 7.1: Basket implied volatility with constant volatilities. $\rho_{ij} = e^{-0.3|i-j|}$, $\sigma_i = 0.1 + 0.1 \times i$, $i, j = 1, 2, 3$, $T = 5$ years.

with

$$\theta_T = \sum_{i=1}^{n} 1(K^i - f_T^i)$$

with f_T^i the forward of asset i at the maturity T. $K_{\min}$ is a global floor and $K_{\max}$ a global cap. By using a static replication, we have

$$\min(\max(\theta_T, K_{\min}), K_{\max}) = K_{\min} + \max(\theta_T - K_{\min}, 0) - \max(\theta_T - K_{\max}, 0)$$

In order to price the CCO, we need to find the fair value for a European call option with the payoff $\max(\theta_T - K, 0)$.
A CCO mimics a Collaterized Debt Obligation (CDO) whose payoff is linked to a call option on the loss default of a basket

$$\theta_T = \sum_{i=1}^{n} 1(T - \tau_i)$$

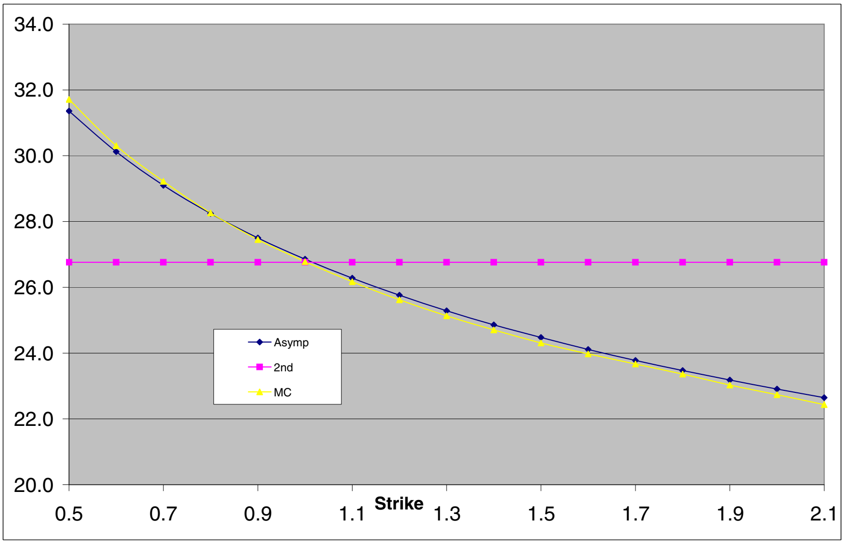

FIGURE 7.2: Basket implied volatility with CEV volatilities. $\rho_{ij} = e^{-0.3|i-j|}$, $\sigma_i = 0.1 + 0.1 \times i$, $i, j = 1, 2, 3$, $\beta_{\mathrm{CEV}} = 0.5$, $T = 5$ years.

where τ_i is the time when the asset i defaults.
Using the fact that θ_T can take only discrete values from 0 to n, we obtain

$$\begin{aligned}\mathbb{E}[\max(\theta_T - K, 0)] &= \int \prod_{i=1}^{n} df^i \max(\theta_T - K, 0) p(T, f|f_0) \\ &= \sum_{i=1}^{n} \max(i - K, 0) \int_{\mathcal{H}^i} \prod_{i=1}^{n} df^i p(T, f|f_0)\end{aligned}$$

where the hyperplane $\mathcal{H}^i$ is defined as

$$\mathcal{H}^i = \{\{f^k\}_{k=1,\cdots,n} | \sum_{j=1}^{n} 1(K^j - f^j) = i\}$$

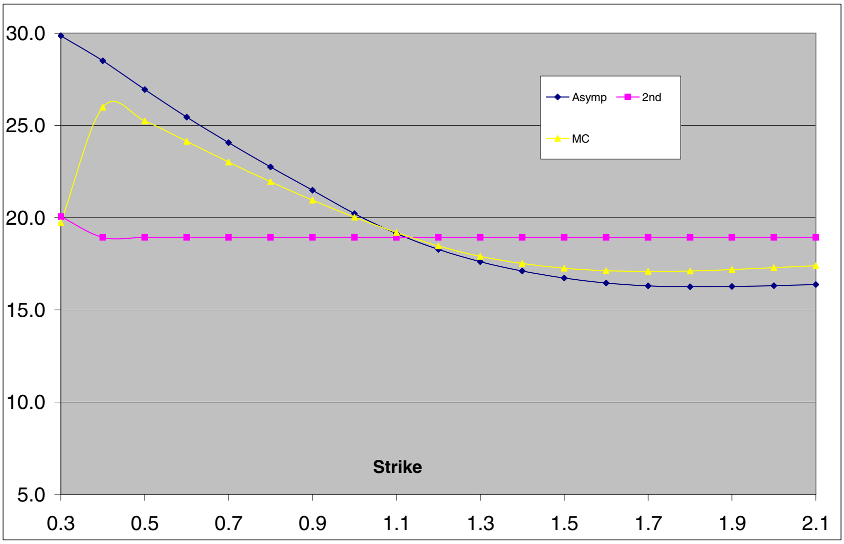

FIGURE 7.3: Basket implied volatility with LV volatilities (NIKKEI-SP500-EUROSTOCK, 24-04-2008). $\rho_{ij} = e^{-0.4|i-j|}$, $T = 3$ years.

It can be partitioned in the following way

$$\mathcal{H}^i = \bigcup_{\text{cyclic}_i} [0, K^1] \times \cdots \times [0, K^i] \times [K^{i+1}, \infty] \times \cdots \times [K^n, \infty]$$

where cyclic_i means that we sum over the configuration such that the number of variables which take their value below their strike is i. For example for two assets, we have

$$\mathcal{H}^1 = [0, K^1] \times [K^2, \infty] \cup [K^1, \infty] \times [0, K^2]$$

Using this characterization of the hyperplane $\mathcal{H}^i$, we have

$$\mathbb{E}[\max(\theta_T - K, 0)] = \sum_{i=1}^{n} \max(i - K, 0) \sum_{\text{cyclic}_i}$$
$$\int_0^{K^1} df^1 \cdots \int_0^{K^i} df^i \int_{K^{i+1}}^{\infty} df^{i+1} \cdots \int_{K^n}^{\infty} df^n p(T, f|f_0)$$

Using the fact that

$$\int_0^{K^i} p(T,f|f_0)df^i = p(T,\bar{f}|\bar{f}_0) - \int_{K^i}^{\infty} p(T,f|f_0)df^i$$

with $\bar{f} \equiv (f^1, \cdots f^{i-1}, f^{i+1}, \cdots f^n)$, we obtain

$$\mathbb{E}[\max(\theta_T - K, 0)] = \sum_{i=1}^{n} \max(i-K,0) \sum_{k=0}^{i} C_{n-k}^{n-i} (-)^{k+i} \sum_{\text{cyclic}_{n-k}} \Phi_{n-k}$$

with

$$\Phi_i(K^1, \cdots, K^i) = \int_{K^1}^{\infty} df^1 \cdots \int_{K^i}^{\infty} df^i p(T, f^1, \cdots f^i | f_0^1, \cdots f_0^i) \quad (7.26)$$

As a next step, we find an approximation of the functions Φ_i using the heat kernel expansion. Before moving on to the derivation, we consider the case of zero correlation for which there is a model-independent exact solution depending on the fair value of digital European call-put options.

7.3.1 Zero correlation

Let us assume that the correlation is zero (i.e., $\rho_{ij} = \delta_{ij}$). The probability density $p(T,f|f_0)$ becomes the product of the individual probability densities for each asset

$$p(T,f|f_0) = \prod_{i=1}^{n} p_i(T, f^i | f_0^i)$$

Moreover, we have

$$p_i(T,K|f_0^i) = \frac{\partial^2 C_i(K,T)}{\partial K^2}$$

with $C_i(K,T)$ a European option of strike K^i and maturity date T for asset i. Then

$$\begin{aligned}
\int_{\mathcal{H}^i} \prod_{i=1}^{n} df^i p(T,f|f_0) &= \sum_{\text{cyclic}} \prod_{j=1}^{i} \left(\int_0^{K^j} d\hat{K}^j \frac{\partial^2 C_j(\hat{K}^j, T)}{\partial \hat{K}^{j\,2}} \right) \\
&\quad \prod_{j=i+1}^{n} \left(\int_{K^j}^{\infty} d\hat{K}^j \frac{\partial^2 C_j(\hat{K}^j, T)}{\partial \hat{K}^{j\,2}} \right) \\
&= \sum_{\text{cyclic}} \prod_{j=1}^{i} \left(1 + \frac{\partial C_j(K^j,T)}{\partial K^j} \right) \prod_{j=i+1}^{n} \left(-\frac{\partial C_j(K^j,T)}{\partial K^j} \right)
\end{aligned}$$

Finally, we have

PROPOSITION 7.3
In the zero correlation case, $\mathbb{E}[\max(\theta_T - K, 0)]$ *is exactly given by*

$$\mathcal{C} = \sum_{i=1}^{n} \max(i - K, 0) \sum_{\text{cyclic}_i} \prod_{j=1}^{i} (1 - DC_j) \prod_{j=i+1}^{n} (DC_j) \tag{7.27}$$

with $DC_i = -\frac{\partial C(K^i, T)}{\partial K^i}$ *a digital call of strike* K_i *and maturity* T *for asset* i.

Using the market-value for European call options, the above-stated formula gives an exact solution to the fair price of the CCO in zero correlation cases. For two assets, the formula (7.27) reduces to

Example 7.1 2 assets

$$\begin{aligned} C &= \max(1 - K, 0)(DC_1 + DC_2 - 2DC_1 DC_2) \\ &\quad + \max(2 - K, 0)(1 - DC_1)(1 - DC_2) \end{aligned}$$

$\square$

7.3.2 Non-zero correlation

In the following, we will assume that $K^i < f_0^i$ for all i as it is the case in practice. The functions Φ_i (7.26) are computed using a saddle-point method. The saddle points f_* are defined by

$$\min_f d^2(u)$$

where d is the geodesic distance (7.4). The distance is minimized for

$$f_* \equiv (f_*^1, \cdots, f_*^i) = (f_0^1, \cdots, f_0^i)$$

which is a global minimum. The functions Φ_i can be computed as the conditional probability at the zero-order is a Gaussian distribution in the u coordinates. Therefore the saddle-point method consists in expanding the non-Gaussian part at the second-order around the point $f^* = f_0$. For this computation, we shall need the non-diagonal heat kernel coefficient $a(0, u)$. As usual, we will approximate the first heat kernel coefficient by its diagonal part given by

$$a(f, f_0) = Q(f_{\text{av}}) - \frac{\mathcal{G}(f_{\text{av}})}{4}$$

with

$$Q(f) = -\frac{1}{8}\rho^{ij}\partial_i C_i(0, f^i)\partial_j C_j(0, f^j) + \frac{1}{4}C_i(0, f^i)\partial_i^2 C^i(0, f^i) \tag{7.28}$$

$$\mathcal{G}(f) = 2(C^i)^{-1}(0, f^i)C^i_{,t}(0, f^i) \tag{7.29}$$

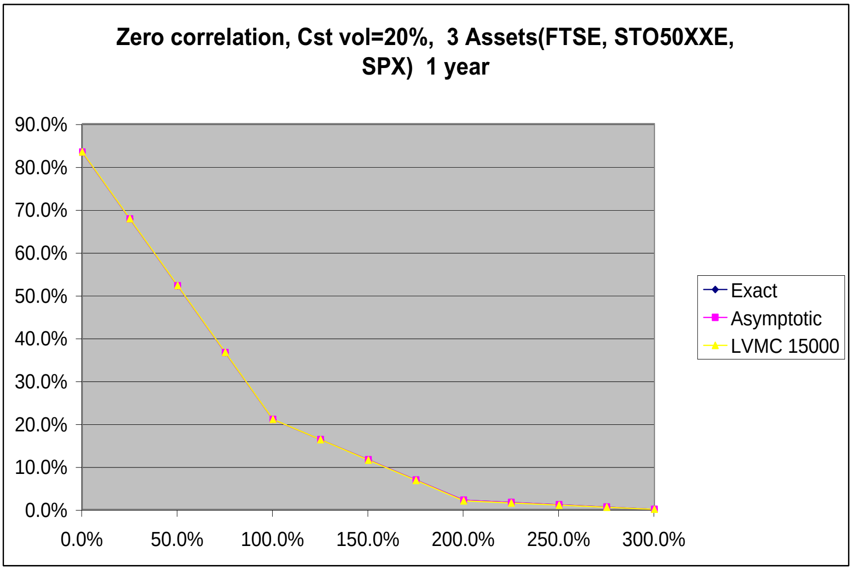

FIGURE 7.4: CCO. Zero-correlation case.

Finally, the asymptotic fair value of the CCO is

PROPOSITION 7.4
At the second-order, $\mathbb{E}[\max(\theta_T - K, 0)]$ *is given by*

$$\mathbb{E}[\max(\theta_T - K, 0)] = \sum_{i=1}^{n} \max(i - K, 0) \sum_{k=0}^{i} \sum_{\text{cyclic}_{n-k}} \Phi_{n-k}(-)^{k+i} C_{n-k}^{n-i}$$

with

$$\Phi_n(K^1, \cdots, K^s) = \det[A\rho^{-1}] N_n \left(\frac{\ln(\frac{f_0^j}{K^j})}{\sigma_{\text{BS}}^j(K^j)\sqrt{T}} - (AB)_j\sqrt{T}, A \right)$$
$$(1 + (Q - \frac{\mathcal{G}}{4} + \frac{B^T AB}{2})T)$$

with

$$N_n(\{X_i\}_i, \rho) \equiv \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{X_1} \cdots \int_{-\infty}^{X_n} \frac{e^{-\rho^{ij}\frac{x_i x_j}{2}}}{\det(\rho)} \prod_i dx_i$$

the n*-th cumulative multi-Gaussian distribution* $\mathcal{N}(\rho)$*, and with*

$$B_i = \frac{1}{2}\rho^{ij}\partial_j C^j(0, f_0^j)$$
$$A_{ij} = \rho_{ij}\left(1 - \frac{T}{4}(C^j(0, f_0^j)\partial^2 C^j(0, f_0^j) + C_i(0, f_0^i)\partial_i^2 C_i(0, f_0^i))\right)$$

To build up our confidence in this asymptotic formula, we apply it in the case of a simple digital call option and a constant volatility $C(f) = \sigma_0 f$. We derive

Example 7.2 Constant volatility

$$DC_K = N\left(\frac{\ln\left(\frac{f^0}{K}\right)}{\sigma_0\sqrt{T}} - \frac{\sigma_0\sqrt{T}}{2}\right)$$

This expression is the exact BS solution. ▯

7.3.3 Implementation

We have tested the validity of this analytical approximation against a Monte-Carlo pricer. As per figures 7.5 and 7.6, the approximation is accurate.

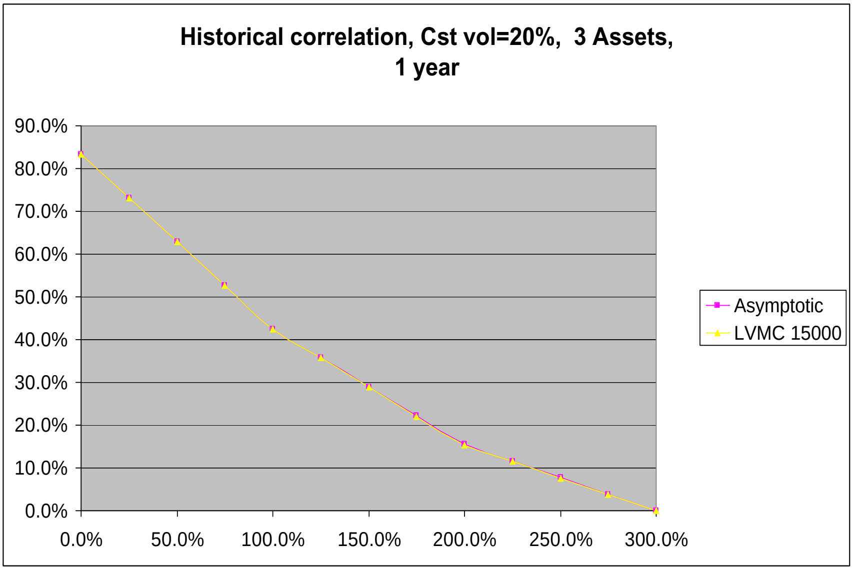

FIGURE 7.5: Exact versus asymptotic prices. 3 Assets, 1 year.

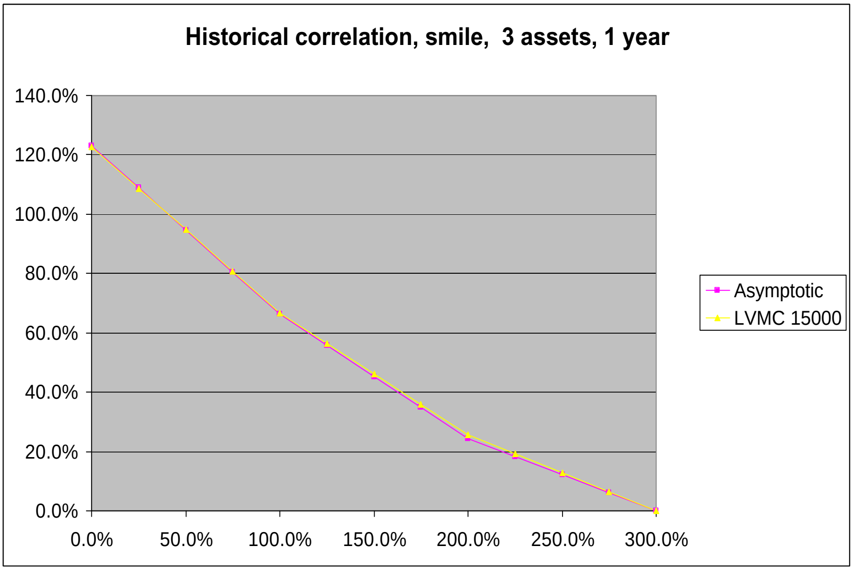

FIGURE 7.6: Exact versus asymptotic prices. 3 Assets, 5 years.

Chapter 8

Stochastic Volatility Libor Market Models and Hyperbolic Geometry

Abstract In this chapter, we focus on the last generation of interest-rate models, the Libor Market Models (LMMs). After a quick review of LMMs, we discuss the calibration of such models. In practice, this model is calibrated to a swaption matrix\cube. As a Monte-Carlo (MC) calibration routine is fairly time-consuming and noisy, the calibration requires an analytical approximation for the swaption implied volatility. For the BGM-LMM, some approximations have been found which are based on the Hull-White-Rebonato *freezing* arguments.
Using our geometrical framework, we give a justification for the *freezing* arguments and derive a more accurate asymptotic swaption implied volatility at the first-order for general stochastic volatility LMMs. We apply this formula to a specific model where the forward rates are assumed to follow a multi-dimensional CEV process correlated to a unique SABR process. The geometry underlying this model is the hyperbolic manifold $\mathbb{H}^{n+1}$ with n the number of Libor forward rates.
Parts of this research were published by the author in Risk magazine [104].

8.1 Introduction

The BGM model [67, 111] has recently been the focus of much attention as it gives a theoretical justification for pricing caps-floors using the classical Black-Scholes formula. The BGM model is a special version of the LIBOR Market Model (LMM), in which all forward rates are log-normal. While the LMM is a very flexible framework as it enables us to specify an individual dynamics for each Libor, the calibration of swaption smiles is difficult as, even in the simple case of log-normal forward rates, swaption prices are not known analytically and straight Monte Carlo calibration is too time-consuming.
In such a model the instantaneous variance of a swap rate is a weighted sum of the covariances of all forward rates, the weights being dependent on the forward rates. A widely used approximation consists in approximating these weights by their value at time $t = 0$ and deriving an effective log-normal

dynamics for the swap rate [138, 108].
By freezing the weights, but without forcing the swap rate to be log-normal, one can get an autonomous dynamics for the swap rate by making the covariances a function of the swap rate itself: this generates an effective local volatility model for the swap rate. This technique has been used in models in which forward rates follow a constant elasticity of variance (CEV) dynamics [45].
Note that, given a swaption smile, we know from Dupire work that there exists a *unique* effective local volatility (ELV) for the swap rate which is consistent with the smile. The techniques mentioned above can be seen as heuristic approximations to this ELV.
While affording more flexibility than the BGM model, a CEV dynamics is still not able to calibrate to both caplet and swaption smiles: it was natural for practitioners to consider stochastic volatility LMMs. The literature on this subject is not particularly large. Andersen *et al.* [45, 46] introduced a LMM where each Libor follows a CEV process, with the same parameter β, all Libors being coupled to an uncorrelated stochastic volatility of the Heston type. Piterbarg has recently extended this framework, allowing this single β to be time-dependent [136]. In both cases the method used for obtaining swaption smiles still relies on deriving an effective autonomous stochastic volatility model for the swap rate.
Our aim in this chapter is two-fold:

- to analyze a LMM model based on a SABR dynamics for forward rates.
- to pursue on the application of the heat kernel expansion on a Riemannian manifold endowed with an Abelian connection for deriving swaption smiles in mixed local/stochastic volatility LMMs, which is generic and equally applicable to all models mentioned above.

This chapter is organized as follows:
We first review LMMs, in particular the calibration and pricing issues.
Then, we present the generic framework that we make use of to compute approximate swaption smiles, which relies on a geometric formulation of the dynamics that underlies the pricing equation. We explicitly carry out the derivation of swaption smiles for the SABR-LMM model. We prove that the geometry underlying this model is the hyperbolic manifold $\mathbb{H}^{n+1}$. Some important properties of this space are then presented. Furthermore, we show that the "freezing" argument is no longer valid when we try to price a swaption in/out the money: The Libors should in fact be frozen to the saddle-point (constrained on a particular hyperplane) which minimizes the geodesic distance on $\mathbb{H}^{n+1}$.
Finally we illustrate the accuracy of our methodology by comparing our approximate smiles with exact Monte Carlo estimates as well as other popular approximations, such as those mentioned in this introduction.

8.2 Libor market models

We adopt the same definition as in section 2.10.3.3 and set

$$L_k(t) \equiv L(t, T_{k-1}, T_k)$$

the forward rate resetting at T_{k-1} with $\tau_k = T_k - T_{k-1}$ the tenor. We recall the link between the Libor $L_k(t)$ and the bonds $P_{tT_{k-1}}$ and P_{tT_k} quoted at t and expiring respectively at T_{k-1} and T_k

$$L_k(t) = \frac{1}{\tau_k}\left(\frac{P_{tT_{k-1}}}{P_{tT_k}} - 1\right)$$

As the product of the bond P_{tT_k} with the forward rate $L_k(t)$ is a difference of two bonds with respective maturities T_{k-1} and T_k and therefore a traded asset, L_k is a (local) martingale under $\mathbb{P}^k$, the (forward) measure associated with the numéraire P_{tT_k}. Therefore, we assume the following driftless dynamics under $\mathbb{P}^k$

$$\begin{aligned} dL_k(t) &= \sigma_k(t)\Phi_k(a, L_k)dW_k \ , \ \forall t \leq T_{k-1} \ , \ k = 1, \cdots, n \\ dW_k dW_l &= \rho_{kl}(t)dt \end{aligned}$$

with the initial conditions $a(t=0) = \alpha$ and $L_k(t=0) = L_k^0$. In order to achieve some flexibility, we assume that the (normal) local volatility $\Phi_k(a, L_k)$ depends on an additional one-dimensional Itô diffusion process a (to be specified later) representing a stochastic volatility. We therefore assume that all the forward rates are coupled with the same stochastic volatility a.

For the sake of comparison we list in Table 8.1 the specification of the models previously mentioned - also in the forward measure $\mathbb{P}^k$. Note that because in such models the stochastic volatility a is uncorrelated with the forward rates, its dynamics remains the same, regardless of the choice of measure, in contrast to our generic stochastic volatility LMM (SV-LMM).

The BGM, (limited) CEV and shifted log-normal models correspond to local volatility models ($a = 1$) and the others to stochastic volatility models with a unique stochastic volatility a driven by a Heston process.

The model parameters in a generic SV-LMM are the volatility $\sigma_i(t)$, the correlation $\rho_{ij}(t)$ and additional parameters coming from the eventual stochastic volatility a, $\Phi_k(a, L_k)$. They are fitted to the prices of caplets and swaptions. In addition, they can also be fitted to the historical or terminal correlations of forward rates.

The calibration of market models has been one of the main focuses in recent research and we review the main idea in the following.

TABLE 8.1: Example of stochastic (or local) volatility Libor market models.

LMM	SDE
BGM	$dL_k = \sigma_k(t) L_k dW_k$
CEV	$dL_k = \sigma_k(t) L_k^\beta dW_k$
LCEV	$dF_k = \sigma_k(t) L_k \min(L_k^{\beta-1}, \epsilon^{\beta-1}) dW_k$ with ϵ a small positive number
Shifted LN	$dX_k = \sigma_k(t) X_k dW_k$ with $L_k = X_k + \alpha_k$
FL-SV	$dL_k = \sigma_k(t)(\beta_k L_k + (1-\beta_k) L_k^0)\sqrt{v} dW_k$ $dv = \lambda(\bar{v} - v)dt + \zeta\sqrt{v} dZ \; ; \; dW_k dZ = 0$
FL-TSS	$dL_k = \sigma_k(t)(\beta_k(t) L_k + (1-\beta_k(t)) L_k^0)\sqrt{v} dW_k$ $dv = \lambda(\bar{v} - v)dt + \zeta\sqrt{v} dZ \; ; \; dW_k dZ = 0$

8.2.1 Calibration

From the definition (3.14), we rewrite the expression for a swap in the following way

$$s_{\alpha\beta} = \sum_{i=\alpha+1}^{\beta} \omega_i(L) L_i$$

with the weight depending on the Libors given by

$$\omega_i(L) = \frac{\tau_i P_{tT_i}}{\sum_{k=\alpha+1}^{\beta} \tau_k P_{tT_k}}$$

Approximating the weight by their initial value

$$\omega_i(L) \approx \omega_i(L_i^0)$$

the swap becomes index-like and a swaption becomes similar to a basket option. It is then clear that the swaption fair value depends on the volatility of the Libors and the correlation matrix $[\rho]_{ij}$. So, in theory, the swaptions carry information about correlation between Libors.

The swaption matrix can be used to calibrate both the Libor volatilities $\sigma_i(t)$ and the correlations $\rho_{ij}(t)$ at the same time. This is the approach explained in [7]. However, calibration techniques in which correlation is an output typically struggle to simultaneously fit the swaptions as well as produce reasonable yield curve correlations [75]. The calibrated correlation matrix is quite noisy and relatively far from an historical correlation. This problem can not be solved by imposing smoothing constraints on the correlation functional form. Therefore, we disregard this methodology and decouple the calibration into two sub-problems: first, the correlation structure is chosen as an input and then this correlation is used in the calibration of the Libor volatility to a swaption matrix [142].

Correlation input

In order to decrease the number of factors and simplify the Monte-Carlo pricing, we can use a piecewise constant reduced rank r correlation: Such a correlation can be written as

$$\rho_{ij}(t) = \sum_{k=1}^{r} b_{ik}(t) b_{jk}(t)$$

where $[b]_{ik}$ is a $n \times r$ matrix. Furthermore, we assume that the coefficients $b_{ik}(t)$ have the following functional form $\forall\, t \in [T_{l-1}, T_l)$

$$b_{ip}(t) = \frac{\sum_{q=1}^{r} \theta_q e^{-k_q (T_{i-1} - T_{l-1})} \bar{b}_{qp}}{\sqrt{\sum_{q,s=1}^{r} \theta_q \theta_s e^{-k_q (T_{i-1} - T_{l-1})} e^{-k_s (T_{i-1} - T_{l-1})} \bar{\rho}_{sq}}} \tag{8.1}$$

depending on the parameters k_p, θ_p and a r-dimensional correlation matrix $[\bar{\rho} \equiv \bar{b}\bar{b}^{\dagger}]_{pq}$ $(p, q = 1, \cdots, r)$. k_r set the time-dependence of the correlation and θ_r,$[\rho_{rp}]$ control the shape of the correlation. Note that the matrix $b_{iq}(t)$ has been normalized in order to have

$$\rho_{ii}(t) = \sum_{k=1}^{r} b_{ik}(t) b_{ik}(t) = 1$$

In practice, we use $r = 2$ or $r = 3$. Below, we have plotted a typical example of correlation shape between Libors 8.2.1.
This low-rank parametrization of the correlation matrix is usually displayed by mean-reverting short rate models as in the Hull-White 2-factor model (HW2).

Correlation structure for the HW2 model

In the HW2 model, the dynamics of the instantaneous-short-rate process under the risk neutral measure $\mathbb{P}$ is given by

$$r_t = x_t + y_t + \varphi(t) \tag{8.2}$$

where x_t and y_t are Ornstein-Uhlenbeck processes

$$\begin{aligned} dx_t &= -a x_t dt + \sigma_a(t) dW_a \\ dy &= -b y_t dt + \sigma_b(t) dW_b \ ,\ dW_a dW_b = \bar{\rho} dt \end{aligned}$$

with a (resp. $\sigma_a(t)$) and b (resp. $\sigma_b(t)$) two constants (resp. two functions). With the HW2 model being an affine model (see exercise 3.2), the value of a bond for the HW2 model quoted at t and expiring at T can be written as

$$P_{tT} = A(t, T) e^{-B(a,t,T) x_t - B(b,t,T) y_t}$$

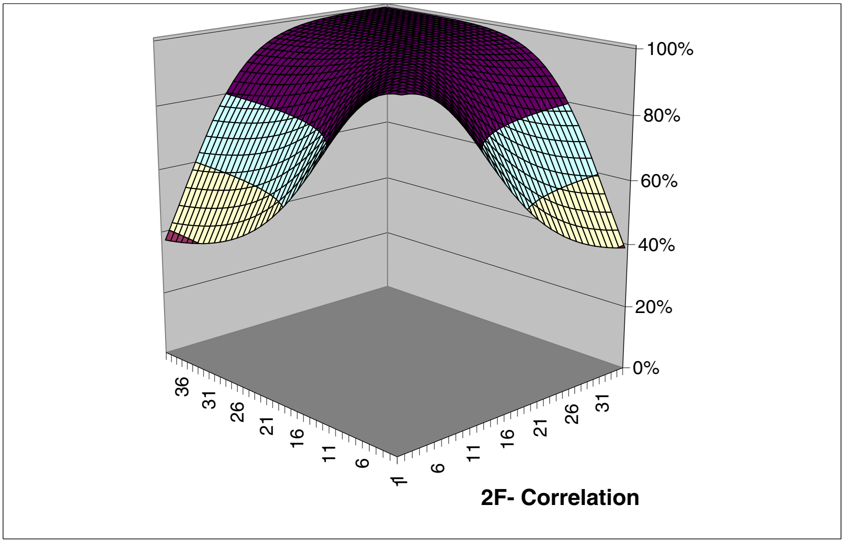

FIGURE 8.1: Instantaneous correlation between Libors in a 2-factor LMM. $k_1 = 0.25$, $k_2 = 0.04$, $\theta_1 = 100\%$, $\theta_2 = 50\%$ and $\rho = -20\%$.

with

$$B(z,t,T) = \frac{1 - e^{-z(T-t)}}{z}$$

$A(t,T)$ is left unspecified as we don't need to know its expression. As

$$L_i(t) = \frac{1}{\tau_i}\left(\frac{P_{tT_{i-1}}}{P_{tT_i}} - 1\right) \tag{8.3}$$

we deduce that the Libor $L_i(t)$ has the following dynamics in the forward measure $\mathbb{P}^i$

$$\begin{aligned} dL_i(t) = \sigma_a(t)(\tau_i L_i(t)+1) \ & \left(\left(\frac{-e^{-a(T_i-t)} + e^{-a(T_{i-1}-t)}}{a\tau_i}\right) dW_a + \right. \\ & \left. \frac{\sigma_b(t)}{\sigma_a(t)}\left(\frac{-e^{-b(T_i-t)} + e^{-b(T_{i-1}-t)}}{b\tau_i}\right) dW_b\right) \end{aligned}$$

This SDE is easily obtained by applying Itô's lemma on (8.3). We don't need to care about the drift term as it must cancel at this end in the forward measure $\mathbb{P}^i$.

Let us assume that the ratio $\frac{\sigma_b(t)}{\sigma_a(t)} \equiv \widehat{\theta}_b$ is a constant. If we choose new parameters $(\{k_\varepsilon, \theta_\varepsilon\}_{\varepsilon=a,b})$ such as $(\widehat{\theta}_a \equiv 1)$

$$\theta_\varepsilon e^{-k_\varepsilon T_{i-1}} e^{k_\varepsilon t} = \widehat{\theta}_\varepsilon e^{\varepsilon t}\left(\frac{-e^{-\varepsilon T_i} + e^{-\varepsilon T_{i-1}}}{\varepsilon \tau_i}\right);$$

equivalent to

$$\begin{aligned} k_\varepsilon &= \varepsilon \\ \theta_\varepsilon &= \widehat{\theta}_\varepsilon\left(\frac{1 - e^{-\varepsilon\tau_i}}{\varepsilon\tau_i}\right) \approx \widehat{\theta}_\varepsilon \end{aligned}$$

HW2 can be rewritten as $(\sigma(t) \equiv \sigma_a(t))$

$$dL_i(t) = \sigma(t)(\tau_i L_i(t)+1)\left(\theta_a e^{-k_a(T_{i-1}-t)} dW_a + \theta_b e^{-k_b(T_{i-1}-t)} dW_b\right)$$

The correlation structure is identical to (8.1) with $r = 2$.

Once the parameters $k_p, \theta_p, \bar{\rho}_{pq}$ have been chosen, we can calibrate the Libor volatilities $\sigma_i(t)$ to at-the-money swaptions.

Calibration to swaptions and Hull-White-Rebonato freezing arguments

The calibration uses an approximated formula for a swaption. Such an approximation has been derived for the LMMs listed in Table 8.1. At this stage, it is useful to recall how this approximation is derived using the Rebonato, Hull-White *freezing* argument [138], [108]. As in the LMMs presented in Table 8.1, we assume in this section that the functional form $\Phi_k(a, L_k)$ is the product of a stochastic volatility a and a local volatility function $C(L_k)$ independent of k

$$\Phi_k(a, L_k) = aC(L_k)$$

and the process a_t is uncorrelated with the Libors $\{L_k\}$

$$\rho_{ka} = 0 \ , \forall \ k = 1, \cdots, n$$

For $a = 1$ and $C(x) = x^\beta$, we have the Andersen-Andreasen CEV-LMM [44] and for $C(L_k) = \beta_k(t)L_k + (1 - \beta_k(t))L_k^0$ and a driven by an uncorrelated Heston process, we have the FL-SV LMM [45].
As the product of the swap rate $s_{\alpha\beta}(t)$,

$$s_{\alpha\beta}(t) = \frac{P_{tT_\alpha} - P_{tT_\beta}}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}}$$

with $\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}$ is the difference between two bonds, $P_{tT_\alpha} - P_{tT_\beta}$, and therefore a traded asset; the forward swap rate is a local martingale in the forward swap measure $\mathbb{P}^{\alpha\beta}$ associated to the numéraire $\sum_{i=\alpha+1}^{\beta} \tau_i P_{tT_i}$. The SDE followed by the swap rate is under $\mathbb{P}^{\alpha\beta}$

$$ds_{\alpha\beta}(t) = \sum_{k=\alpha+1}^{\beta} \frac{\partial s_{\alpha\beta}}{\partial L_k} \sigma_k(t) aC(L_k) dZ_k \tag{8.4}$$

Note that as we use a new measure, we have changed our notation for the Brownian motions from $\{W_k\}$ to $\{Z_k\}$. As a is not correlated with the Libors, its dynamics remains the same in the forward and forward swap measures.
The "freezing" argument consists in assuming that the terms $\frac{\partial s_{\alpha\beta}}{\partial L_k}$ and $\frac{C(s)}{C(L_i)}$ are almost constant and therefore equal to their values at the spot. We note $s_{\alpha\beta}(t = 0) \equiv s^0$ below. Therefore, the SDE (8.4) can be approximated by

$$\begin{aligned} ds_{\alpha\beta} &\simeq \sum_{k=\alpha+1}^{\beta} a \frac{\partial s_{\alpha\beta}}{\partial L_k}(L^0) \sigma_k(t) \frac{C(L_k^0)}{C(s^0)} C(s_{\alpha\beta}) dZ_k \\ &= \sigma_{\alpha\beta}(t) aC(s_{\alpha\beta}) dZ_t \end{aligned}$$

with

$$\sigma_{\alpha\beta}(t)^2 = \sum_{i,j=\alpha+1}^{\beta} \rho_{ij}(t)\sigma_i(t)\sigma_j(t) \left(\frac{C(L_i^0)}{C(s^0)}\right) \left(\frac{C(L_j^0)}{C(s^0)}\right) \frac{\partial s_{\alpha\beta}}{\partial L_i}(L^0) \frac{\partial s_{\alpha\beta}}{\partial L_j}(L^0)$$

The multi-dimensional process driving the Libors and the swap rate has been degenerated into a two-dimensional (resp. one-dimensional) stochastic (resp. local) volatility model (resp. for $a = 1$). Using the asymptotic method introduced in chapters 5 and 6, we can derive an asymptotic swaption implied volatility. We explicitly do the computation for the CEV LMM as an example.

Swaption implied volatility for CEV LMM

By definition, we know that the square of the (Dupire) local volatility associated to the driftless swap process is equal to the mean value (in the swap measure $\mathbb{P}^{\alpha\beta}$) of the square of the swap stochastic volatility conditional to the level of the swap

$$(\sigma_{loc}^{\alpha\beta})^2(t,s) = C(s)^2\sigma_{\alpha\beta}(t)^2\mathbb{E}^{\alpha\beta}[a^2|s_{\alpha\beta}=s]$$

Taking $a=1$ and $C(x)=x^{\beta}$, we obtain that the swap rate follows also a CEV process with the same parameter β

$$\sigma_{\rm loc}^{\alpha\beta}(t,s) = \sigma_{\alpha\beta}(t)s^{\beta}$$

with

$$\sigma_{\alpha\beta}(t)^2 = \sum_{i,j=\alpha+1}^{\beta}\rho_{ij}(t)\sigma_i(t)\sigma_j(t)\left(\frac{L_i^0}{s^0}\right)^{\beta}\left(\frac{L_i^0}{s^0}\right)^{\beta}\frac{\partial s_{\alpha\beta}}{\partial L_i}(L^0)\frac{\partial s_{\alpha\beta}}{\partial L_j}(L^0)$$

Doing a change of local time, the local volatility becomes time-independent and from example 5.6, we have that the swaption implied volatility is given at the first-order by

$$\sigma_{\rm BS}^{\alpha\beta}(T_\alpha,K) = \frac{(1-\beta)\ln\left(\frac{K}{s_{\alpha\beta}^0}\right)}{\left(K^{1-\beta}-s_{\alpha\beta}^{01-\beta}\right)}\frac{\sqrt{\int_0^{T_\alpha}\sigma_{\alpha\beta}(t)^2dt}}{\sqrt{T_\alpha}}$$
$$\left(1+\frac{(\beta-1)^2}{24}\left(\frac{s+s_{\alpha\beta}^0}{2}\right)^{2\beta-2}\int_0^{T_\alpha}\sigma_{\alpha\beta}(t)^2dt\right) \tag{8.5}$$

Swaption implied volatility for FL-SV LMM

In the same way, assuming that

$$C(L_k) = \beta_k L_k + (1-\beta_k)L_k^0$$

and a is driven by an uncorrelated Heston process, we have that the swap rate follows a Heston stochastic volatility model for which we have an exact solution (modulo a Fourier integration) for a swaption.

The "freezing argument" is crucial to obtain an analytical approximation for the swaption implied volatility. However, its theoretical and numerical validity is unclear. Also if we generalize the functional form $\Phi_k(a,L_k)$ (8.2.1), it is not at all obvious how to use the freezing argument. In the following, we explain how to calibrate the Libor volatility $\sigma_k(t)$ using this analytical approximation to at-the-money swaption implied volatilities. Then, we justify in the next section the "freezing" argument and obtain more accurate approximations using the heat kernel expansion technique.

TABLE 8.2: Libor volatility triangle.

1	σ_{10}	0	0	0
2	σ_{20}	σ_{21}	0	0
3	σ_{30}	σ_{31}	σ_{32}	0
4	σ_{40}	σ_{41}	σ_{42}	σ_{43}

Libor volatility triangle

The volatility parameters $\sigma_k(t)$ are calibrated to the at-the-money swaption implied volatilities noted $\{\sigma^{\mathrm{mkt}}_{\alpha\beta}\}$. In order to reduce the complexity of a general time-dependent function, we assume that the volatilities $\sigma_k(t)$ are constant piecewise functions on the tenor intervals:

$$\sigma_k(t) = \sigma_{kq} \quad \forall\, t \in [T_{q-1}, T_q)$$

As the Libor $L_k(t)$ is fixed when $t \geq T_k$, $\sigma_k(t)$ cancels after this date. Therefore the volatilities $[\sigma_{kq}]$ describe a triangle called *Libor volatility triangle* (see Table 8.2). Without any additional specification, we have $\frac{n(n+1)}{2}$ parameters σ_{kq} with n the number of Libors. Using the approximation (8.5), an at-the-money swaption IV is given by

$$\sigma^{\alpha\beta}_{\mathrm{BS}}(T_\alpha, s^0_{\alpha\beta})^2 T_\alpha \simeq (s^0_{\alpha\beta})^{2(\beta-1)} \Sigma_{\alpha\beta}\left(1 + \Sigma_{\alpha\beta}(s^0_{\alpha\beta})^{2(\beta-1)} \frac{(\beta-1)^2}{24}\right)$$

with $\Sigma_{\alpha\beta} = \int_0^{T_\alpha} \sigma_{\alpha\beta}(t)^2 dt$. By inverting this formula, we obtain

$$\Sigma_{\alpha\beta} = \Sigma^{\mathrm{mkt}}_{\alpha\beta}$$

with

$$\Sigma^{\mathrm{mkt}}_{\alpha\beta} = \frac{-1 + \sqrt{1 + \frac{1}{6}(\beta-1)^2 \sigma^{\alpha\beta}_{\mathrm{BS}}(T_\alpha, s^0_{\alpha\beta})^2}}{(s^0_{\alpha\beta})^{2(\beta-1)} \frac{(\beta-1)^2}{12}}$$

The expression $\Sigma_{\alpha\beta}$ is a quadratic function in the variables σ_{iq}. Smoothing constraints must be chosen in order to have a well-defined optimization problem. The objective function for the calibration is

$$\min_{[\sigma_{iq}]} \sum_{\alpha,\beta} \left(\left(\Sigma^{\mathrm{mkt}}_{\alpha\beta}\right)^2 - \left(\Sigma_{\alpha\beta}\right)^2 \right) + \lambda P$$

raised by a penalty function P which can be [142, 137]

$$P = \sum_{i=1}^n \sum_{j=i}^n \sigma_{ij}(-\sigma_{ij-1} - \sigma_{ij+1} - \sigma_{i-1,j} - \sigma_{i+1,j} + 4\sigma_{i,j}) + \epsilon \sum_{i=1}^n \sum_{j=i}^n (\sigma_{i,j} - \sigma^0_{ij})^2$$

with σ_{ij}^0 a fixed Libor volatility triangle or

$$P = \epsilon \sum_{i=1}^{n} \left(\sigma_{ij} - \sigma_{i+1j+1}\right)^2$$

The last penalty function ensures that $\sigma_k(t)$ stays close to a time-homogeneous Libor volatility.
As n is usually large ($n \approx 40$), we face a large optimization problem. In order to decrease the number of parameters to be optimized, we define σ_{kq} with a small number of parameters. In this context, two simple functional forms can be used.

Volatility parametrization

The first one, and the simplest, is called Rebonato parametrization [138]. σ_{kq} depends on $n+3$ parameters ϕ_k $(k = 1, ..., n)$ and a, b, c:

$$\sigma_{iq} = \phi_i \left(1 + (a(T_{i-1} - T_{q-1}) + d)e^{-b(T_{i-1}-T_{q-1})}\right)$$

In the second one [7], σ_{iq} depends on $2n$ parameters $(\phi_i, \varphi_i)_{i=1,\cdots,n}$

$$\sigma_{iq} = \phi_i \varphi_q \ , \ q = 1, \cdots, i$$

In practice, in both parametrizations, the parameters ϕ_i are calibrated exactly to the caplets ATM which are given for the CEV LMM by

$$\phi_i^2 = \frac{\left(\Sigma_{ii+1}^{\text{mkt}}\right)^2}{\sum_{q=1}^{i} \varphi_q^2 \tau_q}$$

For example, in the Rebonato parametrization, we have

$$\phi_i^2 = \frac{\left(\Sigma_{ii+1}^{\text{mkt}}\right)^2}{\sum_{q=1}^{i} \left(1 + (a(T_{i-1} - T_{q-1}) + d)e^{-b(T_{i-1}-T_{q-1})}\right)^2 \tau_q}$$

The closer ϕ_i is to 1, the more $[\sigma_{iq}]$ is time-homogeneous. Note that in the case of a full swaption matrix, an exact calibration can be done which does not require any specific optimization subroutine [7] (see chapter 7). However, in order to get the full swaption matrix, an extrapolation must been done on the market swaption surface. One can observe that some complex Libor volatilities can be obtained if the extrapolation is not done properly. Moreover, the calibration depends on this extrapolation scheme, a feature that it is not appropriate.
In the following, we illustrate the difference between the short rate-model, HW2, and the BGM model with 2-factors as well (Fig. (8.2)). Using a time-dependent volatility, the HW2 model has been calibrated exactly to ATM

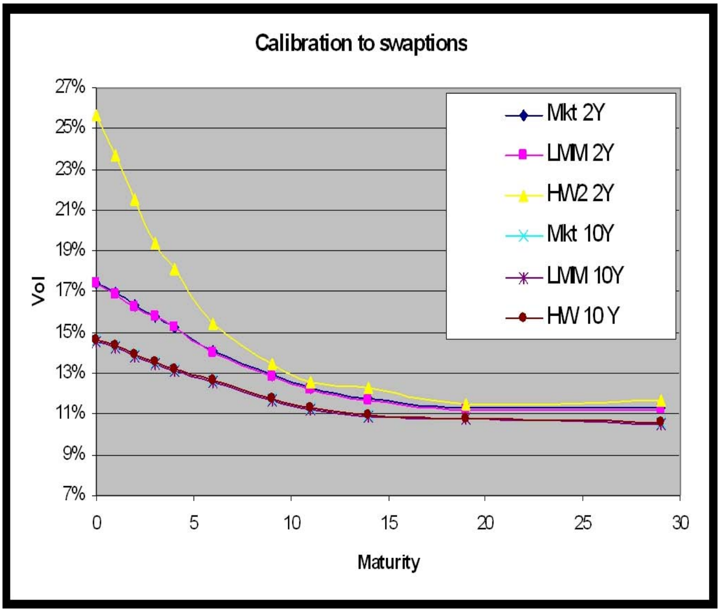

FIGURE 8.2: Comparison between HW2 and BGM models.

swaptions with a 10 year tenor. We have no guarantee on the error for the other swaption IVs (for example swaptions with tenor=2Y). The BGM model has been calibrated to ATM swaptions 1Y-2Y-10Y. Although the calibration is not exact but rather based on a optimization routine, we can see that the error is negligible.

Once the model is calibrated, the pricing of a derivative is performed with a MC pricer (see appendix B for a brief introduction to MC methods).

8.2.2 Pricing with a Libor market model

Monte-Carlo simulation

By choosing an appropriate measure such as the forward (terminal) measure $\mathbb{P}^n$ where T_n coincides with the last coupon date of a derivative product or such as the spot Libor measure as defined below, the Libor dynamics acquires a complicated drift. These drifts which are reminiscent of the non-Markovian nature of the HJM model are too time-consuming in a MC simulation.

In order to determine these drifts, it is necessary to specify a measure. There are usually two alternatives to choose from: the forward measure $\mathbb{P}^n$ associated to the numéraire P_{tT_n} or the spot Libor measure $\mathbb{P}^s$ associated to the numéraire

$$P_s(t) \equiv \frac{P_{tT_{\beta(t)-1}}}{\prod_{j=1}^{\beta(t)-1} P_{T_{j-1}T_j}}$$
$$= P_{tT_{\beta(t)-1}} \prod_{j=1}^{\beta(t)-1} (1 + \tau_j L_j(T_j))$$

The discontinuous index $\beta(t)$ is equal to q if $t \in [T_{q-2}, T_{q-1})$.

Next, we compute the dynamics of Libors under the measure $\mathbb{P}^n$ and $\mathbb{P}^s$. As previously covered, the Libor $L_i(t)$ is driftless under the measure $\mathbb{P}^i$ (associated to the bond P_{tT_i}). To pass from $\mathbb{P}^i$ to $\mathbb{P}^n$ (resp. $\mathbb{P}^s$), we need to compute the volatility of the ratio $\frac{P_{tT_n}}{P_{tT_i}}$ (resp. $\frac{P_s(t)}{P_{tT_i}}$) of numéraires. As the bond P_{tT_k} is by definition

$$P_{tT_k} = P_{tT_{\beta(t)-1}} \prod_{j=\beta(t)}^{k} \frac{1}{1 + \tau_j L_j(t)}$$

we have[1]

$$\frac{P_{tT_n}}{P_{tT_i}} = \prod_{j=i+1}^{n} \frac{1}{1 + \tau_j L_j(t)}$$
$$\frac{P_s(t)}{P_{tT_i}} = \frac{\prod_{j=i+1}^{\beta(t)-1} (1 + \tau_j L_j(T_j))}{\prod_{j=\beta(t)}^{i} \frac{1}{1+\tau_j L_j(t)}}$$

The volatility of the bond P_{tT_n} (resp. $P_s(t)$) minus the volatility of the bond P_{tT_i} is

$$[\sigma_{P_{tT_n}} - \sigma_{P_{tT_i}}].dW = -\sum_{j=i+1}^{n} \frac{\tau_j \sigma_j(t) \Phi_j(a, L_j) dW_j}{1 + \tau_j L_j(t)}$$
$$[\sigma_{P_s(t)} - \sigma_{P_{tT_i}}].dW = \sum_{j=\beta(t)}^{i} \frac{\tau_j \sigma_j(t) \Phi_j(a, L_j) dW_j}{1 + \tau_j L_j(t)}$$

[1]We only explicit the computation for $n > i$.

From proposition 2.1, the Libor $L_i(t)$ in the forward measure $\mathbb{P}^n$ is

$$dL_i(t) = -\sigma_i(t)\Phi_i(a, L_i) \sum_{j=i+1}^{n} \frac{\rho_{ij}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1 + \tau_j L_j(t)} dt + \sigma_i(t)\Phi_i(a, L_i)dZ_i \ , \ i < n \tag{8.6}$$

$$dL_i(t) = \sigma_i(t)\Phi_i(a, L_i) \sum_{j=n+1}^{i} \frac{\rho_{ij}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1 + \tau_j L_j(t)} dt + \sigma_i(t)\Phi_i(a, L_i)dZ_i \ , \ i > n \tag{8.7}$$

$$dL_n(t) = \sigma_n(t)\Phi_n(a, L_n)dZ_n \tag{8.8}$$

Also the Libor dynamics in the spot Libor measure $\mathbb{P}^s$ is

$$dL_i(t) = \sigma_i(t)\Phi_i(a, L_i) \sum_{j=\beta(t)}^{i} \frac{\rho_{ij}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1 + \tau_j L_j(t)} dt + \sigma_i(t)\Phi_i(a, L_i)dZ_i \tag{8.9}$$

REMARK 8.1 Limit $n \to \infty$ We take the continuous limit $n \to \infty$ in the equation (8.9). The Libors converge to the instantaneous forward rate curve $f_t^T = L(t, T, \Delta T)$. Then we have the SDE

$$df_t^T = \sigma(t, T)\Phi_T(a_t, f_t^T) \int_t^T \rho(t, T, T')\sigma(t, T')\Phi_{T'}(a_t, f_t^{T'})dT'dt + \sigma_T(t)\Phi_T(a_t, f_t^T)dZ_{tT}$$

This is the HJM dynamics. Note that in the limit $n \to \infty$, the n Brownian motions converge to a Brownian sheet [33] which is a two-dimensional continuous process parameterized by the variables t and T satisfying

$$dZ_{tT}dZ_{tT'} = \rho(T, T')dt$$

□

When dropped in the MC pricing, the processes (8.6, 8.7, 8.8, 8.9) should be discretized.

Discretization: Log-Euler scheme

As the SDEs above are already fairly complex, it seems unreasonable to use a higher-order discretization scheme such as the Milstein scheme (see appendix B). We will therefore use a log-Euler scheme between two dates t and $t + \Delta t$

$(\Delta \log L_i(t) = \log L_i(t+\Delta t) - \log L_i(t))$

$$\Delta \log L_i(t) = \sigma_i(t)\Phi_i(a, L_i)L_i(t)^{-1}\Delta Z_i \ , \ i < n$$
$$- \left(\sigma_i(t)\Phi_i(a, L_i)L_i(t)^{-1} \sum_{j=i+1}^{n} \frac{\rho_{ij}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1+\tau_j L_j(t)} \right.$$
$$\left. + \frac{\sigma_i(t)^2\Phi_i(a, L_i)^2 L_i(t)^{-2}}{2} \right) \Delta t$$
$$\Delta \log L_i(t) = \sigma_i(t)\Phi_i(a, L_i)L_i(t)^{-1}\Delta Z_i \ , \ i > n$$
$$+ \left(\sigma_i(t)\Phi_i(a, L_i)L_i(t)^{-1} \sum_{j=n+1}^{i} \frac{\rho_{ij}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1+\tau_j L_j(t)} \right.$$
$$\left. - \frac{\sigma_i(t)^2\Phi_i(a, L_i)^2 L_i(t)^{-2}}{2} \right) \Delta t$$
$$\Delta \log L_n(t) = -\frac{\sigma_n(t)^2\Phi_n(a, L_n)^2 L_n(t)^{-2}}{2}\Delta t + \sigma_n(t)\Phi_n(a, L_n)L_n(t)^{-1}\Delta Z_n$$

Assuming that we simulate the n Libors under the forward measure $\mathbb{P}^n$, the computation of the drift terms involves $\frac{n(n-1)}{2}$ operations which can be quite large. In the following, we explain how to reduce the number of operations to $O(r \times n)$ [113].
As we use a low-rank correlation

$$\rho_{ij}(t) = \sum_{k=1}^{r} b_{ik}(t)b_{jk}(t)$$

the drift term becomes

$$-\sum_{j=i+1}^{n} \frac{\rho_{ij}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1+\tau_j L_j(t)} = -\sum_{k=1}^{r} b_{ik}(t) \sum_{j=i+1}^{n} \frac{b_{jk}(t)\tau_j\sigma_j(t)\Phi_j(a, L_j)}{1+\tau_j L_j(t)}$$

In particular, if we precompute the n terms $x_j = \frac{\tau_j\sigma_j(t)\Phi_j(L_j, a)}{1+\tau_j L_j(t)}$, we can define

$$e_{k,i} \equiv -\sum_{j=i+1}^{n} b_{jk}(t)x_j \ , \ i < n$$
$$e_{k,i} \equiv \sum_{j=n+1}^{i} b_{jk}(t)x_j \ , \ i > n$$
$$e_{k,n} \equiv 0$$

We deduce the recurrence equations

$$e_{k,i} = e_{k,i+1} - x_{i+1}b_{i+1,k}(t) \ , \ i < n$$
$$e_{k,i} = e_{k,i-1} + x_i b_{i,k}(t) \ , \ i > n$$

and the SDEs for the Libors can be re-written as a function of the $r \times n$ terms $[e]_{k,i}$

$$\Delta \log(L_i(t)) = -\frac{1}{2}\sigma_i(t)^2\Phi_i(a, L_i)^2 L_i(t)^{-2}\Delta t + \sigma_i(t)\Phi_i(a, L_i)L_i(t)^{-1}\sum_{k=1}^{r} b_{ik}(e_{k,i}\Delta t + \Delta W_k)$$

with $\{\Delta W_k\}_{k=1,\cdots,r}$ r-uncorrelated Brownians. The number of operations involved to compute the terms $\{e_{k,i}\}$ is $r \times n$ and the total number of operations involved to compute the drift in this procedure is $O(r \times n)$. This algorithm considerably reduces the time involved in the computation of the drift.
In order to decrease the discretization error, we can go one step further and use a predictor-corrector scheme as explained below.

Predictor-corrector scheme

The predictor-corrector scheme consists of two steps:
In a first step, we use the log-Euler scheme (with the drift algorithm above) to compute the Libors at the step $t + \Delta t$, noted $\widetilde{L}_i(t + \Delta t)$. This Libor is then used to average the drift by

$$x_j^{\text{new}} = \frac{\tau_j \sigma_j(t)}{2}\left(\frac{\Phi_j(a, L_j)}{1 + \tau_j L_j(t)} + \frac{\Phi_j(a, \widetilde{L}_j)}{1 + \tau_j \widetilde{L}_j(t)}\right)$$

In the second step, using the same Brownian motion as before, we recompute the Libors with this new drift

$$\Delta \log(L_i(t)) = -\frac{1}{2}\sigma_i(t)^2\Phi_i(a, L_i)^2 L_i(t)^{-2}\Delta t + \sigma_i(t)\Phi_i(a, L_i)L_i(t)^{-1}\sum_{k=1}^{r} b_{ik}(e_{k,i}^{\text{new}}\Delta t + \Delta W_k)$$

Note that there also exists a discretization consistent with the discretization of the spot measure [94], where the bonds are arbitrage free in the discretized Libor SDE.

8.3 Markovian realization and Frobenius theorem

The MC pricing for a LMM is too lengthly due to the complicated drifts which are reminiscent of the non-Markovian nature of the HJM model. For example if we use the terminal forward measure $\mathbb{P}^n$, we need to simulate n Libors. For

a derivative product with a maturity of 20 years and semi-annual Libors, n is equal to 40.
As a consequence, a LMM requires to simulate a high number of Markov processes. This is a severe drawback which does not occur in the short-rate model framework where the whole yield curve is modeled by an instantaneous interest rate.
In order to overcome this difficulty, we can try to represent the Libors as functionals of low-dimensional Markov states (typically one or two). In general, such a map between Libors and a low-dimensional Markov process doesn't exist. For example, in exercise 8.1 we show that in the one-factor log-normal BGM model the Libors can not be written as a function of a one-dimensional process.
In this section, we give a necessary and sufficient condition to show that a n-dimensional Itô process (possibly infinite $n = \infty$) can be written as a function of a low-dimensional Itô process $r < n$. This criteria is based on the Frobenius theorem.
Let X_t be a n-dimensional Itô process following the Stratonovich SDE

$$dX_t = V_0 dt + \sum_{i=1}^{d} V_i \diamond dW_t^i \ , \ X_{t=0} = X_0 \tag{8.10}$$

Without any loss of generality, we have assumed that we have a time-homogeneous SDE. A time-inhomogeneous SDE can be put into this normal form (8.10) by including an additional state X_t^{n+1}

$$dX_t^{n+1} = dt$$

DEFINITION 8.1 Markovian representation *We say that the process X_t admits a r-dimensional Markovian representation ($r < n$) if there exists a smooth function $G : \mathbb{R}^r \to \mathbb{R}^n$ such that $X = G(z)$ with z a r-dimensional Itô process.*

THEOREM 8.1 Frobenius theorem [64]
If the free Lie algebra (see B.7 in appendix B) generated by $(V_0, V_1, \cdots, V_d)$ has a constant dimension r (as a vector space) then the SDE (8.10) admits a r-dimensional Markovian representation: $X = G(z)$. We note $f_1, \cdots, f_r$ a basis for the free Lie algebra. The Markovian representation $G(\cdot)$ is given by

$$G(z_1, \cdots z_r) = e^{z_1 f_1} e^{z_2 f_2} \ldots e^{z_r f_r} X_0$$

where $e^{z_i f_i}$ is the flow (see the definition B.6 in appendix B) along the vector f_i at time z_i.

As an example, we classify the one-factor LV LMMs which admits a two-dimensional Markovian representation.

Example 8.1 One-factor LV LMMs
A one-factor LV LMM is defined by the following SDE for each Libor $\{L_i\}_{i=1,\cdots,n}$ in each forward measure $\mathbb{P}^i$

$$dL_i = \sigma_i(t)C(L_i)dW_t$$

In the terminal measure $\mathbb{P}^n$, we have

$$dL_i = \sigma_i(t)C(L_i)\sum_{j=i+1}^{n}\frac{\sigma_j(t)C(L_j)\tau_j}{1+\tau_j L_j}dt + \sigma_i(t)C(L_i)dZ_t$$

Using the Stratonovich calculus, we obtain the time-homogeneous SDE

$$\frac{dL_i}{\sigma_i(u)C(L_i)} = \left(-\frac{1}{2}\sigma_i(u)\partial_i C(L_i) + \sum_{j=i+1}^{n}\frac{\sigma_j(u)C(L_j)\tau_j}{1+\tau_j L_j}\right)dt + \diamond dZ_t \tag{8.11}$$

$$du = dt \tag{8.12}$$

for which we derive the vector fields

$$V_0 = \left(-\frac{1}{2}\sigma_i(u)^2 C(L_i)\partial_i C(L_i) + \sigma_i(u)C(L_i)\sum_{j=i+1}^{n}\frac{\sigma_j(u)C(L_j)\tau_j}{1+\tau_j L_j}\right)\partial_i + \partial_u$$
$$V_1 = \sigma_i(u)C(L_i)\partial_i$$

where we have set $\partial_i \equiv \partial_{L_i}$. We deduce that

$$\begin{aligned}[V_0, V_1] = {} & \left(\partial_u \sigma_i(u)C(L_i) + \frac{\sigma_i(u)^3}{2}C(L_i)^2\partial_i^2 C(L_i)\right)\partial_i \\ & - \sigma_j(u)C(L_j)\sigma_i(u)C(L_i)\sum_{j=i+1}^{n}\sigma_j(u)\partial_j\left(\frac{C(L_j)\tau_j}{1+\tau_j L_j}\right)\partial_i\end{aligned}$$

Therefore, a one-factor LV LMM admits a 2-dimensional Markovian representation if and only if there exists a smooth function $f : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$ such that $[V_0, V_1] = f(u, L)V_1$. This gives the constraint

$$\begin{aligned}\left(\partial_u \ln\sigma_i(u) + \frac{\sigma_i(u)^2}{2}C(L_i)\partial_i^2 C(L_i)\right) - \sigma_j(u)C(L_j)\sum_{j=i+1}^{n}\sigma_j(u)\partial_j\left(\frac{C(L_j)\tau_j}{1+\tau_j L_j}\right) \\ = f(u, L)\ \forall\ i = 1, \cdots, n\end{aligned}$$

For example if we assume that $\tau_i = \tau\ \forall\ i$, we cancel the sum over the index k by taking

$$C(L_i) = (1 + \tau L_i)$$

The constraint above reduces to

$$\partial_u \ln \sigma_i(u) = f(u) \ \forall \ i = 1, \cdots, n$$

which shows that the volatility $\sigma_i(\cdot)$ should be a separable function in T_i and in time t

$$\sigma_i(t) = \Phi_i \nu(t)$$

with $\nu(t) \equiv e^{\int_0^t f(u)du}$. This is the one-factor HW model

$$dL_i = \Phi_i \nu(t) \left(1 + \tau L_i\right) dW_t$$

From the theorem 8.1, the 2d-Markovian representation is given by the flow

$$L(t) = e^{z_0 V_0} e^{z_1 V_1} L_0$$

The flow along the vector V_1 corresponds to solve the following ODE at time $t = z_1$

$$\begin{aligned} \frac{dL_i(t)}{dt} &= \sigma_i(u) \left(1 + \tau L_i\right) \\ \frac{du}{dt} &= 0 \end{aligned}$$

We obtain

$$\begin{aligned} \ln\left(\frac{1 + \tau L_i(z_1)}{1 + \tau L_i(0)}\right) &= \tau \Phi_i \int_0^{z_1} \nu(s) ds \\ u(t) &= u_0 \end{aligned}$$

Similarly, the flow along the vector field V_0 is

$$\begin{aligned} \ln\left(\frac{1 + \tau L_i(z_0)}{1 + \tau L_i(0)}\right) &= \Phi_i \left(-\frac{1}{2}\Phi_i + \sum_{j=i+1}^{n} \Phi_j\right) \tau^2 \int_0^{z_0} \nu(s)^2 ds \\ u(t) &= z_0 \end{aligned}$$

We get our final result thanks to the composition of the two flows described above

$$\begin{aligned} \ln\left(\frac{1 + \tau L_i(z_0, z_1)}{1 + \tau L_i(0)}\right) &= \Phi_i \left(-\frac{1}{2}\Phi_i + \sum_{j=i+1}^{n} \Phi_j\right) \tau^2 \int_0^{z_0} \nu(s)^2 ds \\ &\quad + \tau \Phi_i \int_0^{z_1} \nu(s) ds \\ u(t) &= z_0 \end{aligned}$$

By applying Stratonovich's lemma on both sides of this equation and identifying with the SDEs (8.11) and (8.12), we get

$$dz_0 = dt$$
$$dz_1 = \frac{\nu(t)}{\nu(z_1)} \diamond dW_t$$

The first Markov state is the time t.

□

As seen above, a one-factor LV LMM admits a 2-dimensional Markovian representation if and only if the LV function is a displaced diffusion model

$$dL = (1 + \tau L)\sigma(t) dW_t$$

Modulo several approximations, an (almost exact) Markovian representation can be found for general LV LMMs. This is briefly discussed in the next subsection.

Markovian approximation

The simplest trick to account for these drifts is to freeze the Libor L_i to the spot Libor L_i^0 inside the drifts. The BGM model becomes log-normal and can be mapped to a low-rank Markov model if the volatility $\sigma_i(t)$ is a separable function (see exercise 8.2)

$$\sigma_i(t) = \Phi_i \nu(t)$$

This approximation is efficient for low volatility and small maturities but become quite off otherwise and it should not be used for MC pricing.
After our short review of the calibration and pricing of the LMMs, we make the connection with the main subject of this book and explain how to deduce an accurate approximation for swaption implied volatilities for SV-LMMs. This is particularly useful for the calibration procedure as we saw previously. Although this technique is equally applicable to generic SV-LMMs, we apply it to a LMM based on a SABR dynamics for forward rates.

8.4 A generic SABR-LMM model

While keeping the idea of driving the Libors with a CEV dynamics we let each Libor have its own elasticity parameter β. As the market standard for parameterizing swaption smiles is the SABR formula (see Fig. 8.3), we use a log-normal dynamics for the stochastic volatility - note that references [45, 46], [136] use the Heston dynamics.

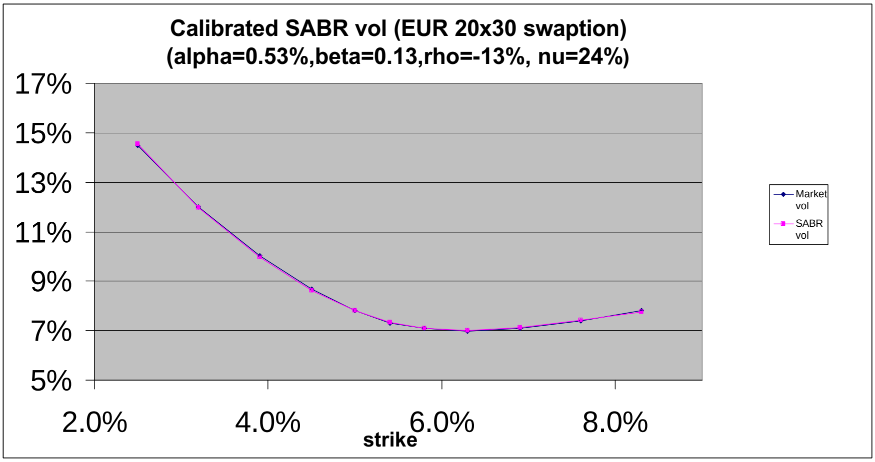

FIGURE 8.3: Calibration of a swaption with the SABR model.

Moreover, with the aim of enhancing the model's ability to fit both caplet and swaption smiles, we let the stochastic volatility process have non-zero correlation with the forward rates.

In our case, because of the presence of stochastic volatility, we prefer to work in the spot measure $\mathbb{P}^s$. In this measure, the dynamics of the forward rates in the SABR-LMM model reads :

$$\begin{aligned} dL_k &= a^2 B_k(t,F)dt + \sigma_k(t) a C_k(L_k) dZ_k \\ da &= \nu a dZ_{n+1};\ ,\ dZ_i dZ_j = \rho_{ij}(t)dt \ \ i,j = 1,\cdots,n+1 \end{aligned}$$

with

$$\begin{aligned} C_k(L_k) &= \phi_k L_k^{\beta_k} \\ B_k(t,F) &= \sum_{j=\gamma(t)}^{k} \frac{\tau_j \rho_{kj} \sigma_k(t) \sigma_j(t) C_k(L_k) C_j(L_j)}{1+\tau_j L_j} \end{aligned}$$

We introduce the constants ϕ_k for normalization purposes, so that $\sigma_k(t = 0) = 1$, for all k.

The forward rate dynamics under the forward measure $\mathbb{P}^k$ is much simpler and is given by:

$$
\begin{aligned}
dL_k(t) &= \sigma_k(t) a C_k(L_k) dW_k \\
da &= -\nu a^2 \sum_{j=\gamma(t)}^{k} \frac{\tau_j \rho_{ja} \sigma_j(t) C_j(L_j)}{1+\tau_j L_j} dt + \nu a dW_{n+1}, \; dW_k dW_{n+1} = \rho_{ka}(t) dt
\end{aligned}
$$

with initial conditions $a(t=0) = a^0$ and $L_k(t=0) = L_k^0$.
Note that as shown in chapter 6, the log-normal SABR model defines a martingale as long as $0 \leq \beta_k < 1$ or $\rho_{ka} \leq 0$ for $\beta_k = 1$. The possibility of moment explosions due to volatility being log-normally distributed in the case of the SABR-LMM model is an open question.

8.5 Asymptotic swaption smile

Our strategy for getting an approximate swaption smile involves the following two main steps:

- Firstly we derive an approximation to the ELV for the swap rate at hand.
- From the expression of this local volatility we derive an approximate expression for the implied volatility.

8.5.1 First step: deriving the ELV

Let $s_{\alpha\delta}$ be the forward swap rate starting at T_α and expiring at T_δ. $s_{\alpha\delta}$ satisfies the following driftless dynamics in the forward swap measure $\mathbb{P}^{\alpha\delta}$ (associated to the numéraire $C_{\alpha\delta}(t) = \sum_{i=\alpha+1}^{\delta} \tau_i P_{tT_i}$):

$$
ds_{\alpha\delta} = \sum_{k=\alpha+1}^{\delta} \frac{\partial s_{\alpha\delta}}{\partial L_k} \sigma_k(t) a C_k(L_k) dZ_k
$$

In order to be thorough, we give the dynamics of the Libors and the stochastic volatility under the forward swap measure $\mathbb{Q}^{\alpha\delta}$:

$$
\begin{aligned}
dL_k &= a^2 b^k(t, L) dt + \sigma_k(t) a C_k(L_k) dZ_k \\
da &= -\nu a^2 b^a(t, L) dt + \nu a dZ_{n+1}; \; , \; dZ_i dZ_j = \rho_{ij}(t) dt \;\; i, j = 1, \cdots, n+1
\end{aligned}
$$

with the drifts

$$b^k(t,L) = \sum_{j=\alpha+1}^{\delta} (2.1_{(j\leq k)} - 1)\tau_j \frac{P_{tT_j}}{C_{\alpha\delta}(t)} \tag{8.13}$$
$$\sum_{i=\min(k+1,j+1)}^{\max(k,j)} \frac{\tau_i \rho_{ki}\sigma_i(t)\sigma_k(t)C_i(L_i)C_k(L_k)}{1+\tau_i L_i}$$
$$b^a(t,L) = \sum_{j=\alpha+1}^{\delta} \tau_j \frac{P_{tT_j}}{C_{\alpha\delta}(t)} \sum_{k=\gamma(t)}^{j} \frac{\tau_k C_k(L_k)\rho_{ka}\sigma_k(t)}{1+\tau_k L_k} \tag{8.14}$$

We know that there exists a unique local volatility function $\sigma_{\mathrm{loc}}^{\alpha\delta}$ – the ELV – which is consistent with $s_{\alpha\delta}$'s smile. It is given by the expectation of the local variance conditional on $s_{\alpha\delta}$'s level. For convenience, we prefer to express the dynamics of $s_{\alpha\delta}$ in the Bachelier – rather than the log-normal – framework:

$$ds_{\alpha\delta} \equiv \sigma_{\mathrm{loc}}^{\alpha\delta}(t, s_{\alpha\delta})dW_t$$

$\sigma_{\mathrm{loc}}^{\alpha\delta}(t, s_{\alpha\delta})$ is given by:

$$(\sigma_{\mathrm{loc}}^{\alpha\delta})^2(t,s) \equiv \mathbb{E}^{\alpha\delta}[\sum_{i,j=\alpha+1}^{\delta} \rho_{ij}(t)\sigma_i(t)\sigma_j(t)aC_i(L_i)aC_j(L_j)\frac{\partial s_{\alpha\delta}}{\partial L_i}\frac{\partial s_{\alpha\delta}}{\partial L_j}|s_{\alpha\delta}=s]$$
$$= \sum_{i,j=\alpha+1}^{\delta} \rho_{ij}(t)\sigma_i(t)\sigma_j(t)\frac{\int_{\mathbb{B}} aC_i(L_i)aC_j(L_j)\frac{\partial s_{\alpha\delta}}{\partial L_i}\frac{\partial s_{\alpha\delta}}{\partial L_j}p(t,a,L|L^0,\alpha)(da\prod_i dL_i)}{\int_{\mathbb{B}} p(t,a,L|L^0,\alpha)(da\prod_i dL_i)} \tag{8.15}$$

where the integration domain is restricted to the set $\mathbb{B} = \{\{L_i\}_i | s_{\alpha\delta} = s\}$ and $p(t,a,L|L^0,\alpha)$ is the joint density for the forward rates L_i and the stochastic volatility a in the forward swap measure $\mathbb{Q}^{\alpha\delta}$.
There are two steps in evaluating this expression:

- We need an approximate expression for $p(t,a,L|L^0,\alpha)$.
- We have to compute the integration over $\mathbb{B}$.

The first step is achieved via the heat kernel expansion technique. It boils down to the calculation of the geodesic distance, the Van-Vleck-Morette determinant and the gauge transport parallel between any two given points, in the metric defined by the SABR-LMM model as explained in chapter 4.

Hyperbolic geometry

While this is generally a non-trivial task, the geodesic distance is analytically known for the special case of the geometry that the SABR-LMM model defines. It is also the case for most popular models (Heston, SABR, 3/2-SVM) which admit a large number of Killing vectors as seen in chapter 6.

By definition, the metric (4.78) (at $t = 0$) is given by

$$ds^2 = \frac{2}{\nu^2 a^2}\left(\sum_{i,j=1}^{n} \rho^{ij} \frac{\nu dL_i}{C_i(L_i)} \frac{\nu dL_j}{C_j(L_j)} + 2\sum_{i=1}^{n} \rho^{ia} \frac{\nu dL_i}{C_i(L_i)} da + \rho^{aa} da^2\right)$$

Here $\rho^{ij} \equiv [\rho^{-1}]_{ij}$, $(i,j) = (1,\cdots,n)$ and $\rho^{ia} \equiv [\rho^{-1}]_{ia}$ are the components of the inverse of the correlation matrix ρ. After some algebraic manipulations, we show that in the new coordinates $[x_k]_{k=1\cdots n+1}$ (L is the Cholesky decomposition of the (reduced) correlation matrix: $[\rho]_{i,j=1\cdots n} = [\hat{L}\hat{L}^\dagger]_{i,j=1\cdots n}$)

$$x_k = \sum_{i=1}^{n} \nu \hat{L}^{ki} \int_{L_i^0}^{L_i} \frac{dL_i'}{C_i(L_i')} + \sum_{i=1}^{n} \rho^{ia} \hat{L}_{ik} a \ , \ k = 1, \cdots, n$$

$$x_{n+1} = (\rho^{aa} - \sum_{i,j}^{n} \rho^{ia} \rho^{ja} \bar{\rho}_{ij})^{\frac{1}{2}} a$$

the metric becomes

$$ds^2 = \frac{2\left(\rho^{aa} - \sum_{i,j}^{n} \rho^{ia}\rho^{ja}\bar{\rho}_{ij}\right)}{\nu^2} \frac{\sum_{i=1}^{n} dx_i^2 + dx_{n+1}^2}{x_{n+1}^2}$$

Here $\overline{\rho}_{ij}$ is the inverse of the reduced matrix $[\rho^{ij}]_{i,j=1,\ldots,n}$.

Geometry $\mathbb{H}^{n+1}$

Written in the coordinates $[x_i]$, the metric is therefore the standard hyperbolic metric on $\mathbb{H}^{n+1}$ modulo a constant factor $\frac{2(\rho^{aa} - \sum_{i,j,k=1}^{n} \rho^{ia}\rho^{ja}\bar{\rho}_{ij})}{\nu^2}$ that we integrate out by scaling the time. By definition, the hyperbolic space is a (unique) simply connected n-dimensional Riemannian manifold with a constant negative sectional curvature -1.
The geodesic distance on $\mathbb{H}^{n+1}$ is given by

THEOREM 8.2 Geodesic distance on** $\mathbb{H}^{n+1}$ **[39]
The geodesic distance $d(x, x')$ on $\mathbb{H}^{n+1}$ is given by

$$d(x, x^0) = \cosh^{-1}\left(1 + \frac{\sum_{i=1}^{n} (x_i - x_i^0)^2}{2x_{n+1}x_{n+1}^0}\right)$$

In particular for $n = 1$, we reproduce the geodesic distance (4.43) on the Poincaré hyperbolic plane.

Using the geodesic distance on $\mathbb{H}^{n+1}$ between the points $x = (\{L\}_k, a)$ and the initial point $x^0 = (\{L^0\}_k, \alpha)$, the geodesic distance associated to the

SABR-LMM model is

$$d(x, x^0) = \cosh^{-1} \left(1 + \frac{\nu^2 \sum_{i,j=1}^{n} \rho^{ij} q_i q_j + 2\nu(a - a^0) \sum_{j=1}^{n} \rho^{ja} q_j + (a - a^0)^2 \rho^{aa}}{2(\rho^{aa} - \sum_{i,j=1}^{n} \rho^{ia} \rho^{ja} \rho_{ij}) a a^0}\right) \tag{8.16}$$

with

$$q_i = \int_{L_i^0}^{L_i} \frac{dL_i'}{C_i(L_i')}$$

$\rho^{\mu\nu}$ are the components of the inverse correlation $[(\rho^{-1})_{\mu\nu}]_{\mu,\nu=1,\ldots,n+1}$.

An important property of the hyperbolic space $\mathbb{H}^n$ is that the heat kernel equation is solvable:

THEOREM 8.3 Heat kernel on $\mathbb{H}^n$ [95]

The heat kernel of the hyperbolic space $\mathbb{H}^n$ is given in even dimensions by

$$p_{2(m+1)}(t, x|y) = \left(\frac{-1}{2\pi}\right)^m \frac{\sqrt{2} e^{\frac{-(2m+1)^2 t}{4}}}{(4\pi t)^{\frac{3}{2}}} \left(\frac{1}{\sinh r} \frac{\partial}{\partial r}\right)^m \left(\int_r^\infty \frac{s e^{\frac{-s^2}{4t}}}{\sqrt{\cosh s - \cosh r}} ds\right)$$

and in odd dimensions by

$$p_{2m}(t, x|y) = \left(\frac{-1}{2\pi}\right)^m \frac{e^{-m^2 t}}{(4\pi t)^{\frac{1}{2}}} \left(\frac{1}{\sinh r} \frac{\partial}{\partial r}\right)^m \left(e^{-\frac{r^2}{4t}}\right)$$

Van-Vleck-Morette determinant

Using the explicit expression for the hyperbolic distance, the Van-Vleck-Morette determinant is

$$\Delta(a, L|\alpha, L^0) = \left(\frac{d(a, L|\alpha, L^0)}{\sinh(d(a, L|\alpha, L^0))}\right)^n$$

8.5.2 Connection

The Abelian connection is given by (4.58)[2]

$$\mathcal{A}_i = \frac{1}{C_i(L_i)}\left(\sum_{j=1}^{n}\rho^{ij}\left(\frac{b^j(t,L)}{C_j(L_j)} - \frac{\partial_j C_j(L_j)}{2}\right) - \nu\rho^{ia}b^a(t,L)\right)$$

$$\mathcal{A}_a = \frac{1}{\nu}\left(\sum_{j=1}^{n}\rho^{aj}\left(\frac{b^j(t,L)}{C_j(L_j)} - \frac{\partial_j C_j(L_j)}{2}\right) - \nu\rho^{aa}b^a(t,L)\right)$$

where we have used that

$$\sqrt{g} = \frac{2^{\frac{n+1}{2}}\det[\rho]^{-\frac{1}{2}}}{\nu a^2\prod_{i=1}^{n} C_i(L_i)}$$

$-\nu a^2 b^a(t,L)$ (8.13) and $a^2 b^i(t,L)$ (8.14) are the drifts in the swap measure. Finally, the Abelian 1-form connection is

$$\mathcal{A} = \frac{1}{\nu}\sum_{j=1}^{n}\left(\frac{b^j(t,L)}{C_j(L_j)} - \frac{\partial_j C_j(L_j)}{2}\right)\left(\nu\sum_{i=1}^{n}\rho^{ij}dq_i + \rho^{aj}da\right) - b^a(t,L)\left(\nu\sum_{i=1}^{n}\rho^{ia}dq_i + \rho^{aa}da\right)$$

In order to compute the log of the parallel gauge transport

$$\ln(\mathcal{P})(a,q|\alpha) = -\int_{\mathcal{C}}\mathcal{A}$$

we need to know a parametrization of the geodesic curve on $\mathbb{H}^{n+1}$. However, we can directly find $\ln(\mathcal{P})(a,q|\alpha)$ if we approximate the drifts $b^k(t,L)$ and $b^a(t,L)$ by their values at the Libor spots (and $t=0$). A similar approximation was done in the Hagan-al formula [100] as shown in remark 6.7. Modulo this approximation (see remark 4.3),

$$\ln(\mathcal{P})(a,q|\alpha) \simeq \frac{1}{\nu}\sum_{j=1}^{n}\left(\frac{b^j(0,L^0)}{C_j(L_j^0)} - \frac{\partial_j C_j(L_j^0)}{2}\right)\left(\nu\sum_{i=1}^{n}\rho^{ij}q_i + \rho^{aj}(a-\alpha)\right) - b^a(0,L^0)\left(\nu\sum_{i=1}^{n}\rho^{ia}q_i + \rho^{aa}(a-\alpha)\right)$$

2

$$\mathcal{A}^i = \frac{a^2}{2}(b^i - \frac{1}{2}C_i\partial_i C_i)$$

$$\mathcal{A}^a = -\frac{\nu^2 a^2 b^a(L,t)}{2}$$

Saddle-point

At this stage, we need to perform the integration over $\mathbb{B}$. This amounts to computing an integral of the following type:

$$\int f(L,a)e^{-\frac{1}{t}\phi(L,a)}\delta(g(L)) \quad da\prod_i dL_i \tag{8.17}$$

where $\delta(g(L))$ is introduced to restrict the integration to the hyperplane defined by $g(x)=0$ – in our case the set of L such that $s_{\alpha\beta}(L)=s$. $\phi(L,a)$ is related to the square of the geodesic distance. As we are interested in the limit $t\to 0$, it is natural to use the saddle-point technique [5, 15] to compute this integral, introducing a Lagrange multiplier λ to enforce the constraint $g(L)=0$. More precisely, the integral (8.17) is approximated in the limit $t\to 0$ by

$$\begin{aligned}\int f(x)e^{-\frac{1}{t}\phi(x)}dx &\sim_{t\to 0} f(x^*)e^{-\frac{1}{t}\phi(x^*)}\,(1-t \\ &\left(-\frac{\partial_{\alpha\beta}f}{2f}A_{\alpha\beta}+\left(\frac{\partial_\alpha f}{2f}\partial_{\beta\gamma\delta}\phi+\frac{1}{8}\partial_{\alpha\beta\gamma\delta}\phi\right)A_{\alpha\beta}A_{\gamma\delta}\right. \\ &\left.\left. -\ 5\frac{\partial_{\alpha\beta\gamma}\phi\partial_{\delta\mu\nu}\phi}{24}A_{\alpha\beta}A_{\gamma\delta}A_{\mu\nu}\right)\right)\end{aligned} \tag{8.18}$$

with $A^{\alpha\beta}=[\partial_{\alpha\beta}\phi]^{-1}$, $dx\equiv\prod_{i=1}^n dx_i$ and x^* the saddle-point (which minimizes $\phi(x)$). This expression can be obtained by developing $\phi(x)$ and $f(x)$ in Taylor series around x^*. The quadratic part in $\phi(x)$ leads to a Gaussian integration over x which can be performed. Details are left in appendix A.
The saddle-point (a^*,L^*) is the point on the hyperplane $s_{\alpha\beta}=s$ which minimizes the geodesic distance that we have previously computed

$$(a^*,\{L_i^*\})\equiv(a,\{L_i\})\mid \min_{a,\{L_i\},\lambda} d^2(x,x^0)+\lambda(s_{\alpha\beta}(L)-s)$$

The saddle-point (a^*,L^*) can then be computed efficiently. It is determined by solving the following non-linear equations ($q_i^*=q_i(L_i^*)$)

$$\frac{(\rho^{ia}\frac{(a^*(s)-a^0)}{\nu}+\sum_{j=1}^n\rho^{ij}q_j^*)d(a^*,\{q_i^*\})}{a^*(s)\sinh(d(a^*,F^*))}=\frac{\lambda}{a^0}\frac{\partial s_{\alpha\delta}}{\partial q_i}|^* \tag{8.19}$$

with $a^*(s)$ fixed by

$$a^*(s)^2\rho^{aa}=(a^0)^2\rho^{aa}-2\nu a^0\sum_{i=1}^n\rho^{ia}q_i^*+\nu^2\sum_{i,j=1}^n\rho^{ij}q_i^*q_j^* \tag{8.20}$$

An approximation (which could be used as a guess solution in a numerical optimization routine) can be obtained by linearizing these equations around the spot Libor rates (i.e., $q_i = 0$):

$$q_i^* \approx \frac{\sum_{j=1}^n \tilde{\rho}_{ij}\omega_j(s-s_0)}{\sum_{p,q=1}^n \omega_p\omega_q\tilde{\rho}_{pq}} \tag{8.21}$$

with $\omega_i \equiv \frac{\partial s_{\alpha\delta}}{\partial q_i}(q_i = 0)$ and $\tilde{\rho}^{ij} = \rho^{ij} - \frac{\rho^{ia}\rho^{ja}}{\rho^{aa}}$. Note that when the strike is close to the money, the saddle-point coordinates are close to the spot Libors and $a^* = a^0$. In our experience, the solution of the linearized equations is sufficiently accurate – this is what we have used to generate the swaption smiles shown in the following section.

By plugging the asymptotic expression (4.67) for the density $p(t, L, a|L^0, \alpha)$ into (8.15) and doing the integration over $\mathbb{B}$ using the saddle-point approximation, we finally get the following expression for the effective local volatility $\sigma_{\rm loc}^{\alpha\delta}$ at first order in time [104]: ($\partial_{n+1} \equiv \partial_a$, $\partial_i \equiv \partial_{L_i}$ $i = 1,\ldots,n$)

$$\begin{aligned}(\sigma_{\rm loc}^{\alpha\delta})^2(t,s) = \sum_{i,j=1}^n \rho_{ij}\sigma_i(t)\sigma_j(t)f_{ij}(a^*,L^*)(1+2t'\sum_{\mu,\nu=1}^{n+1}A^{\mu\nu}\{\frac{\partial_{\mu\nu}f_{ij}(a^*,L^*)}{f_{ij}(a^*,L^*)} \\ +2\frac{\partial_\mu f_{ij}(a^*,L^*)}{f_{ij}(a^*,L^*)}\frac{\partial_\nu\psi(a^*,L^*)}{\psi(a^*,L^*)} - \sum_{\gamma,\delta=1}^{n+1}A^{\gamma\delta}\frac{\partial_\mu f_{ij}(a^*,L^*)}{f_{ij}(a^*,L^*)}\partial_{\nu\gamma\delta}d^2(a^*,L^*)\})\end{aligned} \tag{8.22}$$

with $({a^*}^2(s), \{F_i^*\}_i(s))$ the saddle-point satisfying the equations (8.19,8.20) approximated by (8.21) and

$$\begin{aligned}
f_{ij}(a,F) &= a^2C_i(L_i)C_j(L_j)\frac{\partial s_{\alpha\delta}}{\partial L_i}\frac{\partial s_{\alpha\delta}}{\partial L_j}\ ,\ \psi(a,F) = \sqrt{g\Delta}\mathcal{P} \\
A^{\alpha\beta} &= [\partial_{\alpha\beta}d^2]^{-1} \\
\ln(\mathcal{P})(a,q|\alpha) &\simeq \frac{1}{\nu}\sum_{j=1}^n\left(\frac{b^j(0,L^0)}{C_j(L_j^0)} - \frac{\partial_jC_j(L_j^0)}{2}\right)\left(\nu\sum_{i=1}^n\rho^{ij}q_i + \rho^{aj}(a-\alpha)\right) \\
&\quad - b^a(0,L^0)\left(\nu\sum_{i=1}^n\rho^{ia}q_i + \rho^{aa}(a-\alpha)\right) \\
\Delta(a,F) &= \left(\frac{d(a,F)}{\sinh d(a,F)}\right)^n \\
\sqrt{g} &= \frac{2^{\frac{n+1}{2}}\det[\rho]^{-\frac{1}{2}}}{\nu a^{1+n}\prod_{i=1}^n C_i(L_i)} \\
t' &= \frac{\nu^2}{2(\rho^{aa} - \sum_{i,j=1}^n \rho^{ia}\rho^{ja}\overline{\rho}_{ij})}t;\ \overline{\rho}_{ij} \equiv [\rho^{ij}]^{-1}_{i,j=1,\ldots,n}
\end{aligned}$$

Here, by definition, $a^2b^j(t,F)$ and $-va^2b^a(t,F)$ are the drifts of the Libors (8.13) and the volatility (8.14) in the swap numéraire. The terms $b^j(0,F^0)$

and $b^a(0, F^0)$ are small and can be neglected. As great an expression (8.22) may appear, it is analytical and the result of the straightforward application of methods which can be used in a similar fashion for all stochastic volatility models listed in Table 8.1. It is important to point out that an additional benefit of using the techniques is that the expression (8.22) for the ELV is *exact* in the limit $t \to 0$.

The effective local volatility for Libors L_k is given by plugging in a straightforward manner $\delta = \alpha + 1 \equiv k$ in equation (8.22):

$$(\sigma^k_{\mathrm{loc}})^2(t, L_k) = \sigma_k(t)^2 f_k(a^*, L^*_k)\left(1 + 2\nu^2 tA \left(-3 + 2a^*\partial_a \ln \mathcal{P}^2(a^*, L_k) - a^* A\partial^3_a d^2(a^*, L_k)\right)\right) \tag{8.23}$$

with

$$\begin{aligned} f_k(a, L_k) &= a^2 C_k(L_k)^2\ , \ A = (\partial_{aa} d^2)^{-1}(a^*, L_k) \\ \ln(\mathcal{P})(a, F) &= -\frac{\partial_k C_k(L^0_k)}{2\nu(1-\rho^2)}(\nu q_k - \rho(a - a^0)) \\ d(a, L_k) &= \cosh^{-1}[1 + \frac{\nu^2 q_k^2 - 2\nu(a - a^0)\rho q_k + (a - a^0)^2}{2(1-\rho^2)aa^0}] \\ a^{*2} &= (a^0)^2 + 2\nu a^0 \rho q_k + \nu^2 q_k^2 \end{aligned} \tag{8.24}$$

Note that as explained above, we have neglected the drift of the stochastic volatility a.

8.5.3 Second step: deriving an implied volatility smile

Given the effective local volatility for $s_{\alpha\delta}$ we now need to compute its smile. The great benefit of using the ELV is that we have shrunk an $(n+1)$-dimensional problem down to a one-dimensional problem. Solving numerically the one-dimensional forward PDE is a natural method for obtaining the whole smile. We prefer to derive an accurate analytical approximation, using the method explained in chapter 5:

To start with, we re-write the swap rate dynamics under $\mathbb{P}^{\alpha\delta}$ as

$$ds_{\alpha\delta} = \frac{\sigma^{\mathrm{loc}}_{\alpha\delta}(t, s_{\alpha\delta})}{\sigma^{\mathrm{loc}}_{\alpha\delta}(t, s^0_{\alpha\delta})}\sigma^{\mathrm{loc}}_{\alpha\delta}(t, s^0_{\alpha\delta})dW_t$$

Doing a change of time, we obtain

$$ds_{\alpha\delta} = C(t', s_{\alpha\delta})dW_{t'}$$

with $t' = \int_0^t \left(\sigma^{\mathrm{loc}}_{\alpha\delta}(u, s^0_{\alpha\delta})\right)^2 du$ and $C(t', f) = \frac{\sigma^{\mathrm{loc}}_{\alpha\delta}(t, s_{\alpha\delta})}{\sigma^{\mathrm{loc}}_{\alpha\delta}(t, s^0_{\alpha\delta})}$. Finally, by applying the formula (5.40), this yields the following asymptotic formula for $\sigma^{\alpha\delta}_{\mathrm{BS}}(K, T_\alpha)$

in the short-time limit:

$$\sigma_{\rm BS}^{\alpha\delta}(K,T_\alpha) = \sqrt{\frac{\int_0^{T_\alpha}(\sigma_{\rm loc}^{\alpha\delta})^2(u,s_0^{\alpha\delta})du}{T_\alpha}}\sigma_{\rm BS}^{\alpha\delta}(K)_0 \left(1+\frac{1}{2}\int_0^{T_\alpha}(\sigma_{\rm loc}^{\alpha\delta})^2(u,s_0^{\alpha\delta})du\sigma_{\rm BS}^{\alpha\delta}(K)_1\right) \tag{8.25}$$

with

$$\sigma_{\rm BS}^{\alpha\delta}(K)_0 = \frac{\ln\left(\frac{K}{s_0^{\alpha\delta}}\right)}{\int_{s_0^{\alpha\beta}}^K \frac{dx}{C(x)}}$$

$$\sigma_{\rm BS}^{\alpha\delta}(K)_1 = -\frac{1}{\left(\int_{s_0^{\alpha\delta}}^K \frac{dx}{C(x)}\right)^2}\ln\left(\frac{(\sigma_{\rm BS}^{\alpha\delta}(K)_0)^2 K s_0^{\alpha\delta}}{C(K)}\right) + \frac{1}{(\sigma_{\rm loc}^{\alpha\delta})(0,s_0^{\alpha\beta})^2}\frac{\partial_t\left(\frac{\sigma_{\rm loc}^{\alpha\delta}(0,f_{\rm av})}{\sigma_{\rm loc}^{\alpha\delta}(0,s_0^{\alpha\beta})}\right)}{C(f_{\rm av})}$$

with $C(f) \equiv \frac{\sigma_{\rm loc}^{\alpha\delta}(0,K)}{\sigma_{\rm loc}^{\alpha\delta}(0,s_0^{\alpha\delta})}$, $f_{\rm av} \equiv \frac{s_0^{\alpha\delta}+K}{2}$ and $\sigma_{\rm loc}^{\alpha\delta}(t,s)$ given by (8.22). Note that this expression for $\sigma_{\rm BS}^{\alpha\delta}$ is exact when $T_\alpha \to 0$ and becomes identical to the familiar BBF formula (5.39) involving the harmonic average of local volatility.

8.5.4 Numerical tests and comments

We have tested our asymptotic swaption formula (8.25) in various scenarios. In the following graphs, an $x \times y$ swaption means an option maturity of x years, a swap length of y years and a tenor of one year. The SABR-LMM models used semi-annual Libors. We have chosen a two factor correlation structure as in [44]

$$dW_k = \sum_{r=1}^{2}\theta_r e^{-k_r(T_{k-1}-t)}dZ_r;\ dZ_1 dZ_2 = \rho dt \tag{8.26}$$

with $\theta_1 = 1$, $\theta_2 = 0.5$, $k_1 = 0.25$, $k_2 = 0.04$, $\rho = -40\%$.

▷ Firstly, by assuming that the CEV parameters have the same $\beta_k = \beta$ and the volatility of the volatility $\nu = 0$, the model degenerates into the CEV LMM model: we can thus compare our smiles against the Andersen-Andreasen asymptotic results [44]. The expressions above degenerate into

$$
\begin{aligned}
f_{ij}(L) &= C_i(L_i)C_j(L_j)\frac{\partial s_{\alpha\delta}}{\partial L_i}\frac{\partial s_{\alpha\delta}}{\partial L_j} \\
d(L) &= \sqrt{2\sum_{i,j=1}^{n}\rho^{ij}q_iq_j} \\
\ln(\mathcal{P})(q) &= \sum_{j=1}^{n}(\frac{b^j(L^0,0)}{C_j(L_j^0)} - \frac{\partial_j C_j(L_j^0)}{2})\sum_{i=1}^{n}\rho^{ij}q_i \\
\Delta(L,L^0) &= 1 \\
\sqrt{g} &= \frac{2^{\frac{n}{2}}\det[\rho]^{-\frac{1}{2}}}{\prod_{i=1}^{n}C_i(L_i)}
\end{aligned}
$$

with the saddle-points satisfying the non-linear equations (modulo the constraint $s_{\alpha\delta} = s$)

$$\sum_{j=1}^{n}\rho^{ij}q_j^* = -\frac{\lambda}{4}\frac{\partial s_{\alpha\delta}}{\partial q_i}|^*$$

The CEV model is calibrated (using an optimization routine) against ATM caplets (with $\beta = 0.2$) and to all ATM swaptions $x \times 2$ and $x \times 10$ in order to set the Libor volatility term structure (i.e., $\sigma_i(t)$). Calibration was done using the JPY curve on February 17th 2007. The two approaches yield very close results (see Fig. 8.4).

$\triangleright$ Secondly, we take $\nu = 20\%$ with a zero correlation between the volatility and the Libors (i.e., $\rho_{ia} = 0$). We calibrate the SABR-LMM model to all caplet smiles (by adjusting ϕ_i for the ATM volatility and β_i for the skew) and we set the Libor volatility term structure (i.e., $\sigma_i(t)$) by matching (best fit) all ATM swaptions $x \times 10$ and $x \times 2$. Calibration was done using the EUR curve on February 17th 2007. Our formula is tested against a Monte-Carlo simulation with a time step $\Delta t = 0.02$ and 2^{17} paths[3] (see Fig. 8.5). We also plot the calibrated values of beta (see Fig. 8.6).

$\triangleright$ Thirdly, we choose ρ_{ia} to be non-zero and run similar tests as above (see Fig. 8.7, 8.8).

$\triangleright$ In the generic SABR-LMM model, once caplet smiles are calibrated, we can use the parameters ϕ_i and β_i to generate different swaption smiles using the parameters v and ρ_{ia}. We illustrate this capability of the model we propose with the example of the 10×10 swaption smile (caplets calibrated to the EUR curve-February 17th 2007) (see Fig. 8.9). Note that by construction, the 10×10 ATM swaption is calibrated.

[3] We have used a predictor-corrector scheme with a Brownian bridge.

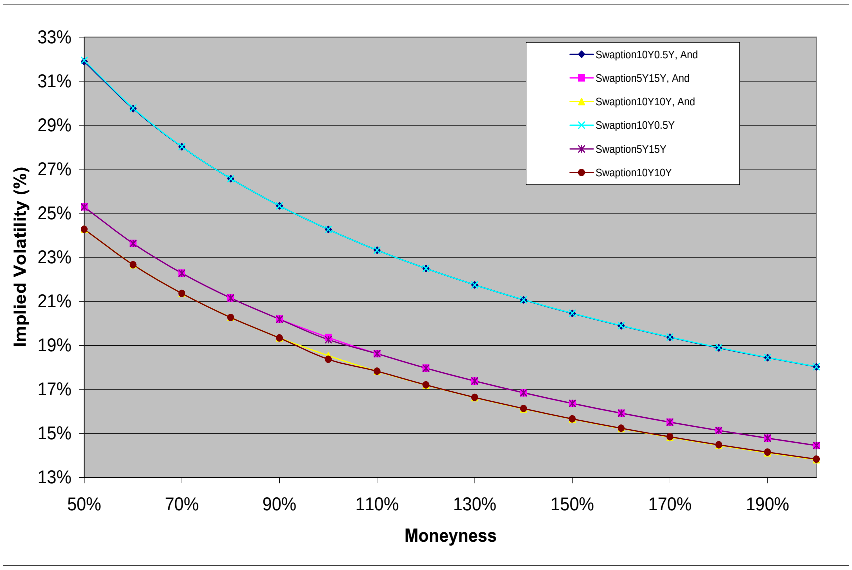

FIGURE 8.4: The figure shows implied volatility smiles for swaption 10×0.5, 5×15 and 10×10 using our asymptotic formula and the Andersen-Andreasen expression. Here $\nu = 0$ and $\beta = 0.2$.

Conclusion

We have investigated a LMM model coupled to a SABR stochastic volatility process. Not only does this model allow each Libor to have its own value of the skew parameter β, but it also makes it possible to correlate the Libors with the stochastic volatility process.

By using the heat kernel expansion technique, we have derived accurate expressions for swaption implied volatilities. We have tested the accuracy of our swaption asymptotic formula against an exact (MC) pricing and other known analytical approximations.

We have shown that the additional degree of freedom afforded by the correlation between the Libors and the stochastic volatility makes it possible to adjust swaption smiles once caplet smiles have been individually calibrated. This decoupling will assist in the joint calibration of caplet and swaption smiles.

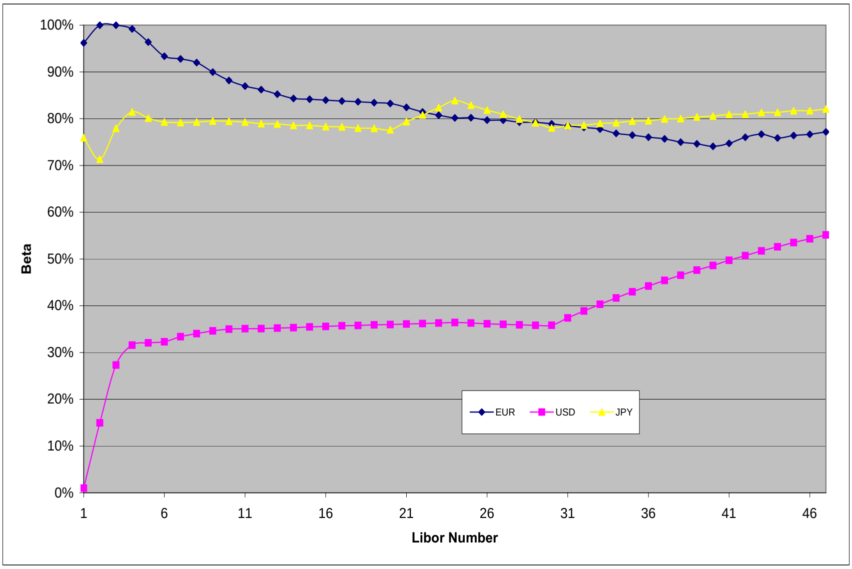

FIGURE 8.5: The figure shows implied volatility smiles for swaption 10×0.5, 5×15 and 10×10 using our asymptotic formula and a MC simulation. Here $\nu = 20\%$ and $\rho_{ia} = 0\%$.

8.6 Extensions

The formula can be slightly extended to include a mean-reverting drift for the stochastic volatility. Under the spot Libor measure, we assume that the volatility follows the process

$$da = -\nu a^2 \psi^a(a)dt + \nu a dZ_{n+1}$$

with $\psi^a(a)$ a general analytical function of a (the scaling a^2 in front of $\psi^a(a)$ has been put for convenience). After some algebraic computations, we derive

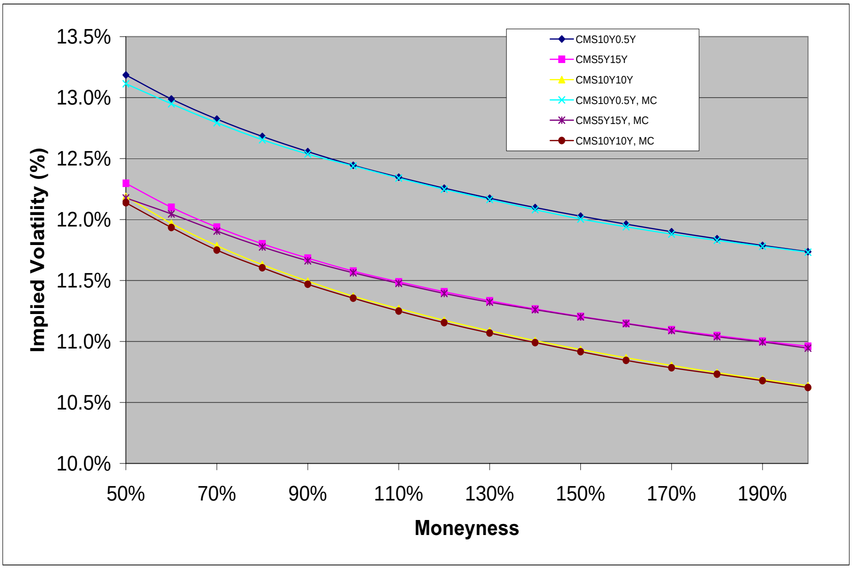

FIGURE 8.6: CEV parameters β_k calibrated to caplet smiles for EUR-JPY-USD curves. The x coordinate refers to the Libor index.

the new Abelian 1-form connection

$$\mathcal{A} = \frac{1}{\nu}\sum_{j=1}^{n}\left(\frac{b^j(t,L)}{C_j(L_j)} - \frac{\partial_j C_j(L_j)}{2}\right)\left(\nu\sum_{i=1}^{n}\rho^{ij}dq_i + \rho^{aj}da\right)$$
$$- \left(b^a(t,L) + \psi(a)\right)\left(\nu\sum_{i=1}^{n}\rho^{ia}dq_i + \rho^{aa}da\right)$$

Using a similar approximation as before, i.e., $C_j(L_j) \sim C_j(L_j^0)$ and $\psi^a(a) \sim \psi^a(\alpha)$, we obtain for the parallel gauge transport (see remark 4.3)

$$\ln(\mathcal{P})(a,q|\alpha) \sim \quad \frac{1}{\nu}\sum_{j=1}^{n}\left(\frac{b^j(0,L^0)}{C_j(L_j^0)} - \frac{\partial_j C_j(L_j^0)}{2}\right)\left(\nu\sum_{i=1}^{n}\rho^{ij}q_i + \rho^{aj}(a-\alpha)\right)$$
$$- \left(b^a(0,L^0)\psi(\alpha)\right)\left(\nu\sum_{i=1}^{n}\rho^{ia}q_i + \rho^{aa}(a-\alpha)\right)$$

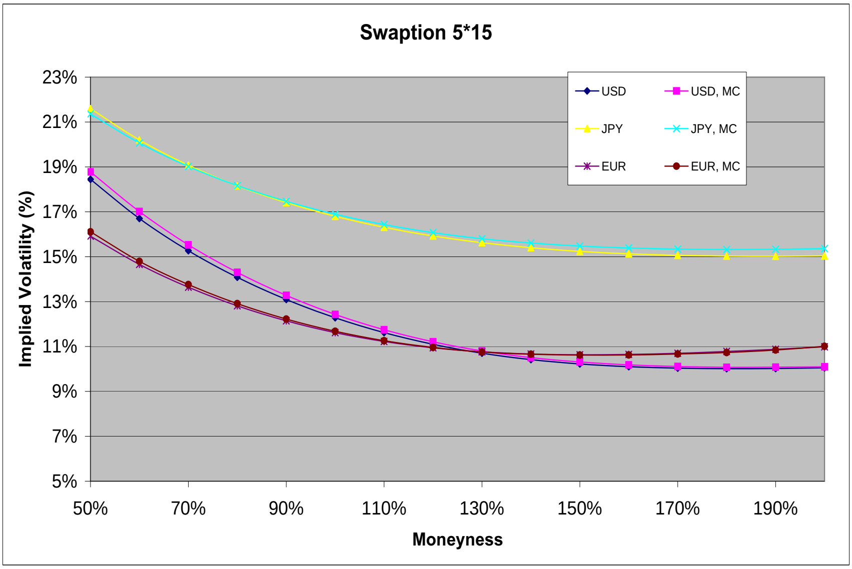

FIGURE 8.7: The figure shows implied volatility smiles for swaption 5×15 (EUR, JPY, USD curves-February 17th 2007) using our asymptotic formula and a MC simulation. Here $\nu = 20\%$ and $\rho_{ia} = -30\%$.

Finally, the implied volatility is obtained using our general formula (8.25). Note that the metric and the geodesic equations remain unchanged when we only modify drift terms.

8.7 Problems

Exercises 8.1 Markov LMM
As seen in this chapter, the MC simulation of a LMM is time-consuming. The main problem comes from the drifts which are reminiscent of the non-Markovian nature of the HJM model. To circumvent this difficulty, one can

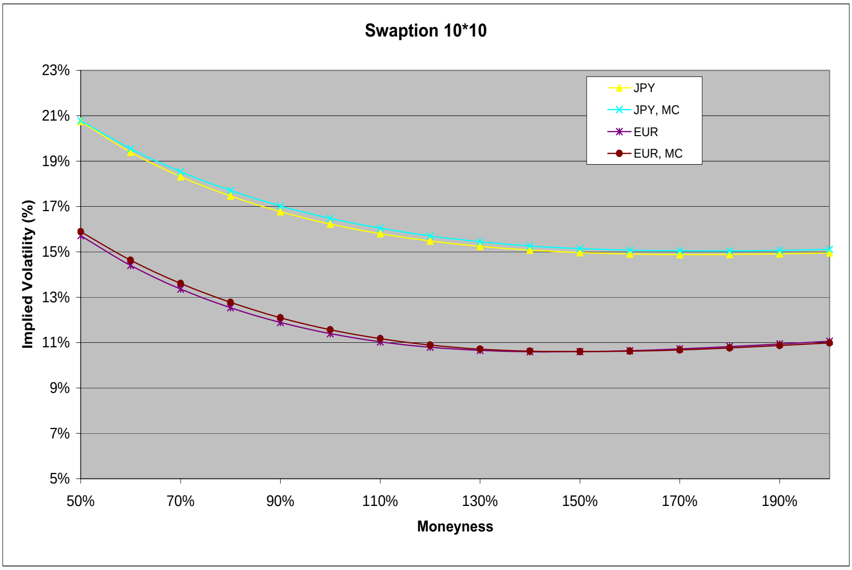

FIGURE 8.8: The figure shows implied volatility smiles for swaption 10×10 (EUR, JPY-February 17th 2007) using our asymptotic formula and a MC simulation. Here $\nu = 20\%$ and $\rho_{ia} = -30\%$.

try to find a *low-dimensional Markov representation* of the LMM. By this, we assume that each Libor can be written as a functional of the time t and a low-dimensional Itô diffusion process Y_t

$$L_i \equiv L_i(t, Y_t) \ , \ i = 1, \cdots, n$$

In this problem, we make the additional assumption that Y_t is a one dimensional process.
The process follows a general Itô SDE

$$dY = \mu(t, Y)dt + \sigma(t, Y)dW_t$$

1. Prove that it is always possible to assume that the process Y_t can be written as $Y_t = f(t, X_t)$ with X_t a driftless process. By redefining the functional L_i, we have $L_i = L_i(t, X_t)$.

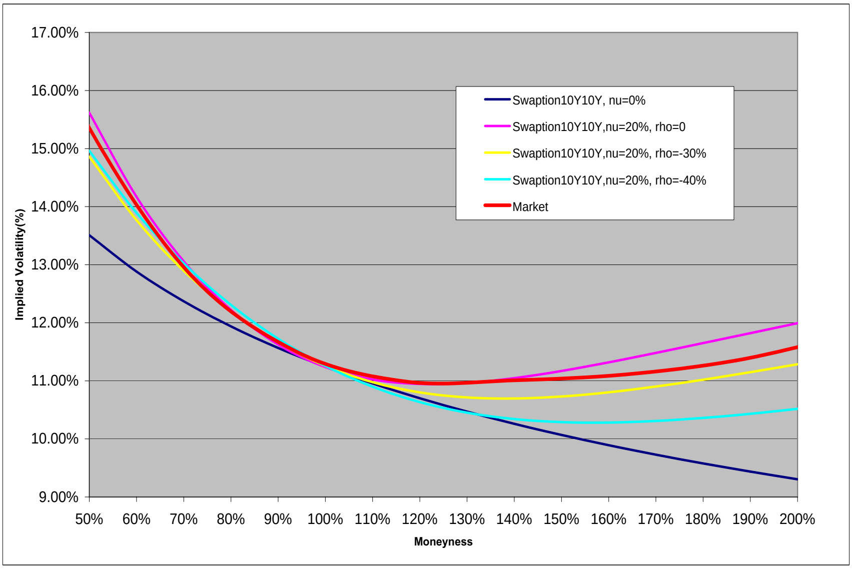

FIGURE 8.9: The figure shows implied volatility smiles for swaption 10×10 (EUR-February 17th 2007) using different value of (ν, ρ). The LMM has been calibrated to caplet smiles and the ATM swaptions $x \times 2$, $x \times 10$.

Note: The process X_t has no financial interpretation. However, for a short rate model, X_t can be seen as the instantaneous interest rate r_t.
2. For an affine short-rate model such as the HW2 model (see 8.2), obtain the explicit map $L_i(t, x_t, y_t)$.
3. Prove that the Markovian property imposes from (8.6) the following constraint

$$\sigma_i = \sigma(t, X)\partial_X L_i \tag{8.27}$$

$$-\sigma(t, X)^2 \partial_X L_i \sum_{j=i+1}^{n} \frac{\tau_j \partial_X L_j}{1 + \tau_j L_j} = \partial_t L_i + \frac{1}{2}\sigma(t, X)^2 \partial_X^2 L_i \tag{8.28}$$

In the following, we explicitly check that the BGM model **does not** have a Markov representation. The Libors are driven by a log-normal process $\sigma_i = \psi_i(t) L_i$.

The equations (8.27,8.28) reduce to

$$\psi_i(t) = \sigma(t, X)\partial_X \ln L_i \tag{8.29}$$

$$-\sigma(t, X)^2 \partial_X L_i \sum_{j=i+1}^{n} \frac{\tau_j \partial_X L_j}{1 + \tau_j L_j} = \partial_t L_i + \frac{1}{2}\sigma(t, X)^2 \partial_X^2 L_i \tag{8.30}$$

Without loss of generality, we take $L_n = X$.
4. Prove that the first constraint (8.29) is equivalent to

$$L_i = \xi_i(t) L_n^{\frac{\psi_n(t)}{\psi_i(t)}}$$

5. Using this solution, prove that the second constraint (8.30) reduces

$$-\psi_n^2 \xi_i(t) \sum_{j=i+1}^{n} \frac{\tau_j \xi_j(t) L_n^{\frac{\psi_n(t)}{\psi_j(t)}}}{1 + \tau_j \xi_j(t) L_n^{\frac{\psi_n(t)}{\psi_j(t)}}} = \xi_i(t)' + \xi_i(t) \left(\frac{\psi_n(t)}{\psi_i(t)}\right)' L_n^{-1} + \frac{1}{2}\psi_n(t)^2 \xi_i(t) \left(\frac{\psi_n(t)}{\psi_i(t)}\right)^2$$

6. Specifying this equation for $i = n - 1$, show that this equation has only a solution in the degenerate case $\xi_{n-1}(t) = 0$.
So, although every short-rate models admit a Markovian representation by construction, it is not the case for a particular LMM. To be able to do that, we must assume that the volatility of each Libor is a general Dupire local volatility $\sigma_i(t, L_i)$. In the following, we explain how to construct a MF introduced first in [110].
In order to be able to simulate exactly the process X_t, we take X_t as a Gaussian process

$$dX = \sigma(t) dW_t \tag{8.31}$$

with the initial condition $X_0 = 0$.
We introduce the deflated bond

$$\zeta_i(t) = \frac{P_{tT_i}}{P_{tT_n}} \; i = 1, \cdots n$$

We recall that a Libor $L_i(t)$ is given by

$$\tau_i L_i(t) = \frac{\zeta_{i-1}(t)}{\zeta_i(t)} - 1 \tag{8.32}$$

7. Prove that $\zeta_i \equiv \zeta_i(t, X)$ satisfies the PDE

$$\partial_t \zeta_i(t, X) + \frac{1}{2}\sigma(t)^2 \partial_X^2 \zeta_i(t, X) = 0$$

8. Doing a change of local time, prove that this is equivalent to the heat kernel on $\mathbb{R}$ and the fundamental solution is

$$\zeta_i(t', X) = \int_{-\infty}^{\infty} \zeta_i(T_i', Y) p_G(T_i', Y|t', X) dY \tag{8.33}$$

where $t' < T_i'$.
Therefore the process $\zeta_i(t, X)$ is completely characterized if we specify its terminal value at $t = T_i$. Its terminal value will be determined in order to calibrate exactly the implied volatility of a unique swaption starting at T_i.
9. Prove that the fair value of a digital swaption option with an underlying swap $s_{\alpha\beta}$ is in the measure $\mathbb{P}^n$

$$\mathcal{D}_{\alpha\beta}(K) = -P_{tT_n} \mathbb{E}^{\mathbb{P}^n}[\sum_{i=\alpha+1}^{\beta} \tau_i \zeta_i(T_\alpha, X_\alpha) 1(s_{\alpha\beta}(T_\alpha, X_\alpha) > K)|\mathcal{F}_t] \tag{8.34}$$

10. Assuming that $s_{\alpha\beta}(T_\alpha, X)$ is a monotone function in X, prove that the digital fair value (8.34) becomes

$$\mathcal{D}_{\alpha\beta}(K) = -P_{tT_n} \sum_{i=\alpha+1}^{\beta} \tau_i \int_{X^*}^{\infty} \zeta_i(T_\alpha, X) p_G(T_\alpha', X|0) dX \tag{8.35}$$

11. In particular for $\alpha = n-1$, $\beta = n$, prove that (8.35) becomes

$$\mathcal{D}_{n-1,n}(K) = -P_{tT_n} \tau_n (1 - \mathcal{N}(X^*))$$

with $\mathcal{N}(X)$ the cumulative Gaussian distribution. Given the caplet implied volatility $\sigma_{n-1,n}^{\mathrm{BS}}(K)$, we can determine X^* as a function of K by

$$L_n(T_{n-1}, X^*) = K^*$$

12. Now, let us assume that $\zeta_i(t, X)$ is known for $\forall i > \alpha$. Then, identifying in (8.35) $\mathcal{D}_{\alpha\beta}(K)$ with the market fair value, we determine K^* for a given X^*. We obtain

$$s_{\alpha\beta}(T_\alpha, X^*) = K^*$$

13. Prove the relation

$$\zeta_\alpha(T_\alpha, X^*) = \zeta_\beta(T_\alpha, X^*) + K \sum_{i=\alpha+1}^{\beta} \tau_i \zeta_i(T_\alpha, X^*) \tag{8.36}$$

By integrating (8.33) with the terminal condition $\zeta_\alpha(T_\alpha, X^*)$ given by (8.36), we obtain $\zeta_\alpha(t, X)$ for all $t \leq T_\alpha$ and the Libors via (8.32).

14. Using a Taylor expansion at the first-order around the swap spot, prove that the terminal correlation of two swap rates is approximatively

$$< \ln s_{i\beta}(T_i), \ln s_{i\delta}(T_j) > \approx \sqrt{\frac{\int_0^{T_i} \sigma(s)^2 ds}{\int_0^{T_j} \sigma(s)^2 ds}}$$

By taking $\sigma(t) = e^{at}$ where a is called the *mean-reverting* coefficient (see question 2.), we have in particular

$$< \ln s_{i\beta}(T_i), \ln s_{i\delta}(T_j) > \approx \sqrt{\frac{e^{2aT_i} - 1}{e^{2aT_j} - 1}}$$

Exercises 8.2 Almost Markov LMM

For a generic LV LMM, the Libor dynamics in the spot Libor measure $\mathbb{P}^s$ is

$$dL_i(t) = \sigma_i(t)\Phi_i(L_i) \sum_{j=\beta(t)}^{i} \frac{\rho_{k,j}(t)\tau_j\sigma_j(t)\Phi_j(L_j)}{1+\tau_j L_j(t)} dt$$
$$+\sigma_i(t)\Phi_i(L_i)dZ_i$$

1. To get rid of the local volatility $\Phi_i(L_i)$, we define the new variables $q_i(t) = \int_{L_i^0}^{L_i} \frac{dx}{\Phi_i(x)}$. Show that in these new variables, the SDEs above become

$$dq_i(t) = \sigma_i(t) \left(\mu_i(t, L)dt - \frac{1}{2}\sigma_i(t)\Phi_i'(L_i)dt + dZ_i \right)$$

The difficulty in making this SDE Markovian comes from the drift term. The simplest way to deal with this problem is to replace it with its value at the spot Libor

$$dq_i(t) = \sigma_i(t) \left(\mu_i(t, L^0)dt - \frac{1}{2}\sigma_i(t)\Phi_i'(L_i^0)dt + dZ_i \right) \tag{8.37}$$

This is equivalent in spirit to the Hull-White-Rebonato freezing argument.
2. Integrate (8.37) to

$$q_i(t) = d_i(t) + \int_0^t \sigma_i(t)dZ_i$$

with $d_i(t)$ a deterministic function.
3. Deduce that a Libor L_i can be written as a function of a time-changed Brownian motion

$$L_i(t) = L_i(t, Z_{i,t'})$$

4. Assuming the separability condition of the volatility $\sigma_i(t)$

$$\sigma_i(t) = \nu_i \sigma(t)$$

prove that the Libors $\{L_i\}$ can be mapped to a unique time-changed Brownian motion

$$L_i(t) = L_i(t, W_{t'})$$

Chapter 9

Solvable Local and Stochastic Volatility Models

Abstract In the previous chapters, we have been focusing on (geometric) approximation methods particularly useful for the calibration of local and stochastic volatility models. However, in the case of particular models, we can have an exact solution for the Kolmogorov and the Black-Scholes equations. In this chapter we provide an extensive classification of one and two dimensional diffusion processes which admit an exact solution to the Kolmogorov (and hence Black-Scholes) equation (in terms of hypergeometric functions). By identifying the one-dimensional solvable processes with the class of *integrable superpotentials* introduced recently in *supersymmetric Quantum Mechanics*, we obtain new analytical solutions. In particular, by applying *supersymmetric transformations* on a known solvable diffusion process (such as the Natanzon process for which the solution is given by a hypergeometric function), we obtain a hierarchy of new solutions. These solutions are given by a sum of hypergeometric functions. For two-dimensional processes, more precisely stochastic volatility models, the classification is achieved for a specific class called gauge-free models including the Heston model, the 3/2-model and the geometric Brownian model. We then present a new exact stochastic volatility model belonging to this class.
In addition to our geometrical framework, we will use tools from functional analysis in particular the spectral decomposition of (unbounded) linear operators.

9.1 Introduction

For most mathematical models of asset dynamics, an exact solution for the corresponding Kolmogorov & Black-Scholes equation is usually not available; there are, however, a few notable exceptions. The known solutions for local volatility models are the constant elasticity of variance (CEV) [76] including the classical log-normal Black-Scholes process. For the instantaneous short rate models, there are the CIR process [77] (Bessel process) and the Vasicek-Hull-White process [108] (Ornstein-Uhlenbeck process). For stochastic volatility models, the known exact solutions are for the Heston model [106], the 3/2-model [29] and the geometric Brownian model [29]. These analytical solutions can be used for calibrating a model quickly and efficiently or can serve as a benchmark for testing the implementation of more realistic models requiring intensive numerical computations (Monte-Carlo, PDE). For example, the existence of a closed-form solution for the fair value of a European call option in the Heston model allows us to quickly calibrate the model to the implied volatilities observed on the market. The calibrated model can then be used to value path-dependent exotic options using, for example, a Monte-Carlo methodology.

In this chapter, we show how to obtain new analytic solutions to the Kolmogorov & Black-Scholes equation, which we refer to as KBS throughout the rest of this chapter, for 1d & 2d diffusion processes. In order to get to our classification, we first present a general reduction method to simplify the multi-dimensional KBS equation. Rewriting the KBS equation as a heat kernel equation on a Riemannian manifold endowed with an Abelian connection, we show that this covariant equation can be simplified using both the group of diffeomorphisms (i.e., change of variables) and the group of Abelian gauge transformations. In particular for the models admitting a *flat Abelian connection*, there always exists a gauge transformation that eliminates the Abelian connection of the diffusion operator.

In the second part we apply the reduction method, previously presented, to one-dimensional, time-homogeneous diffusion processes. Modulo a change of variable, the metric becomes flat and the Abelian connection is an exact one-form for which a gauge transformation can always be applied. Using these two transformations, the resulting KBS equation becomes an *Euclidean Schrödinger equation* with a scalar potential. Extensive work has already been done to classify the set of scalar potentials which admit an exact solution. In particular, using a supersymmetric formulation of the Schrödinger equation which consists in doubling the KBS equation with another equation, we show how to generate a hierarchy of new solvable diffusion processes starting from a known solvable diffusion process (for, e.g., a Natanzon potential [129]). In this context, the local volatility function is identified with a *superpotential*. Applying *supersymmetric transformations* on the *Natanzon potential* (which is the most general potential for which the Schrödinger equation can be reduced to either a hypergeometric or a confluent equation), we obtain a new class of solvable one-dimensional diffusion processes which are characterized by six parameters.

The classification of one-dimensional time-homogeneous solvable diffusion processes for which the solution to the KBS equation can be written as a hypergeometric function has been achieved in [40, 41, 42, 120] using the well-known Natanzon classification. The application of supersymmetric techniques to the classification of solvable potentials for the Schrödinger equation has been reviewed in [10] where a large number of references can be found. For the Kolmogorov & Fokker-Planck equations, one can consult [118, 115].

In the last part we pursue this classification for SVMs which admit a flat Abelian connection: we refer to these as *gauge-free models*. Surprisingly, this class includes all the well-known exact SVMs (i.e., the Heston model, the 3/2-model and the geometric Brownian model). For these gauge-free models, we reduce the two-dimensional KBS equation to an Euclidean Schrödinger equation with a scalar potential. Then, we present a new exact SVM which is a combination of the Heston and 3/2-models.

9.2 Reduction method

In this section, we explain how to simplify the KBS equation. This reduction method will be used in the next section to classify the solvable one and two dimensional time-homogeneous processes. This method is already well known for one-dimensional processes and is presented in [70, 30, 126] amongst others. However, the extension of this method to multi-dimensional diffusion processes requires the introduction of differential geometric objects such as a metric and an Abelian connection on a Riemannian manifold, as we have already seen in chapter 4.

Let us assume that our time-homogeneous multi-dimensional market model depends on n Itô processes x_i which can either be traded assets or market-unobservable Itô processes (such as a stochastic volatility a or an instantaneous short rate r). Let us denote $x = (x_i)_{i=1,\cdots,n}$, with the initial conditions $\alpha = (\alpha_i)_{i=1,\cdots,n}$. These variables x_i satisfy the following time-homogeneous SDE

$$\begin{aligned} dx_t^i &= b^i(x_t)dt + \sigma^i(x_t)dW_i \\ dW_i dW_j &= \rho_{ij}dt \end{aligned}$$

with the initial condition $x_i(t = 0) = \alpha_i$. The no-arbitrage condition implies that there exists an equivalent measure $\mathbb{P}$ such that the traded assets are (local) martingales under this measure. For $\mathbb{P}$, the drifts b_i are consequently zero for the traded assets (i.e., forwards). Note that the measure $\mathbb{P}$ is not unique as the market is not necessarily complete. Finally, the fair value of a (European) option, with payoff $f(x^i)$ at maturity T, is given by the discounted mean value of the payoff f conditional on the filtration $\mathcal{F}_t$ generated by the Brownian motions $\{W^i_{s\leq t}\}_{i=1,\cdots,n}$

$$\mathcal{C}(\alpha, t, T) = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T r_s ds} f|\mathcal{F}_t]$$

with r_s the instantaneous short rate. This mean-value depends on the probability density $p(T, x|\alpha)$ which satisfies the backward Kolmogorov equation

($\tau = T - t$, $\partial_i = \frac{\partial}{\partial \alpha_i}$)

$$\frac{\partial p}{\partial \tau} = b^i \partial_i p + \frac{1}{2} \rho_{ij} \sigma^i \sigma^j \partial_{ij} p \tag{9.1}$$

with the initial condition $p(\tau = 0) = \delta(x - \alpha)$. As usual we have used the Einstein convention meaning that two repeated indices are summed. Using the Feynman-Kac theorem 2.5, one can show that the fair value $\mathcal{C}$ of the option satisfies the Black-Scholes equation

$$\frac{\partial \mathcal{C}}{\partial \tau} = b^i \partial_i \mathcal{C} + \frac{1}{2} \rho_{ij} \sigma^i \sigma^j \partial_{ij} \mathcal{C} - r(\alpha) \mathcal{C} \tag{9.2}$$

with the initial condition $\mathcal{C}(\tau = 0, \alpha) = f(\alpha)$. In chapter 4, the PDE 9.1 has been interpreted as a heat kernel on a smooth n-dimensional manifold M endowed with a metric g_{ij} and an Abelian connection $\mathcal{A}$

$$\frac{\partial p(\tau, x|\alpha)}{\partial \tau} = D p(\tau, x|\alpha) \tag{9.3}$$

and D given by (4.55). Similarly, the Black-Scholes equation (9.2) can be rewritten as

$$\frac{\partial \mathcal{C}(\tau, \alpha)}{\partial \tau} = (D - r) \mathcal{C}(\tau, \alpha)$$

The heat kernel equation can now be simplified by applying the actions of the following groups:

▷ The group of diffeomorphisms $\mathrm{Diff}(\mathcal{M})$ which acts on the metric g_{ij} and the connection $\mathcal{A}_i$ by

$$\begin{aligned}(f^* g)_{ij} &= g_{pk} \partial_i f^p(x) \partial_j f^k(x) \\ (f^* \mathcal{A})_i &= \mathcal{A}_p \partial_i f^p(x) \ , \ f \in \mathrm{Diff}(\mathcal{M})\end{aligned}$$

▷ The group of gauge transformations $\mathcal{G}$ (see section 4.5 for details) which acts on the conditional probability (and the fair value $\mathcal{C}$) by

$$\begin{aligned}p'(\tau, x|\alpha) &= e^{\chi(\tau, x) - \chi(0, \alpha)} p(\tau, x|\alpha) \\ \mathcal{C}'(\tau, \alpha) &= e^{\chi(\tau, \alpha)} \mathcal{C}(\tau, \alpha)\end{aligned}$$

Then p' ($\mathcal{C}'$) satisfies the same equation as p ($\mathcal{C}$) (9.3) only with

$$\begin{aligned}\mathcal{A}'_i &\equiv \mathcal{A}_i - \partial_i \chi \\ Q' &\equiv Q + \partial_\tau \chi\end{aligned}$$

If the connection $\mathcal{A}$ is an exact form (meaning that there exists a smooth function Λ such that $\mathcal{A}_i = \partial_i \Lambda$), then by applying a gauge transformation, we can eliminate the connection so that the heat kernel equation for p' (or $\mathcal{C}'$)

has a connection equal to zero. It can be shown that for a simply-connected manifold, the statement "$\mathcal{A}$ is exact" is equivalent to $\mathcal{F} = 0$, where $\mathcal{F}$ is the 2-form curvature given by

$$\mathcal{F} = (\partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i) d\alpha^i \wedge d\alpha^j$$

It is straightforward to prove that if $\mathcal{A}_i = \partial_i \Lambda$, then $\mathcal{F} = 0$ as $\partial_{ij}\Lambda = \partial_{ji}\Lambda$ for a smooth function. In the following, we will restrict our classification to those processes for which $\mathcal{F} = 0$, meaning there exists a gauge transformation such that the transformed connection vanishes. The operator D reduces in this case to the symmetric operator $D = \Delta + Q$ for which we can use an *eigenvector expansion*. In this context, we review necessary materials from the (linear) operator theory and the spectral decomposition of (unbounded) self-adjoint linear operators on a (infinite-dimensional) Hilbert space. The Itô generator of the Brownian $D = \partial_x^2$ on an interval $I \subset \mathbb{R}$ will be our guide to illustrate the notions we introduce in the next section.

9.3 Crash course in functional analysis

Classical references for this section are [35] and [36].
Here $\mathcal{H}$ denotes a real *separable Hilbert space* with the scalar product (x, y). For completeness sake, a real (complex) Hilbert space is a real (complex) linear space equipped with an inner product $(x, y) \in \mathbb{R}$ ($\mathbb{C}$) such that

$$\begin{aligned}
(x, y) &= \overline{(y, x)} \\
(\lambda x_1 + x_2, y) &= \lambda(x_1, y) + (x_2, y) \\
(x, \lambda y_1 + y_2) &= \overline{\lambda}(x, y_1) + (x, y_2)
\end{aligned}$$

where $\lambda \in \mathbb{R}$ ($\mathbb{C}$) and $\bar{\cdot}$ stands for complex conjugation.

$$\begin{aligned}
&(x, x) \geq 0 \ , \ x \in \mathcal{H} \\
&(x, x) = 0 \text{ if and only if } x = 0
\end{aligned}$$

Such an inner product defines a *norm* $\parallel x \parallel \equiv \sqrt{(x, x)}$ and by definition $\mathcal{H}$ is complete with respect to this norm. An Hilbert space is *separable* if it admits a countable basis meaning that every vector x can be decomposed over an orthonormal basis $\{e_i\}_{i=1,\cdots,\infty}$[1]

$$x = \sum_{i=1}^{\infty} x_i e_i \ \ x_i \in \mathbb{R} \ (\mathbb{C})$$

[1]Note that we have necessarily $\parallel x \parallel^2 = \sum_{i=1}^{\infty} x_i^2 < \infty$.

9.3.1 Linear operator on Hilbert space

An operator D is a linear mapping $D : \mathrm{Dom}(D) \to \mathcal{H}$ where $\mathrm{Dom}(D)$, called the domain of D, is a linear subspace of $\mathcal{H}$.
For any two operators D_1, D_2, their sum $D_1 + D_2$ and the product $D_1 D_2$ are defined as follows

$$(D_1 + D_2)(x) = D_1 x + D_2 x \ , \ \forall \ x \in \mathrm{Dom}(D_1) \cap \mathrm{D}om(D_2)$$
$$(D_1 D_2)(x) = D_1(D_2(x)) \ , \ \forall \ x \in \mathrm{Dom}(D_1 D_2)$$

with $\mathrm{Dom}(D_1 D_2) \equiv \{x \in \mathrm{Dom}(D_2) \mid D_2 x \in \mathrm{Dom}(D_1)\}$.

DEFINITION 9.1 Operator densely defined *An operator D with domain $\mathrm{Dom}(D)$ is said to be densely defined if the subset $\mathrm{Dom}(D)$ is dense in $\mathcal{H}$, i.e., for any $x \in \mathcal{H}$ one can find a sequence $\{x_n\}$ in $\mathrm{D}om(D)$ which converges in norm to x.*

We define the norm of the operator D by

DEFINITION 9.2 Norm

$$||D|| = \sup_{x \in \mathcal{H} \ ; \ x \neq 0} \frac{||Dx||}{||x||} \tag{9.4}$$

If $||D|| < \infty$, then D is a bounded operator otherwise unbounded. The finiteness of the norm $||D||$ is equivalent to the continuity of D. Hence if D is a bounded operator densely defined on $\mathcal{H}$, it can be extended by continuity to the whole Hilbert space $\mathcal{H}$.

In the following, all the operators D are densely defined operators on a (separable) Hilbert space $\mathcal{H}$.

DEFINITION 9.3 Adjoint operator *We call the subspace $\mathrm{Dom}^\dagger$, the space of vectors x such that the linear form $x \to (y, Dx)$ is continuous for the norm of $\mathcal{H}$ for all $y \in \mathcal{H}$. Hence using Riesz's theorem [36], there exists a unique x' such that*

$$(y, Dx) = (x', x) \tag{9.5}$$

By definition, we set $D^\dagger y = x'$. $D^\dagger$ is called the adjoint operator of D and its domain is $\mathrm{Dom}(D^\dagger) \equiv \mathrm{Dom}^\dagger$.

Note that an adjoint operator D is only defined for an operator densely defined on $\mathcal{H}$. Indeed, if $\mathrm{Dom}(D)$ is not dense in $\mathcal{H}$, there exists $z \neq 0$ such that $(z, y) = 0 \ \forall y \in \mathcal{H}$. If x' and y are vectors satisfying (9.5), then $x' + \lambda z$ for all

$\lambda \in \mathbb{C}$ satisfies (9.5) for the same y. The adjoint $D^\dagger$ is not uniquely defined in this case.

DEFINITION 9.4 Symmetric operator *An operator D is called symmetric if for all $x, y \in \mathrm{Dom}(D)$, we have*

$$(Dx, y) = (x, Dy)$$

It is equivalent to the condition

$$D^\dagger = D \ \text{ on } \mathrm{Dom}(D) \subset \mathrm{Dom}(D^\dagger)$$

DEFINITION 9.5 Self-adjoint operator *D is self-adjoint if additionally*

$$\mathrm{Dom}(D^\dagger) = \mathrm{Dom}(D)$$

Before illustrating these notions on the Itô generator of a Brownian motion, we review the definition of a *Sobolev space* H^m and list few useful properties.

Sobolev space H^m

Let I be an interval of $\mathbb{R}$. As a recall, a function $g \in L^2(I)$ is said to be the *weak derivatives* of f if

$$\int_I g(x)\phi(x)dx = -\int_I f(x)\phi'(x)dx$$

for all $\phi \in C_0^\infty(I)$ the space of C^∞ function with compact support on I. In this case, we write $g \equiv f'$. Similarly, we define the k^{th} weak derivative $f^{(k)}$. For any integer m, the *Sobolev space* $H^m(I)$ is

$$H^m(I) = \{f \in L^2(I) \ : \ f^{(k)} \in L^2(I) \ , \ \forall \, k = 1, \cdots, m\}$$

Endowed with the norm

$$||f||_{H^m} = \left(\sum_{k=1}^{m} ||f^k||_{L^2}\right)^{\frac{1}{2}}$$

$H^m(I)$ is a Hilbert space. Below, we list some useful properties that characterize a function in $H^m(I)$

- $f \in H^m(I)$ if and only if $f \in L^2(I) \cap C(\overline{I})$ is continuously differentiable up to order $m-1$, $f^{(m)}$ exists almost everywhere and $f^{(m)} \in L^2(I)$. Moreover, all the functions $f, f^{(1)}, \cdots, f^{(m-1)}$ are *absolutely continuous*.[2]

[2] A function F is said to be absolutely continuous on I if there exists a function $f \in L^2(I)$ such that $F(x) = F(a) + \int_a^x f(y)dy \ , \ \forall x \in I$.

- $\forall f, g \in H^1(I)$, the integration by parts formula holds

$$\int_a^b f(x)g'(x)dx = [fg]_a^b - \int_a^b f'(x)g(x)dx \ , \ \forall a, b \in I$$

and $\forall f, g \in H^2(I)$,

$$\int_a^b f(x)g''(x)dx = [fg']_a^b - [f'g]_a^b + \int_a^b f''(x)g(x)dx \ , \ \forall a, b \in I \tag{9.6}$$

- In the case of unbounded I, if $f \in H^1(I)$ then

$$\lim_{|x|\to\infty} f(x) = 0 \tag{9.7}$$

To illustrate the notions above, we look at the *self-adjoint extensions* of the Itô generator of a one-dimensional Brownian motion $D = \partial_x^2$ on an interval I (not necessarily bounded). Here $\mathcal{H} = L^2(I)$.

DEFINITION 9.6 Self-adjoint extension *An operator D' is a self-adjoint extension of an operator D if*

$$D' = D \text{ on } \mathrm{Dom}(D) \subset \mathrm{Dom}(D^\dagger)$$

and

$$D'^\dagger = D' \text{ on } \mathrm{Dom}(D) \subset \mathrm{Dom}(D') = \mathrm{Dom}(D'^\dagger) \subset \mathrm{Dom}(D^\dagger)$$

Self-adjoint extensions: Examples

Using the integration by parts formula (9.6), we have

$$(Df, g) = (f, Dg) \ \ \forall f, g \in H^2(I)$$

with $D = \partial_x^2$ and $I = (a, b)$ if and only if

$$[fg']_a^b = [f'g]_a^b \tag{9.8}$$

Example 9.1 $I = \mathbb{R}$
We set $\mathrm{Dom}(D) = C_0^\infty(\mathbb{R})$. As the space $C_0^\infty(\mathbb{R})$ is dense in $\mathcal{H}$, the adjoint can be defined. From the condition (9.8) and the property (9.7), we obtain that the adjoint is $D^\dagger = \partial_x^2$ with the domain $\mathrm{Dom}(D^\dagger) = H^2(\mathbb{R})$. $(D, \mathrm{Dom}(D))$ is therefore symmetric but not self-adjoint. It admits a unique self-adjoint extension: $(D, H^2(\mathbb{R}))$ is self-adjoint. $\square$

Example 9.2 $I = \mathbb{R}_+$
$(D, C_0^\infty(\mathbb{R}_+))$ is symmetric but not self-adjoint. Note that because of the property (9.7), the condition (9.8) reduces to

$$\frac{f'(0)}{f(0)} = \frac{g'(0)}{g(0)}$$

We deduce that D admits self-adjoint extensions parameterized by the group $U(1) = \{e^{i\theta} \ , \ \theta \in [0, 2\pi)\}$:

$$\mathrm{Dom}(D_\theta) = \{f \in H^2(\mathbb{R}_+) \, , \, (f(0) - \imath f'(0)) = e^{i\theta}(f(0) + \imath f'(0)) \ , \ \theta \in [0, 2\pi)\}$$

which is equivalent to

$$\mathrm{Dom}(D_\theta) = \{f \in H^2(\mathbb{R}_+) \, , \, f(0) = \lambda f'(0) \ , \ \lambda \in \mathbb{R} \cup \{\infty\}\}$$

with $\lambda = -\tan(\frac{\theta}{2})$.
The Neumann (resp. Dirichlet) boundary $f'(0) = 0$ (resp. $f(0) = 0$) corresponds to $\lambda = \infty$ (resp. $\lambda = 0$). □

Example 9.3 $I = [0, 1]$
$(D, C_0^\infty(I))$ is symmetric but not self-adjoint. It admits self-adjoint extensions parameterized by the group G of two-dimensional unitary matrices $U(2)$ satisfying the condition:

$$G = \{U \in U(2) \ , \ \det(id - \overline{U}U) = 0\} \tag{9.9}$$

where $\bar{\cdot}$ is the complex conjugation. The domain of all of these extensions is

$$\mathrm{Dom}(D_G) = \{f \in H^2(I) \, , \, \begin{pmatrix} f'(0) - \imath f(0) \\ f'(1) + \imath f(1) \end{pmatrix} = U \begin{pmatrix} f'(0) + \imath f(0) \\ f'(1) - \imath f(1) \end{pmatrix} \ , \ U \in G\}$$

The reality of the function f implies the condition (9.9). □

9.3.2 Spectrum

For any operator on a complex Hilbert space $\mathcal{H}$, we say that a complex number λ is a *regular value* of D if the operator $D - \lambda id$ has a bounded inverse, e.g.,

$$\ker(D - \lambda id) = \{0\} \ , \ \mathrm{ran}(D - \lambda id) = \mathcal{H} \ , \ ||(D - \lambda id)^{(-1)}|| < \infty$$

Any $\lambda \in \mathbb{C}$ which is not a regular value is called a *singular value*. The set of all regular values is called the *resolvent* of D and the set of all singular values is called the *spectrum* of D and is denoted $\mathrm{spec}(D)$. For example, if λ is an eigenvalue of D, that is $Dx = \lambda x$ for some $x \neq 0$, then λ belongs to the spectrum of D.

If D is an operator in a real Hilbert space $\mathcal{H}$ then we define first complexified space $\mathcal{H}^{\mathbb{C}} = \mathcal{H} + \imath\mathcal{H}$ and extend D by linearity to an operator $D^{\mathbb{C}}$ in $\mathcal{H}^{\mathbb{C}}$. We set $\mathrm{spec}(D) = \mathrm{spec}(D^{\mathbb{C}})$. The spectrum of a self-adjoint operator can be characterized as

PROPOSITION 9.1
The spectrum of a self-adjoint operator D is a non-empty closed subset of $\mathbb{R}$. Also, we have

$$||D|| = \sup_{\lambda \in \mathrm{spec}(D)} |\lambda|$$

9.3.3 Spectral decomposition

Let us assume that $(D, \mathrm{Dom}(D))$ is self-adjoint. Before presenting the spectral decomposition in the general case of a self-adjoint unbounded linear operator on an (infinite-dimensional) Hilbert space, we recall the finite-dimensional case in a form generalizable to the infinite-dimensional case. For a self-adjoint operator D on a finite-dimensional Hilbert space, we can find a basis where the operator D is diagonalizable. If we denote $(\lambda_i)_{i=1,\cdots,m}$ the set of different eigenvalues and P_i the projector operator on the vector subspace generated by the eigenvectors associated to the eigenvalue λ_i, then D can be decomposed as

$$D = \sum_{i=1}^{m} \lambda_i P_i \tag{9.10}$$

We recall that the projectors P_i satisfy the algebra

$$P_i P_j = P_i \delta_{ij} \ , \ P_i^{\dagger} = P_i \ , \ \sum_{i=1}^{m} P_i = id$$

Introducing the operator $\mathbb{E}(\lambda) = \sum_{i:\lambda_i \leq \lambda} P_i$, we prove that $\mathbb{E}(\lambda)$ satisfies the following properties

$$\begin{aligned}
\mathbb{E}(\lambda)\mathbb{E}(\lambda') &= \mathbb{E}(\min(\lambda, \lambda')) \\
\lim_{\lambda \to \lambda_i^-} \mathbb{E}(\lambda) &= \mathbb{E}(\lambda_i) \\
\lim_{\lambda \to -\infty} \mathbb{E}(\lambda) &= 0 \\
\lim_{\lambda \to \infty} \mathbb{E}(\lambda) &= 1
\end{aligned}$$

with 1 the identity map on $\mathcal{H}$. The sign lim should be understood as the strong convergence meaning that $E_n \to E$ strongly if

$$\lim_{n \to \infty} ||E_n - E|| = 0$$

The operators $\mathbb{E}(\lambda)$ allow to define a specific measure where the sign sum in the formula (9.10) can be written as a Lebesgue-Stieltjes integral. This spectral decomposition formula is generalizable to unbounded self-adjoint operators. Beforehand, let us recall briefly the construction of the Lebesgue-Stieltjes integral.

DEFINITION 9.7 Stieltjes function *Let $F(\lambda)$ be a monotone, increasing, left-continuous function satisfying* $\lim_{\lambda\to-\infty} F(\lambda) = 0$ *and* $\lim_{\lambda\to\infty} F(\lambda) < \infty$*.*

From the properties satisfied by $\mathbb{E}(\lambda)$, $||\mathbb{E}(\lambda)x||$ is an example of Stieltjes functions.
From this function, we can define a measure μ_F on the Borel σ-algebra on $\mathbb{R}$: On the semi-open interval $[a, b)$, we set

$$\mu_F([a, b)) = F(b) - F(a)$$

As μ_F is a σ-additive function on the semi-ring formed by all semi-open intervals, it can be extended to all Borel sets by the Carathéodory extension theorem (2.1).
Hence we can integrate a measurable function ϕ against the measure μ_F:

$$\int_{\mathbb{R}} \phi(\lambda) d\mu_F(\lambda)$$

This integral is called the Lebesgue-Stieltjes integral. If a function F can be decomposed as the difference of two Stieltjes functions F_1 and F_2 ($F = F_1 - F_2$), we set

$$\int_{\mathbb{R}} \phi(\lambda) d\mu_F(\lambda) \equiv \int_{\mathbb{R}} \phi(\lambda) d\mu_{F_1}(\lambda) - \int_{\mathbb{R}} \phi(\lambda) d\mu_{F_2}(\lambda) \tag{9.11}$$

Note that for any two vectors $f, g \in \mathcal{H}$, we have

$$(\mathbb{E}(\lambda)f, g) = (\mathbb{E}(\lambda)f, \mathbb{E}(\lambda)g) = \frac{1}{4}\left(||\mathbb{E}(\lambda)(f+g)||^2 - ||\mathbb{E}(\lambda)(f-g)||^2\right)$$

Hence, $(\mathbb{E}(\lambda)f, g)$ is a difference of two Stieltjes functions and therefore defines a measure according to (9.11). In this context, the sum in the formula (9.10) can be rewritten as a Lebesgue-Stieltjes integral:

$$(g, Df) = \int_{\mathbb{R}} \lambda d(g, \mathbb{E}(\lambda)f) \tag{9.12}$$

Here $(g, \mathbb{E}(\lambda)f)$ is a discrete measure with support the eigenvalues $\{\lambda_i\}_{i=1,\cdots,m}$. The formula (9.12) can be written formally as

$$D = \int_{\mathbb{R}} \lambda d\mathbb{E}(\lambda) \tag{9.13}$$

For a self-adjoint operator in an infinite-dimensional Hilbert space, a similar spectral decomposition is still valid. However, the spectrum is not necessarily discrete and can have a continuous part. The spectrum can belong to a Borel subset of $\mathbb{R}$, and $\mathbb{E}(\lambda)$ defines a resolution of the identity:

DEFINITION 9.8 Resolution of the identity *Let $\mathcal{H}$ be a separable real Hilbert space and let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$. A family of bounded linear operators $\{\mathbb{E}(B)\ : B \in \mathcal{B}(\mathbb{R})\}$ in $\mathcal{H}$ such that*

- *Each $\mathbb{E}(B)$ is an orthogonal projection (i.e., $\mathbb{E}(B)^2 = \mathbb{E}(B)$ and $\mathbb{E}(B)^\dagger = \mathbb{E}(B)$).*
- $\mathbb{E}(\varnothing) = 0$, $\mathbb{E}(\mathbb{R}) = 1$ *(1 is the identity operator in $\mathcal{H}$).*
- *If $B = \cup_{n=1}^{\infty} B_n$ with $B_n \cap B_m = 0$ if $n \neq m$, then $\mathbb{E}(B) = \sum_{n=1}^{\infty} \mathbb{E}(B_n)$ (where the limit involved in the infinite series is taken in the strong convergence topology).*
- $\mathbb{E}(B_1)\mathbb{E}(B_2) = \mathbb{E}(B_1 \cap B_2)$

is called a resolution of the identity.

The definition above ensures that $(f, \mathbb{E}(B)f)$ is a Stieltjes function.
For a self-adjoint operator, one can show that there exists a resolution of the identity and the operator can be decomposed as (9.13). More precisely, we have

THEOREM 9.1 Spectral decomposition
Let D be a self-adjoint operator in a real Hilbert space $\mathcal{H}$. Then there exists a unique spectral resolution $\{\mathbb{E}(B)\ : B \in \mathcal{B}(\mathbb{R})\}$ in $\mathcal{H}$ such that the following spectral decomposition holds

$$(g, Df) = \int_{\mathbb{R}} \lambda d(g, \mathbb{E}(\lambda)f)$$

Moreover, the domain of D is given by

$$\mathrm{Dom}(D) = \{f \in \mathcal{H} | \int_{\mathbb{R}} \lambda d(f, \mathbb{E}(\lambda)f) < \infty\}$$

This formula is written formally as (9.13).

The spectral decomposition is valid only if the symmetric operator D is an (unbounded) self-adjoint operator or admits a self-adjoint extension. This will depend strongly on the boundary conditions as seen in the examples above. In order to show that D is self-adjoint or admits self-adjoint extensions, we can use the *deficiency indices* technique introduced by Von Neumann [36] (see [66] for a pedagogical introduction). The *deficiency indices* are defined by

$$n_{\pm} = \dim \ker(D^{\dagger} \mp \imath id)$$

THEOREM 9.2
For an operator with deficiency indices (n_-, n_+), there are three possibilities:

1. *If $n_- = n_+ = 0$, then D is self-adjoint (necessary and sufficient condition).*

2. *If $n_+ = n_- = n \geq 1$, then D has infinitely many self-adjoint extensions, parameterized by the unitary group $U(n)$ satisfying the real condition (9.9).*

3. *If $n_- \neq n_+$, then D has no self-adjoint extension.*

REMARK 9.1 Note that if D is a linear differential operator of order n and the deficiency indices are (n, n) then the spectrum is purely discrete [35].
□

If D admits a self-adjoint extension, the resulting conditional probability is not unique but depends on the boundary conditions which are parameterized by the unitary group $U(n)$. We look back at the self-adjoint extensions of the Itô generator of a one-dimensional Brownian motion $D = \partial_x^2$ on an interval I using the Von-Neumann criteria. Here $D^{\dagger} = \partial_x^2$. The eigenvectors for the eigenvalues $\pm\imath$ are

$$\phi_{\pm}(x) = e^{\pm\sqrt{\pm\imath}x}$$

We examine if $\phi_{\pm}$ belong to $L^2(I)$.

Example 9.4 $I = \mathbb{R}$
$(n_-, n_+) = (0, 0)$. D is self-adjoint. □

Example 9.5 $I = \mathbb{R}_+$
$(n_-, n_+) = (1, 1)$. D admits a self-adjoint extension parameterized by $U(1)$.
□

Example 9.6 $I = [0, 1]$
$(n_-, n_+) = (2, 2)$. D admits a self-adjoint extension parameterized by $U(2)$. Moreover, the spectrum is discrete from the remark 9.1. □

Having proved that $-D \equiv \Delta + Q - r$ is a self-adjoint operator or admits a self-adjoint extension, the gauge-transform option value $\mathcal{C}'$ satisfying the equation

$$\frac{\partial \mathcal{C}'(\tau,\alpha)}{\partial \tau} = (\Delta + Q - r)\mathcal{C}'(\tau,\alpha)$$

admits a unique solution in $\mathcal{H}$ given by

$$\mathcal{C}'(\tau,\alpha) = \int_{\mathrm{spec}(D)} e^{-\lambda\tau} d\mathbb{E}(\lambda) f \tag{9.14}$$

We have applied the spectral decomposition to D and not $-D$ only for the purpose of convenience. As we will see in the next section, the spectrum of $-D$ is included in $\mathbb{R}_+$.

DEFINITION 9.9 Semi-group *In the equation (9.14), the family D_τ $(\tau \geq 0)$ defined as*

$$D_\tau \equiv e^{-\tau D} = \int_{\mathrm{spec}(D)} e^{-\lambda\tau} d\mathbb{E}(\lambda)$$

is called the heat semi-group associated to D and satisfies the following properties:

1. $D_{t+s} = D_t D_s$ $t, s > 0$ *and* $D_0 = 1$. *(semi-group axiom)*

2. $||D_t|| \leq 1$ *(contraction property)*

3. *The mapping $t \rightarrow D_t$ is strongly continuous on $[0,\infty)$. That is, for any $t \geq 0$ and $f \in \mathcal{H}$,*

$$\lim_{s \rightarrow t} D_s f = D_t f$$

where the limit is understood in the norm of $\mathcal{H}$. In particular, for any $f \in \mathcal{H}$,

$$\lim_{t \rightarrow 0} D_t f = f$$

4. *For all $f \in \mathcal{H}$ and $t > 0$, we have $D_t f \in \mathrm{Dom}(\mathcal{H})$ and*

$$\frac{d}{dt}(D_t f) = -D(D_t f)$$

where $\frac{d}{dt}$ is the Fréchet derivative.

How can we construct a resolution of identity in practice?

We consider the *generalized* eigenvectors

$$D\phi_\lambda = \lambda\phi_\lambda$$

By "generalized," we mean that it is not required that ϕ_λ belongs to $\mathcal{H}$ and therefore λ does not necessarily belong to the discrete spectrum. We normalize the function ϕ_λ (ϕ_λ is a distribution) such that

$$\int_{\mathrm{spec}(D)} \phi_\lambda(s)\phi_\lambda(s_0)d\lambda = \delta(s - s_0) \tag{9.15}$$

Then, the resolution of identity of D is

$$(g, \mathbb{E}(\lambda)f) = \int_{-\infty}^{\lambda} (g, \phi_\lambda)(\phi_\lambda, f)d\lambda$$

and the unique solution to the gauge-transform Kolmogorov equation is

$$\begin{aligned} p'(t, s|s_0) &= \int_{\mathrm{spec}(D)} e^{-\lambda t}\phi_\lambda(s)\phi_\lambda(s_0)d\lambda \qquad (9.16) \\ &= \sum_k e^{-\lambda_k t}\phi_{\lambda_k}(s)\phi_{\lambda_k}(s_0) + \int_{\mathrm{spec}/\mathrm{spec}_{\mathrm{d}}(D)} e^{-\lambda t}\phi_\lambda(s)\phi_\lambda(s_0)d\lambda \end{aligned}$$

where we have split the spectrum into a discrete $\mathrm{spec}_d(D)$ and continuous parts $\mathrm{spec}/\mathrm{spec}_{\mathrm{d}}(D)$.

Example 9.7 $D = \partial_s^2, I = \mathbb{R}$

The eigenvector $\phi_\lambda(s)$ associated to the eigenvalue $-\lambda^2$ satisfies the ODE

$$D\phi_\lambda(s) = -\lambda^2\phi_\lambda(s)$$

for which the solutions are

$$\phi_\lambda(s) = e^{\pm i\lambda s} \ , \ \lambda \in \mathbb{R}$$

The (real) eigenvector which satisfies the completude relation (9.15) is

$$\phi_\lambda(s) = \frac{1}{\sqrt{2\pi}} (\cos(\lambda s) - \sin(\lambda s))$$

From (9.16), we get the normal distribution

$$\begin{aligned} p'(t, s|s_0) &= \int_{-\infty}^{\infty} e^{-\lambda^2 t}\phi_\lambda(s)\phi_\lambda(s_0)d\lambda \\ &= \frac{1}{\sqrt{4\pi t}} e^{-\frac{(s-s_0)^2}{4t}} \end{aligned}$$

□

9.4 1D time-homogeneous diffusion models

In the next section, we apply the general reduction method to the one-dimensional KBS equation. Though similar reductions to a Schrödinger equation with a scalar potential (without any references to differential geometry however) can be found in [70, 30, 126], we rederive this classic result using our general reduction method. This method will prove to be useful when we discuss the classification of SVMs. What is more, we find the supersymmetric partner of this Schrödinger equation and show how to generate new exact solutions for European call options.

9.4.1 Reduction method

Let us consider a one-dimensional, time-homogeneous diffusion process with drift[3]

$$df_t = \mu(f_t)dt + \sigma(f_t)dW_t$$

This process has been used as a basis for various mathematical models in finance. If f is a traded asset (i.e., a forward), the drift vanishes in the forward measure and we end up with a local volatility model where we assume that the volatility is only a function of f. This has been explored in detail in chapter 5. The one-dimensional process is not necessarily driftless as the random variable is not always a traded asset as it is the case for an instantaneous short rate model, or a model of stochastic volatility.

In our framework, this process corresponds to a (one-dimensional) real curve endowed with the metric $g_{ff} = \frac{2}{\sigma(f)^2}$. For the new coordinate

$$s(f) = \sqrt{2}\int_{.}^{f} \frac{df'}{\sigma(f')}$$

the metric is flat: $g_{ss} = 1$. It follows that the Laplace-Beltrami operator (4.57) becomes $\triangle = \partial_s^2$. Using the definition (4.58), (4.59), we find that the Abelian connection $\mathcal{A}$ and the function Q are given by

$$\mathcal{A} = \left(-\frac{1}{2}\partial_f \ln\sigma(f) + \frac{\mu(f)}{\sigma^2(f)}\right) df \tag{9.17}$$

$$Q = \frac{1}{4}\left(\sigma(f)\sigma''(f) - \frac{1}{2}\sigma'(f)^2\right) - \frac{\mu'(f)}{2} + \frac{\mu(f)\sigma'(f)}{\sigma(f)} - \frac{\mu^2(f)}{2\sigma^2(f)} \tag{9.18}$$

[3] The time-dependent process $df_t = \mu(f_t)A^2(t)dt + \sigma(f_t)A(t)dW_t$ is equivalent to this process under the change of local time $t' = \int_0^t A(s)^2 ds$.

$\mathcal{A}$ is an exact form (it is always the case in one dimension) with

$$\Lambda(f) = -\frac{1}{2}\ln\left(\frac{\sigma(f)}{\sigma(f_0)}\right) + \int_{f_0}^{f} \frac{\mu(f')}{\sigma^2(f')} df'$$

By applying a gauge transformation on the conditional probability $p(t, f|f_0)$

$$P(t, s) = \frac{\sigma(f_0)}{\sqrt{2}} e^{\Lambda(f)} p(t, f|f_0) \tag{9.19}$$

the connection vanishes and P satisfies the equation (in the s flat coordinate)

$$\partial_\tau P(\tau, s) = \left(\partial_s^2 + Q(s)\right) P(\tau, s) \tag{9.20}$$

This is known as an *Euclidean Schrödinger equation* with a scalar potential $Q(s)$. This equation is considerably simpler than the Quantum Mechanics Schrödinger equation seeing that all the terms are real-valued.
The solution P has been scaled by the (constant) factor $\frac{\sigma(f_0)}{\sqrt{2}}$ in order to obtain the initial condition

$$\lim_{\tau \to 0} P(\tau, s) = \delta(s)$$

Moreover, Q is given in the s coordinate by

$$Q = \frac{1}{2}(\ln\sigma)^{''}(s) - \frac{1}{4}((\ln\sigma)^{'}(s))^2 - \frac{\mu'(s)}{\sqrt{2}\sigma(s)} + \frac{\sqrt{2}\mu(s)\sigma'(s)}{\sigma(s)^2} - \frac{\mu(s)^2}{2\sigma(s)^2} \tag{9.21}$$

where the prime $'$ indicates a derivative according to s.

Example 9.8 Quadratic volatility process
Let us assume that f satisfies a driftless process (i.e., $\mu(f) = 0$). The Black-Scholes equation reduces to the heat kernel on $\mathbb{R}$ if $Q(s) = \text{cst}$ (i.e., $Q(s)$ is zero modulo a time-dependent gauge transformation) which is equivalent to

$$\sigma(f)\sigma''(f) - \frac{1}{2}\sigma'(f)^2 = 4\text{cst}$$

By differentiating with respect to f, we get

$$\sigma(f)\sigma'''(f) = 0$$

and finally

$$\sigma(f) = \alpha f^2 + \beta f + \gamma$$

(i.e., the quadratic volatility model studied in detail in [145]) with cst $= \frac{\alpha\gamma}{2} - \frac{\beta^2}{8}$. The solution $P(t,s|s_0)$ to the Schrödinger equation is

$$P(t,s|s_0) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{(s-s_0)^2}{4t} + \mathrm{cst}t}$$

By using the explicit map (9.19) between $p(t,f|f_0)$ and its gauge transform $P(t,s|s_0)$ with $s = \sqrt{2}\int^f \frac{df'}{\sigma(f')}$, we obtain (see remark 4.3) the solution to the Kolmogorov equation

$$p(t,f|f_0) = \frac{1}{\sigma(f)\sqrt{2\pi t}} \frac{\sigma(f_0)}{\sigma(f)} e^{-\frac{(\int_{f_0}^f \frac{df'}{\sigma(f')})^2}{2t} + \mathrm{cst}t}$$

□

Example 9.9 CEV process
For the CEV process $df_t = f_t^\beta dW_t$, the potential is

$$Q(s) = \frac{\beta(\beta-2)}{4(1-\beta)^2 s^2}\ ,\ 0 < \beta < 1$$
$$Q(s) = -\frac{1}{8}\ ,\ \beta = 1$$

Note that in order to ensure that f_t is a true martingale, we have assumed that $\beta \leq 1$. The Schrödinger equation is

$$D\phi_\lambda(s) = \lambda\phi_\lambda(s)$$

with $D = -\partial_s^2 - gs^{-2}$ and $g = \frac{\beta(\beta-2)}{4(1-\beta)^2}$.

(Generalized) eigenvectors

We will now attempt to find the (generalized) eigenvectors $\phi_\lambda(s)$. Let us take the ansatz

$$\phi_\lambda(s) = \sqrt{s}\Phi_\lambda(s)$$

We obtain

$$s^2\Phi_\lambda''(s) + s\Phi_\lambda'(s) + \left(-\frac{1}{4} + g + \lambda s^2\right)\Phi_\lambda(s) = 0$$

Thanks to the scaling $u = \sqrt{\lambda}s$; $\lambda \neq 0$, the PDE becomes a Bessel differential equation[4] [1]

$$u^2\Phi_\lambda''(u) + u\Phi_\lambda'(u) + \left(-\frac{1}{4} + g + u^2\right)\Phi_\lambda(u) = 0$$

[4] Bessel $I_\alpha(x)$: $x^2 I_\alpha''(x) + xI_\alpha'(x) + (x^2 - \alpha^2)I_\alpha(x) = 0$.

The eigenvectors are

$$\phi_\lambda(s) = \sqrt{s}\left(I_\nu(\sqrt{\lambda}s) + aI_{-\nu}(\sqrt{\lambda}s)\right) \ , \ \lambda \neq 0$$

with

$$\nu \equiv \frac{1}{2|(1-\beta)|}$$

For the zero eigenvalue state, we take the ansatz

$$\phi(s) = s^\alpha$$

This implies that $g = -\alpha(\alpha - 1)$. The solution to this equation is $\alpha_\pm = \frac{1}{2} \pm \frac{\sqrt{1-4g}}{2}$. This gives two independent solutions and by completeness all the solutions. These solutions do not lead to normalizable states since they diverge at the origin or at infinity. As a result, there is no normalizable $\lambda = 0$ solution and the (continuous) spectrum is $(0, \infty)$.

Deficiency indices

In order to determine the deficiency indices, we search the square integrable solutions of

$$D^\dagger \phi_\pm(s) = \pm\imath\phi_\pm(s)$$

The solutions are given by

$$\begin{aligned}\phi_+ &= \sqrt{s}\left(I_\nu(se^{i\frac{\pi}{4}}) + aI_{-\nu}(se^{i\frac{\pi}{4}})\right)\\ \phi_- &= \sqrt{s}\left(I_\nu(se^{3i\frac{\pi}{4}}) + AI_{-\nu}(se^{3i\frac{\pi}{4}})\right)\end{aligned}$$

with a, A two constants. Note that $I_\nu(z)$ and $I_{-\nu}(z)$ ($z \in \mathbb{C}$) are linearly independent except when ν is an integer.
When ν is fixed and $|z| \to \infty$ ($|\arg z| < \pi$)

$$I_\nu(z) = \sqrt{\frac{2}{\pi z}}\left(\cos(z - \frac{\nu\pi}{2} - \frac{\pi}{4}) + e^{|\Im z|}O(|z|^{-1})\right)$$

We deduce that ϕ_+ and ϕ_- are renormalizable at $s = \infty$ if and only if $a = A = -e^{i\nu\pi}$. By using that $I_\nu(ze^{-\pi i}) = e^{-\pi i}I_\nu(z)$, we get

$$\begin{aligned}\phi_+ &= \sqrt{s}\left(I_\nu(se^{i\frac{\pi}{4}}) - I_{-\nu}(se^{-i\frac{3\pi}{4}})\right)\\ \phi_- &= \sqrt{s}\left(I_\nu(se^{i\frac{3\pi}{4}}) - I_{-\nu}(se^{-i\frac{\pi}{4}})\right)\end{aligned}$$

When ν is fixed and $|z| \to 0$

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu\left(\frac{1}{\Gamma(\nu+1)} + o(z^2)\right)$$

We deduce that ϕ_+ and ϕ_- are not square integrable at $s=0$ if and only if $\nu \geq 1$ equivalent to $\beta \in [\frac{1}{2}, 1)$. In this case D has deficiency indices $(0,0)$ and is (essentially) self-adjoint with the domain

$$\mathrm{Dom}(D) = \left(\phi \in L^2(\mathbb{R}_+) \ : \ \phi(0) = \phi'(0) = 0\right) \tag{9.22}$$

Both $\phi_\pm$ are square integrable when $\nu < 1$ equivalent to $\beta \in (0, \frac{1}{2})$. In these ranges, D has deficiency indices $(1,1)$. D is not self-adjoint on the domain Dom but admits self-adjoint extensions parameterized by $\theta \in U(1)$. The domain is

$$\mathrm{Dom}(D_\theta) = \left(\phi \in L^2(\mathbb{R}_+) \ : \ \phi(0) = -\tan(\frac{\theta}{2})\phi'(0)\right) \tag{9.23}$$

We have reproduced properties 2 and 5 from lemma 5.1.

Conditional probability

By assuming the absorbing condition $\Phi_\lambda(0) = 0$, we obtain

$$\Phi_\lambda(s) = \sqrt{s}\frac{\mathrm{I}_\nu(\sqrt{\lambda}s)}{\sqrt{2}\lambda^{\frac{1}{4}}}$$

The scale factor $\frac{1}{\sqrt{2}\lambda^{\frac{1}{4}}}$ has been included in order to satisfy the completude relation (9.15) as we have the identity

$$\int_0^\infty \mathrm{I}_\nu(\lambda s) I_\nu(\lambda s_0) = \frac{1}{s_0}\delta(s - s_0)$$

From (9.16), the conditional probability is

$$\begin{aligned} p'(t, s|s_0) &= \sqrt{s s_0}\int_0^\infty e^{-\lambda^2 t}\mathrm{I}_\nu(\lambda s)\mathrm{I}_\nu(\lambda s_0) d\lambda \\ &= \frac{\sqrt{s s_0}}{\sqrt{2}t} e^{-\frac{s^2 + s_0^2}{2t}} \mathrm{I}_\nu\left(\frac{s s_0}{t}\right) \end{aligned}$$

By using the explicit map (9.19) between $p(t, f|f_0)$ and its gauge transform $p'(t, s|s_0)$ with $s = \frac{f^{1-\beta}}{1-\beta}$, we obtain (see remark 4.3)

$$p(t, f|f_0) = \frac{f^{\frac{1}{2}-2\beta}}{(1-\beta)t}\sqrt{f_0}e^{-\frac{f^{2(1-\beta)} + f_0^{2(1-\beta)}}{2(1-\beta)^2 t}} \mathrm{I}_{\frac{1}{2(1-\beta)}}\left(\frac{(f f_0)^{1-\beta}}{(1-\beta)^2 t}\right)$$

□

Boundary conditions If the operator $D = (-\partial_s^2 - Q)$ is self-adjoint or admits self-adjoint extensions, the spectral decomposition can be used and $P(\tau, s)$ is decomposed as (9.16). For a one-dimensional diffusion process, one doesn't

TABLE 9.1: Feller boundary classification for one-dimensional Itô processes.

Boundary type	$S(e,d)$	$M(e,d)$	$\Sigma(e)$	$N(e)$
Regular	$< \infty$	$< \infty$	$< \infty$	$< \infty$
Exit	$< \infty$	$= \infty$	$< \infty$	$< \infty$
Entrance	$= \infty$	$< \infty$	$< \infty$	$< \infty$
Natural	$< \infty$	$= \infty$	$= \infty$	$= \infty$
	$= \infty$	$< \infty$	$= \infty$	$= \infty$
	$= \infty$	$= \infty$	$= \infty$	$= \infty$

need to use the deficiency indices technique as the complete classification of boundary conditions is given by the *Feller classification* [27]. More precisely, for a 1D time-homogeneous diffusion process, the boundary falls into one of the four following types: *regular, entrance, exit or natural.* For entrance, exit or natural, no boundary conditions are needed. As for a regular boundary, the conditional probability is not unique but depends on the boundary conditions. It corresponds to the case when the deficiency indices are (n, n). The boundary classification depends on the behavior of the following functions

$$S(c,d) = \int_c^d s(f)df$$

$$M(c,d) = \int_c^d m(f)df$$

$$\Sigma(e) = \lim_{c \to e} \int_c^d S(c,f)m(f)df$$

$$N(e) = \lim_{c \to e} \int_c^d S(x,d)m(x)dx$$

with $s(f) = e^{-2\int^f \frac{\mu(x)dx}{\sigma(x)^2}}$ and $m(f) = \frac{1}{\sigma^2(f)s(f)}$ (see Table 1 below).

Example 9.10 European call option with deterministic interest rate
We have that the forward f satisfies a driftless process (i.e., $\mu(f) = 0$)

$$df_t = \sigma(f_t)dW_t$$

The value at t of a European call option (with strike K and expiry date T) is given by ($\tau = T - t$)

$$\mathcal{C}(K, \tau | f_t) = P_{tT} \int_K^\infty \max(f - K, 0) p(\tau, f | f_t) df$$

Doing an integration by parts, or equivalently, applying the Itô-Tanaka formula on the payoff $\max(f_t - K, 0)$ (see equation (5.29)), the option $\mathcal{C}$ can be rewritten as

$$\frac{\mathcal{C}(K,\tau|f_t)}{P_{tT}} = \max(f_t - K, 0) + \frac{\sigma(K)^2}{2}\int_0^\tau dt' p(t', K|f_t)$$

Using the relation between the conditional probability $p(t', f|f_t)$ and its gauge-transform $P(t', s(f)|s_t)$ (9.19), we obtain (see remark 4.3)

$$\frac{\mathcal{C}(K,\tau|f_t)}{P_{tT}} = \max(f_t - K, 0) + \frac{\sqrt{\sigma(K)\sigma(f_t)}}{\sqrt{2}}\int_0^\tau P(t', s(K)|s_t)dt' \quad (9.24)$$

Plugging the expression for $P(t', s|s_t)$ (9.16) into (9.24) and doing the integration over the time t, we obtain

$$\begin{aligned}\frac{\mathcal{C}(K,\tau|f_t)}{P_{tT}} = & \frac{\sqrt{\sigma(K)\sigma(f_t)}}{\sqrt{2}}\int_{\mathrm{spec}(D)} \phi_\lambda(s(K))\phi_\lambda(s(f_t))\frac{(1-e^{-\lambda\tau})}{\lambda}d\lambda \\ & + \max(f_t - K, 0) \qquad (9.25)\end{aligned}$$

□

Example 9.11 Quadratic model, $\sigma(f) = \alpha f^2 + \beta f + \gamma$

$$\frac{\mathcal{C}(\tau,K|f_0)}{P_{0T}} = \max(f_0 - K, 0) + \frac{\sqrt{\sigma(K)\sigma(f_0)}}{\sqrt{2}}\int_0^\tau \frac{1}{\sqrt{4\pi t'}}e^{-\frac{s^2}{2t'}+Qt'}dt'$$

with $s = |\int_{f_0}^K \frac{dx}{\sigma(x)}|$ and $Q = \frac{\alpha\gamma}{2} - \frac{\beta^2}{8}$. By doing the integration over the time, we obtain

$$\frac{\mathcal{C}(\tau,K|f_0)}{P_{0T}} = \max(f_0 - K, 0) + \frac{\sqrt{\sigma(K)\sigma(f_0)}}{2\sqrt{-2Q}}$$

$$\left(e^{s\sqrt{-Q}}N\left(-\frac{s}{\sqrt{2\tau}} - \sqrt{-2Q\tau}\right) - e^{-s\sqrt{-Q}}N\left(-\frac{s}{\sqrt{2\tau}} + \sqrt{-2Q\tau}\right)\right)$$

□

A specific local volatility model admits an exact solution for a European call option if we can find the eigenvalues and eigenvectors for the corresponding Euclidean Schrödinger equation with a scalar potential. As examples of solvable potentials, one can cite the harmonic oscillator, Coulomb, Morse, Pöschl-Teller I&II, Eckart and Manning-Rosen potentials. The classification of solvable scalar potentials was initiated by Natanzon [129]. This work provides the most general potential for which the Schrödinger equation can be reduced to either a hypergeometric or confluent equation. We show that the Schrödinger equation can be doubled into a set of two independent Schrödinger equations with two different scalar potentials which transform into each other under a *supersymmetric* transformation. Moreover, if one scalar potential is solvable, the other one is. Applying this technique to the *Natanzon potential* which depends on 6 parameters, we obtain a new class of solvable potentials corresponding to a new class of solvable diffusion processes. For these models, the solution to the KBS equation is given by a sum of hypergeometric functions.

9.4.2 Solvable (super)potentials

In this part, we show that the Schrödinger equation can be formulated using supersymmetric techniques (see [10] for a nice review). In particular, the local volatility is identified as a superpotential. Using this formalism, we show how to generate a hierarchy of solvable diffusion models starting from a known solvable superpotential, for example, the hypergeometric or confluent hypergeometric Natanzon superpotential.

Superpotential and local volatility

We have seen that the Kolmogorov equation can be cast into (9.16) where $\phi_\lambda(s)$ are formal eigenvectors of the Schrödinger equation

$$-\left(\partial_s^2 + Q(s)\right)\phi_\lambda(s) = \lambda\phi_\lambda(s) \tag{9.26}$$

The eigenvectors $\phi_\lambda(s)$ must satisfy the identity (9.15). Let us write (9.26) as

$$\lambda\phi_\lambda^{(1)} = A_1^\dagger A_1 \phi_\lambda^{(1)} \tag{9.27}$$

where we have introduced the first-order operator A_1 and its adjoint $A_1^\dagger$ (called *supercharge operators*)

$$A_1 = \partial_s + W^{(1)}(s)\ ,\ A_1^\dagger = -\partial_s + W^{(1)}(s)$$

$W^{(1)}$ is called a *superpotential* which satisfies the Riccati equation

$$Q^{(1)}(s) = \partial_s W^{(1)}(s) - W^{(1)}(s)^2 \tag{9.28}$$

Surprisingly, this equation is trivially solved for our specific expression for $Q^{(1)}$ (9.21) (even with a drift $\mu(f)$!)

$$W^{(1)}(s) = \frac{1}{2}\frac{d\ln\sigma^{(1)}(s)}{ds} - \frac{\mu^{(1)}(s)}{\sqrt{2}\sigma^{(1)}(s)} \tag{9.29}$$

Example 9.12 Coulomb superpotential and CEV process
The CEV process corresponding to $\sigma(f) = f^\beta$ and $\mu(f) = 0$ has the *Coulomb superpotential*

$$W(s) = \frac{\beta}{2(1-\beta)s}$$

$\square$

Equation (9.27) implies that $\lambda \in \mathbb{R}_+$ as

$$\lambda ||\phi_\lambda^{(1)}||^2 = ||A_1 \phi_\lambda^{(1)}||^2$$

For a zero drift, the local volatility function is directly related to the superpotential by

$$\sigma(s) = e^{2\int_\cdot^s W(z)dz}$$

A similar correspondence between superpotentials and driftless diffusion processes has been found in [118, 115]. In addition, if we have a family of solvable superpotentials $W_{\mathrm{solv}}^{(1)}(s)$, we can always find an analytic solution to the Kolmogorov equation for any diffusion term $\sigma^{(1)}(s)$ by adjusting the drift with the relation (9.29)

$$\mu^{(1)}(s) = \frac{\sigma^{(1)'}(s)}{\sqrt{2}} - \sqrt{2}\sigma^{(1)}(s) W_{\mathrm{solv}}^{(1)}(s)$$

Note that D admits a zero eigenvalue if and only if the Kolmogorov equation admits a stationary distribution. By observing that $A_1^\dagger A_1 \phi_0^{(1)} = 0$ is equivalent to $A_1 \phi_0^{(1)} = 0$ as

$$(\phi_0^{(1)}, A_1^\dagger A_1 \phi_0^{(1)}) = \parallel A_1 \phi_0^{(1)} \parallel^2$$

we obtain the stationary distribution

$$\phi_0^{(1)}(s) = C e^{-\int^s W^{(1)}(z)dz} \tag{9.30}$$

with C a normalization constant. Therefore, the stationary distribution exists if the superpotential is normalisable (i.e., $\phi_0^{(1)}(s) \in L^2$).
Next, we define the *Scholes-Black equation* by intervening the operator A_1 and $A_1^\dagger$

$$\begin{aligned} \lambda \phi_\lambda^{(2)}(s) &= A_1 A_1^\dagger \phi_\lambda^{(2)}(s) \\ &= -\left(\partial_s^2 + Q^{(2)}(s)\right) \phi_\lambda^{(2)}(s) \end{aligned} \tag{9.31}$$

This corresponds to a new Schrödinger equation with the partner potential

$$Q^{(2)}(s) = -\partial_s W^{(1)} - (W^{(1)})^2 \tag{9.32}$$

By plugging our expression for the superpotential (9.29) in (9.32), we have

$$Q^{(2)} = -\frac{1}{2}(\ln \sigma^{(1)})''(s) - \frac{1}{4}(\ln \sigma^{(1)})'(s)^2 + \frac{\mu^{(1)'}(s)}{\sqrt{2}\sigma^{(1)}(s)} - \frac{\mu^{(1)}(s)^2}{2\sigma^{(1)}(s)^2} \tag{9.33}$$

In the same way as before, H_2 admits a zero eigenvector (i.e., stationary distribution) if

$$\phi_0^{(2)}(s) = Ce^{\int^s W^{(1)}(z)dz}$$

is normalisable.

REMARK 9.2 In physics, the supersymmetry (SUSY) is said to be broken if at least one of the eigenvectors $\phi_0^{(1,2)}(s)$ exists. ☐

Now we want to show that provided that we can solve the equation (9.27), then we have automatically a solution to (9.31) and vice versa. The SUSY-partner Hamiltonians $H_1 = A_1^\dagger A_1$ and $H_2 = A_1 A_1^\dagger$ obey the relations $A_1^\dagger H_2 = H_1 A_1^\dagger$ and $H_2 A_1 = A_1 H_1$. As a consequence H_1 and H_2 are isospectral. It means that the strictly positive eigenvalues all coincide and the corresponding eigenvectors are related by the supercharge operators A_1 and $A_1^\dagger$:

▷ If H_1 admits a zero eigenvalue (i.e., broken supersymmetry), we have the relation

$$\begin{aligned} &\phi_0^{(1)}(s) = Ce^{-\int^s W^{(1)}(z)dz} \ , \ \lambda^{(1)} = 0 \\ &\phi_\lambda^{(2)}(s) = (\lambda)^{-\frac{1}{2}} A_1 \phi_\lambda^{(1)}(s) \ , \ \lambda^{(2)} = \lambda^{(1)} = \lambda \neq 0 \\ &\phi_\lambda^{(1)}(s) = (\lambda)^{-\frac{1}{2}} A_1^\dagger \phi_\lambda^{(2)}(s) \end{aligned} \tag{9.34}$$

▷ If H_1 (and H_2) doesn't admit a zero eigenvalue (i.e., unbroken supersymmetry)

$$\begin{aligned} &\lambda^{(2)} = \lambda^{(1)} = \lambda \neq 0 \\ &\phi_\lambda^{(2)}(s) = (\lambda)^{-\frac{1}{2}} A_1 \phi_\lambda^{(1)}(s) \\ &\phi_\lambda^{(1)}(s) = (\lambda)^{-\frac{1}{2}} A_1^\dagger \phi_\lambda^{(2)}(s) \end{aligned}$$

The eigenvectors have been scaled with the factor $(\lambda)^{-\frac{1}{2}}$ to get

$$\| \phi_\lambda^{(1)} \| = \| \phi_\lambda^{(2)} \|$$

In the unbroken SUSY case, there are no zero modes and consequently the spectrum of H_1 and H_2 is the same. One can obtain the solution to the Scholes-Black (resp. Black-Scholes) equation if the eigenvalues/eigenvectors of the Black-Scholes (resp. Scholes-Black) are known. We clarify this correspondence by studying a specific example: the CEV process $df = f^\beta dW$. In particular, we show that for $\beta = \frac{2}{3}$, the partner superpotential vanishes. It is therefore simpler to solve the Scholes-Black equation as Scholes-Black (rather than Black-Scholes) reduces to the heat kernel on $\mathbb{R}_+$. Applying a supersymmetric transformation on the Scholes-Black equation, we can then derive the solution to the Black-Scholes equation. For example, in the case of unbroken SUSY, we have

$$p(t,s|s_0) = \int_{\mathrm{spec}(H^{(2)})} e^{-t\lambda}(\lambda)^{-1}\left(A_1^\dagger\phi_\lambda^{(2)}\right)(s)\left(A_1^\dagger\phi_\lambda^{(2)}\right)(s_0)d\lambda$$

where the generalized eigenvectors $\phi_\lambda^{(2)}(s)$ satisfy

$$-(\partial_s^2 + Q^{(2)}(s))\phi_\lambda^{(2)}(s) = \lambda\phi_\lambda^{(2)}(s)$$

Example 9.13 CEV with $\beta = 2/3$ and Bachelier process
We saw previously that the superpotential associated with the CEV process is given by

$$W^{(1)}(s) = \frac{\beta}{2s(1-\beta)}$$

with the flat coordinate $s = \frac{\sqrt{2}f^{1-\beta}}{(1-\beta)} \in [0,\infty)$. The potential (9.28) is

$$Q^{(1)}(s) = \frac{\beta(\beta-2)}{4(1-\beta)^2 s^2}$$

from which we deduce the partner potential (9.32)

$$Q^{(2)}(s) = \frac{\beta(2-3\beta)}{4(1-\beta)^2 s^2}$$

This partner potential corresponds to the potential of a CEV process $df = f^B dW$ with B depending on β by

$$\frac{B(B-2)}{(1-B)^2} = \frac{\beta(2-3\beta)}{(1-\beta)^2}$$

Surprisingly, we observe that for $\beta = \frac{2}{3}$, $Q^{(2)}(s)$ cancels and the corresponding partner local volatility model is the Bachelier model $df = dW$ for which the heat kernel is given by the normal distribution. The eigenvectors of the supersymmetric Hamiltonian partner $H_2 = -\partial_s^2$ to H_1 are given by (with the absorbing boundary condition $\phi_\lambda(0) = 0$)

$$\phi_\lambda^{(2)}(s) = \frac{\sin(\sqrt{\lambda}s)}{\sqrt{4\pi}\lambda^{\frac{1}{4}}}$$

with a continuous spectrum $\mathbb{R}_+$. Applying the supersymmetric transformation (9.34), we obtain the eigenvectors for the Hamiltonian $H_1 = -\partial_s^2 + \frac{2}{s^2}$ corresponding to the CEV process with $\beta = \frac{2}{3}$

$$\begin{aligned}\phi_\lambda^{(1)}(s) &= \lambda^{-\frac{1}{2}}\left(-\partial_s + \frac{1}{s}\right)\phi_\lambda^{(2)}(s)\\ &= \frac{1}{\sqrt{4\pi}\lambda^{\frac{3}{4}}}\left(-\sqrt{\lambda}\cos(\sqrt{\lambda}s) + \frac{\sin(\sqrt{\lambda}s)}{s}\right)\end{aligned}$$

By plugging this expression in (9.25), we obtain the fair value for a European call option

$$\frac{\mathcal{C}(\tau, f_t, K)}{P_{tT}} = \max(f_t - K, 0) + \frac{(f_t K)^{\frac{1}{3}}}{\sqrt{2}} \int_0^\infty d\lambda \frac{(1 - e^{-\lambda\tau})}{\lambda} \phi_\lambda^{(1)}(s)\phi_\lambda^{(1)}(s_0)$$

This expression can be integrated out and written in terms of the cumulative distribution [30]. The fact that the CEV model with $\beta = \frac{2}{3}$ depends on the cumulative normal distribution and is therefore related to the heat kernel on $\mathbb{R}_+$ has been observed empirically by [30]. Here we have seen that it corresponds to the fact that the supersymmetric partner potential vanishes for this particular value of β.

$\Box$

9.4.3 Hierarchy of solvable diffusion processes

In the previous section we saw that the operators A_1 and $A_1^\dagger$ can be used to factorize the Hamiltonian H_1. These operators depend on the superpotential $W^{(1)}$ which is determined once we know the first eigenvector $\phi_0^{(1)}(s)$ of H_1 (see eq. (9.30)):

$$W^{(1)}(s) = -\frac{d \ln \phi_0^{(1)}(s)}{ds}$$

We have assumed that H_1 admits a zero eigenvalue. By shifting the eigenvalue $\lambda^{(1)}$ it is always possible to achieve this condition. The partner Schrödinger equation (9.31) can then be recast into a Schrödinger equation with a zero eigenvalue

$$H_{(2)} = A_1 A_1^\dagger = A_2^\dagger A_2 + E_1^{(1)}$$

where $A_2 \equiv \partial_s + W_2(s)$ and $A_2^\dagger \equiv -\partial_s + W_2(s)$,

$$W^{(2)}(s) \equiv \frac{1}{2}\frac{d \ln \sigma^{(2)}(s)}{ds} - \frac{\mu^{(2)}(s)}{\sqrt{2}\sigma^{(2)}(s)} \tag{9.35}$$

We have introduced the notation $E_n^{(m)}$ where n denotes the first eigenvalue and (m) refers to the m$^{\text{th}}$ Hamiltonian H_m. By construction, this new Hamiltonian $H_2 = A_2^\dagger A_2 + E_1^{(1)}$ is solvable as $A_1 A_1^\dagger$ is so and the associated diffusion process with volatility $\sigma^{(2)}$ and drift $\mu^{(2)}$ satisfying (9.35) is solvable. The superpotential $W^{(2)}(s)$ is determined by the first eigenvector of H_2, $\phi_0^{(2)}(s)$ associated to the eigenvalue $E_1^{(1)}$,

$$W^{(2)}(s) = -\frac{d \ln(\phi_0^{(2)})(s)}{ds}$$

At this stage, we can apply a supersymmetric transformation on H_2. The new Hamiltonian H_3 can be refactorised exactly in the same way as was done for H_2. Finally, it turns out that if H_1 admits p (normalisable) eigenvectors, then one can generate a family of solvable Hamiltonians H_m (with a zero-eigenvalue by construction)

$$H_m = A_m^\dagger A_m + E_{m-1}^{(1)} = -\partial_s^2 + Q_m(s)$$

where $A_m = \partial_s + W_m(s)$. This corresponds to the solvable diffusion process with a drift and a volatility such that

$$W_m(s) = -\frac{d \ln \phi_0^{(m)}}{ds} = \frac{1}{2}\frac{d \ln \sigma^{(m)}(s)}{ds} - \frac{\mu^{(m)}(s)}{\sqrt{2}\sigma^{(m)}(s)}$$

The eigenvalues/eigenvectors of H_m are related to those of H_1 by

$$\begin{aligned}
E_n^{(m)} &= E_{n+1}^{(m-1)} = \cdots = E_{n+m-1}^{(1)} \\
\phi_n^{(m)} &= (E_{n+m-1}^{(1)} - E_{m-2}^{(1)})^{-\frac{1}{2}} \cdots (E_{n+m-1}^{(1)} - E_0^{(1)})^{-\frac{1}{2}} A_{m-1} \cdots A_1 \phi_{n+m-1}^{(1)}
\end{aligned}$$

In particular, the superpotential of H_m is determined by the $(m-1)^{th}$ eigenvector of H_1, $\phi_{m-1}^{(1)}(s)$,

$$\begin{aligned}
W_m(s) &= -\frac{d \ln(A_{m-1} \cdots A_1 \phi_{m-1}^{(1)}(s))}{ds} \\
&= \frac{1}{2}\frac{d \ln \sigma^{(m)}(s)}{ds} - \frac{\mu^{(m)}(s)}{\sqrt{2}\sigma^{(m)}(s)}
\end{aligned}$$

Consequently, if we know all the m discrete eigenvalues and eigenvectors of H_1, we immediately know all the energy eigenvalues and eigenfunctions of the hierarchy of $m-1$ Hamiltonians. In the following we apply this procedure with a known solvable superpotential, *the Natanzon* superpotential, as a starting point.

9.4.4 Natanzon (super)potentials

The Natanzon potential [129] is the most general potential which allows us to reduce the Schrödinger equation (9.20) to a Gauss or confluent hypergeometric equation (GHE or CHE).

Gauss hypergeometric potential

The potential is given by

$$Q(s) = \frac{S(z)-1}{R(z)} - \left(\frac{r_1 - 2(r_2 + r_1)z}{z(1-z)} - \frac{5}{4}\frac{(r_1^2 - 4r_1 r_0)}{R(z)} + r_2 \right) \frac{z^2(1-z^2)}{R(z)^2}$$

with $R(z) = r_2 z^2 + r_1 z + r_0 \ > 0$ and $S(z) = s_2 z^2 + s_1 z + s_0$ (two second order polynomials). The z coordinate, lying in the interval $[0, 1]$, is implicitly defined in terms of s by the differential equation

$$\frac{dz(s)}{ds} = \frac{2z(1-z)}{\sqrt{R(z)}}$$

Example 9.14
The hypergeometric Natanzon potential includes as special cases the Pöschl-Teller potential II

$$Q(s) = A + Bsech\left(\frac{2s}{\sqrt{r_1}}\right)^2 + Ccsch\left(\frac{2s}{\sqrt{r_1}}\right)^2$$

for $r_0 = r_2 = 0$ and the Rosen-Morse potential

$$Q(s) = A + B\tanh(\frac{2s}{\sqrt{r_0}}) + Csech(\frac{2s}{\sqrt{r_0}})^2$$

$r_1 = r_2 = 0$. $\square$

By construction, the solution to the Schrödinger equation with a GHE potential is given in terms of the Gauss hypergeometric function $F(\alpha, \beta, \gamma, z)$

$$z(s)' \equiv \phi_\lambda(s) = (z')^{-\frac{1}{2}} z^{\frac{\gamma}{2}} (1-z)^{\frac{-\gamma+\alpha+\beta+1}{2}} F(\alpha, \beta, \gamma, z)$$

where $F(\alpha, \beta, \gamma, z)$ satisfies the differential equation [1]

$$z(1-z)\frac{d^2F}{dz^2} + (\gamma - (\alpha+\beta+1)z)\frac{dF}{dz} - \alpha\beta z = 0 \tag{9.36}$$

The most general solution to this equation (9.36) is generated by a two-dimensional vector space

$$F(\alpha, \beta, \gamma, z) = c_{12}\, F_1(\alpha, \beta, \gamma, z) + c_2 z^{1-\gamma}{}_2\, F_1(\alpha - \gamma + 1, \beta - \gamma + 1, 2 - \gamma, z)$$

with ${}_2\, F_1(\alpha, \beta, \gamma, z)$ satisfying ${}_2\, F_1(\alpha, \beta, \gamma, 0) = 1$ and c_1 and c_2 two arbitrary coefficients. The parameters α, β, γ depend explicitly on the eigenvalue λ by

$$\begin{aligned} 1 - (\alpha - \beta)^2 &= r_2\lambda + s_2 \\ 2\gamma(\alpha + \beta - 1) - 4\alpha\beta &= r_1\lambda + s_1 \\ \gamma(2 - \gamma) &= r_0\lambda + s_0 \end{aligned}$$

From these equations, one can show that λ satisfies a fourth-order polynomial and find λ as an explicit function of α, β, γ [97]. By imposing the condition that the eigenvectors are normalisable (i.e., belong to $L^2([0, 1])$) we obtain

TABLE 9.2: Condition at $z = 0$.

$r_0 \neq 0$	$c_1 = 0, \gamma < 1$ or $c_2 = 0, \gamma > 1$
$r_0 = 0,\ r_1 \neq 0$	$c_1 = 0, \gamma < 2$ or $c_2 = 0, \gamma > 0$
$r_0 = 0,\ r_1 = 0$	$c_1 = 0, \gamma < 3$ or $c_2 = 0, \gamma > -1$

TABLE 9.3: Condition at $z = 1$.

$c_1 = 0$	$\alpha - 1 \in \mathbb{N}^*$, $\alpha + \beta - \gamma < 0$ or $-1 - \alpha + \gamma \in \mathbb{N}^*$, $\alpha + \beta - \gamma > 0$
$c_2 = 0$	$-\alpha \in \mathbb{N}^*$, $\alpha + \beta - \gamma > 0$ or $\alpha - \gamma \in \mathbb{N}^*$, $\alpha + \beta - \gamma < 0$

the discrete spectrum λ_n and can determine the coefficients c_1 and c_2. We impose conditions on α, β, γ, c_1 and c_2 such that

$$\int_{-\infty}^{\infty} ds \phi_\lambda(s)^2 = \int_0^1 dz \frac{R(z)}{4} z^{\gamma-2}(1-z)^{-\gamma+\alpha+\beta-1} F(\alpha,\beta,\gamma,z)^2 < \infty$$

Looking at the asymptotic behavior of ${}_2F_1(\alpha,\beta,\gamma,z)$ near $z = 0$ and $z = 1$ [1][5], we obtain the following conditions (see table 9.2, 9.3)
We have the discrete eigenvalues ($\alpha_n = -n\ ; n \in \mathbb{N}^*$)

$$\begin{aligned}
2n+1 &= -\sqrt{1 - r_0\lambda_n - s_0} + \sqrt{1 - r_2\lambda_n - s_2} \\
&\quad -\sqrt{1 - (r_0 + r_1 + r_2)\lambda_n - (s_0 + s_1 + s_2)} \\
\gamma_n &= 1 + \sqrt{1 - r_0\lambda_n - s_0} \\
\alpha_n - \beta_n &= -\sqrt{1 - r_2\lambda_n - s_2} \\
\alpha_n + \beta_n &= \gamma_n + \sqrt{1 - (r_0 + r_1 + r_2)\lambda_n - (s_0 + s_1 + s_2)}
\end{aligned}$$

and the (normalised) eigenvectors

$$\phi_n(s) = B_n(z')^{-\frac{1}{2}} z^{\frac{\gamma_n}{2}} (1-z)^{\frac{-\gamma_n - n + \beta_n + 1}{2}} P_n^{(\gamma_n - 1, -n + \beta_n - \gamma_n + 1)}(1 - 2z)$$

with

$$B_n = \left(\left(\frac{R(1)}{\alpha + \beta - \gamma} + \frac{r_0}{\gamma - 1} - \frac{r_2}{\beta - \alpha} \right) \frac{\Gamma(\gamma + n + 1)\Gamma(\alpha + \beta - \gamma)}{n!\Gamma(\beta - \alpha - n)} \right)^{-\frac{1}{2}}$$

and $P_n^{(\gamma-1,\alpha+\beta-\gamma)}(2z-1)$ the Jacobi polynomials [1].

Confluent hypergeometric potential

A similar construction can be achieved for the class of *scaled confluent hypergeometric functions*. The potential is given by

$$Q(s) = \frac{S(z) - 1}{R(z)} - \left(\frac{r_1}{z} - \frac{5}{4} \frac{(r_1^2 - 4r_2r_0)}{R(z)} - r_2 \right) \frac{z^2}{R(z)^2}$$

[5] ${}_2F_1(\alpha,\beta,\gamma,z) \sim_{z\to 1} \Gamma(\gamma) \left((-1+z)^{-\alpha-\beta+\gamma} + \frac{\Gamma(-\alpha-\beta+\gamma)}{\Gamma(-\alpha+\gamma)\Gamma(-\beta+\gamma)} \right)$.

with $R(z) = r_2 z^2 + r_1 z + r_0 > 0$ and $S(z) = s_2 z^2 + s_1 z + s_0$. The z coordinate, lying in the interval $[0, \infty)$, is defined implicitly in terms of s by the differential equation

$$\frac{dz(s)}{ds} = \frac{2z}{\sqrt{R(z)}}$$

Example 9.15
The confluent Natanzon potential reduces to the Morse potential

$$Q(s) = \frac{-1 + s_0 + s_1 e^{\frac{2s}{\sqrt{r_0}}} + s_2 e^{\frac{4s}{\sqrt{r_0}}}}{r_0}$$

for $r_1 = r_2 = 0$, to the 3D oscillator

$$Q(s) = \frac{-\frac{3}{4} + s_0}{s^2} + \frac{s_1}{r_1} + \frac{s_2 s^2}{r_1^2}$$

for $r_0 = r_2 = 0$ and to the Coulomb potential

$$Q(r) = \frac{-r_2 s_0 - 2s\sqrt{r_2} s_1 - 4s^2 s_2}{4 r_2 s^2}$$

for $r_0 = r_1 = 0$. □

By construction, the eigenvector solutions to the CHE potential are given in terms of the confluent hypergeometric function $F(\alpha, \beta, \gamma, z)$

$$\phi_\lambda(s) = z(s)^{\frac{\gamma}{2}} e^{-\frac{\omega z(s)}{2}} (z'(s))^{-\frac{1}{2}} F(\alpha, \beta, \gamma, \omega z(s))$$

By definition, $\phi(z) \equiv F(\alpha, \beta, \gamma, \omega z(s))$ satisfies the differential equation [1]

$$z\phi''(z) + (\gamma - \omega z)\phi'(z) - \omega\alpha\phi(z) = 0 \tag{9.37}$$

The parameters ω, γ, α depend explicitly on the eigenvalue λ by

$$\begin{aligned} &\omega^2 = -r_2\lambda - s_2 \\ &2\omega(\gamma - 2\alpha) = r_1\lambda + s_1 \\ &\gamma(2 - \gamma) = r_0\lambda + s_0 \end{aligned}$$

The most general solution to (9.37) is generated by a two-dimensional vector space

$$F(\alpha, \gamma, \omega z) = c_1 M(\alpha, \gamma, \omega z) + c_2 (\omega z)^{1-\gamma} M(1 + \alpha - \gamma, 2 - \gamma, \omega z)$$

with $M(\alpha, \gamma, \omega z)$ the M-Whittaker function [1] and c_1 and c_2 two arbitrary coefficients. By imposing that the eigenvectors are normalisable (i.e., belong

TABLE 9.4: Condition at $z = \infty$.

$\alpha > 2$	no condition
$\alpha \leq 2$	$c_1 = 0, -1-\alpha+\gamma \in \mathbb{N}^*$ or $c_2 = 0, -\alpha \in \mathbb{N}^*$

TABLE 9.5: Example of solvable superpotentials and local volatility models.

Superpotential	$W(s)$	$\frac{\sigma(s)}{\sigma_0}$
Shifted oscillator	$as+b$	e^{as^2+2bs}
Coulomb	$a+\frac{b}{s}$	$s^b e^{2as}$
Morse	$a+be^{-\alpha s}$	$e^{2\left(-\left(\frac{b}{\alpha e^{\alpha s}}\right)+as\right)}$
Eckart	$a\coth(\alpha s)+b$	$e^{2\left(bs+\frac{a\log(\sinh(\alpha s))}{\alpha}\right)}$
Rosen-Morse	$a\tanh(\alpha s)+b$	$e^{2bs+\frac{2a}{\alpha}\ln(\cosh(\alpha))}$
3D oscillator	$as+\frac{b}{s}$	$e^{as^2+2b\ln(s)}$
P-T I $\alpha > 2b$	$a\tan(\alpha s)+bcotg(\alpha s)$	$e^{2\left(-\left(\frac{a\log(\cos(\alpha s))}{\alpha}\right)+\frac{b\log(\sin(\alpha s))}{\alpha}\right)}$
P-T II $\alpha > 2b$	$a\tanh(\alpha s)+bcoth(\alpha s)$	$e^{2\left(\frac{a\log(\cosh(\alpha s))}{\alpha}+\frac{b\log(\sinh(\alpha s))}{\alpha}\right)}$

to $L^2(\mathbb{R}_+)$), we obtain the following conditions (see Table 9.4) which give the discrete spectrum λ and the coefficients c_1 and c_2.[6] We have the discrete eigenvalues $\alpha_n = -n$; $n \in \mathbb{N}$

$$\gamma_n = 1+\sqrt{1-r_0\lambda_n - s_0}$$
$$\omega_n = \sqrt{-r_2\lambda_n - s_2}$$
$$2n+1 = \frac{r_1\lambda_n + s_1}{2\sqrt{-r_2\lambda_n - s_2}} - \sqrt{1-r_0\lambda_n - s_0}$$

and the (normalized) eigenvectors

$$\phi_n(s) = \frac{n!}{(\gamma_n)_n} z(s)^{\frac{\gamma_n}{2}} e^{-\frac{\omega_n z(s)}{2}} (z'(s))^{-\frac{1}{2}} L_n^{\gamma_n - 1}(\omega_n z(s))$$

with $L_n^{\gamma_n-1}(z)$ the (generalized) Laguerre polynomial [1]. In tables 9.5, 9.6, we have listed classical solvable superpotentials and the corresponding solvable local volatility models and solvable instantaneous short-rate models.

Natanzon hierarchy and new solvable processes

We know that the Natanzon superpotential is related to the zero-eigenvector

$$W_{\text{nat}} = -\partial_s \ln \phi_0(s)$$

[6] $M(\alpha,\beta,z) \sim_{z\to\infty} \frac{\Gamma(\beta)}{\Gamma(\alpha)} e^z z^{\alpha-\beta}(1+O(|z|^{-1}))$.

TABLE 9.6: Example of solvable one-factor short-rate models.

one-factor short-rate model	SDE	Superpot.
Vasicek-Hull-White	$dr = k(\theta - r)dt + \sigma dW$	Shifted Osc.
CIR	$dr = k(\theta - r)dt + \sigma\sqrt{r}dW$	3D Osc.
Dooleans	$dr = krdt + \sigma r dW$	Constant
EV-BK	$dr = r(\eta - \alpha\ln(r))dt + \sigma r dW$	Shifted Osc.

and the corresponding supercharge A is

$$\begin{aligned} A &= \partial_s + W_{\mathrm{nat}}(s) \\ &= \frac{2z(1-z)}{\sqrt{R}}\left(\partial_z - \frac{\gamma_0}{2z} - \frac{(1+\alpha_0+\beta_0-\gamma_0)}{2(z-1)}\right. \\ &\left. - \frac{\alpha_0\beta_0 F(1+\alpha_0, 1+\beta_0, 1+\gamma_0, z)}{\gamma_0 F(\alpha_0,\beta_0,\gamma_0,z)} + \frac{z'^{-\frac{3}{2}}z''(s)}{2}\right) \end{aligned}$$

With a zero drift, the Natanzon superpotential corresponds to the diffusion process (9.29)

$$\sigma^{(1)}(s) = \frac{\sigma_0^{(1)}}{\phi_0^{(1)}(s)^2}$$

with $\sigma_0^{(1)}$ a constant of integration. Applying the results of section 9.4.3, we obtain that the driftless diffusion processes

$$\sigma^{(m)}(s) = \frac{\sigma_0^{(m)}}{A_{m-1}\cdots A_1\phi_{m-1}^{(1)}(s)^2}$$

are solvable. Using the fact that the eigenvector $\phi_{m-1}^{(1)}(s)$ associated to a discrete eigenvalue is an hypergeometric function and that the derivative of a hypergeometric function is a new hypergeometric function,[7] the action of $A_{m-1}\cdots A_1$ on $\phi_{m_1}^{(1)}(s)$ results in a sum of $(m-1)$ hypergeometric functions, thus generalizing the solution found in [40].

9.5 Gauge-free stochastic volatility models

In this section, pursuing our application of the reduction method, we try to identify the class of time-homogeneous stochastic volatility models which leads

[7] ${}_2F_1'(\alpha,\beta,\gamma,z) = \frac{\alpha\beta}{\gamma}{}_2F_1(\alpha+1,\beta+1,\gamma+1,z)$ and $M'(\alpha,\beta,z) = \frac{\alpha}{\beta}M(\alpha+1,\beta+1,z)$.

to an exact solution of the KBS equation. We assume that the forward f and the volatility a are driven by two correlated Brownian motions in the forward measure $\mathbb{P}^T$

$$\begin{aligned}
df_t &= a_t C(f_t) dW_t \\
da_t &= b(a_t, f_t)dt + \sigma(a_t, f_t)dZ_t \; ; \; dW_t dZ_t = \rho dt
\end{aligned}$$

with the initial conditions $a_0 = a$ and $f_0 = f$.
Using the definition for the Abelian connection (4.58), we obtain[8]

$$\begin{aligned}
\mathcal{A}_f &= -\frac{1}{2(1-\rho^2)} \partial_f \ln\left(\frac{C}{\sigma}\right) - \frac{\rho}{(1-\rho^2)} \left(\frac{b}{aC\sigma} - \frac{1}{2C}\partial_a \frac{\sigma}{a}\right) \\
\mathcal{A}_a &= \frac{1}{(1-\rho^2)} \left(\frac{b}{\sigma^2} - \frac{1}{2}\partial_a \ln\left(\frac{\sigma}{a}\right)\right) + \frac{\rho}{2(1-\rho^2)} a\partial_f \left(\frac{C}{\sigma}\right)
\end{aligned}$$

Then, the field strength is

$$\begin{aligned}
\mathcal{F} &= (\partial_a \mathcal{A}_f - \partial_f \mathcal{A}_a)\, da \wedge df \\
&= \frac{1}{(1-\rho^2)} \left(\left(\partial_{af} \ln(\sigma) - \partial_f \frac{b}{\sigma^2} \right) \right. \\
&\qquad \left. -\rho \left(\frac{1}{C}\partial_a \frac{b}{a\sigma} - \frac{1}{2C}\partial_a^2 \frac{\sigma}{a} + \frac{a}{2}\partial_f^2 \frac{C}{\sigma} \right) \right) da \wedge df
\end{aligned}$$

We assume that the connection is flat,

$$\mathcal{F}_{af} = 0$$

meaning that the connection can be eliminated modulo a gauge transformation. In this case, the stochastic volatility model is called a *gauge-free model*. This condition is satisfied for every correlation $\rho \in (-1, 1)$ if and only if

$$\begin{aligned}
&\partial_{af} \ln \sigma = \partial_f \frac{b}{\sigma^2} \\
&\partial_a \frac{b}{a\sigma} - \frac{1}{2}\partial_a^2 \frac{\sigma}{a} + \frac{aC}{2}\partial_f^2 \frac{C}{\sigma} = 0
\end{aligned}$$

Moreover, if we assume that $\sigma^a(a)$ is only a function of a (this hypothesis is equivalent to assuming that the metric admits a Killing vector), the model is gauge-free if and only if

$$\frac{b}{\sigma} = \frac{a}{2}\partial_a \frac{\sigma}{a} + a\phi(f) - \frac{aC(f)}{2}\partial_f^2 C(f) \int \frac{a' da'}{\sigma(a')}$$

8

$$\begin{aligned}
\mathcal{A}^f &= -\frac{a^2 C\sigma}{4}\partial_f \frac{C}{\sigma} \\
\mathcal{A}^a &= \frac{1}{2}\left(b - \frac{a\sigma}{2}\partial_a \frac{\sigma}{a}\right)
\end{aligned}$$

TABLE 9.7: Example of Gauge free stochastic volatility models with $df = \delta(\mu + \nu f)\sqrt{v}dW'_1$.

name	$\sigma(a)$	SDE
Heston	$\sigma(a) = \eta$	$dv = \sqrt{\delta}(2v\gamma)dt + 2\eta\sqrt{\delta}\sqrt{v}dZ_t$
Geometric Brownian	$\sigma(a) = \eta a$	$dv = \sqrt{\delta}(2\eta\gamma v^{\frac{3}{2}} + \eta^2 v)dt + 2\sqrt{\delta}\eta v dZ_t$
3/2-model	$\sigma(a) = \eta a^2$	$dv = 2\sqrt{\delta}\eta(\eta + \gamma)v^2 dt + 2\sqrt{\delta}\eta v^{\frac{3}{2}} dZ_t$

with $\phi(f)$ satisfying

$$\partial_f \phi(f) = \frac{\partial_f (C \partial_f^2 C)}{2} \int \frac{a' da'}{\sigma(a')}$$

This last equation is equivalent to $C(f)\partial_f^2 C(f) = \beta$ with β a constant and $\phi = \gamma$ a constant function. For $\beta = 0$, the last equation above simplifies and we obtain

$$C(f) = \mu f + \nu$$
$$b(a) = a\sigma(a)\left(\gamma + \frac{1}{2}\partial_a \frac{\sigma(a)}{a}\right)$$

with μ, ν , γ three integration constants. The gauge-free condition has therefore imposed the functional form of the drift term:

$$dv = 2v\sigma(v)\left(\gamma + \partial_v \sigma(v)\right)dt + 2\sqrt{v}\sigma(v)dZ_t$$

with $v \equiv a^2$. When the volatility function is fixed respectively to a constant (Heston model), a linear function (geometric Brownian model) and a quadratic function (3/2-model) in the volatility, one obtains the correct (mean-reverting) drift[9] (see table 9.7). Note that the gauge-free Heston model has a zero-mean reverting volatility (this corrects a typo in [105]).

The gauge transformation eliminating the connection is then

$$\Lambda(f, a) = \frac{1}{1 - \rho^2}\left(-\frac{1}{2}\ln C(f) - \gamma\rho \int^f \frac{df'}{C(f')} + \left(\gamma + \frac{\rho}{2}\partial_f C\right) \int^a \frac{a' da'}{\sigma(a')}\right)$$

Finally, by plugging the expression for $C(f)$ and $b(a)$ into (4.59), we find that

[9]In order to obtain the correct number of parameters, one needs to apply a change of local time $dt = \delta dt'$, $dW_{1,2} = \sqrt{\delta}dW'_{1,2}$.

the function Q is[10]

$$Q = Aa^2 + B\sigma^2 \partial_a \left(\frac{a}{\sigma} \right)$$

The Black-Scholes equation for a European call option $\mathcal{C}(\tau = T - t, a, f)$ (with strike K and maturity T) satisfied by the gauge transformed function

$$\mathcal{C}'(\tau, a, f) = e^{\Lambda(f,a)} \mathcal{C}(\tau, a, f)$$

is

$$\partial_\tau \mathcal{C}'(\tau, a, f) = \Delta \mathcal{C}'(a, f, \tau) + Q(a) \mathcal{C}'(\tau, a, f)$$

with the initial condition $\mathcal{C}'(\tau = 0, a, f) = e^{\Lambda(f,a)} \max(f - K, 0)$. In the coordinates $q(f) = \int_\cdot^f \frac{df'}{C(f')}$ and a, the Laplace-Beltrami operator is given by

$$\Delta = \frac{a\sigma}{2} \left(\frac{a}{\sigma} \partial_q^2 + 2\rho \partial_{aq} + \partial_a \frac{\sigma}{a} \partial_a \right)$$

Applying a Fourier transformation according to q,

$$\begin{aligned}
\hat{\mathcal{C}}'(\tau, k, a) &= \mathcal{F}\mathcal{C}'(\tau, q, a) \\
&\equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikq} \mathcal{C}'(\tau, q, a) dq
\end{aligned}$$

we obtain

$$\partial_\tau \hat{\mathcal{C}}'(\tau, k, a) = \frac{a\sigma}{2} \left(-k^2 \frac{a}{\sigma} + 2ik\rho \partial_a + \partial_a \frac{\sigma}{a} \partial_a \right) \hat{\mathcal{C}}'(\tau, k, a) + Q(a) \hat{\mathcal{C}}'(\tau, k, a)$$

with the initial condition

$$\mathcal{C}'(\tau = 0, k, a) = \mathcal{F}[e^{\Lambda(f,a)} \max(f - K, 0)]$$

This is a Schrödinger-type equation. Using the spectral decomposition

$$\mathcal{C}'(\tau, k, a) = \sum_n \phi_{nk}(a) e^{-E_{nk}\tau}$$

[10] A and B are two constants given by

$$\begin{aligned}
A &= \frac{1}{2}\left(-\frac{1}{2}\rho^{ff}\partial_f C + \gamma\rho^{af}\right)^2 + \frac{1}{2}\rho^{ff} C^2 \partial_f^2 \ln(C) \\
&\quad + \rho\left(-\frac{1}{2}\rho^{ff}\partial_f C + \gamma\rho^{af}\right)\left(\gamma\rho^{ff} - \frac{\rho^{af}}{2}\partial_f C\right) + \frac{1}{2}\left(\rho^{ff}\gamma - \frac{\rho^{af}}{2}\partial_f C\right)^2 \\
B &= -\frac{1}{2}\left(\rho^{ff}\gamma - \rho^{af}\frac{\partial_f C}{2}\right)
\end{aligned}$$

the eigenvectors $\phi_{nk}(a)$ satisfy the equation

$$-E_{nk}\phi_{nk}(a) = \frac{a\sigma}{2}\left(-k^2\frac{a}{\sigma} + 2ik\rho\partial_a + \partial_a\frac{\sigma}{a}\partial_a\right)\phi_{nk}(a) + Q\phi_{nk}(a) \tag{9.38}$$

This equation (9.38) can be further simplified by applying a *Liouville transformation* consisting in a gauge transformation and a change of variable [128]

$$\psi_{nk}(s) = \left(\frac{\sigma}{a}\right)^{\frac{1}{2}} e^{ik\rho\int^a \frac{a'}{\sigma(a')}da'}\phi_{nk}(a)$$
$$\frac{ds}{da} = \frac{\sqrt{2}}{\sigma(a)}$$

$\psi_{nk}(s)$ satisfies a Schrödinger equation

$$\psi''_{nk}(s) + (E_{nk} - J(s))\psi_{nk}(s) = 0$$

with the scalar potential

$$J(s) = Q(a) - \frac{k^2a^2}{2} + \frac{1}{2}\{a,s\} + \frac{4\,a^4\,k^2\,\rho^2 - 3\,\sigma(a)^2 + a^2\,\sigma'(a)^2 + 2\,a\,\sigma(a)\,\left(\sigma'(a) - a\,\sigma''(a)\right)}{8\,a^2}$$

and where the curly bracket denotes the Schwarzian derivative of a with respect to s

$$\{a,s\} = \left(\frac{a''(s)}{a'(s)}\right)' - \frac{1}{2}\left(\frac{a''(s)}{a'(s)}\right)^2$$

The two-dimensional PDE corresponding to our original KBS equation for our stochastic volatility model has therefore been reduced via a change of coordinates and two gauge transformations to a Schrödinger equation with a scalar potential $J(s)$. The stochastic volatility model is therefore solvable in terms of hypergeometric functions if the potential $J(s)$ belongs to the Natanzon class. Finally, the solution is given (in terms of the eigenvectors ψ_{nk} with the appropriate boundary condition at $\tau = 0$) by

$$\mathcal{C}(\tau,a,f) = e^{-\Lambda(a,f)}\mathcal{F}^{-1}[\left(\frac{\sigma}{a}\right)^{-\frac{1}{2}} e^{-ik\rho\int^a \frac{a'}{\sigma(a')}da'}\sum_n \psi_{nk}(s(a))e^{-E_{nk}\tau}]$$

Let us examine classical examples of solvable stochastic volatility models and show that the potentials $J(s)$ correspond to the Natanzon class (see table 9.8).

We also present a new example of solvable stochastic volatility model which corresponds to the Posh-Teller I potential.

TABLE 9.8: Stochastic volatility models and potential $J(s)$.

name	potential	$J(s)$
Heston	3D osc.	$J(s) = \frac{-3+4Bs^2\eta+s^4\eta^2\left(2A+k^2\left(-1+\rho^2\right)\right)}{4s^2}$
Geometric Brownian	Morse	$J(s) = \frac{-\eta^2+4\,e^{\sqrt{2}s\,\eta}\left(2\,A+k^2\left(-1+\rho^2\right)\right)}{8}$
3/2-model	Coulombian	$J(s) = \frac{8\,A+\eta\,(-8\,B+\eta)+4\,k^2\left(-1+\rho^2\right)}{4\,s^2\,\eta^2}$

Example 9.16 Posh-Teller I
For a volatility function given by $\sigma(a) = \alpha + \eta a^2$, the potential is given by a Posh-Teller I potential[11]

$$J(s) = \frac{\alpha}{8\eta} \; (4\left(-2A + k^2 + 4B\eta - k^2\rho^2\right) - 3\eta^2 \mathrm{csc}\left(\frac{s\sqrt{\alpha}\sqrt{\eta}}{\sqrt{2}}\right)^2 + \left(8A + \eta\left(-8B + \eta\right) + 4k^2\left(-1 + \rho^2\right)\right) \sec\left(\frac{s\sqrt{\alpha}\sqrt{\eta}}{\sqrt{2}}\right)^2)$$

▯

9.6 Laplacian heat kernel and Schrödinger equations

In chapter 6, we have seen that the Kolmogorov PDE for the log-normal SABR model can be mapped to the Laplacian heat kernel on $\mathbb{H}^3$. In this section, we generalize this result and exhibit a strong relation between a Schrödinger equation on an a n-dimensional Riemannian manifold M with a negative potential and a pure Laplacian heat kernel equation on a $(n+1)$-dimensional Riemannian manifold $\mathbf{M}$ with a Killing vector.

Reduction

$\mathbf{M}$ can be considered as a $\mathbb{R}$-principal bundle over the manifold M. In short, a G-principal bundle is a vector bundle with a Lie group G acting transitively on the fibers. In our case, the Lie algebra of G is generated by the Killing vector and is isomorphic to $\mathbb{R}$. In such a space $\mathbf{M}$, the data of a metric are equivalent to the data of

- a metric on the base M
- an Abelian connection $\mathcal{A}$

[11] $\mathrm{csc}(z) \equiv \frac{1}{\sin(z)}$ and $\sec(z) \equiv \frac{1}{\cos(z)}$.

- a scalar function ϕ (i.e., a scalar potential)

Indeed, an Abelian connection on $\mathbf{M}$ corresponds to the orthogonal decomposition of the tangent space $T_p\mathbf{M}$ to $\mathbf{M}$ at a point p into a vertical part $T_p\mathbf{V}$ (i.e., vectors co-linear to the Killing vector) and an horizontal part $T_p\mathbf{H}$. The one-form on $\mathbf{M}$ which cancels on these horizontal vectors is an Abelian connection.
Using our metric $\mathbf{g}$ on $\mathbf{M}$, we can define an horizontal vector space as the orthogonal complement to $T_p\mathbf{V}$. This gives an Abelian connection $\mathcal{A}$. Moreover, the metric restricted to $T_p\mathbf{V}$ gives rise to a positive function e^{ϕ} and the metric pullback to the base gives a metric on M.
Finally, the metric $\mathbf{g}_{\mu,\nu}$ with $\mu, \nu = 1, \cdots, n+1$ can be written locally as

$$\begin{aligned} \mathbf{ds}^2 &= \mathbf{g}_{\mu\nu} dx^\mu dx^\nu \\ &= g_{ij} dx^i dx^j + e^{\phi(x)} (d\theta + \mathcal{A}_i dx^i)(d\theta + \mathcal{A}_i dx^i) \end{aligned} \tag{9.39}$$

with the Latin indices $i, j = 1, \cdots, n$.
In the following, to get our connection with a Schrödinger equation on M, we assume that $\mathcal{A} = 0$. $\mathbf{M}$ becomes a so-called *wrapped product*, noted $M \bowtie \mathbb{R}$, of M with $\mathbb{R}$. We consider the (Laplacian) heat kernel equation on $\mathbf{M}$:

$$\partial_t \mathbf{p}(t, x, \theta | x', \theta') = \triangle_{\mathbf{M}} \mathbf{p}(t, x, \theta | x', \theta')$$

with $\triangle_{\mathbf{M}} = \mathbf{g}^{-\frac{1}{2}} \partial_i \mathbf{g}^{ij} \mathbf{g}^{\frac{1}{2}} \partial_j$ the Beltrami-Laplace operator and with the initial condition $\lim_{t \to 0} \mathbf{p}(t, x, \theta | x', \theta') = \delta(x - x')\delta(\theta - \theta')$. From (9.39), we obtain by direct computation

$$\triangle_{\mathbf{M}} = \triangle_M + \frac{1}{2} (\partial_i \phi) g^{ij} \partial_j + e^{-\phi} \partial_\theta^2$$

By assuming that $p(t, x, \cdot | x', \theta')$ is in $L^2(\mathbb{R})$, we can apply a Fourier transform over the variable θ

$$\mathbf{p}(t, x, \theta | x', \theta') \equiv \frac{1}{2\pi} \int_{\mathbb{R}} p_k(t, x | x', \theta') e^{\imath k \theta} dk$$

or equivalently

$$p_k(t, x | x', \theta') = \int_{\mathbb{R}} \mathbf{p}(t, x, \theta | x', \theta') e^{-\imath k \theta} d\theta \tag{9.40}$$

We take $\theta' = 0$ on order to impose the initial condition

$$\lim_{t \to 0} p(t, x, k | x', \theta' = 0) = \delta(x - x')$$

$p_k(t, x | x', \theta')$ satisfies then a heat kernel equation on M:

$$\partial_t p_k(t, x | x', \theta' = 0) = D p_k(t, x | x', \theta' = 0)$$

where $D = g^{-\frac{1}{2}}\left(\partial_i + \mathcal{A}_i\right) g^{\frac{1}{2}} g^{ij}\left(\partial_j + \mathcal{A}_j\right) + Q$ with

$$\mathcal{A}_i = \frac{1}{4}\partial_i\phi$$
$$Q = -k^2 e^{-\phi}$$

Then by doing a gauge transform

$$\begin{aligned} p'_k(t,x|x') &= e^{\frac{1}{4}\left(\phi(x)-\phi(x')\right)} p_k(t,x,k|x',\theta'=0) \\ &= e^{\frac{1}{4}\left(\phi(x)-\phi(x')\right)} \int_{\mathbb{R}} \mathbf{p}(t,x,\theta|x',\theta'=0) e^{-\imath k\theta} d\theta \end{aligned} \qquad (9.41)$$

we obtain that $p'_k(t,x|x')$ satisfies a Schrödinger equation on M with a negative potential

$$\partial_t p'_k(t,x|x') = \Delta_M p'_k(t,x|x') - k^2 e^{-\phi(x)} p'_k(t,x|x')$$

Analytical solution for gauge-free SVMs

The Kolmogorov equation reduces to a Schrödinger equation with a negative potential on a non-compact Riemann surface for the so-called gauge-free SVMs. In this chapter, modulo a Fourier transform, we have shown that this $2d$ PDE can be converted into a $1d$-Schrödinger equation for which we can apply a spectral decomposition if the potential belongs to the Natanzon class. From the previous section, this is equivalent to solve a pure heat kernel equation on a three-dimensional (non-compact) manifold. Once we have a solution, we only need to apply a Fourier transform without relying on spectral expansion. To our knowledge, the known manifolds for which we have an analytical heat kernel are the spaces of constant sectional curvature. As $\mathbf{M}$ has at least two Killing vectors, ∂_y and ∂_θ, M can be $\mathbb{H}^3$, $\mathbb{H}^2 \times \mathbb{R}$ and $\mathbb{R}^3$. The log-normal (resp. normal) SABR model corresponds to $\mathbb{H}^3$ (resp. $\mathbb{H}^2 \times \mathbb{R}$) (see chapter 6).

Conclusion

From all this, we have shown how to use supersymmetric methods to generate new solutions to the Kolmogorov & Black-Scholes equation (KBS) for one-dimensional diffusion processes. In particular, by applying a supersymmetric transformation on the Natanzon potential, we have generated a hierarchy of new solvable processes. Then, we have classified the stochastic volatility models which admit a flat Abelian connection (with one Killing vector). The two-dimensional KBS equation has been converted into a Schrödinger equation with a scalar potential. The models for which the scalar potential belongs to the Natanzon class are solvable in terms of hypergeometric functions. This is the case for the Heston model, the geometric Brownian model and the 3/2-model. A new solution with a volatility of the volatility $\sigma(a) = \alpha + \eta a^2$, corresponding to the Posh-Teller I, has been presented.

9.7 Problems

Exercises 9.1 Symmetry

1. For a LVM defined by $df_t = C(f_t)dW_t$, prove that the intrinsic fair value of a European call option defined as

$$G(\tau, f_0, K) \equiv \mathcal{C}(\tau, K) - \max(f_0 - K, 0)$$

is

$$G(\tau, f_0, K) = \frac{\sqrt{C(K)C(f_0)}}{\sqrt{2}} \int_0^\tau P(t', s(K))dt'$$

where the fundamental solution $P(t', s(K))$ satisfies an Euclidean Schrödinger equation.

2. Deduce the symmetry

$$G(\tau, f_0, K) = G(\tau, K, f_0)$$

Chapter 10

Schrödinger Semigroups Estimates and Implied Volatility Wings

Abstract We study the small/large strike behavior of the implied volatility for local and stochastic volatility models. The derivation uses two-sided estimates for Schrödinger equations on Riemannian manifolds with scalar potentials belonging to different classes.

10.1 Introduction

Since Black-Scholes, several alternative models have emerged such as local volatility models (LVMs) and stochastic volatility models (SVMs) which have been reviewed in chapters 5 and 6. Although all these models are able to fit reasonably the market implied volatilities across a (liquid) range of strikes and maturities, the overall properties of each model are quite different.

Notable differences come from the dynamics of the implied volatility. More precisely, although matching the initial market implied volatility surface, two models, belonging to different model classes, can give different prices when pricing exotic options such as forward-starting options.

Similarly, the different (illiquid) large-strike behaviors, not quoted on the market, produce different prices for volatility derivatives, such as variance swaps which strongly depend on the implied volatility wings.

In order to choose the best model for pricing and capturing the risk, it is therefore important to understand the general properties of these models. Below is a list of common properties that can characterize a model:

- Is the model exploding in a finite time? A study for some particular stochastic volatility models was achieved in [47, 29] and reviewed in chapter 6 using the Feller criteria.

- What is the short-time behavior of the implied volatility? Can we use this short-time asymptotics to calibrate the model if no analytical solutions are available? A general short-time asymptotics for LVMs and SVMs at the first-order in the maturity is proposed in chapters 5 and 6 using the heat kernel expansion on a Riemann manifold endowed with an Abelian connection.

- Is it solvable? In particular, can we obtain analytical prices for liquid options and calibrate efficiently the model? An extensive classification of solvable LVMs and SVMs is achieved in chapter 9.

- What is the large-strike behavior of the implied volatility? A partial answer to this question was given by the Lee moment formula [122] which translates the behaviors of the wings in the existence of higher moments for the forward. Recently, Benaim and Friz gave a tail-wing formula [53] sharpening the moment formula which relates the tails of the risk neutral returns with the wings behavior of the implied volatility.

- What is the large-time behavior of the implied volatility? A study of the large-time behavior of the implied volatility for analytical SVMs, mainly Heston, 3/2 and geometric Brownian models was achieved in [29].

- What is the forward implied volatility and skew implied by a model? An asymptotic analytical answer was given in [59] for the Heston model for small/long maturities.

In this chapter, we study the large-strike behavior of the implied volatility for LVMs and SVMs. The motivation is to obtain elegant expressions to describe the left/right-tail behavior in the regime where numerical PDE or Monte Carlo methods break down.
In this context, we use sharp bounds for heat kernel equations. In order to obtain accurate estimates, we restrain ourselves to SVMs for which the Itô generator can be transformed into a self-adjoint unbounded operator. Then, the Kolmogorov equation reduces to an (Euclidean) Schrödinger equation on a Riemannian manifold with a time-dependent scalar potential as explained in detail in chapter 9. For LVMs, this reduction is always possible. Although the reading of chapter 9 is recommended, we have included a brief reminder.

10.2 Wings asymptotics

For the sake of completeness, we recall in this section a few basic definitions and results. The implied volatility is the (unique) value of the volatility $\sigma_{\mathrm{BS}}(\tau, k)$ that, when put in the Black-Scholes formula, reproduces the market price for a European call option with log-strike $k = \ln\left(\frac{K}{f_0}\right)$ and maturity τ

$$P_{0\tau}^{-1}\mathcal{C}(\tau, k, \sigma) = f_0 N(d_+) - K N(d_-)$$

with $d_\pm = -\frac{k}{\sigma\sqrt{\tau}} \pm \frac{\sigma\sqrt{\tau}}{2}$. Below, we denote the dimensionless parameter $V_{\mathrm{BS}}(\tau,k) \equiv \sigma_{\mathrm{BS}}(\tau,k)\sqrt{\tau}$ and let the strictly decreasing function $\Psi : [0,\infty] \to [0,2]$ be

$$\Psi(x) = 2 - 4(\sqrt{x^2+x} - x)$$

A first step in the study of the implied volatility wings is achieved by the Lee moment formula [122] which translates the behavior of the wings in the existence of higher moments for the forward. This formula states that the implied volatility is at-most linear in the moneyness k as $|k| \to \pm\infty$ with a slope depending on the existence of higher moments. More precisely, we have

THEOREM 10.1 Lee's theorem [122]
Let $q^* = \sup\{q : \mathbb{E}[f_T^{-q}] < \infty\}$. *Then*

$$\limsup_{k\to\infty} \frac{V_{\mathrm{BS}}(\tau,-k)^2}{|k|} = \Psi(q^*)$$

Also let $p^* = \sup\{p : \mathbb{E}[f_T^{1+p}] < \infty\}$. *Then*

$$\limsup_{k\to\infty} \frac{V_{\mathrm{BS}}(\tau,k)^2}{k} = \Psi(p^*)$$

REMARK 10.1 Small-strike asymptotics from large-strike asymptotics We can deduce the small-strike asymptotics from the large-strike asymptotics using the put-call duality (see section 6.2 for details) which means that

$$\mathbb{E}^{\mathbb{P}^f}[\max(X_T - K, 0)|\mathcal{F}_t] = \frac{K}{f_0}\mathbb{E}^{\mathbb{P}}[\max(K^{-1} - f_T, 0)\mathcal{F}_t] \qquad (10.1)$$

with $X_T \equiv f_T^{-1}$. Here $\mathbb{P}^f$ is the measure associated to the forward f_T defined by the Radon-Nikodym derivative

$$\frac{d\mathbb{P}^f}{d\mathbb{P}}|\mathcal{F}_T = f^T$$

From (10.1) and the definition of the implied volatility, we have

$$\limsup_{k\to\infty} \frac{V_{\mathrm{BS}}^X(\tau,k)^2}{|k|} = \limsup_{k\to\infty} \frac{V_{\mathrm{BS}}^f(\tau,-k)^2}{|k|}$$

By using the large-strike asymptotics for the process X_T, we have

$$\limsup_{k\to\infty} \frac{V_{\mathrm{BS}}^X(\tau,k)^2}{k} = \Psi(p_X^*)$$

with $p_X^* = \sup\{p : \mathbb{E}^{\mathbb{P}^f}[X_T^{p+1}] < \infty\}$. As

$$\mathbb{E}^{\mathbb{P}^f}[X_T^{p+1}] = \mathbb{E}^{\mathbb{P}}[\frac{d\mathbb{P}^f}{d\mathbb{P}} f_T^{-p-1}] = \mathbb{E}^{\mathbb{P}}[f_T^{-p}]$$

we deduce the small-strike asymptotics formula. ∎

For example, when all the moments exist, we have

$$\limsup_{k\to\pm\infty} \frac{V_{\mathrm{BS}}(\tau,k)^2}{|k|} = 0$$

If we apply this result to the Black-Scholes log-normal process, the lim sup is rough as the implied volatility is flat. Moreover, the Lee moment formula is quite hard to use as we generally don't have at hand an analytical conditional probability, hence the difficulty to examine the existence of higher moments. The moment formula was recently sharpened by Benaim and Friz [53] who show how the tail asymptotics of the log stock price translate *directly* to the large-strike behavior of the implied volatility. Their tail-wing formula reliably informs us when the lim sup in Lee moment formula [122] can be strengthened to a true limit. More specifically, it links the tail asymptotics of the distribution function of the log return to the implied volatility behavior at extreme strikes. Moreover, Benaim and Friz [54] develop criteria based on Tauberian theorems to establish when the lim sup in the Lee moment formula can be replaced by a true limit.
We define the function class $\mathbb{R}_\alpha$ of *regularly varying function* of index α by

DEFINITION 10.1 Class $\mathbb{R}_\alpha$ *A positive real-valued measurable function g is regularly varying with index α, in symbols $g \in \mathbb{R}_\alpha$, if*

$$\lim_{x\to\infty} \frac{g(\lambda x)}{g(x)} = \lambda^\alpha$$

Assumption (IR): We assume the integrability condition on the right tail denoted by (IR)

$$\exists\, \epsilon > 0 \ : \ \mathbb{E}[f_\tau^{(1+\epsilon)}] < \infty$$

Then we have the following theorem [53]

THEOREM 10.2 [53]
Assume condition (IR). Then if $-\ln p(\tau,k) \in \mathbb{R}_{\alpha>0}$ with $p(\tau,k)$ the probability of the log-forward, we have[1]

$$\frac{V_{\mathrm{BS}}(\tau,k)}{k}^2 \sim \Psi\left(-1 - \frac{\ln p(\tau,k)}{k}\right)$$

[1] $f \sim g$ means $\lim_{k\to\infty} \frac{f(k)}{g(k)} = 1$.

In the following, we try to derive the wings asymptotics directly from the behavior of the local or stochastic volatility function. Our technique is applicable when neither the moment generating function nor the distribution function is known.
Note that for commonly used LVMs such as the Constant Elasticity of Variance model (CEV), the probability density (5.23) is not a regularly varying function and the theorem above is not applicable. Our study will rely on the theorem 10.3 which relates the asymptotics of the wings to the behavior of the call option price. We note $\mathcal{C}(\tau, k)$ the fair value of a European call option with log-strike k and maturity τ and we assume
Assumption (IR'): The condition (IR) is satisfied and $-\ln \mathcal{C}(\tau, k)$ is a regularly varying function in the moneyness k (or in the strike K) with a strictly positive index.

THEOREM 10.3 [55]
Let us assume (IR'). Then

$$\frac{V_{\mathrm{BS}}(\tau, k)^2}{k} \sim \Psi\left(-\frac{\ln \mathcal{C}(\tau, k)}{k}\right) \tag{10.2}$$

In particular, if $-\frac{\ln(\mathcal{C}(\tau,k))}{k}$ goes to infinity as $k \to \infty$, the expression above becomes

$$\frac{V_{\mathrm{BS}}(\tau, k)^2}{k} \sim \frac{k}{-2\ln \mathcal{C}(\tau, k)}$$

Note that a similar result can be obtained for the left wing using the put-call duality (see remark 10.1). This is why we only focus on the right wing in the next sections.
Let us now explain how to find Gaussian estimates for the conditional probability for LVMs. From these Gaussian estimates, we obtain lower and upper estimates of the wings using the theorem 10.3. In order to find these Gaussian estimates, we first reduce the Kolmogorov equation associated to a LVM to an (Euclidean) Schrödinger equation with a (time-dependent) scalar potential.

10.3 Local volatility model and Schrödinger equation

10.3.1 Separable local volatility model

Before moving on to the general case, we recall briefly the reduction method explained in chapter 9 with a simpler example, namely a separable LV model:

We assume that under $\mathbb{P}$, the forward f follows a one-dimensional, time-homogeneous regular diffusion on an interval $I \subset \mathbb{R}_+$

$$df_t = A(t)C(f_t)dW_t$$

with $f_{t=0} = f_0$. The so-called local volatility $C(f)$ is a positive continuous function on I and $A(t)$ a strictly positive continuous function on $\mathbb{R}_+$. The Itô infinitesimal generator $\mathcal{H}$ associated to the backward Kolmogorov equation for the conditional probability $p(t, f|f_0)$ is

$$\mathcal{H}p(t, f|f_0) = \frac{1}{2}A(t)^2 C(f_0)^2 \partial_{f_0}^2 p(t, f|f_0)$$

We now convert the backward Kolmogorov equation into a simpler form. For this purpose, we assume that $C'(f)$ and $C''(f)$ exist and are continuous on I in order to perform our (Liouville) transformation. By introducing the new coordinate $s = \sqrt{2}\int_{\cdot}^{f_0} \frac{df'}{C(f')}$ which can be interpreted as the geodesic distance and the new time $t^{'} = \int_0^t A(s)^2 ds$, one can show that the new function $P(t', s)$ defined by

$$p(t, f|f_0) = P(t', s)\frac{\sqrt{2C(f_0)}}{C(f)^{\frac{3}{2}}} \tag{10.3}$$

satisfies a (Euclidean) one-dimensional Schrödinger equation

$$(\partial_s^2 + Q(s))P(t', s) = \partial_{t'} P(t', s) \tag{10.4}$$

The time-homogeneous potential is[2]

$$Q(s) = \frac{1}{2}(\ln C)^{''}(s) - \frac{1}{4}((\ln C)^{'}(s))^2 \tag{10.5}$$

where the prime $'$ indicates a derivative according to s. Table 10.1 is a list of examples of potentials for a few particular time-homogeneous LV models.
By applying Itô-Tanaka's formula on the payoff $\max(f_t - K, 0)$ (5.29), we obtain that a European call option $\mathcal{C}(\tau, k)$ with maturity τ and log-strike k can be rewritten as a local time (without loss of generality, we assume a zero interest rate)

$$\mathcal{C}(\tau, k) = \max(f_0 - K, 0) + \int_0^\tau \frac{C(K)^2}{2} p(t', K|f_0)dt'$$

Using the relation (10.3) between the conditional probability $p(t, f|f_0)$ and its gauge-transform $P(t, s)$, we obtain

$$\mathcal{C}(\tau, k) = \max(f_0 - K, 0) + \frac{\sqrt{C(K)C(f_0)}}{\sqrt{2}} \int_0^\tau P(t', s(K))dt' \tag{10.6}$$

[2]In the f-coordinate, we have $Q(f) = \frac{1}{8}\{2C(f)C''(f) - C'(f)^2\}$.

TABLE 10.1: Example of potentials associated to LV models.

LV Model	$C(f)$	Potential
Black-Scholes	f	$Q(s) = -\frac{1}{8}$
Quadratic	$af^2 + bf + c$	$Q(s) = -\frac{1}{8}(b^2 - 4ac)$
CEV	f^β, $0 \leq \beta < 1$	$Q_{CEV}(s) = \frac{\beta(\beta-2)}{4(1-\beta)^2 s^2}$
LCEV	$f \min\left(f^{\beta-1}, \epsilon^{\beta-1}\right)$ with $\epsilon > 0$	$s_\epsilon \equiv \frac{\sqrt{2}\epsilon^{1-\beta}}{(1-\beta)}$ $Q_{LCEV}(s) = Q_{CEV}(s)\ \forall\ s \geq s_\epsilon$ $Q_{LCEV}(s) = -\frac{1}{8}\epsilon^{2(\beta-1)}\ \forall\ s < s_\epsilon$

10.3.2 General local volatility model

In the Dupire LV model [81], it is assumed that under $\mathbb{P}$ the forward follows a one-dimensional regular diffusion on an interval $I \subset \mathbb{R}_+$

$$df_t = C(t, f_t)dW_t$$

The so-called Dupire local volatility $C(t, f)$ is a strictly positive continuous function on $\mathbb{R}_+ \times I$. The Itô infinitesimal generator $\mathcal{H}$ is

$$\mathcal{H}p(t, f[f_0) = \frac{1}{2}C(t, f_0)^2 \partial_{f_0}^2 p(t, f|f_0)$$

Throughout this chapter, we adopt the reduced variable $s = \ln \frac{f}{f_0}$, $\sigma(t, s) = \frac{C(t, f_0 e^s)}{f_0 e^s}$ and $p(t, s) = p(t, f|f_0)$. The transformed forward Kolmogorov equation satisfies

$$\partial_t p(t, s) = \frac{1}{2}\partial_s\left(\sigma(t, s)^2 \partial_s p(t, s)\right) + \frac{1}{2}\partial_s\left(\sigma(t, s)^2 p(t, s)\right)$$

We assume also
Assumption 1: $\sigma(t, s)$, $\partial_s\sigma(t, s)$, are uniformly continuous on $\mathbb{R}_+ \times I$.
Note that these conditions are not at all restrictive and are satisfied for (reasonable) market conditions. The new function $P(t, s)$ defined by

$$p(t, s) = e^{-\frac{s}{2}} P(t, s)$$

satisfies a one-dimensional (Euclidean) Schrödinger equation

$$\frac{1}{2}\partial_s(\sigma(t, s)^2 \partial_s)P(t, s) + Q(t, s)P(t, s) = \partial_t P(t, s) \tag{10.7}$$

with a time-dependent scalar potential given by

$$Q(t, s) = -\frac{1}{8}\sigma(t, s)^2 + \frac{1}{2}\sigma(t, s)\partial_s\sigma(t, s) \tag{10.8}$$

Proceeding as in the previous section, the fair value of a European call option $\mathcal{C}(\tau, k)$ can be written as

$$\mathcal{C}(\tau, k) = \max(f_0 - K, 0) + \sqrt{\frac{f_0}{K}} \int_0^\tau \frac{C(t', K)^2}{2} P(t', k)dt' \tag{10.9}$$

Still proceeding as previously discussed, we convert the PDE (10.7) into a more conventional Schrödinger equation on $\mathbb{R}$ by doing a time-dependent change of coordinates $s_t = \sqrt{2}\int_{f_0}^{f} \frac{df'}{C(t,f')}$. Thus, the new function $P(t,s)$ defined by

$$p(t,f|f_0) = \frac{\sqrt{2}}{C(t,f)}\sqrt{\frac{C(t,f_0)}{C(t,f)}}P(t,s)e^{\int_{f_0}^{f} \frac{df'}{C(t,f')}\partial_t \int_{f_0}^{f'} \frac{df''}{C(t,f'')}} \tag{10.10}$$

satisfies a one-dimensional (Euclidean) Schrödinger equation

$$(\partial_s^2 + Q(t,s))P = \partial_t P \tag{10.11}$$

with a time-dependent scalar potential given by

$$Q(t,s) = -\partial_s \mu(t,s) - \mu(t,s)^2 - \int_0^s \partial_t \mu(t,s')ds' \tag{10.12}$$

with $\mu(t,s) = \partial_t(\int_{f_0}^{f(s)} \frac{df'}{C(t,f')}) - \frac{1}{2}\partial_s \ln(C(t,s))$. Note that for a time-homogeneous local volatility $C(t,f) \equiv C(f)$, the potential $Q(t,s)$ reduces to (10.5).
After the reduction of the Kolmogorov equation into a symmetric semigroup, we explain how to find Gaussian lower and upper bounds.

10.4 Gaussian estimates of Schrödinger semigroups

The history of *Gaussian estimates* of parabolic PDEs is quite rich starting with the works of Nash [130] and Aronsov [48] on Gaussian estimates for Laplacian heat kernel equation. By Gaussian estimates, we mean in short that the fundamental solution $P(t,s)$ can be bounded by two Gaussian distributions

$$c_1 p_G(c_2 t, s|s_0) \le P(t,s) \le C_1 p_G(C_2 t, s|s_0)$$

with $p_G(t,y|x) = \frac{1}{(4\pi t)^{\frac{1}{2}}}\exp(-\frac{(y-x)^2}{4t})$ the Gaussian heat kernel and $\{C_i, c_i\}_{i=1,2}$ some constants.

10.4.1 Time-homogenous scalar potential

The fundamental solution of (10.4) satisfies Gaussian bounds provided that the potential $Q(s)$ belongs to the Kato class:

DEFINITION 10.2 Autonomous Kato class *We say that $Q(\cdot)$ is in the Kato class K if*

$$\lim_{\delta\to 0}\sup_{x\in\mathbb{R}} \int_0^\delta dt \int_{\mathbb{R}} dy |Q(y)| p_G(t,y|x) = 0$$

Moreover, we say that $Q \in K_{\text{loc}}$ *if* $\forall N \geq 1$, $Q(y)1_N(y) \in K$.[3]

Additional properties for potentials belonging to the Kato class can be found in [37]. Then, we have

THEOREM 10.4 [139]
Let $Q^+ \equiv \max(Q, 0) \in K$ *and* $Q^- = \max(-Q, 0) \in K_{\text{loc}}$. *Then we have an upper bound*

$$P(t, y|x) \leq C_1 e^{C_2 t} p_G(t, y|x), \; t > 0, \; x, y \in \mathbb{R}$$

with two constants C_1, C_2. *Note that the constant* $C_2 = 0$ *if* $Q^+ = 0$. *By assuming that* Q^+ *and* Q^- *are both in the Kato class* K, *we have also a lower bound*

$$c_1 e^{c_2 t} p_G(t, y|x) \leq P(t, x|y), \; t > 0, \; x, y \in \mathbb{R}$$

with two constants c_1 *and* c_2.

One can check that the models listed in (Table 10.1) except the CEV model (due to the singularity at $f = 0$) belong to the Kato class.

Example 10.1
For the short-rate Vasicek model, $Q(y)$ is the harmonic potential $Q(y) = y^2$. This potential does not belong to the Kato class. □

REMARK 10.2 Lower bound from upper bound It is a classical fact that for Schrödinger operators $H = \Delta + Q$ on $L^2(\mathbb{R}^d)$, Gaussian lower bounds follow from upper bounds as reviewed in [133]. The derivation relies on the Feynman-Kac formula.
Let us denote W_t the d-dimensional Brownian motion and Δ the Laplacian on $\mathbb{R}^d$. We have for every non-negative $f \in C_c^\infty(\mathbb{R}^d)$

$$\begin{aligned} e^{-t\Delta} f(x) &= \mathbb{E}_x^{\mathbb{P}}[f(W_t)] \\ &\leq \sqrt{\mathbb{E}_x^{\mathbb{P}}[f(W_t) e^{-\int_0^t Q(W_s) ds}]} \sqrt{\mathbb{E}_x^{\mathbb{P}}[f(W_t) e^{\int_0^t Q(W_s) ds}]} \\ &= \sqrt{e^{-t(\Delta - Q)} f(x)} \sqrt{e^{-t(\Delta + Q)} f(x)} \end{aligned}$$

where we have used the Cauchy-Schwartz inequality. Then, we have

$$e^{-t\Delta} f(x) \leq \frac{1}{\epsilon} e^{-t(\Delta + Q)} f(x) + \epsilon e^{-t(\Delta - Q)} f(x)$$

[3] $1_N(y) = 1$ if $y \leq N$, zero otherwise.

This inequality holds for all $\epsilon > 0$. We deduce

$$p_G(t, x|x_0) \leq \frac{1}{\epsilon} p_{\Delta+Q}(t, x|x_0) + \epsilon p_{\Delta-Q}(t, x|x_0)$$

where $p_{\Delta\pm Q}(t, x|x_0)$ is the conditional probability associated to the Itô generator $H = \Delta \pm Q$. By assuming that Q^- is in the Kato class, we have an upper bound for $p_{\Delta-Q}$

$$\epsilon p_G(t, x|x_0) \leq \epsilon^2 C_1 e^{C_2 t} p_G(t, x|x_0) + p_{\Delta+Q}(t, x|x_0)$$

and

$$p_{\Delta+Q}(t, x|x_0) \geq p_G(t, x|x_0) \left(\epsilon - \epsilon^2 C_1 e^{C_2 t}\right)$$

We optimize over ϵ and we obtain the Gaussian lower bound

$$p_{\Delta+Q}(t, x|x_0) \geq \frac{1}{4C_1} e^{-C_2 t} p_G(t, x|x_0)$$

□

10.4.2 Time-dependent scalar potential

A similar result [127] can be obtained for Schrödinger equation (10.11) with a time-dependent scalar potential $Q(t, s)$ provided that the Kato class K is extended to the non-autonomous Kato class $\hat{K}$ defined by

DEFINITION 10.3 Non-autonomous Kato class *We say that $Q(\cdot, \cdot)$ is in the non-autonomous Kato class $\hat{K}$ if*

$$\lim_{\delta \to 0} N_\delta^{\pm}(Q) = 0$$

where

$$N_\delta^{\pm}(Q) = \sup_{s, x \in \mathbb{R}} \int_0^\delta dt \int_{\mathbb{R}} dy |Q(s \pm t, y)| p_G(t, x|y)$$

Unfortunately, even for simple models such as $C(t, f) = a(t)f + b(t)$, the potential $Q(t, s)$ (10.12) does not belong to the non-autonomous Kato class. Therefore, as explained by [87], we assume a stronger boundness condition on the potential Q defined by (10.8) so that we can apply a sharp Gaussian estimate found by Norris and Stroock [131] for the PDE (10.7).
We define the control distance

$$d(t, s)^2 = \inf_{\gamma \in \Gamma(t, s)} \int_0^t \frac{2}{\sigma(u, \gamma(u))^2} \left(\frac{d\gamma(u)}{du}\right)^2 du$$

where $\Gamma(t,s) = \{\gamma \in C^1([0,t],\mathbb{R}) : \gamma_0 = 0, \gamma_t = s, \int_0^t \left(\frac{d\gamma(u)}{du}\right)^2 du < \infty\}$. Note that for time homogeneous LVMs, the control distance reduces to the classical geodesic distance.
Then, we have (as reported in [87])

THEOREM 10.5
Assume the condition **Assumption 1** *and that there are constants* $\lambda \in [1,\infty)$ *and* $\Lambda \in [0,\infty)$ *such that, uniformly on* $\mathbb{R}_+ \times I$,

$$\lambda^{-1} \leq \frac{1}{2}\sigma(t,s)^2 \leq \lambda \text{ and } |Q(t,s)| \leq \Lambda$$

Let $\alpha \in (\frac{1}{2},1)$ *satisfying* $\frac{\alpha^2}{2\alpha-1} > \lambda^2$ *be given. Then for all* $\lambda \in [1,\infty), \Lambda \in [0,\infty)$ *and all* $T \in (0,\infty)$*, we have the following asymptotic result for the fundamental solution* $P(t,s)$ *associated with (10.7)*

$$\begin{aligned} \lim_{M\to\infty} \inf_{0<t\leq T,\, s\in,\mathbb{R}, E\geq M} \frac{\log P(t,s)+E}{E^{\frac{1}{4\alpha-1}}} &= 0 \\ \limsup_{M\to\infty} \sup_{0<t\leq T,\, s\in,\mathbb{R}, E\geq M} \frac{\log P(t,s)+E}{\log E} &\leq \frac{N}{4\alpha-2} \end{aligned} \tag{10.13}$$

where we have set $E \equiv \frac{d(t,s)^2}{4t}$.

REMARK 10.3 Norris & Stroock [131, 87] also mention that we have the estimate

$$\frac{|s|^2}{8\lambda t} - \frac{1}{4}\Lambda t < E < \frac{\lambda |s|^2}{2t} + \frac{1}{2}\Lambda t \tag{10.14}$$

in terms of the coefficient bounds λ and Λ. ∎

For fixed t, s, by (10.14), we see that E can be made arbitrarily large if s is sufficiently large. In particular for all $\epsilon > 0$, there exists a s such that

$$e^{-\epsilon E^{\frac{1}{4\alpha-1}}}\, e^{-E} < P(t,s) < E^{(\frac{N}{4\alpha-2}+\epsilon)}\, e^{-E}$$

Thus, the theorem (10.5) implies the following corollary of the Norris-Stroock result [87]

COROLLARY 10.1

$$-\ln P(t,s) \sim \frac{d(t,s)^2}{4t} \qquad (|s| \to \infty)$$

Note that in [87], this corollary was obtained in terms of the energy functional which is more difficult to compute than the *control distance*.
In the subsequent pages, we apply the theorems (10.2, 10.3) to obtain the large-strike behavior of the implied volatility.

10.5 Implied volatility at extreme strikes

10.5.1 Separable local volatility model

Assuming that the scalar potential (10.5) associated to a separable local volatility function belongs to the Kato class, we have the Gaussian bounds on the function $P(t,s)$ from the theorem (10.4)

$$c_1 e^{c_2 t} p_G(t,s) \le P(t,s) \le C_1 e^{C_2 t} p_G(t,s)$$

This inequality directly translates on an estimation of the fundamental solution $p(t,f|f_0)$ using the relation between $p(t,f|f_0)$ and its gauge-transform $P(t,s)$ (10.3)

$$c_1 e^{c_2 t} p_G(t,s) \le \frac{C(f)^{\frac{3}{2}}}{\sqrt{2C(f_0)}} p(t,f|f_0) \le C_1 e^{C_2 t} p_G(t,s) \tag{10.15}$$

In the following example, we will check the validity of these Gaussian estimates for the CEV model for which we know analytically the conditional probability.

Example 10.2 CEV model
The CEV model is a LV model for which the local volatility function is a power of the forward: $C(f) = f^\beta$, with $0 < \beta < 1$. A closed-form expression for the risk-neutral conditional probability (5.23) is given by

$$p(t,f|f_0) = \frac{f^{\frac{1}{2}-2\beta}}{(1-\beta)t}\sqrt{f_0}\, e^{-\frac{f^{2(1-\beta)}+f_0^{2(1-\beta)}}{2(1-\beta)^2 t}} \mathrm{I}_{\frac{1}{2(1-\beta)}}\left(\frac{(f f_0)^{1-\beta}}{(1-\beta)^2 t}\right)$$

where I is the modified Bessel function of the first kind. As

$$\mathrm{I}_{\frac{1}{2(1-\beta)}}(x) \sim_{x\to\infty} \frac{1}{\sqrt{2\pi x}} e^x$$

we deduce that the large forward limit exhibits a Gaussian behavior

$$\frac{\sqrt{2C(f_0)}}{C(f)^{\frac{3}{2}}} p(t,f|f_0) \sim \frac{1}{\sqrt{4\pi t}} e^{-\frac{s(f)^2}{4t}} \tag{10.16}$$

with $s(f) = \int_{\cdot}^{f} x^{-\beta} dx$. The potential associated to the CEV model is

$$Q(s) = \frac{\beta(\beta-2)}{4(1-\beta)^2 s^2}$$

Unfortunately, the potential doesn't belong to Kato class K due to the singularity at $s = 0$. Therefore, we modify the CEV model by the limited CEV model (LCEV) 5.5 for which the potential belongs to K and we have a Gaussian estimate for the heat kernel. From the LCEV definition, the large-forward behavior of the CEV and LCEV conditional probabilities are the same as shown in [44]. In the limit $f \to \infty$, from (10.16), the inequalities (10.15) are trivially satisfied for the LCEV model. Note that the lower and upper bound are *exact* in the limit $f \to \infty$ for $C_1 = 1$, $c_1 = 1$, $C_2 = 0$ and $c_2 = 0$. $\square$

By plugging our Gaussian estimates for the conditional probability (10.15) into the expression (10.6) and by doing the integration over the time t, we obtain lower and upper bounds for the fair value of a European call option. Then, assuming that the condition (IR') is fulfilled and using theorem (10.3), we can directly translate the Gaussian bounds on the call option into bounds on the implied volatility. For a time-homogeneous local volatility model, we obtain

THEOREM 10.6
By assuming that the quantity $-\frac{1}{2}\ln(C(K)) + \frac{s(K)^2}{4\tau}$ belongs to $\mathbb{R}_{\alpha>0}$ (in k or K) and that the potential (10.5), associated to a time-homogeneous LVM, belongs to the Kato class K, the large strike behavior of the implied volatility is given by

$$\frac{V_{\mathrm{BS}}(\tau,k)^2}{k} \sim_{k\to\infty} \Psi\left(\frac{-\frac{1}{2}\ln(C(K)) + \frac{s(K)^2}{4\tau}}{k}\right)$$

with $s(K) = \sqrt{2}\int_{f_0}^{K} \frac{df'}{C(f')}$. Moreover, if $s(K)$ is the leading term, we have that the large-strike behavior of the implied volatility involves the harmonic average of the local volatility function

$$\sigma_{\mathrm{BS}}(\tau,k) \sim_{k\to\infty} \frac{\sqrt{2}k}{s(K)}$$

REMARK 10.4 Note that this limit is the BBF formula (5.39) and was obtained for the short-time limit of the implied volatility in chapter 5: In the limit $\tau \to 0$, the implied volatility is the harmonic mean of the local volatility, namely

$$\lim_{\tau\to 0}\sigma_{\mathrm{BS}}(\tau,k) = \frac{\sqrt{2}k}{s(K)}$$

Therefore assuming that Q belongs to the Kato class, the large-strike and short-time behaviors coincide. □

As an example, we obtain the following tail estimates for the implied volatility in the CEV model

Example 10.3 CEV model
For $0 \leq \beta < 1$,we have

$$\sigma_{\rm BS}(k,\tau) \sim_{k\to\infty} \frac{k(1-\beta)}{K^{1-\beta}} \tag{10.17}$$

and for $\beta = 1$, we have $\sigma_{\rm BS}(\tau,k) \sim_{k\to\infty} 1$. □

This relationship is derived in [87, 55] using the Freidlin-Wentzell theory. This result should be compared with the result obtained using the Lee's moment formula

REMARK 10.5 By applying the Lee's moment formula, we obtain for $0 \leq \beta \leq 1$

$$\limsup_{k\to\infty} \frac{V_{\rm BS}(k,\tau)^2}{k} = 0$$

as all the moments exist. □

10.5.2 Local volatility model

For the Dupire LV model, we proceed as in the previous section by plugging the Gaussian estimates for $P(t,f|f_0)$ given by corollary (10.1) into (10.9). We obtain

THEOREM 10.7
By assuming that the control distance $d(t,s) \in R_{\alpha>0}$ (in k or in K) and that the local volatility $C(t,f)$ and the potential (10.8), associated to a LV model, satisfies the assumption in the theorem (10.5), the large strike behavior of the implied volatility is given by

$$\frac{V_{\rm BS}(k,\tau)^2}{k} \sim_{k\to\infty} \Psi\left(-\frac{1}{2} + \frac{d(\tau,k)^2}{4k\tau}\right)$$

10.6 Gauge-free stochastic volatility models

We assume that the forward f and the volatility a are driven by two correlated Brownian motions in the forward measure $\mathbb{P}^T$

$$df_t = a_t C(f_t) dW_t \tag{10.18}$$
$$da_t = b(a_t)dt + \sigma(a_t)dZ_t \tag{10.19}$$
$$dW_t dZ_t = \rho dt$$

with the initial conditions $a = \alpha$ and $f = f_0$. We assume that the diffusion process is non-explosive. The resulting Itô generator is generally not a symmetric operator. However, by doing a gauge transform on the conditional probability $p(\tau, f, a|f_0, \alpha)$ with a specific time-homogeneous function $\Lambda(f, a)$:

$$P(\tau, f, a|f_0, \alpha) = e^{\Lambda(f,a)} p(\tau, f, a|f_0, \alpha)$$

the Itô generator can be reduced to a symmetric operator for certain SVMs, called *gauge free stochastic volatility models* (GFSVM) in section 9.5. By construction for these models, $P(\tau, f, a|f_0, \alpha)$ satisfies a self-adjoint heat kernel equation ($x \equiv (f, a)$, $x_0 \equiv (f_0, \alpha)$)

$$\partial_\tau P(\tau, x|x^0) = \triangle P(x, \tau|x^0) + Q(x)P(\tau, x|x^0) \tag{10.20}$$

where $\triangle$ is the Laplace-Beltrami operator on a (non-compact) Riemann surface. Here Q is an (unspecified) time-homogeneous potential. In section 9.5, these (GFSVM) models have been completely characterized and are defined by

$$C(f) = \mu f + \nu$$
$$b(a) = a\sigma(a) \left(\gamma + \frac{1}{2} \partial_a \frac{\sigma(a)}{a} \right)$$

with μ, ν, γ three constants. The symmetric condition has therefore imposed the functional form of the drift term $b(a)$ and the volatility $C(f)$ of the forward.

The gauge transformation $\Lambda(a, f)$ which reduces the backward Kolmogorov equation to (10.20) is

$$(1 - \rho^2)\Lambda(f, a) = -\frac{1}{2} \ln \left(\frac{C(f)}{C(f_0)} \right) - \gamma\rho \int_{f_0}^{f} \frac{df'}{C(f')} + (\gamma + \frac{\rho}{2} \partial_f C(f)) \int_{\alpha}^{a} \frac{a' da'}{\sigma(a')}$$

The derivation of Gaussian estimates for a Schrödinger equation on a Riemannian manifold (10.20) is a difficult problem. A review of approaches and results can be found in [96]. Note that it is not possible to use the Norris-Stroock result as the metric g_{ij} is not uniformly elliptic and the potential is not bounded (it does not even belong to the Kato class in general). For example for the log-normal SABR model, g_{ij} is the metric on the non-compact complete hyperbolic plane and the potential Q is unbounded: $Q = -\frac{a^2}{8(1-\rho^2)}$. If the potential Q cancels and the Ricci curvature is bounded for below, there exists Gaussian estimates found by Li-Yau [125] and improved in [78, 141]:

THEOREM 10.8

Let Σ be a complete (non-compact) Riemannian manifold without boundary. If the Ricci curvature of M is bounded from below by $-K$, i.e.,

$$R_{\mu\nu}(x)X_\mu X_\nu \geq -(n-1)\kappa X_\mu X_\nu \quad (\forall x \in M, \forall X \in T_x M)$$

with a constant $\kappa \geq 0$, then the heat kernel $P(\tau, x|x^0)$ satisfies an upper bound [78]: For any $\epsilon > 0$, there exists a constant $C_1 = C_1(\kappa, d, \epsilon)$ such that for all $t > 0$ and $x, y \in M$:

$$p(t, x|y) \leq C_1 m^{-\frac{1}{2}}(B_{\sqrt{t}}(x)) m^{-\frac{1}{2}}(B_{\sqrt{t}}(y)) e^{\frac{-(1-\epsilon)d^2(x,y)}{4t}} e^{(\epsilon - \lambda(M))t}$$

$d(x, y)$ is the geodesic distance on Σ between the points x and y. $\lambda(M) \geq 0$ is the bottom of the L^2-spectrum of the operator $-\frac{1}{2}\Delta$ on M. $B_{\sqrt{t}}(x)$ is the geodesic volume of a ball of radius $\sqrt{t}$ and center x.
We also have the lower bound [141]

$$p(t, x|y) \geq (2\pi t)^{-\frac{d}{2}} e^{-(1-\epsilon)\frac{d^2(x,y)}{4t} - (n-1)2^{-\frac{3}{2}}\sqrt{\kappa} d(x,y)} e^{-\lambda^{\kappa,n} t}$$

with $\lambda^{\kappa,n} = \frac{(n-1)^2}{8}\kappa$.

Note that the geodesic distance associated to the SVM defined by (10.18,10.19) has been computed in chapter 6 and is given by (6.17).
Warning: In the following, we will assume that our Schrödinger equation (10.20) satisfies Gaussian bounds:

$$\frac{1}{4\pi t} c_1 e^{\frac{-d^2(x,y)}{4t}} \leq p(t, x|y) \leq \frac{1}{4\pi t} C_1 e^{\frac{-d^2(x,y)}{4t}} \tag{10.21}$$

From (10.21), we derive an upper (lower) bound for the value of a call Vanilla option with strike K and maturity τ

$$\mathcal{C}(\tau, k) \leq \max(f_0 - K, 0) + \frac{C_1 C(K)}{4\pi\sqrt{1-\rho^2}} \int_0^\infty \frac{a da}{\sigma(a)} e^{\Lambda(K,a)} E_1\left(\frac{d(K,a)^2}{4\tau}\right)$$

Note the sign =+ in front of Λ as we have switched f_0 (resp. α) with K (resp. a) (see remark 4.3). $E_1(x) = \int_x^\infty \frac{e^{-u}}{u} du$ is the exponential integral function. By assuming that $\lim_{K\to\infty} d(K, a) = \infty$ and as

$$E_1(x) \sim_{x\to\infty} \frac{e^{-x}}{x}\left(1 - \frac{1}{x} + \frac{2}{x^2}\right)$$

we obtain from the lower and upper bounds

$$\ln(\mathcal{C}(\tau,k)) \sim_{k\to\infty} \ln(C(K)) + \ln\left(\int_0^\infty \frac{a da}{\sigma(a)} \frac{e^{-\frac{d(K,a)^2}{4\tau}+\Lambda(K,a)}}{d(K,a)^2}\right)$$
$$\sim_{k\to\infty} \ln(C(K)) - \frac{d(K,a_{\min})^2}{4\tau} + \Lambda(K,a_{\min})$$

where $a_{\min} = a_{\min}(K)$ is the saddle-point which minimizes the function

$$-\frac{d(K,a)^2}{4\tau} + \Lambda(K,a)$$

From the theorem 10.3, we deduce

$$\frac{V_{\mathrm{BS}}(\tau,k)^2}{k} \sim_{k\to\infty} \Psi\left(\frac{-\ln(C(K)) + \frac{d(K,a_{\min})^2}{4\tau} - \Lambda(K,a_{\min})}{k}\right)$$

Specifying the function Λ, we have our final result

$$\frac{V_{\mathrm{BS}}(\tau,k)^2}{k} \sim_{k\to\infty} \Psi\left(-\frac{\ln C(K)}{k} + \frac{d(K,a_{\min})^2}{4k\tau}\right. \tag{10.22}$$
$$\left.+\frac{1}{2(1-\rho^2)k}\left(\frac{1}{2}\ln\frac{C(K)}{C(f_0)} + \gamma\rho\int_{f_0}^K \frac{df'}{C(f')} - \left(\gamma + \frac{\rho}{2}\partial_f C\right)\int_\alpha^{a_{\min}} \frac{a' da'}{\sigma(a')}\right)\right)$$

Example 10.4 Normal SABR Model
For the normal SABR model (i.e., $C(f) = 1$), the volatility is $\sigma(a) = \nu a$ and $\gamma = 0$. The metric associated to this model corresponds to the two-dimensional hyperbolic manifold $\mathbb{H}^2$. Besides, the gauge function Λ and the potential cancel: $\Lambda = 0$, $Q = 0$. In this case, the Schrödinger equation reduces to the Laplacian heat kernel on $\mathbb{H}^2$ for which the Li-Yau-Davies Gaussian estimate is valid.
We found that the effective distance is explicitly given by

$$d(a_{\min}) = \sqrt{\frac{2}{\nu^2}}\cosh^{-1}\left(\frac{-q\nu\rho - \alpha\rho^2 + a_{\min}}{\alpha(1-\rho^2)}\right)$$

with $q = K - f_0$ and $a_{\min}^2 = \alpha^2 + 2\alpha\nu\rho q + \nu^2 q^2$. We obtain

$$d(a_{\min}) \sim \sqrt{\frac{2}{\nu^2}}\ln q$$

From (10.22), we get

$$\sigma_{\mathrm{BS}}(\tau,k) \sim_{k\to\infty} \nu$$

□

Example 10.5 Log-normal SABR Model
For the log-normal SABR model, the potential is unbounded $Q(a) = -\frac{a^2}{8(1-\rho^2)}$ and we note that at the saddle-point it diverges when $k \to \infty$

$$Q(a_{\min}) \sim -\frac{\nu^2 k^2}{8(1-\rho^2)}$$

We suspect that our Gaussian estimates (10.21) are not correct. Indeed, from (10.22), we get

$$\frac{V_{\mathrm{BS}}(\tau,k)^2}{k} \sim_{k\to\infty} \Psi\left(-\frac{1+2\rho}{2(1+\rho)}\right)$$

This result is incorrect as the argument for Ψ should be positive. In [55], the Lee's moment formula has been applied

$$\limsup_{k\to\infty} \frac{V_{\mathrm{BS}}(\tau,k)^2}{k} = \Psi\left(\frac{\rho^2}{1-\rho^2}\right)$$

This result should also be compared with the (uncorrect) result obtained using the short-time asymptotics of the implied volatility

$$\lim_{k\to\infty} \frac{\sigma_{\mathrm{BS}}(\tau,k)^2}{k} = \lim_{k\to\infty} \frac{\nu^2 k}{\ln^2(k)} = \infty!$$

This illustrates the fact that the limits $\tau \to 0$ and $k \to \infty$ do not commute in general. □

We conclude this section with two conjectures.

Some conjectures

The potential for the SABR model with $\beta < 1$ evaluated at the saddle-point is constant

$$Q(a_{\min}(f)) = \frac{\nu^2}{4(1-\beta)^2}\left(\beta(\beta-1) - \frac{\beta^2}{2(1-\rho^2)}\right)$$

In this case, we conjecture that the Gaussian estimate (10.21) is valid and we get
Conjecture 1:
For the SABR model with $\beta < 1$, we have

$$\sigma_{\mathrm{BS}} \sim \frac{\nu}{1-\beta}$$

This conjecture is proved in [55] in the case $\rho = 0$.
Conjecture 2:
If the potential $Q(a_{\min})$ is bounded, the relation (10.22) is valid.

10.7 Problems

Exercises 10.1 SVI parametrization
Gatheral [89] presents the following "Stochastic Volatility Inspired" (SVI) parametrization of the implied volatility

$$\sigma_{\mathrm{BS}}(\tau, k)^2\tau = a + b\left(\rho(k-m) + \sqrt{(k-m)^2+\sigma^2}\right)$$

where the five parameters a, b, ρ, σ and m depend on the maturity τ. This parametrization agrees with the Lee moment formula, i.e., the large-strike behavior is at most linear in the moneyness k.

1. Prove that the no-arbitrage condition $\frac{\partial^2 \mathcal{C}(\tau,K)}{\partial K^2} > 0$ is preserved when $k >> 1$ if and only

$$|b(1+\rho)| \leq 2$$

Chapter 11

Analysis on Wiener Space with Applications

Abstract We review the main notions and results in the stochastic calculus of variations, commonly called Malliavin calculus and developed by Malliavin in 1976 in order to give a probabilistic proof of the "sums of squares" Hörmander theorem. In the last section, we focus on various applications: The convexity adjustment of CMS option, the probabilistic representation of sensitivities and the calibration of stochastic volatility models.

11.1 Introduction

Since the works of Fournié et al. [85, 86] on the probabilistic representation of sensitivities of derivatives products (i.e., Greeks), various lecture notes on Malliavin calculus have appeared. Let us cite [31, 33, 132, 52, 88]. From our perspective, all these lectures are highly technical and do not give a motivation to get familiar with such a method. Moreover, the mathematical proofs of the main theorems are hidden by various technical probability results that are often out of reach of the quantitative analysts.

In this chapter, we explain the main notions and sketch the *proofs* in the Malliavin calculus, highlighting the link with the *formal* (simpler) functional integration invented by the physicists in the fifties to find a Lagrangian formulation of a Quantum Field Theory (QFT). In particular, we connect the Malliavin derivative to the functional derivative, the Skorohod integral to the Wick product and the Wiener chaos decomposition to the Fock space. In spirit, our presentation follows [73].

We consider three distinct applications:

Firstly, we compute the convexity adjustment for CMS options in the case of short-rate affine models, mainly the Hull-White model. Our tool will be the Wick calculus and the Wiener chaos decomposition.

Secondly, we obtain a probabilistic representation of sensitivities of derivative products allowing an accurate computation via Monte-Carlo simulation.

Finally, we obtain a probabilistic representation of conditional expectations such as the instantaneous variance of a stock process conditional to its forward.

TABLE 11.1: A dictionary from Malliavin calculus to QFT.

Malliavin Derivative	Functional Derivative
Skorohod integration	Wick product
Ornstein-Uhlenbeck semigroup	Harmonic oscillator
Wiener chaos	Fock space

This gives the local volatility associated to a SVM. This representation paves the way for the calibration of a mixed local and stochastic volatility model.

11.2 Functional integration

11.2.1 Functional space

Let $\mathbb{W}$ be the *path space* of pointed continuous paths ω on the interval $[0,1]$

$$\mathbb{W} = \{\omega : [0,1] \to \mathbb{R} \ : \ \omega(0) = 0\}$$

In QFT, ω is a $(0+1)$-field. As usual, we endow this space with the Wiener measure $(\Omega, \mathcal{F}, \mathbb{P})$. For further use, we introduce the *Cameron-Martin space* $\mathcal{H}$:

$$\mathcal{H} = \{h \in L^2([0,1]) \ : \ \tilde{h}(\cdot) = \int_0^{\cdot} h(s) ds \in \mathbb{W}\}$$

The derivative of $\tilde{h}$, $\tilde{h}(s)' = h(s)$ (where the $'$ indicates a derivative according to s) exists ds-almost surely. We note below $< \cdot, \cdot >$ the scalar product on $L^2([0,1])$.
On this space, we define $\mathbb{R}$-valued $\mathcal{F}$-measurable functions $F(\cdot) : \mathbb{W} \to \mathbb{R}$. From the functional integration point of view, $F(\cdot)$ is called a (Wiener) functional.
We introduce integration and derivation operations on this space of functionals. These operations are first defined on some simple functions $F(\cdot)$ called *cylindrical functions* and then extended to a large class using an integration by parts formula.

11.2.2 Cylindrical functions

DEFINITION 11.1 Cylindrical functions *A functional $F : \mathbb{W} \to \mathbb{R}$ is called a cylindrical function if it has the form*

$$F = f\left(\omega(t_1), \cdots, \omega(t_n)\right)$$

where $0 < t_1 < \cdots < t_n \leq 1$ *and* f *is a smooth function on* $\mathbb{R}^n$ *such that all its derivatives have at most polynomial growth. The set of cylindrical functions is denoted* $\mathcal{C}_1$ *and forms an algebra.*

REMARK 11.1 Note that we have $\mathcal{C}_1 \subset L^p(\Omega)$ for all $p \geq 1$ as the polynomial growth of f ensures the existence of all the moments.
We can also define the algebra $\mathcal{C}_2$ of functionals of the form

$$F = f\left(W(h_1), \cdots, W(h_n)\right)$$

where $W(h_i) = \int_0^1 h_i(s) dW_s$ with $h_i \in L^2([0,1])$.
In particular, replacing each $W(h_i)$ by the Riemann sums

$$\sum_{j=1}^{N} h_{\frac{j-1}{N}} \left(\omega_{\frac{j}{N}} - \omega_{\frac{j-1}{N}}\right)$$

the resulting function belongs to $\mathcal{C}_1$. We have therefore that $\mathcal{C}_1$ is dense in $\mathcal{C}_2$.
□

In the following, we introduce the Wiener measure $\mathbb{P}$ using the *Feynman path integral.*

11.2.3 Feynman path integral

We endow the space $\mathbb{W}$ with the Wiener measure $d\gamma(\omega)$ (i.e., Feynman quantification procedure) which is characterized by its generating function [73]

$$\int_{\mathbb{W}} d\gamma(\omega) e^{-\imath < \omega', \omega >} = e^{-\frac{1}{2}\int_0^1 \int_0^1 d\omega'(t) d\omega'(t') \min(t,t')} \tag{11.1}$$

where ω' is an element of the topological dual[1] $\mathbb{W}'$ of $\mathbb{W}$, i.e., a bounded measure on the semi-open interval $(0,1)$ where

$$< \omega', \omega >= \int_0^1 d\omega'(t)\omega(t)$$

Formally, we would like to represent the measure $d\gamma(\omega)$ as

$$d\gamma(\omega) = \mathcal{D}(\omega) e^{-\frac{1}{2}\int_0^1 (\frac{d\omega}{dt})^2 dt} \tag{11.2}$$

This equation is formal as the term $\mathcal{D}$ cannot be defined as a measure on $\mathbb{W}$ and moreover the derivative according to the time $\frac{d\omega(t)}{dt}$ does not exist.

[1]The set of all continuous linear functionals, i.e., continuous linear maps from $\mathbb{W}$ into $\mathbb{R}$. A topology on the dual can be defined to be the coarsest topology such that the dual pairing $\mathbb{W}' \times \mathbb{W} \to \mathbb{R}$ is continuous.

This derivative could be put on a rigorous ground if we introduced the Hida distribution [132] but we do not go down this route as we prefer to use formal computations. We call the (formal) derivative $\frac{d\omega(t)}{dt}$ a *white noise*.
Although $\mathcal{D}$ is not a measure on $\mathbb{W}$, we try to give a meaning to (11.2) and the following expression, called a *functional integration*

$$\mathbb{E}^{\mathbb{P}}[F] \equiv \int \mathcal{D}(\omega)F(\omega)e^{-\frac{1}{2}\int_0^1\left(\frac{d\omega}{dt}\right)^2 dt} \tag{11.3}$$

Firstly, the "Riemann" integral is replaced by its discrete sum

$$\int_0^1 \left(\frac{d\omega}{dt}\right)^2 dt = \frac{1}{\Delta t}\sum_{i=1}^{n}\left(\omega(t_i) - \omega(t_{i-1})\right)^2$$

where the interval $[0,1]$ is subdivided into $0 < t_1 < \cdots < t_i < \cdots < t_n = 1$ with length Δt. Then, we view the functional integral (11.3) as the limit $n \to \infty$ of a (finite) n-dimensional Gaussian integration:

DEFINITION 11.2 Functional integration

$$\mathbb{E}^{\mathbb{P}}[F] \equiv \lim_{n\to\infty}\int_{\mathbb{R}^n}\left(\prod_{i=1}^{n}\frac{d\omega_i}{\sqrt{2\pi\Delta t}}\right)F(\omega_1,\cdots,\omega_n)\left(\prod_{i=1}^{n}e^{-\frac{1}{2\Delta t}(\omega_i-\omega_{i-1})^2}\right)$$

with $\omega_0 = 0$.

Using this definition, we will to reproduce our previous characterization (11.1) of the Wiener measure $d\gamma$. To start with, using an integration by parts, we write the characteristic function

$$I \equiv \int_{\mathbb{W}} d\gamma(\omega)e^{-\imath\int_0^1 \dot{\omega}'(t)\omega(t)dt}$$

as

$$I = \int_{\mathbb{W}} \mathcal{D}(\omega)e^{-\frac{1}{2}\int_0^1 dt\int_0^1 dt'\omega(t')K(t,t')\omega(t)}e^{-\imath\int_0^1\dot{\omega}'(t)\omega(t)dt}$$

with the kernel operator $K(t,t')$ defined as

$$K(t,t') = -\delta(t'-t)\frac{d^2}{dt^2}$$

Using the definition 11.2, the expression above becomes

$$\lim_{n\to\infty}\int_{\mathbb{R}^n}\left(\prod_{i=1}^{n}\frac{d\omega_i}{\sqrt{2\pi\Delta t}}\right)e^{-\frac{(\Delta t)^2}{2}\sum_{i,j=1}^n\omega_iK_{ij}\omega_i-\imath\sum_{i=1}^n(\omega_i'-\omega_{i-1}')\omega_i}$$
$$= \lim_{n\to\infty}\frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2(\Delta t)^2}\sum_{i,j=1}^n(\omega_i'-\omega_{i-1}')_iG_{ij}(\omega_j'-\omega_{j-1}')}$$

with G the inverse of the n-dimensional matrix K:

$$K_{ik}G_{kj} = \delta_{ij}$$

In the limit $n \to \infty$, we obtain

$$I = e^{-\frac{1}{2}\int_0^1 dt \int_0^1 dt' \dot{\omega}'(t') G(t',t)\dot{\omega}'(t)}$$

with $G(t, t')$, the so-called Green function, solution of

$$\int_0^1 dsK(t,s)G(s,t') = -\frac{d^2}{dt^2}G(t,t') = \delta(t-t')$$

The solution is given by $G(t, t') = \min(t, t')$ and we reproduce (11.1).
Via the definition 11.2, it is clear that the functional integration is linear

$$\mathbb{E}^{\mathbb{P}}[F + G] = \mathbb{E}^{\mathbb{P}}[F] + \mathbb{E}^{\mathbb{P}}[G]$$

We obtain as an exercise

$$\mathbb{E}^{\mathbb{P}}[\omega(t_1)\omega(t_2)] = \min(t_1, t_2)$$

This can be generalized for any product of ω. It is called the *Wick identity*

THEOREM 11.1 Wick identity

$$\mathbb{E}^{\mathbb{P}}[\omega(t_1)\omega(t_2)\cdots\omega(t_{2n})] = \min(t_1,t_2)\cdots\min(t_{2n-1},t_{2n}) + \text{cyclic perm.}$$
$$\mathbb{E}^{\mathbb{P}}[\omega(t_1)\omega(t_2)\cdots\omega(t_{2n-1})] = 0$$

Below, we note

$$(F, G) \equiv \int_{\mathbb{W}} F(\omega)G(\omega)d\gamma(\omega)$$

11.3 Functional-Malliavin derivative

Having defined how to integrate functionals on $\mathbb{W}$, we introduce a derivation, called in the mathematical literature a Malliavin derivative and in the physics literature a functional derivation. The derivative is first defined for cylindrical functions belonging to $\mathcal{C}_1$ (or equivalently $\mathcal{C}_2$) . Then using an integration by parts formula, we show how to extend its domain $\subset L^2(\Omega)$.

DEFINITION 11.3 Malliavin derivative 1 *For $F \in \mathcal{C}_1$, we define the Malliavin derivative D_hF along the direction $h \in \mathcal{H}$ by*

$$D_hF(\omega) = \frac{d}{d\epsilon}F\left(\omega + \varepsilon \int_0^{\cdot} h(s)ds\right)|_{\epsilon=0}$$

$$= \lim_{\varepsilon \to 0} \frac{f\left(\omega(t_1) + \varepsilon \int_0^{t_1} h(s)ds, \cdots\right) - f\left(\omega(t_1), \cdots\right)}{\varepsilon}$$

The limit is under $L^2(\Omega)$.

Formally, the Malliavin derivative corresponds to the bumping of the path $\omega(\cdot)$ by the function $\int_0^{\cdot} h(s)ds$. This definition can be simplified and we have

DEFINITION 11.4 Malliavin derivative 2 *For $F \in \mathcal{C}_1$, the Malliavin derivative D_hF along the direction $h \in \mathcal{H}$ is defined by*

$$D_hF(\omega) = \sum_{i=1}^{n} \partial_i f(\omega) \int_0^{t_i} h(s)ds \tag{11.4}$$

Below are the main properties of the Malliavin derivative derived from the formula above for all $F, G \in \mathcal{C}_1$:

1. Linearity:

$$D_h\left(F(\omega) + G(\omega)\right) = D_hF(\omega) + D_hG(\omega)$$

2. Leibnitz rule:

$$D_h\left(F(\omega)G(\omega)\right) = D_hF(\omega)G(\omega) + F(\omega)D_hG(\omega)$$

For a fixed ω, the map $h \to D_hF(\omega)$ is a linear bounded functional on $L^2([0,1])$ (see 11.4). By the Riesz representation theorem, there exists a unique element $DF(\omega) \in L^2([0,1])$ such that

$$D_hF(\omega) =< DF(\omega), h >\equiv \int_0^1 (DF(\omega))(s)h(s)ds$$

Below, we note $DF(\omega)(s) \equiv D_sF(\omega)$. D_tF can be considered as a $L^2([0,1])$-valued stochastic process.
The assumption that the functional f is a smooth function on $\mathbb{R}^n$ such that all its derivatives have at most polynomial growth ensures that for all $p \geq 1$, we have $DF \in L^p(\Omega, L^2([0,1]))$ (we take $p = 2$):

$$D : \mathcal{C}_1 \to L^2(\Omega, L^2([0,1])) \simeq L^2(\Omega \times [0,1])$$

From (11.4), we obtain

$$D_s F(\omega) = \sum_{i=1}^{n} \partial_i f(\omega) 1(s \in [0, t_i])$$

The Malliavin derivative can be formally obtained with the following definition

DEFINITION 11.5 ***Functional derivative***

$$D_s F(\omega) = \lim_{\varepsilon \to 0} \frac{F(\dot{\omega}(\cdot) + \varepsilon \delta(\cdot - s)) - F(\dot{\omega}(\cdot))}{\varepsilon} \quad \text{with } F \in \mathcal{C}_1$$

and

$$D_h F(\omega) = \int_0^1 h(s) D_s F(\omega) ds$$

where $h \in L^2([0,1])$*. In order to differentiate the use of the functional derivative from the Malliavin derivative, we note below the functional derivative as* $D_s \equiv \frac{\delta}{\delta \dot{\omega}(s)}$*.*

Applying this definition, we reproduce the result

$$\begin{aligned} \frac{\delta \omega(t)}{\delta \dot{\omega}(s)} &= \frac{\delta \int_0^t \dot{\omega}(u) du}{\delta \dot{\omega}(s)} = \int_0^t \delta(u - s) du \\ &= 1(s \in [0, t]) \end{aligned} \tag{11.5}$$

Since $\mathcal{C}_1$ is dense in $L^2(\Omega)$, we could extend the domain of the Malliavin derivative in $L^2(\Omega)$ by means of taking limits. For this purpose, we define the class $\mathbb{D}_{1,2}$.

DEFINITION 11.6 $\mathbb{D}_{1,2}$ *We define* $\mathbb{D}_{1,2}$ *to be the closure of the family* $\mathcal{C}_1$ *with respect to the norm* $|| \cdot ||_{1,2}$*:*

$$||F||_{1,2} = ||F||_{L^2(\Omega)} + ||D_t F||_{L^2([0,1] \times \Omega)}$$

It means that $\mathbb{D}_{1,2}$ *consists of all* $F \in L^2(\Omega)$ *such that there exists* $F_n \in \mathcal{C}_1$ *with the property that* $F_n \to F$ *in* $L^2(\Omega)$ *as* $n \to \infty$ *and* $\{DF_n\}_{n=1,\cdots,\infty}$ *is convergent in* $L^2([0,1] \times \Omega)$*.*

Then, we could extend the Malliavin derivative for $F \in L^2(\Omega)$ by

$$DF \equiv \lim_{n \to \infty} DF_n$$

However, there is no guarantee that if F_n and $\hat{F}_n$ are two sequences in $\mathcal{C}_1$ both converging to F under L^2, then DF_n and $D\hat{F}_n$ converge to the same limit.

When this property is satisfied, D is called a *closable* operator on $\mathbb{D}_{1,2}$ with *core* $\mathcal{C}_1$. In the following, having introduced the integration by parts formula on $\mathcal{C}_1$, we show that D is closable.

LEMMA 11.1 Integration by parts formula
Suppose $F, G \in \mathcal{C}_1$ and $h \in \mathcal{H}$. Then

$$(D_h F, G) = (F, D_h^* G) \tag{11.6}$$

with

$$D_h^* = -D_h + \int_0^1 h(s) d\omega(s)$$

PROOF By the means of the Girsanov theorem, we have

$$\mathbb{E}^{\mathbb{P}}[F(\omega(\cdot) + \epsilon \int_0^{\cdot} h(s)ds)G(\omega)] = \mathbb{E}^{\mathbb{P}}[e^{-\frac{1}{2}\epsilon^2 \int_0^1 h(s)^2 ds + \epsilon \int_0^1 h(s)d\omega(s)} \\ F(\omega(\cdot))\, G\left(\omega(\cdot) - \epsilon \int_0^{\cdot} h(s)ds\right)]$$

Note that the Girsanov transform is well defined as $\int_0^1 h(s)^2 ds < \infty$ and the Novikov condition is satisfied. Differentiating both sides with respect to ϵ at $\epsilon = 0$, we obtain

$$\mathbb{E}^{\mathbb{P}}[D_h F(\omega) G(\omega)] = \mathbb{E}[\int_0^1 h(s)d\omega(s) F(\omega)G(\omega)] - \mathbb{E}[F D_h G]$$

We give another proof using the formal calculus provided by the Feynman path integral and the functional derivative. The proof is based on the following identity

$$\int \mathcal{D}\omega \frac{\delta}{\delta \dot{\omega}(s)} F(\omega) = 0 \tag{11.7}$$

Then, we have

$$\begin{aligned}
\left(\frac{\delta F}{\delta \dot{\omega}(s)}, G\right) &= \int \mathcal{D}\omega \left(\frac{\delta}{\delta \dot{\omega}(s)}[F(\omega)G(\omega)e^{-\frac{1}{2}\int_0^1 \dot{\omega}(t)^2 dt}] \right. \\
&\quad \left. + \left(-F(\omega)\frac{\delta G(\omega)}{\delta \dot{\omega}(s)} + \dot{\omega}(s)F(\omega)G(\omega)\right) e^{-\frac{1}{2}\int_0^1 \dot{\omega}(t)^2 dt}\right) \\
&= \int \mathcal{D}\omega \left(-F(\omega)\frac{\delta G(\omega)}{\delta \dot{\omega}(s)} + \dot{\omega}(s)F(\omega)G(\omega)\right) e^{-\frac{1}{2}\int_0^1 \dot{\omega}(t)^2 dt}
\end{aligned}$$

where we have used

$$\frac{\delta}{\delta \dot{\omega}(s)} e^{-\frac{1}{2}\int_0^1 \dot{\omega}(t)^2 dt} = -\dot{\omega}(s) e^{-\frac{1}{2}\int_0^1 \dot{\omega}(t)^2 dt}$$

For the smooth derivative, $D_h(F) \equiv \int_0^1 h(s)\frac{\delta F(\omega)}{\delta \dot{\omega}(s)}\,ds$, we obtain

$$(D_h F, G) = (F, A_h G) - (F, D_h G)$$

with

$$A_h = \int_0^1 h(s)\dot{\omega}(s)ds = \int_0^1 h(s)d\omega(s)$$

∎

THEOREM 11.2
D is closable.

PROOF Considering the difference $H_n = F_n - \hat{F}_n$, we see that D is closable if and only if $H_n \to 0$ under $L^2(\Omega)$, then $DH_n \to 0$ under $L^2(\Omega \times [0,1])$. By the lemma 11.1, we get

$$(D_h H_n, G) = (H_n, D_h^* G) \to_{n\to\infty} 0$$

with $G \in \mathcal{C}$. Since $D_h H_n$ converges by definition and $\mathcal{C}$ is dense in $L^2(\Omega)$, we conclude that $D_h H_n \to_{n\to\infty} 0$ in $L^2(\Omega \times [0,1])$. ∎

Example 11.1

$$D_s \int_0^T f(u)d\omega(u) = f(s)1(s \in [0,T])$$
$$D_s e^{\omega(t_0)} = e^{\omega(t_0)}1(s \in [0,t_0])$$

∎

11.4 Skorohod integral and Wick product

11.4.1 Skorohod integral

The Malliavin derivative is a closed and unbounded operator valued in $L^2(\Omega \times [0,1])$ and defined on a dense subset $\mathbb{D}_{1,2}$ of $L^2(\Omega)$:

$$D : \mathbb{D}_{1,2} \subset L^2(\Omega) \to L^2(\Omega \times [0,1])$$

One can then define an adjoint operator

$$\partial : \mathrm{Dom}(\partial) \subset L^2(\Omega \times [0,1]) \to L^2(\Omega)$$

The domain of ∂, denoted $\mathrm{Dom}(\partial)$, is the set of $L^2([0,1])$-valued square integrable r.v. $u \in L^2(\Omega \times [0,1])$ for which there exists an unique $u^* \in L^2(\Omega)$ such that

$$\mathbb{E}^{\mathbb{P}}[< DF, u >] = \mathbb{E}^{\mathbb{P}}[F u^*]$$

As ∂ is densely defined, we set $u^* = \partial(u)$ and we have

$$\mathbb{E}^{\mathbb{P}}[< DF, u >] = \mathbb{E}^{\mathbb{P}}[F \partial(u)] \tag{11.8}$$

PROPOSITION 11.1
For $F \in \mathcal{C}_1$ and $h \in L^2([0,1])$, the following formula holds

$$\partial(Fh) = F \int_0^1 h(s) d\omega(s) - D_h F$$

PROOF For $F, G \in \mathcal{C}_1$ and $h \in L^2([0,1])$, we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[\partial(Fh)G] &= \mathbb{E}^{\mathbb{P}}[< Fh, DG >] \text{ using (11.8)} \\
&= \mathbb{E}^{\mathbb{P}}[F < h, DG >] \equiv (F, D_h G) \\
&= \mathbb{E}^{\mathbb{P}}[-G D_h F] + \mathbb{E}^{\mathbb{P}}[FG \int_0^1 h(s) d\omega(s)] \text{ using (11.6)}
\end{aligned}$$

□

REMARK 11.2 Skorohod reduces to an Itô for adapted process
If $u(s,\omega)$ is a $\mathcal{F}_s$-adapted process, then the Skorohod integral reduces to an Itô integral:

$$\partial(u) = \int_0^T u(\omega, s) d\omega(s)$$

□

LEMMA 11.2
For $F \in \mathbb{D}_{1,2}$, $u \in \mathrm{Dom}(\partial)$ such that $Fu \in L^2(\Omega \times [0,1])$. Then F belongs to the domain of ∂ and the following equality is true

$$\partial(Fu) = F\partial(u) - < DF, u > \tag{11.9}$$

PROOF

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[< Fu, DG >] &= \mathbb{E}^{\mathbb{P}}[< u, FDG >] = \mathbb{E}^{\mathbb{P}}[< u, D(FG) - GDF >] \\
&= \mathbb{E}^{\mathbb{P}}[\partial(u) FG] - \mathbb{E}^{\mathbb{P}}[< DF, u > G]
\end{aligned}$$

□

In the following, we give a second definition of the Skorohod integral using the so-called *Wick calculus*. The Wick product (also called normal ordering) was introduced as a first step to renormalize a QFT, in particular to eliminate the infinite energy of the vacuum. The Wick calculus will lead us to the Wiener chaos expansion.

11.4.2 Wick product

This product can be defined using the *Hermite polynomials* $\{H_n(x)\}_{n\in\mathbb{N}}$ of order n

$$\begin{aligned} H_0(x) &= 1 \\ H_n(x) &= \frac{(-)^n}{n!} e^{-\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}} \;,\; n \geq 1 \end{aligned}$$

These polynomials are the coefficients of the expansion in powers of t of the function

$$F(t,x) = e^{tx-\frac{t^2}{2}} = \sum_{n=0}^{\infty} t^n H_n(x)$$

We scale these functions by

$$He_n(x) = 2^{-\frac{n}{2}} H_n\left(\frac{x}{\sqrt{2}}\right)$$

In particular, we have

$$\begin{aligned} He_1(q) &= q \\ He_2(q) &= q^2 - 1 \\ He_3(q) &= q^3 - 3q \\ He_4(q) &= q^4 - 6q^2 + 3 \end{aligned}$$

DEFINITION 11.7 Wick product 1 *The Wick product, denoted* $: \cdot :$, *is defined by*

$$: W(h)^n := He_n(W(h))$$

with $n \in \mathbb{N}$, $||h||_{L^2([0,1])} = 1$ *and* $W(h) \equiv \int_0^1 h(s) d\omega(s)$.

REMARK 11.3 Summing this expression over n, we obtain that $: e^{W(h)} :$ is an exponential martingale

$$: e^{W(h)} := e^{W(h)-\frac{1}{2}}$$

□

Example 11.2

$$
\begin{aligned}
: W(h) : &= W(h) \\
: W(h)^2 : &= W(h)^2 - 1 \\
: W(h)^3 : &= W(h)^3 - 3W(h)
\end{aligned}
\tag{11.10}
$$

□

An equivalent (useful although formal) definition for the Wick product is directly given on the white noise $\dot{\omega}(t)$:

DEFINITION 11.8 Wick product 2

$$
\begin{aligned}
: \dot{\omega}(t_1) \cdots \dot{\omega}(t_n) : \quad &= \quad \dot{\omega}(t_1) \cdots \dot{\omega}(t_n) \\
&- \quad (\dot{\omega}(t_3) \cdots \dot{\omega}(t_n) \delta(t_1 - t_2) + \text{perm.}) \\
&+ \quad (\delta(t_1 - t_2)\delta(t_3 - t_4)\dot{\omega}(t_5) \cdots \dot{\omega}(t_n) + \text{perm.}) \\
&- \quad \cdots \\
&(-)^k \, (\delta(t_1 - t_2)\delta(t_3 - t_4) \cdots \delta(t_{2k-1} - t_{2k}) + \text{perm.}) \\
&+ \quad \cdots
\end{aligned}
$$

For example, we have

$$
: \dot{\omega}(t)\dot{\omega}(s)\dot{\omega}(u) : = \omega(t)\omega(s)\omega(u) - (\delta(t-s)\omega(u) + \text{perm.})
$$

We prove the equivalence of these two definitions with a simple example (11.10):

$$
\begin{aligned}
: W(h)^2 : &= \int_0^1 ds \int_0^1 du h(s)h(u) : \dot{\omega}(s)\dot{\omega}(u) : \\
&= \int_0^1 ds \int_0^1 du h(s)h(u)\dot{\omega}(s)\dot{\omega}(u) - \int_0^1 ds \int_0^1 du h(s)h(u)\delta(s-u) \\
&= \left(\int_0^1 h(s)\dot{\omega}(s) ds \right)^2 - \int_0^1 h(s)^2 ds
\end{aligned}
$$

Here is a list of a few useful properties:

1. Orthogonality :

$$
\mathbb{E}[: W(f)^n :: W(g)^m :] = \delta_{nm} \int_0^1 f(s)g(s) ds
$$

2. Commutativity of the Wick product and Malliavin derivative:

$$D_s : F(\omega) :=: D_s F(\omega) :$$

3. Commutativity law: $: FG :=: GF :$
4. Associativity law: $:: FG : H :=: F : GH ::$
5. Distributive law: $: H(Y+Z) :=: HY : + : HZ :$
6. $: u(t,\omega)\dot{\omega}(t) := \left(\dot{\omega}(t)u(t,\omega) - \frac{\delta u(t,\omega)}{\delta \dot{\omega}(t)}\right)$

Here are some examples:

Example 11.3

1. $: e^{W_t} := e^{W_t - \frac{t}{2}}$. In particular, $: e^{W_t} :$ is a martingale.
2. $: W(f)W(g) := W(f)W(g) - \int_0^1 f(s)g(s)ds$
3. $: W_{t_1} W_{t_0} := W_{t_1} W_{t_0} - \min(t_1, t_0)$

$\square$

Everything is now in place to give a second definition of the Skorohod integral.

DEFINITION 11.9 Skorohod integral 2

$$\partial(u) = \int_0^1 : u(s,\omega)\dot{\omega}(s) : ds$$

The integral should be understood in the "Riemann sense." We prove below that this definition satisfies the duality pairing (11.8):

PROOF

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}[< DF, u >] = \mathbb{E}^{\mathbb{P}}[\int_0^1 D_t F u(t,\omega) dt] \\
&= \int \mathcal{D}\omega e^{-\frac{1}{2}\int_0^1 (\dot{\omega})^2 ds} \int_0^1 \frac{\delta F}{\delta \dot{\omega}(t)} u(t,\omega) dt \\
&= \int \mathcal{D}\omega e^{-\frac{1}{2}\int_0^1 (\dot{\omega})^2 ds} F \int_0^1 \left(\dot{\omega}(t)u(t,\omega) - \frac{\delta u(t,\omega)}{\delta \dot{\omega}(t)}\right) dt \ \text{ using an IPP} \\
&= \int \mathcal{D}\omega e^{-\frac{1}{2}\int_0^1 (\dot{\omega})^2 dt} F \int_0^1 : u(t,\omega)\dot{\omega}(t) : dt \ \text{ using the property above (6)} \\
&= \mathbb{E}^{\mathbb{P}}[F \partial u]
\end{aligned}$$

□

Here are two examples of computation of a Skorohod integral using the Wick calculus:

Example 11.4

1.

$$\partial\left(\omega(t)(\omega(T)-\omega(t))\right) = \int_0^T :\omega(t)(\omega(T)-\omega(t))\dot{\omega}(t): dt$$
$$= \frac{1}{6} :\omega(T)^3 := \frac{1}{6}(\omega(T)^3 - 3T\omega(T))$$

2. Let $t_0 < T$.

$$\begin{aligned}\partial(\omega(t_0)^2) &= \int_0^T :\omega(t_0)^2\dot{\omega}(t): dt \\ &= :\omega(t_0)^2\omega(T): \\ &= ::\omega(t_0)^2:\omega(T): + t_0\omega(T) \\ &= :\omega(t_0)^2:\omega(T) - 2t_0\omega(T) + t_0\omega(T) \\ &= \omega(t_0)^2\omega(T) - 2t_0\omega(T)\end{aligned}$$

□

11.5 Fock space and Wiener chaos expansion

DEFINITION 11.10 Fock space *A Fock structure is defined by a quadruple*

$$\{\mathcal{H}, a(h_i), a^\dagger(h_i), |0>\}$$

with h_i an orthonormal basis of $L^2([0,1])$. $\mathcal{H}$ is a separable Hilbert space, $|0>\in\mathcal{H}$ a unit vector called the (Fock) vacuum and $a(h)$, $a^\dagger(h)$ are operators defined on a domain Dom *of vectors dense in $\mathcal{H}$. Furthermore,* Dom *is a linear set containing $|0>$ and the vectors $a(h)$ carry vectors in* Dom *into vectors in* Dom. *If $\Phi, \Omega \in$ Dom, then $(\Phi, a(h)^\dagger, \Omega)$ is a tempered distribution regarded as a functional of h. The Fock space $\mathcal{F}(\mathcal{H})_n$ is the Hilbert space generated by the vectors in* Dom

$$a^\dagger(h_1)\cdots a^\dagger(h_n)|0>$$

and we set

$$\mathcal{F}(\mathcal{H}) = \oplus_{n=1}^{\infty} \mathcal{F}(\mathcal{H})_n$$

DEFINITION 11.11 cyclic vector $|0>$ *is called a cyclic vector if*

$$\mathcal{H} \simeq \mathcal{F}(\mathcal{H})$$

Fock space realization on $L^2(\mathbb{W}, \mathbb{P})$

In the following, we show how to define a Fock space on the Hilbert space $\mathcal{H} = L^2(\mathbb{W}, \mathbb{P})$. We recall that the Malliavin derivative D_h is

$$D_h = \int_0^1 h(s) ds \frac{\delta}{\delta \dot{\omega}(s)}$$

and we set

$$A_h = \int_0^1 h(s) d\omega(s)$$

These operators acting on $\mathbb{D}_{1,2}$ satisfy the Heisenberg Lie algebra

$$\begin{aligned} [D_{h_1}, D_{h_2}] &= 0 \\ [A_{h_1}, A_{h_2}] &= 0 \\ [D_{h_1}, A_{h_2}] &= < h_1, h_2 > 1 \end{aligned}$$

with the commutator $[A, B] \equiv AB - BA$. Setting

$$\begin{aligned} a(h) &= D_h \\ a^\dagger(h) &= A_h - D_h \end{aligned}$$

the "annihilation-creation" algebra is given by

$$\begin{aligned} [a(h_1), a(h_2)] &= 0 \\ [a^\dagger(h_1), a^\dagger(h_2)] &= 0 \\ [a(h_1), a^\dagger(h_2)] &= < h_1, h_2 > 1 \end{aligned} \tag{11.11}$$

$a(h)$ and $a^\dagger(h)$ are adjoint in the Hilbert space $L^2(\mathbb{W}, \mathbb{P})$. Indeed from the integration by parts formula (11.6), we have

$$(F, a^\dagger(h) G) = (a(h) F, G)$$

The vacuum $|0 >\in \mathcal{F}$ is the constant function equal to 1. In particular, $a(h) 1 = 0$.

$(L^2(\mathbb{W},\mathbb{P}), a(h_i), a^\dagger(h_i), 1)$ defines a Fock structure. Moreover, one can prove that $|0>=1$ is a cyclic vector: It gives the Itô-Wiener-Segal chaos decomposition

THEOREM 11.3 Wiener chaos

$$L^2(\mathbb{W},\mathbb{P}) \simeq \mathcal{F}(L^2(\mathbb{W},\mathbb{P}))$$

In particular, $\mathcal{F}(\mathcal{H})_n$ is generated by $\prod_{i=1}^n a(h_i)^\dagger|0>$ (iterated integrals):

- $\mathcal{F}(\mathcal{H})_0$ is generated by 1.
- $\mathcal{F}(\mathcal{H})_1$ is generated by $\int_0^1 h_1(s)d\omega(s)$.
- $\mathcal{F}(\mathcal{H})_2$ is generated by $\int_{0\leq s\leq t\leq 1} h_1(s)h_2(t)d\omega(s)d\omega(t) - < h_1, h_2 >$.

PROOF In this part, only a sketch of the proof is provided. Let $\mathcal{J}$ denote the set of indices $I = \{n_i\}$ such that $n_i \in \mathbb{N}^*$ and almost all of them are equal to zero. Denote $|I| = n_1 + n_2 + \cdots$. For each $I \in \mathcal{J}$, define

$$H_I = \prod_{i=1} a^\dagger(h_i)^{n_i} 1$$

where $a^\dagger(h_i)^{n_i}$ means the product of $a^\dagger(h_i)$ n_i times. Following the definition of $a^\dagger(h_i)$, each H_I belongs to $\mathcal{C}_2 \subset L^2(\mathbb{W},\mathbb{P})$. The fact that the Hermite polynomials form an orthonormal basis of $L^2([0,1])$ implies that $\{H_I\}_{I\in\mathcal{J}}$ gives a basis of the algebra $\mathcal{C}_2$. Using that $\mathcal{C}_2$ is dense in $L^2(\mathbb{P},\mathbb{W})$, the result follows immediately. ∎

11.5.1 Ornstein-Uhlenbeck operator

We introduce the (quantum) number operator N

$$N = \sum_{i=0}^{\infty} a^\dagger(h_i)a(h_i)$$

As N is a self-adjoint operator on $L^2(\mathbb{W},\mathbb{P})$, it can be diagonalized over an orthogonal basis. We next show that H_I is an eigenvector of N with eigenvalue $|I|$. This can be achieved by recurrence. For example,

$$\begin{aligned} Na(h_i)^\dagger|0> &= a^\dagger(h_i)a(h_i)a(h_i)^\dagger|0> \\ &= a^\dagger(h_i)(1 + a(h_i)^\dagger a(h_i))|0> \text{ using (11.11)} \\ &= a^\dagger(h_i)|0> \quad \text{using } a(h_i)|0>=0 \end{aligned}$$

Note that N can be represented on the Hilbert space $L^2(\mathbb{W}, \mathbb{P})$ as

$$N = -\partial D$$

Indeed it is easy to check that the functions $H_I(\omega)$ are precisely the eigenvectors of N.

11.6 Applications

We illustrate the use of Malliavin calculus in finance with three applications: Firstly, we apply the Wiener chaos expansion by computing convexity adjustment for CMS rate (see section 3.4.4 for definitions). We focus on affine short-rate models, mainly the HW2 model. This part follows from [57]. Secondly, we obtain a probabilistic representation of Greeks and finally we compute the local volatility associated to SVMs.

11.6.1 Convexity adjustment

The fair value of a CMS in the forward measure $\mathbb{P}^{T_\alpha}$ is

$$\mathrm{CMS}_{\alpha\beta}(t) = P_{tT_\alpha} \mathbb{E}^{\mathbb{P}^{T_\alpha}}[s_{\alpha\beta,T_\alpha}|\mathcal{F}_t]$$

By definition of the swap rate, the expression above becomes

$$\mathrm{CMS}_{\alpha\beta}(t) = P_{tT_\alpha} \mathbb{E}^{\mathbb{P}^{T_\alpha}}[\frac{P_{T_\alpha T_\alpha} - P_{T_\alpha T_\beta}}{\sum_{i=\alpha+1}^{\beta} \tau_i P_{T_\alpha T_i}}|\mathcal{F}_t] \tag{11.12}$$

For an n-factor affine model, the forward bond P_{tT_i} satisfies in the risk-neutral measure

$$\frac{dP_{tT_i}}{P_{tT_i}} = r_t dt + \sum_{k=1}^{n} \sigma_k(t, T_i) dW_t^k$$

with $\sigma(t, T_i)$ function of the time t and the maturity T_i and $dW_t^k dW_t^p = \rho_{kp} dt$. In the forward measure $\mathbb{P}^{T_\alpha}$, we have

$$\frac{dP_{tT_i}}{P_{tT_i}} = (r_t - \sigma(t, T_\alpha).\sigma(t, T_i))dt + \sum_{k=1}^{n} \sigma_k(t, T_i) dZ_t^k$$

where we have set $\sigma(t, T_\alpha).\sigma(t, T_i) \equiv \sum_{k,p=1}^{n} \rho_{kp} \sigma_k(t, T_\alpha) \sigma_p(t, T_i)$. The solution is

$$P_{tT_i} = P_{0T_i} e^{\int_0^t r_s ds} e^{-\int_0^t \sigma(s,T_\alpha).\sigma(s,T_i) ds} : e^{\int_0^t \sum_{k=1}^{n} \sigma_k(s,T_i) dZ_s^k} : \tag{11.13}$$

where we have used the Wick product to represent the exponential martingale term. By plugging the solution (11.13) into (11.12), we obtain

$$\mathrm{CMS}_{\alpha\beta}(t) = P_{tT_\alpha}\mathbb{E}^{\mathbb{P}^{T_\alpha}}[\frac{\phi_\alpha : e^{\Phi_\alpha} :}{\sum_{i=\alpha+1}^{\beta}\tau_i\phi_i : e^{\Phi_i} :} - \frac{\phi_\beta : e^{\Phi_\beta} :}{\sum_{i=\alpha+1}^{\beta}\tau_i\phi_i : e^{\Phi_i} :}|\mathcal{F}_t] \tag{11.14}$$

where $\Phi_i \equiv \int_0^{T_\alpha}\sum_{k=1}^n \sigma_k(s,T_i)dZ_s^k$ and $\phi_i \equiv P_{0T_i}e^{-\int_0^{T_\alpha}\sigma(s,T_\alpha).\sigma(s,T_i)ds}$. We consider the first term; the second term can be treated similarly.
The exponential martingale can be decomposed as

$$: e^{\Phi_i} := 1 + : \Phi_i : + : \Phi_i^2 : + \cdots$$

where $\cdots$ indicates higher-order volatility terms. This corresponds to a Wiener chaos expansion. The first term in (11.14) can therefore be written at this order as

$$\phi_\alpha \frac{1 + : \Phi_\alpha : + : \Phi_\alpha^2 :}{N}\left(1 - \frac{1}{N}\sum_{i=\alpha+1}^{\beta}\tau_i\phi_i : \Phi_i : - \frac{1}{N}\sum_{i=\alpha+1}^{\beta}\tau_i\phi_i : \Phi_i^2 : \right.$$
$$\left. + \frac{1}{N^2}\sum_{i=\alpha+1}^{\beta}\tau_i\tau_j\phi_i\phi_j : \Phi_i\Phi_j : \right)$$

where $N = \sum_{i=\alpha+1}^{\beta}\tau_i\phi_i$. Using that

$$\mathbb{E}^{P^{T_\alpha}}[: \Phi_i^n :] = 0 \ , \ \forall\, n \in \mathbb{N}$$
$$\mathbb{E}^{P^{T_\alpha}}[: \Phi_i :: \Phi_j :] = c_{ij} \equiv \int_0^{T_\alpha}\sigma(s,T_i).\sigma(s,T_j)ds$$

we obtain finally the convexity adjustment for the CMS rate

$$P_{tT_\alpha}^{-1}\mathrm{CMS}_{\alpha\beta}(t) = \frac{(\phi_\alpha - \phi_\beta)}{N} - \frac{1}{N^2}\sum_{i=\alpha+1}^{\beta}\tau_i\phi_i(\phi_\alpha c_{i\alpha} - \phi_\beta c_{i\beta})$$
$$+ \frac{(\phi_\alpha - \phi_\beta)}{N^3}\sum_{i,j=\alpha+1}^{\beta}\tau_i\tau_j\phi_i\phi_j c_{ij}$$

In the 2-factor affine short-rate model, HW2 model, we have $\sigma(s,T_i) = \sigma(t)\left(\frac{e^{-a(T_i-t)}-1}{a}, \frac{e^{-b(T_i-t)}-1}{b}\right)$. Therefore the correlation matrix is

$$c_{ij} = \int_0^{T_\alpha}\sigma(s)^2\, ds\left(\frac{e^{-a(T_i-t)}-1}{a}\frac{e^{-a(T_j-t)}-1}{a} + \frac{e^{-b(T_i-t)}-1}{b}\frac{e^{-b(T_j-t)}-1}{b}\right.$$
$$\left. + \rho\frac{e^{-a(T_i-t)}-1}{a}\frac{e^{-b(T_j-t)}-1}{b} + \rho\frac{e^{-b(T_i-t)}-1}{b}\frac{e^{-a(T_j-t)}-1}{a}\right)$$

11.6.2 Sensitivities

Suppose that our market model depends on n parameters $\{\lambda_i\}_{i=1,\cdots,n}$ such as the volatility, the spot price of each asset, the interest yield curve, the correlation matrix between assets, $\cdots$. The fair value $\mathcal{C}$ of an option will depend on these parameters: $\mathcal{C} = \mathcal{C}(\lambda)$. For hedging purpose, we need to compute the sensitivities (the so-called *Greeks*) of the option with respect to the model parameters:

$$\frac{\partial^n \mathcal{C}}{\partial \lambda_{i_1} \cdots \lambda_{i_n}}$$

For example, the *Delta*, denoted $\{\Delta_i\}_{i=1,\cdots,n}$, is the sensitivity with respect to spot prices $\{S_0^i\}_{i=1,\cdots,n}$:

$$\Delta_i = \frac{\partial \mathcal{C}}{\partial S_0^i}$$

and the (Black-Scholes) *Vega*, denoted $\mathcal{V}_i$, is the sensitivity with respect to the Black-Scholes volatility σ^i of asset i:

$$\mathcal{V}_i = \frac{\partial \mathcal{C}}{\partial \sigma^i}$$

We restrict ourselves to first-order sensitivities although a full treatment can be obtained straightforwardly using the techniques presented below. We consider also payoffs $\phi = \phi(S_{t_1}, \cdots, S_{t_n})$ which are functions of a single underlying S_t at date $t_1, \cdots, t_n$. We recall that the fair value $\mathcal{C}$ at time $t < t_1$ is given by the mean value of the payoff conditional to the filtration $\mathcal{F}_t$:

$$\mathcal{C}(t, \lambda) = \mathbb{E}^{\mathbb{P}}[\phi(S_{t_1}, \cdots, S_{t_n})|\mathcal{F}_t]$$

No discount factor has been included as we assume zero interest rate for the sake of simplicity.

When the number of underlying assets is large, the option $\hat{C}(t, \lambda)$ is estimated using a Monte-Carlo simulation (MC). The simplest method to compute sensitivities with respect to λ_i is based on a finite difference scheme:

$$\partial_{\lambda_i} \mathcal{C}(t, \lambda) = \lim_{\epsilon \to 0} \frac{\hat{C}(t, \lambda_i + \epsilon) - \hat{C}(t, \lambda_i - \epsilon)}{2\epsilon}$$

In practice $\hat{C}(t, \lambda + \epsilon)$ and $\hat{C}(t, \lambda - \epsilon)$ are computed using the same normal random variables. Note that when ϵ is small, the more accurate the difference scheme is, the more the variance of the MC increases.

In the following, we explain how to obtain probabilistic representation of Greeks as

$$\partial_{\lambda_i} \mathcal{C}(t, \lambda) = \mathbb{E}^{\mathbb{P}}[\phi\omega|\mathcal{F}_t]$$

where the *universal* weight ω is a r.v. The weight is universal as it does not depend on the payoff $\phi(\cdot)$.

Tangent process

We assume that the process S_t satisfies the one-dimensional SDE

$$dS_t = b(t, S_t)dt + \sigma(t, S_t)dW_t \tag{11.15}$$

DEFINITION 11.12 Tangent process *The tangent process* $Y_t \equiv \frac{\partial S_t}{\partial S_0}$ *satisfies the SDE*

$$\frac{dY_t}{Y_t} = \frac{\partial b(t, S_t)}{\partial S}dt + \frac{\partial \sigma(t, S_t)}{\partial S}dW_t \tag{11.16}$$

with the initial condition $Y_0 = 1$.

Example 11.5 Black-Scholes
For the Black-Scholes log-normal process, the tangent process is $Y_t = \frac{S_t}{S_0}$. □

The next theorem gives the Malliavin derivative of S_t as a function of the tangent process:

THEOREM 11.4 Malliavin derivative of S_t

$$D_s S_t = \sigma(s, S_s)\frac{Y_t}{Y_s}1(s \in [0, t]) \tag{11.17}$$

PROOF The SDE (11.15) reads as

$$S_t = S_0 + \int_0^t b(u, S_u)du + \int_0^t \sigma(u, S_u)dW_u$$

We obtain that (recall that formally $D_s\dot{\omega}(u) = \delta(u - s)$)

$$D_s S_t = \sigma(s, S_s) + \int_s^t D_s b(u, S_u)du + \int_s^t D_s \sigma(u, S_u)dW_u$$

Using the chain rule (11.4), we have

$$D_s S_t = \sigma(s, S_s) + \int_s^t b'(u, S_u)D_s S_u du + \int_s^t \sigma'(u, S_u)D_s S_u dW_u$$

where the prime means a derivative with respect to S. Then, the process $Z_t = D_s S_t$ satisfies the SDE

$$\frac{dZ_t}{Z_t} = b'(t, S_t)dt + \sigma'(t, S_t)dW_t$$

with the initial condition $Z_s = \sigma(s, S_s)$. Note that this SDE is identical (modulo the initial condition) to (11.16). By scaling $Z_t = \lambda Y_t 1(s \in [0,t])$, we obtain (11.17). $\square$

Now, everything is in place to compute Greeks: we focus on the Delta and the (local) Vega.

Delta

Intervening the derivative ∂_{S_0} and the mean value $\mathbb{E}^{\mathbb{P}}[\cdot|\mathcal{F}_t]$ operators, we obtain

$$\Delta = \partial_{S_0}\mathbb{E}^{\mathbb{P}}[\phi|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\sum_{i=1}^{n} \partial_i \phi(S_{t_1}, \cdots, S_{t_n}) Y_{t_i} |\mathcal{F}_t] \tag{11.18}$$

We would like to represent this sensitivity price as

$$\Delta = \mathbb{E}[\phi\omega|\mathcal{F}_t]$$

As an ansatz, we take $\omega = \partial(\pi)$ with ∂ the Skorohod operator. Using the duality formula (11.8), we obtain

$$\mathbb{E}^{\mathbb{P}}[\phi\partial(\pi)|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\int_0^T D_s\phi\pi_s ds|\mathcal{F}_t]$$

Using the chain rule (11.4) and the formula (11.17), we have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[\phi\partial(\pi)|\mathcal{F}_t] &= \sum_{i=1}^{n} \mathbb{E}^{\mathbb{P}}[\partial_i\phi \int_0^T D_s S_{t_i} \pi_s ds|\mathcal{F}_t] \\ &= \sum_{i=1}^{n} \mathbb{E}^{\mathbb{P}}[\partial_i\phi \int_0^{t_i} \frac{Y_{t_i}}{Y_s}\sigma(s, S_s)\pi_s ds|\mathcal{F}_t] \end{aligned} \tag{11.19}$$

By identifying the equation (11.18) with the equation (11.19) for any smooth payoff, we have

$$\sum_{i=1}^{n} \mathbb{E}^{\mathbb{P}}[Y_{t_i}\left(1 - \int_0^{t_i} \frac{\sigma(s, S_s)}{Y_s}\pi_s ds\right)|S_{t_1}, \cdots, S_{t_n}] = 0$$

By setting $\pi_s = \frac{Y_s u_s}{\sigma(s,S_s)}$, this reduces to the constraint

$$\int_0^{t_i} \mathbb{E}^{\mathbb{P}}[u_s|S_{t_1}, \cdots, S_{t_n}]ds = 1 \ , \ \forall\, i = 1, \cdots, n \tag{11.20}$$

and the Delta is given by

$$\Delta = \mathbb{E}^{\mathbb{P}}[\phi\partial\left(\frac{Y_s u_s}{\sigma(s, S_s)}\right)|\mathcal{F}_t]$$

The simplest solution to (11.20) is

$$u_s = \sum_{k=1}^{n} a_k 1(s \in [t_{k-1}, t_k)) \ , \ \sum_{k=1}^{i} a_k = \frac{1}{t_i}$$

and we obtain

$$\Delta = \sum_{k=1}^{n} a_k \mathbb{E}^{\mathbb{P}}[\phi \int_{t_{k-1}}^{t_k} \frac{Y_s}{\sigma(s, S_s)} dW_s | \mathcal{F}_t]$$

where we have used that $\frac{Y_s}{\sigma(s,S_s)}$ is an adapted process and therefore the Skorohod integral reduces to an Itô integral (see remark 11.2).

Example 11.6 Black-Scholes

$$\Delta = \frac{1}{S_0 \sigma} \sum_{k=1}^{n} a_k \mathbb{E}^{\mathbb{P}}[\phi \left(W_{t_k} - W_{t_{k-1}} \right) | \mathcal{F}_t]$$

□

Note that the weight $\omega = \partial \pi$ which satisfies the condition (11.20) is not unique. Among all the weights such that (11.20) holds, the one which yields the minimum variance is given by the following proposition:

PROPOSITION 11.2
The weight with minimal variance denoted by ω_0 is the conditional expectation of any weight with respect to the r.v.s $(S_{t_1}, \cdots, S_{t_n})$

$$\omega_0 = \mathbb{E}^{\mathbb{P}}[\omega | S_{t_1}, \cdots, S_{t_n}]$$

PROOF The variance is

$$\sigma^2 = \mathbb{E}^{\mathbb{P}}[(\phi\omega - \Delta)^2]$$

We introduce the conditional expectation ω_0

$$\begin{aligned}\sigma^2 &= \mathbb{E}^{\mathbb{P}}[(\phi(\omega - \omega_0) + \phi\omega_0 - \Delta)^2] \\ &= \mathbb{E}^{\mathbb{P}}[\phi^2(\omega - \omega_0)^2] + \mathbb{E}^{\mathbb{P}}[(\phi\omega_0 - \Delta)^2] + \mathbb{E}^{\mathbb{P}}[\phi(\omega - \omega_0)(\phi\omega_0 - \Delta)]\end{aligned}$$

We have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[\phi(\omega - \omega_0)(\phi\omega_0 - \Delta)] &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\phi(\omega - \omega_0)(\phi\omega_0 - \Delta) | S_{t_1}, \cdots, S_{t_n}]] \\ &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[(\omega - \omega_0) | S_{t_1}, \cdots, S_{t_n}] \phi(\phi\omega_0 - \Delta)] \\ &= 0\end{aligned}$$

where we have used $\mathbb{E}^{\mathbb{P}}[\phi(\omega - \omega_0)|S_{t_1}, \cdots, S_{t_n}] = 0$. ∎

As a further example, we compute the sensibility according to a (local) deformation of the Dupire local volatility.

Local Vega

The asset S_t^0 is assumed to follow a local volatility model

$$dS_t^0 = \sigma(t, S_t^0)dW_t$$

We deform locally the local volatility by

$$dS_t^\epsilon = \sigma^\epsilon(t, S_t^\epsilon)dW_t$$

where $\sigma^\epsilon(t, S) = \sigma(t, S) + \epsilon\delta(t, S)$ and the initial condition $S_0^\epsilon = S_0$. Then, the process $Z_t^\epsilon \equiv \partial_\epsilon S_t^\epsilon$ satisfies the SDE

$$dZ_t^\epsilon = (\partial_S \sigma^\epsilon(t, S_t^\epsilon)Z_t^\epsilon + \delta(t, S_t^\epsilon))dW_t \tag{11.21}$$

with the initial condition $Z_0^\epsilon = 0$. Similarly, the tangent process $Y_t^\epsilon \equiv \frac{S_t^\epsilon}{S_0}$ satisfies the SDE

$$dY_t^\epsilon = \partial_S \sigma^\epsilon(t, S_t^\epsilon) Y_t^\epsilon dW_t$$

with the initial condition $Y_0^\epsilon = 1$. Using Itô's lemma, the solution of (11.21) is

$$Z_t^\epsilon = \left(- \int_0^t \frac{1}{Y_s^\epsilon} \partial_S \sigma^\epsilon(s, S_s^\epsilon)\delta(s, S_s^\epsilon)ds + \int_0^t \frac{1}{Y_s^\epsilon}\delta(s, S_s^\epsilon)dW_s \right) Y_t^\epsilon \tag{11.22}$$

The payoff sensitivity with respect to ϵ (i.e., *local Vega*) is

$$\partial_\epsilon \mathcal{C}^\epsilon|_{\epsilon=0} = \sum_{i=1}^n \mathbb{E}^{\mathbb{P}}[\partial_i \phi Z_{t_i}^0 | \mathcal{F}_t]$$

that we write as

$$\partial_\epsilon \mathcal{C}^\epsilon|_{\epsilon=0} = \sum_{i=1}^n \mathbb{E}^{\mathbb{P}}[\partial_i \phi Y_{t_i} \hat{Z}_{t_i}^0 | \mathcal{F}_t] \tag{11.23}$$

with

$$\hat{Z}_t^0 = \frac{Z_t^0}{Y_t^0}$$

As in the previous section, we would like to represent the Vega as

$$\mathbb{E}[\phi \partial(\pi)|\mathcal{F}_t] = \sum_{i=1}^n \mathbb{E}[\partial_i \phi \int_0^{t_i} \frac{Y_{t_i}}{Y_s} \sigma(s, X_s)\pi_s ds | \mathcal{F}_t] \tag{11.24}$$

Identifying equation (11.23) with equation (11.24) for any smooth payoff, we have

$$\sum_{i=1}^{n} \mathbb{E}[Y_{t_i} \left(\int_0^{t_i} \frac{\sigma(s, S_s)}{Y_s} \pi_s ds - \hat{Z}^0_{t_i} \right) |S_{t_1}, \cdots, S_{t_n}] = 0 \tag{11.25}$$

The simplest solution to (11.25) is

$$\pi_s = \frac{Y_s}{\sigma(s, S_s)} \sum_{j=1}^{n} a(s)(\hat{Z}^0_{t_j} - \hat{Z}^0_{t_{j-1}}) 1(s \in [t_{j-1}, t_j])$$

with $a(\cdot)$ such that

$$\int_{t_{j-1}}^{t_j} a(s) ds = 1 \ , \ \forall j = 1, \cdots, n$$

Using (11.9), we obtain finally

$$\begin{aligned}\partial(\pi) = \sum_{j=1}^{n} \Bigg(&(\hat{Z}^0_{t_j} - \hat{Z}^0_{t_{j-1}}) \int_{t_{j-1}}^{t_j} \frac{Y_s a(s)}{\sigma(s, S_s)} dW_s \\ &- \int_{t_{j-1}}^{t_j} \frac{Y_s a(s)}{\sigma(s, S_s)} (D_s \hat{Z}^0_{t_j} - D_s \hat{Z}^0_{t_{j-1}}) ds \Bigg)\end{aligned}$$

Note that the Skorohod integral involves an infinite number of processes parameterized by s: $\{D_s \hat{Z}^0_t\}_{s \in [0,T]}$. Other weights for higher-order Greeks can be found in [58].

11.6.3 Local volatility of stochastic volatility models

The SV models introduced in chapter 6 are defined by the following SDEs

$$\begin{aligned} df_t &= a_t f_t \Phi(t, f_t)(\rho dZ_t + \sqrt{1-\rho^2} dB_t) \\ da_t &= b(a_t) dt + \sigma(a_t) dZ_t \end{aligned} \tag{11.26}$$

Here W_t, B_t are two uncorrelated Brownian motions. In order to be able to calibrate exactly the implied volatility surface, we have decorated the volatility of the forward by a local volatility function $\Phi(t, f)$. By definition, the model is exactly calibrated to the implied volatility if and only if

$$\sigma_{\mathrm{loc}}(T, f)^2 = \Phi(T, f)^2 \mathbb{E}^{\mathbb{P}}[a_T^2 | f_T = f]$$

with $\sigma_{\mathrm{loc}}(T, f)$ the Dupire local volatility. We use the following result [86, 84] to find a simpler probabilistic representation of the conditional mean value.

PROPOSITION 11.3
Let $f_T, a_T^2 \in \mathbb{D}_{1,2}$. Assume that $D_t a_T^2$ is non-degenerate $\mathbb{P}$-almost sure for almost all $t \in [0,T]$ and there exists a process $u_t \in \mathrm{Dom}(\partial)$ such that

$$\mathbb{E}^{\mathbb{P}}[\int_0^T (D_s f_T).u_s ds|\sigma(f,a)] = 1 \tag{11.27}$$

$$\mathbb{E}^{\mathbb{P}}[\int_0^T (D_s a_T^2).u_s ds|\sigma(f,a)] = 0 \tag{11.28}$$

where $\sigma(f,a)$ is the σ-algebra generated by the processes $\{f_t, a_t\}_{0\leq t\leq T}$. Then the following formula holds

$$\mathbb{E}^{\mathbb{P}}[a_T^2|f_T = f] = \frac{\mathbb{E}^{\mathbb{P}}[a_T^2 1(f_T - f)\partial(u)]}{\mathbb{E}^{\mathbb{P}}[1(f_T - f)\partial(u)]} \tag{11.29}$$

The proof follows closely [86, 84].

PROOF Denote with $\delta(\cdot)$ the Dirac function and $1(x)$ the Heaviside function. Formally, we have

$$\mathbb{E}^{\mathbb{P}}[a_T^2|f_T = f] \equiv \frac{\mathbb{E}^{\mathbb{P}}[a_T^2 \delta(f_T - f)]}{\mathbb{E}^{\mathbb{P}}[\delta(f_T - f)]}$$

Using condition (11.27), the chain rule (11.4) and an integration by parts, we obtain

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[a_T^2 \delta(f_T - f)] &= \mathbb{E}^{\mathbb{P}}[a_T^2 \delta(f_T - f)\mathbb{E}[\int_0^T (D_s f_T) u_s ds|\sigma(f,a)]] \\
&= \mathbb{E}^{\mathbb{P}}[a_T^2 \int_0^T (D_s 1(f_T - f)) u_s ds] \\
&= \mathbb{E}^{\mathbb{P}}[\int_0^T D_s \left(a_T^2 1(f_T - f)\right) u_s] \\
&- \mathbb{E}^{\mathbb{P}}[1(f_T - f) \int_0^T (D_s a_T^2) u_s ds] \\
&= \mathbb{E}^{\mathbb{P}}[a_T^2 1(f_T - f)\partial(u)] - \mathbb{E}^{\mathbb{P}}[1(f_T - f) \int_0^T (D_s a_T^2) u_s ds]
\end{aligned}$$

The second term cancels thanks to the condition (11.28) and, finally, we obtain (11.29). ∎

The SDEs satisfied by the Malliavin derivatives of the forward are

$$\begin{aligned}
dD_s^Z f_t &= \left(f_t \Phi(t, f_t) D_s^Z a_t + a_t (f_t \Phi(t, f_t))' D_s^Z f_t\right)\left(\rho dZ_t + \sqrt{1-\rho^2} dB_t\right) \\
dD_s^B f_t &= \left(a_t (f_t \Phi(t, f_t))' D_s^B f_t\right)\left(\rho dZ_t + \sqrt{1-\rho^2} dB_t\right)
\end{aligned}$$

with the initial conditions

$$D_s^Z f_s = \rho a_s f_s \Phi(s, f_s)$$
$$D_s^B f_s = \sqrt{1-\rho^2} a_s f_s \Phi(s, f_s)$$

Note that as a_t is not driven by the Brownian B_t, we have

$$D_s^B a_t = 0$$

Let the functions u^B, u^Z be

$$u_s^B = \frac{1}{T D_s^B f_T} 1(s \in [0, T])$$
$$u_s^Z = 0$$

They satisfy trivially the conditions (11.27) and (11.28). Note that

$$D_s^B f_T = \sqrt{1-\rho^2} a_s f_s \Phi(s, f_s) \frac{\Xi_T}{\Xi_s} 1(s \in [0, T])$$

with the Itô process Ξ_t given by

$$d\Xi_t = a_t (f_t \Phi(t, f_t))' \Xi_t dN_t \tag{11.30}$$
$$\Xi_0 = 1$$

where we have set $dN_t \equiv (\rho dZ_t + \sqrt{1-\rho^2} dB_t)$.
Using (11.9), we obtain

$$T\sqrt{1-\rho^2}\Xi_T \partial(u^B) = \int_0^T \frac{\Xi_s}{a_s f_s \Phi(s, f_s)} dB_s + \Xi_T^{-1} \int_0^T \frac{\Xi_s}{a_s f_s \Phi(s, f_s)} D_s^B \Xi_T ds \tag{11.31}$$

where we have

$$dD_s^B \Xi_t = a_t \left((f_t \Phi(t, f_t))'' \sqrt{1-\rho^2} a_s f_s \Phi(s, f_s) \frac{\Xi_t^2}{\Xi_s} + (f_t \Phi(t, f_t))' D_s^B \Xi_t \right) dN_t \tag{11.32}$$

$$D_s^B \Xi_s = \sqrt{1-\rho^2} a_s (f_s \Phi(s, f_s))' \Xi_s \tag{11.33}$$

Effective vector field

We prove below that the process $D_s^B \Xi_t$ is an *effective vector field* [31]: it can be written as a sum of products of a smooth function by an adapted Itô process

$$D_s^B \Xi_t = A_s \Xi_t + B_s \zeta_t \tag{11.34}$$

PROOF By plugging equation (11.34) into (11.32), we obtain

$$B_s d\zeta_t = a_t \left((f_t\Phi(t,f_t))'' \sqrt{1-\rho^2} a_s f_s \Phi(s,f_s) \frac{\Xi_t^2}{\Xi_s} + B_s (f_t\Phi(t,f_t))'\zeta_t \right) dN_t$$

Setting $B_s = \sqrt{1-\rho^2} a_s \frac{f_s\Phi(s,f_s)}{\Xi_s}$, we get

$$d\zeta_t = a_t \left((f_t\Phi(t,f_t))''\Xi_t^2 + (f_t\Phi(t,f_t))'\zeta_t \right) dN_t \tag{11.35}$$

where we set $\zeta_0 = 1$. The initial condition (11.33) is satisfied if

$$A_s\Xi_s + B_s\zeta_s = \sqrt{1-\rho^2} a_s (f_s\Phi(s,f_s))'\Xi_s$$

This gives $D_s^B \Xi_t$ as an effective vector field

$$D_s^B\Xi_t = \sqrt{1-\rho^2} a_s \left(\left((f_s\Phi(s,f_s))' - \frac{f_s\Phi(s,f_s)\zeta_s}{\Xi_s^2} \right) \Xi_t + \frac{f_s\Phi(s,f_s)}{\Xi_s}\zeta_t \right) \tag{11.36}$$

∎

As $D_s^B\Xi_t$ is an effective vector field, equation (11.31) simplifies and reduces to

$$T\sqrt{1-\rho^2}\Xi_T \partial(u^B) = \int_0^T \frac{\Xi_s}{a_s f_s \Phi(s,f_s)} dB_s + \sqrt{1-\rho^2} \left(\int_0^T \left(\Xi_s \ln(f_s\Phi(s,f_s))' - \frac{\zeta_s}{\Xi_s} \right) ds + T\zeta_T \Xi_T^{-1} \right)$$

Setting

$$d\Theta_t = \frac{\Xi_t}{a_t f_t \Phi(t,f_t)} dB_t + \sqrt{1-\rho^2} \left(\partial_f \ln(f_t\Phi(t,f_t))\Xi_t \right) dt \ , \ \Theta_0 = 0$$

the local volatility can be written as

$$\sigma_{\mathrm{loc}}(T,f)^2 = \Phi(T,f)^2 \frac{\mathbb{E}^{\mathbb{P}}[1(f_T - f)V_T \left(\Xi_T^{-1}\Theta_T + \sqrt{1-\rho^2}\Xi_T^{-1} \int_0^T s d(\zeta_s \Xi_s^{-1}) \right)]}{\mathbb{E}^{\mathbb{P}}[1(f_T - f) \left(\Xi_T^{-1}\Theta_T + \sqrt{1-\rho^2}\Xi_T^{-1} \int_0^T s d(\zeta_s \Xi_s^{-1}) \right)]}$$

The processes $Y_t^{(1)} \equiv f_t\Phi(t,f_t)\Xi_t^{-1}\Theta_t$ and $Y_t^{(2)} \equiv \Xi_t^{-1}\zeta_t$ satisfy the following SDEs

$$dY_t^{(1)} = \frac{dB_t}{a_t} + \partial_t \ln(\Phi) Y_t^{(1)} dt + \frac{1}{2}(f\Phi)(f\Phi)'' a_t^2 Y_t^{(1)} dt + \sqrt{1-\rho^2}(f\Phi)' dt \tag{11.37}$$

$$dY_t^{(2)} = (f_t\Phi)''\Xi_t \left(a_t dN_t - a_t^2 (f_t\Phi)' dt \right) \tag{11.38}$$

with the initial conditions $Y_0^{(1)} = 0$ and $Y_0^{(2)} = 1$. The local volatility becomes

$$\sigma_{\mathrm{loc}}(T,f)^2 = \Phi(T,f)^2$$
$$\frac{\mathbb{E}^{\mathbb{P}}[(f_T\Phi(T,f_T))^{-1}1(f_T-f)a_T^2\left(Y_T^{(1)} + \sqrt{1-\rho^2}\Xi_T^{-1}f_T\Phi(T,f_T)\int_0^T sdY_s^{(2)}\right)]}{\mathbb{E}^{\mathbb{P}}[(f_T\Phi(T,f_T))^{-1}1(f_T-f)\left(Y_T^{(1)} + \sqrt{1-\rho^2}\Xi_T^{-1}f_T\Phi(T,f_T)\int_0^T sdY_s^{(2)}\right)]}$$

The term $\int_0^T sdY_s^{(2)}$ should be written as $TY_T^{(2)} - \int_0^T Y_s^{(2)}ds$. Replacing $\delta(f_T - f)$ in the conditional expectation by $\delta(\ln f_T - \ln f)$, the expression above can be replaced by a simpler formula and we get

PROPOSITION 11.4 Local Volatility
The local volatility associated to the SVMS as defined by (11.26) is

$$\sigma_{\mathrm{loc}}(T,f)^2 = \Phi(T,f)^2\frac{\mathbb{E}^{\mathbb{P}}[1\left(\ln\frac{f_T}{f}\right)a_T^2\Omega]}{\mathbb{E}^{\mathbb{P}}[1\left(\ln\frac{f_T}{f}\right)\Omega]}$$

with

$$\Omega = Y_T^{(1)} - T\sqrt{1-\rho^2}(f\Phi(T,f_T))' + \sqrt{1-\rho^2}\Xi_T^{-1}(f_T\Phi)\int_0^T sdY_s^{(2)}$$

which depends on the Itô processes Ξ_t (11.30), $Y_t^{(1)}$ (11.37) and $Y_t^{(2)}$ (11.38).

REMARK 11.4 Log-normal SVM In the particular case $\Phi(t,f_t) = 1$, the equation above for the local volatility reduces to [84]

$$\sigma_{loc}(T,f)^2 = \frac{\mathbb{E}^{\mathbb{P}}[1(\ln f_T - \ln f)a_T^2\int_0^T\frac{dB_s}{a_s}]}{\mathbb{E}^{\mathbb{P}}[1(\ln f_T - \ln f)\ \int_0^T\frac{dB_s}{a_s}]}$$

We can go one step further and reduce the complexity of this equation using a mixing solution (see problem 6.1). We have

$$\mathbb{E}^{\mathbb{P}}[1(\ln f_T - \ln f)a_T^2\int_0^T\frac{dB_s}{a_s}] = \mathbb{E}^{\mathbb{P}}[a_T^2$$
$$\mathbb{E}^{\mathbb{P}}[1\left(\sqrt{1-\rho^2}\int_0^T a_s dB_t - \ln\frac{f}{f_0} - \frac{1}{2}\int_0^T a_s^2 ds + \rho\int_0^T a_s dZ_s\right)\int_0^T\frac{dB_s}{a_s}|\mathcal{F}^Z]]$$

The r.v. $X \equiv \sqrt{1-\rho^2}\int_0^T a_s dB_s$ and $Y \equiv \int_0^T\frac{dB_s}{a_s}$ are normally distributed (conditional to $\mathcal{F}^Z$) with variance matrix

$$\begin{pmatrix} (1-\rho^2)\int_0^T a_s^2 ds & \sqrt{1-\rho^2}T \\ \sqrt{1-\rho^2}T & \int_0^T\frac{ds}{a_s^2} \end{pmatrix}$$

We obtain that

$$\mathbb{E}^{\mathbb{P}}[1(X-K)Y|\mathcal{F}^Z] = \mathcal{N}\frac{e^{-\frac{K^2}{2(1-\rho^2)\int_0^T a_s^2 ds}}}{\sqrt{\int_0^T a_s^2 ds}}$$

where $\mathcal{N}$ is a constant independent of $\{a_s^2\}_s$. We deduce that the local volatility is given by

PROPOSITION 11.5 Mixing solution local volatility

$$\sigma_{\mathrm{loc}}(T,f)^2 = \frac{\mathbb{E}^{\mathbb{P}}[a_T^2 \frac{e^{-\frac{K^2}{2(1-\rho^2)\int_0^T a_s^2 ds}}}{\sqrt{\int_0^T a_s^2 ds}}]}{\mathbb{E}^{\mathbb{P}}[\frac{e^{-\frac{K^2}{2(1-\rho^2)\int_0^T a_s^2 ds}}}{\sqrt{\int_0^T a_s^2 ds}}]}$$

with $K = \ln \frac{f}{f_0} + \frac{1}{2}\int_0^T a_s^2 ds - \rho \int_0^T a_s dZ_s$. *Note that this expression only requires the simulation of the stochastic volatility* a_t.

□

11.7 Problems

Exercises 11.1 Brownian bridge
Let $T \in [0,1]$.

1. Using the Malliavin calculus, prove that the conditional mean-value

$$\mathbb{E}^{\mathbb{P}}[W_T|W_1 = x]$$

   can be written as

$$\mathbb{E}^{\mathbb{P}}[W_T|W_1 = x] = \frac{\mathbb{E}^{\mathbb{P}}[W_T(W_1 - W_T)1(W_1 - x)]}{\mathbb{E}^{\mathbb{P}}[(W_1 - W_t)1(W_1 - x)]}$$

2. By noting that W_T and $W_1 - W_T$ are independent r.v., deduce from the previous question that we have

$$\mathbb{E}^{\mathbb{P}}[W_T|W_1 = x] = \frac{\mathbb{E}^{\mathbb{P}}[W_T e^{-\frac{(x-W_T)^2}{2(1-T)}}]}{\mathbb{E}^{\mathbb{P}}[e^{-\frac{(x-W_T)^2}{2(1-T)}}]}$$

3. By doing the integration over W_T, prove that

$$\mathbb{E}^{\mathbb{P}}[W_T|W_1 = x] = Tx$$

Chapter 12

Portfolio Optimization and Bellman-Hamilton-Jacobi Equation

Abstract Pricing and Hedging derivatives products is essentially a problem of portfolio optimization. Once a measure of risk has been chosen, the price can be defined as the mean value of the profit and loss (P&L) and the best hedging strategy is the optimal control which minimizes the risk.
In the Black-Scholes model, the only source of risk is the spot process and the optimal control is the delta-strategy which cancels the risk. However, under the introduction of stochastic volatility, the market model becomes incomplete. The resulting risk is finite and the delta-strategy is not optimal. A portfolio optimization problem appears also naturally if we assume that the market is illiquid and the trading strategy affects the price movements. In the following, we will focus on these optimal control problems when the market is incomplete and the market is illiquid. Our study involves the use of perturbation methods for non-linear PDEs.

12.1 Introduction

Since the famous papers of Black-Scholes on option pricing [65], some progress has been made in order to extend these results to more realistic arbitrage-free market models. As a reminder, the Black-Scholes theory consists in following a (hedging) strategy to decrease the risk of loss given a fixed amount of return. This theory is based on three important hypotheses which are not satisfied under real market conditions:

- The traders can revise their decisions continuously in time. This first hypothesis is not realistic for obvious reasons. A major improvement was recently introduced in [6] in their time-discrete model. They introduce an elementary time τ after which a trader is able to revise his decisions again. The optimal strategy is fixed by the minimization of the risk defined by the variance of the portfolio. The resulting risk is no longer zero and in the continuous-time limit where τ goes to zero, one recovers the classical result of Black-Scholes: the risk vanishes.

- The spot dynamics is a log-normal process (with a constant volatility). As a consequence, the market is complete and the risk cancels. This second hypothesis doesn't truthfully reflect the market as indicated by the existence of an implied volatility. In chapters 5 and 6, we have seen how local and stochastic volatility models (SVMs) can account for an implied volatility. The SVMs are incomplete as it is not possible to trade the volatility. The resulting risk is non-zero.

- Markets are assumed to be completely elastic. This third hypothesis means that small traders don't modify the prices of the market by selling or buying large amounts of assets. The limitation of this last assumption lies with the fact it is only justified when the market is liquid.

While so far we have mainly focused on relaxing the second hypothesis described above by introducing local and stochastic volatility models, in this chapter we shall explain how to obtain optimal hedging strategy when the market is incomplete and is illiquid. Our main tool, we use, is the stochastic Hamilton-Bellman-Jacobi that we review.
In the first part, we analyze the hedging in an incomplete market. The aim of the second part is to analyze the feedback effect of hedging in portfolio optimization. In this section, we model the market as composed of small traders and a large one whose demand is given by a hedging strategy. We derive from this toy-model the dynamics of the asset price with the influence of the large trader. Taking into account the influence on the volatility and the return, we compute the hedging strategy of the large trader in order to optimize dynamically a portfolio composed of a risky asset and a bond. This leads to a well-defined stochastic optimization problem.

12.2 Hedging in an incomplete market

Let us assume that a trader holds at time t a certain number of shares Δ_t of which the price at time t is S_t. The price S_t satisfies the SVM in the historical measure $\mathbb{P}^{\text{hist}}$

$$\begin{aligned} dS_t &= \mu_t S_t dt + a_t C(S_t) dW_t \\ da_t &= b(a_t)dt + \sigma(a_t)dZ_t \end{aligned}$$

with W_t and Z_t two correlated Brownian processes ($dW_t dZ_t = \rho dt$) and with the initial conditions $S_{t=0} = S_0$, $a_{t=0} = \alpha$. The trader holds also a number b_t of bonds the price of which satisfies the following equation

$$dB_t = r_t B_t dt$$

with r_t the interest rate of the bond. For the sake of simplicity, r_t is assumed to be a deterministic process. The value of his portfolio at time t is

$$\pi_t = \Delta_t S_t + b_t B_t$$

Assuming a self-financing portfolio, the variation of its value between t and $t + dt$ is

$$d\pi_t = \Delta_t dS_t + b_t dB_t$$

and its discounted variation

$$dD_{0t}\pi_t = \Delta_t D_{0t} \left(dS_t - r_t S_t dt\right)$$

At $t = 0$, the trader sells a (European) option with payoff $f(S_T)$ at maturity T at a price $\mathcal{C}$. The discounted change in his total portfolio from $t = 0$ to $t = T$ where the holder can exercise his option is

$$\Pi_T = -D_{0T} f(S_T) + \mathcal{C} + \int_0^T D_{0u}\Delta_u (dS_u - r_u S_u du) \tag{12.1}$$

Using the martingale representation theorem, we have

$$D_{0T} f(S_T) = \mathbb{E}^{\mathbb{P}^{\text{hist}}}[D_{0T} f(S_T)] + \int_0^T D_{0u}[aC(S)\partial_S f dW_u + \sigma(a)\partial_a f dZ_u]$$

By plugging this expression into (12.1), we obtain

$$\begin{aligned}\Pi_T = \mathcal{C} - \mathbb{E}^{\mathbb{P}^{\text{hist}}}[D_{0T} f(S_T)] + \int_0^T D_{0u}[-\sigma(a)\partial_a f dZ_u \\ + aC(S_u)(\Delta_u - \partial_S f)dW_u] + \int_0^T D_{0u}\Delta_u S_u(\mu_u - r_u)du\end{aligned}$$

We impose that the option price should be fixed such that the mean value of the portfolio $\mathbb{E}^{\mathbb{P}^{hist}}[\Pi_T]$ should cancel. We obtain

$$\mathcal{C} = \mathbb{E}^{\mathbb{P}^{\text{hist}}}[D_{0T} f(S_T) + \int_0^T D_{0u}\Delta_u S_u(\mu_u - r_u)du]$$

and therefore the portfolio Π_t follows the SDE

$$d\Pi_t = -D_{0t}\sigma(a)\partial_a f dZ_t + D_{0t} aC(S)(\Delta_t - \partial_S f)dW_t$$

Note that the mean-value is under the historical measure. As seen in chapter 2, if we impose the non-arbitrage condition, the mean-value will be under a risk-neutral measure.

Once the option price has been fixed, we must focus on the optimal hedging strategy. The hedging function Δ_t is obtained by requiring that the risk, measuring by the variance of the portfolio, is minimal

$$\Delta \text{ such as } \min_{\Delta}[\mathbb{E}_t^{\mathbb{P}^{\text{hist}}}[\Pi_T^2] - \mathbb{E}_t^{\mathbb{P}^{hist}}[\Pi_T]^2] = \min_{\Delta}[\mathbb{E}_t^{\mathbb{P}^{\text{hist}}}[\Pi_T^2]]$$

where we have used that by construction $\mathbb{E}_t^{\mathbb{P}^{\text{hist}}}[\Pi_T] = 0$. We introduce the function

$$J(t, S_t, \Pi_t, a) = \min_{\Delta} \mathbb{E}_t^{\mathbb{P}^{hist}}[\Pi_T^2]$$

A small digression: The Hamilton-Jacobi-Bellman equation

The Hamilton-Jacobi-Bellman theorem states that the function J satisfies the parabolic PDE

$$\partial_t J + \min_{\Delta} \left(DJ + \frac{1}{2} D_{0t}^2 \, \partial_\Pi^2 J \right. \tag{12.2}$$

$$\left(\left((\Delta - \partial_S f) \, aC(S) - \rho\sigma(a)\partial_a f \right)^2 + (1-\rho^2)\sigma(a)^2(\partial_a f)^2 \right)$$

$$+ \, D_{0t}\partial_{\Pi a}\mathcal{J}\sigma(a) \left((\Delta - \partial_S f) \, aC(S)\rho - \sigma(a)\partial_a f \right)$$

$$+ \, D_{0t}\partial_{\Pi S}\mathcal{J} aC(S) \left((\Delta - \partial_S f) \, aC(S) - \rho\sigma(a)\partial_a f \right)) = 0 \tag{12.3}$$

with the boundary condition $J(T, S, \Pi, S) = \Pi^2$ and D the Itô generator

$$D = \frac{1}{2} \left(a^2 C(S)^2 \partial_{SS} + \sigma(a)^2 \partial_{aa} \right) + \rho a C(S) \sigma(a) \partial_{aS}$$

Taking the functional derivative according to Δ in the equation above, we obtain the optimal control

$$\Delta = \partial_S f + \rho \frac{\sigma(a)}{aC(S)} \partial_a f - \frac{\partial_{\Pi a} J \frac{\sigma(a)}{aC(S)} \rho + \partial_{\Pi S} \mathcal{J}}{D_{0t} \partial_\Pi^2 \mathcal{J}}$$

We try as an ansatz for the solution of the HJB equation

$$J(t, S, \Pi, a) = \Pi^2 + A(t, S, a)$$

with $A(T, S, a) = 0$. By plugging this ansatz in (12.3), we obtain easily that

$$\partial_t A + DA + \frac{1}{2} D_{0t}^2 (1-\rho)^2 \sigma(a)^2 (\partial_a f)^2 = 0$$

and therefore

$$\Delta = \partial_S f + \rho \frac{\sigma(a)}{aC(S)} \partial_a f$$

A similar expression was obtained in [61]. When the volatility of the volatility $\sigma(a)$ cancels, we reproduce the Black-Scholes Delta-hedging strategy and the risk cancels.

In the next section, we focus on the problem of optimal hedging when the market is illiquid.

12.3 The feedback effect of hedging on price

By definition, a market is liquid when the *elasticity* parameter is small. The elasticity parameter ϵ is given by the ratio of relative change in price S_t to change in the net demand D:

$$\frac{dS_t}{S_t} = \epsilon dD_t \tag{12.4}$$

We observe empirically that when the demand increases (resp. decreases), the price rises (decreases). The parameter ϵ is therefore positive and we will assume that it is also constant. Another interesting (because more realistic) assumption would be to define ϵ as a stochastic variable.
In a crash situation, the small traders who tend to apply the same hedging strategy can be considered as a large trader and the hedging feedback effects become very important. This can speed up the crash. In the following, we use the simple relation (12.4) to analyze the influence of dynamical trading strategies on the prices in financial markets. The dynamical trading $\phi(t, S_t)$ which represents the number of shares of a given stock that a large trader holds will be determined in order to optimize his portfolio. An analytic solution will not be possible and we will resolve the non-linear equations by expanding the solution as a formal series in the parameter ϵ. The first order correction will be obtained.
Let us call $D(t, W, S_t)$ the demand of all the traders in the market which depends on time t, a Brownian process W_t and the price S_t. S_t is assumed to satisfy the local volatility model in $\mathbb{P}^{\text{hist}}$

$$\frac{dS_t}{S_t} = \mu(t, S_t)dt + \sigma(t, S_t)dW_t \tag{12.5}$$

with $\mu(t, S_t)$ the (historical) return and $\sigma(t, S_t)$ the volatility. The process W_t models both the information the traders have on the demand and the fluctuation of the price S_t. Applying Itô's lemma on the function $D(t, S_t, W)$, we obtain

$$dD_t = (\partial_t D + \frac{1}{2}\sigma^2 S^2 \partial_S^2 D + \frac{1}{2}\partial_W^2 D + \sigma S \partial_{SW} D)dt + \partial_S D dS_t + \partial_W D dW_t \tag{12.6}$$

As we have also

$$dD_t = \frac{1}{\epsilon}\frac{dS_t}{S_t} = \frac{1}{\epsilon}\left(\mu(t, S_t)dt + \sigma(t, S_t)dW_t\right) \tag{12.7}$$

the identification of the coefficients for dt and dW_t in (12.6) and (12.7), one obtains μ and σ as a function of the derivatives of D:

$$\sigma(t, S_t) = \frac{\epsilon \partial_W D}{(1 - \epsilon S \partial_S D)} \tag{12.8}$$

$$\mu(t, S_t) = \frac{\epsilon[\partial_t D + \frac{1}{2}\sigma^2 S^2 \partial_S^2 D + \sigma S \partial_{SW} D + \frac{1}{2}\partial_W^2 D]}{(1 - \epsilon S \partial_S D)} \tag{12.9}$$

We will now assume that the market is composed of a group of *small traders* who don't modify the prices of the market by selling or buying large amounts of assets on the one hand and a *large trader* one on the other hand. One can consider a large trader as an aggregate of small traders following the same strategy given by his hedging position ϕ. We will choose the demand D_{small} of small traders in order to reproduce the Black-Scholes log-normal process for S_t (12.5) with a constant return $\mu = \mu_0$ and a constant volatility $\sigma = \sigma_0$. The solution is given by

$$D_{\text{small}} = \frac{1}{\epsilon}(\mu_0 t + \sigma_0 W_t) \tag{12.10}$$

Now, we include the effect of a large trader whose demand D_{large} is generated by his trading strategy

$$D_{\text{large}} = \phi(t, S_t)$$

It is implicitly assumed that ϕ only depends on S and t and not explicitly on W_t. The hedging position ϕ is then added to the demand of the small traders D_{small} and the total net demand is given by

$$D_t = D_{\text{small}} + \phi(t, S_t) \tag{12.11}$$

By inserting (12.11) in (12.8)-(12.9), one finds the volatility and return as a function of ϕ:

$$\sigma(t, S) = \frac{\sigma_0}{(1 - \epsilon S \partial_S \phi)} \tag{12.12}$$

$$\mu(t, S) = \frac{\mu_0 + \epsilon(\partial_t \phi + \frac{1}{2}\sigma^2 S^2 \partial_S^2 \phi)}{(1 - \epsilon S \partial_S \phi)} \tag{12.13}$$

This relation describes the feedback effect of dynamical hedging on volatility and return. The hypothesis that the hedging function only depends on S and t allows to obtain a volatility and a return independent of historic effects and therefore a Markovian process for S_t. As we will see in the next section, for derivatives and portfolio hedging strategies, the feedback effect on the volatility will not be the same.

12.4 Non-linear Black-Scholes PDE

The Black-Scholes analysis can be trivially modified to incorporate the hedging feedback effect. Let Π be the self-financing portfolio and $\mathcal{C}(t, S_t)$ the value of the option at time t:

$$\Pi_t = -\mathcal{C}(t, S_t) + \phi(t, S_t) S_t \tag{12.14}$$

As a result of Itô's lemma, $d\Pi$ is given by

$$d\Pi_t = -d\mathcal{C} + \phi dS_t \tag{12.15}$$

$$= -(\partial_t \mathcal{C} + \frac{1}{2}\sigma^2 S^2 \partial_S^2 \mathcal{C})dt + (\phi - \partial_S \mathcal{C})dS_t \tag{12.16}$$

A free-arbitrage condition gives $d\Pi_t = r(t)\Pi_t$ with $r(t)$ the interest rate. The risk is therefore zero if $\phi = \partial_S \mathcal{C}$ which is the usual hedging position. The *non-linear Black-Scholes* PDE (modulo adequate boundary conditions) is then

$$\partial_t \mathcal{C} + \frac{1}{2} \frac{\sigma_0^2}{(1 - \epsilon S \partial_S^2 \mathcal{C})^2} S^2 \partial_S^2 \mathcal{C} + r(-\mathcal{C} + S\partial_S \mathcal{C}) = 0 \tag{12.17}$$

This non-linear PDE can be solved by expanding $\mathcal{C}$ (and so ϕ) as an asymptotic series in ϵ.

Note that if the option's gamma $\partial_S^2 \mathcal{C}$ is positive, then $\partial_S \phi(t, S) > 0$ meaning that the trader buys additional shares when the price rises. The volatility σ is then greater than σ_0. A destabilizing effect is then obtained.

On the other hand, as we will see in the next section, in dynamical portfolio optimization, we have the expression $\partial_S \phi(t, S) < 0$ according to which a trader should sell stocks when his price increases and buy more stocks when his price decreases. As a result $\sigma < \sigma_0$ and the price increases. In this situation, a trader could buy large amounts of shares at a price S and the price would then move to a higher price S'. By selling his shares, it would make a free-risk profit $S' - S$ per share. The key feature that allows this manipulation is that the price reacts with a delay that allows the trader to buy at low price and sell at a higher price before the price goes down.

12.5 Optimized portfolio of a large trader

Let us assume that a large trader holds at time t a certain number of shares $\phi(t, S_t)$ of which the price at time t is S_t. The price S_t satisfies the SDE (12.5) which depends implicitly on the hedging position $\phi(t, S_t)$ (see the volatility

(12.12) and the return (12.13)). The trader holds also a number b_t of bonds the price of which satisfies the following equation $dB_t = r(t)B_t dt$ with $r(t)$ the deterministic interest rate of the bond. The change in his portfolio $\Pi_t = \phi(t, S_t)S_t + b_t B_t$ during an infinitesimal time dt is then

$$\begin{aligned} d\Pi_t &= \phi(t, S_t)dS_t + r(t)b_t B_t dt \\ &= r(t)\Pi_t dt + \phi(t, S_t)S_t \left((\mu(t, S_t) - r(t))dt + \sigma(t, S_t)dW_t\right) \end{aligned} \quad (12.18)$$

One should note that ϕ can be negative (which is equivalent to a short position). At a maturity date T, the value of the portfolio Π_T is

$$\Pi_T = \Pi_0 + \int_0^T d\Pi$$

The basic strategy of the trader is to find the optimal strategy $\phi^*(t, S)$ so that the risk $\mathcal{R}$ is minimized for a given fixed value of the profit $\mathbb{E}[\Pi_T] = \mathcal{G}$. By definition, the risk is given by the variance of the portfolio Π_T:

$$\mathcal{R} \equiv \mathbb{E}^{\mathbb{P}^{\text{hist}}}[\Pi_T^2] - \mathbb{E}^{\mathbb{P}^{\text{hist}}}[\Pi_T]^2 = \mathbb{E}^{\mathbb{P}^{\text{hist}}}[\Pi_T^2 - \mathcal{G}^2]$$

It is clear that the result of this section depends on the above mentioned definition for the risk, its main advantage being that it gives simple computations. The mean-variance portfolio selection problem is then formulated as the following optimization problem parameterized by $\mathcal{G}$:

$$\min_{\phi} \ \mathbb{E}^{\mathbb{P}^{\text{hist}}}[\Pi_T^2 - \mathcal{G}^2]$$

subject to

$$\mathbb{E}^{\mathbb{P}^{\text{hist}}}[\Pi_T] = \mathcal{G} \ , \ (S_t, \Pi_t) \ \text{satisfying} \ (12.5) - (12.18)$$

This problem is equivalent to the following one in which we have introduced a Lagrange multiplier ζ:

$$\min_{\phi} \ \mathbb{E}^{\mathbb{P}^{\text{hist}}}[\Pi_T^2 - \mathcal{G}^2 + \zeta(\Pi_T - \mathcal{G})]$$

subject to

$$(S_t, \Pi_t) \ \text{satisfying} \ (12.5) - (12.18)$$

For the sake of simplicity, we assume in the following that the interest rate is negligible (i.e., $r(t) = 0$) although this hypothesis can be easily relaxed.
To derive the optimality equations, the problem above can be restated in a dynamic programming form so that the Bellman principle of optimality can be applied [32]. To do this, let us define

$$\mathcal{J}(t, S, \Pi) = \min \ \mathbb{E}_t^{\mathbb{P}^{\text{hist}}}[\Pi_T^2 - \mathcal{G}^2 + \zeta(\Pi_T - \mathcal{G})]$$

where $\mathbb{E}_t^{\mathbb{P}^{\text{hist}}}$ is the conditional expectation operator at time t with $S_t = S$ and $\Pi_t = \Pi$. Then, J satisfies the HJB equation:

$$\min_{\phi,\zeta}\{\partial_t \mathcal{J} + \mu S\left(\partial_S \mathcal{J} + \phi \partial_\Pi \mathcal{J}\right) + \frac{1}{2}\sigma^2 S^2\left(\partial_S^2 \mathcal{J} + \phi^2 \partial_\Pi^2 \mathcal{J} + 2\phi \partial_{S\Pi}\mathcal{J}\right) = 0\} \tag{12.19}$$

subject to

$$\mathcal{J}(T, S, \Pi) = \Pi^2 - \mathcal{G}^2 + \zeta(\Pi - \mathcal{G}) \tag{12.20}$$

$$= (\Pi + \frac{\zeta}{2})^2 - \mathcal{G}^2 - \zeta\mathcal{G} - \frac{\zeta^2}{4} \tag{12.21}$$

The equation (12.19) is quite complicated due to the explicit dependence of the volatility and the return in the control ϕ. A simple method to solve the H-J-B equation (12.19) perturbatively in the liquidity parameter ϵ is to use a mean-field approximation well known in statistical physics: one computes the optimal control $\phi^*(t, S)$ up to a given order ϵ^i, say $\phi_i^*(t, S)$. This gives a mean volatility $\sigma_i(t, S) = \sigma(t, S, \phi_i^*(t, S))$ and mean return $\mu_i(t, S) = \mu(t, S, \phi_i^*(t, S))$.
The parameters μ, σ can be written as a formal series in ϵ and we will see in the following section that this is also the case for the hedging function

$$\phi = \sum_{i=0} \phi_i(t, S)\epsilon^i$$

We give the first order expansion of σ (12.12) and μ (12.13):

$$\sigma = \sigma_0(1 + \epsilon S \partial_S \phi) + o(\epsilon^2) \tag{12.22}$$

$$\mu = \mu_0 + \epsilon(\partial_t \phi + \frac{1}{2}\sigma_0^2 S^2 \partial_S^2 \phi + \mu_0 S \partial_S \phi) + o(\epsilon^2) \tag{12.23}$$

The optimal control, at the order ϵ^{i+1}, is then easily found by taking the functional derivative of the expression (12.19) according to ϕ:

$$\phi_{i+1}^* = -\frac{(\mu_i \partial_\Pi \mathcal{J} + \sigma_i^2 S \partial_{S\Pi}\mathcal{J})}{\sigma_i^2 S \partial_\Pi^2 \mathcal{J}} \tag{12.24}$$

By inserting this expression in the HJB equation (12.19), one obtains:

$$\partial_t \mathcal{J} + \mu_i S \partial_S \mathcal{J} + \frac{1}{2}\sigma_i^2 S^2 \partial_S^2 \mathcal{J} - \frac{(\mu_i \partial_\Pi \mathcal{J} + \sigma_i^2 S \partial_{S\Pi}\mathcal{J})^2}{2\sigma_i^2 \partial_\Pi^2 \mathcal{J}} = 0 \tag{12.25}$$

The resolution of the equation above gives $\phi(t, S)_{i+1}$ and the procedure can be iterated. We apply this method up to order one. The form of (12.25) and the boundary conditions (12.21) suggest that $\mathcal{J}$ takes the following form:

$$\mathcal{J}(t, S, \Pi) = A(t, S)(\Pi + \frac{\zeta}{2})^2 - \mathcal{G}^2 - \zeta\mathcal{G} - \frac{\zeta^2}{4} \tag{12.26}$$

with ζ such that $\partial_\zeta \mathcal{J} = 0$ and with the boundary condition $A(T, S) = 1$. By inserting this expression in (12.25) and (12.24), one then obtains the equations:

$$\phi^*_{i+1}(t, u) = -S^{-1}(\frac{\mu_i}{\sigma_i^2} + \partial_u X)(\Pi + \frac{\zeta}{2}) \tag{12.27}$$

$$\partial_t X - (\mu_i + \frac{1}{2}\sigma_i^2)\partial_u X + \frac{1}{2}\sigma_i^2(\partial_u^2 X - (\partial_u X)^2) = \frac{\mu_i^2}{\sigma_i^2} \tag{12.28}$$

with $u = \ln(S)$ and $A(t, S) = e^{X(t,S)}$. The Lagrange condition $\partial_\zeta \mathcal{J} = 0$ gives

$$\Pi + \frac{\zeta}{2} = \frac{\mathcal{G} - \Pi}{e^X - 1} \tag{12.29}$$

Let us define the parameter $\lambda \equiv \frac{\mu_0^2}{\sigma_0^2}$. In the zero-order approximation, the volatility and return are constant and in this case, A will be a function of t only given by

$$A(t, S) = e^{\lambda(t-T)}$$

By inserting this expression in (12.27) and using (12.29), one finds the optimal control (at zero-order):

$$\phi^*(t, S, \Pi) = \frac{\lambda}{\mu_0} S^{-1} \frac{\Pi - \mathcal{G}}{e^{\lambda(t-T)} - 1}$$

This expression allows to derive a SDE for the portfolio Π subject to the condition $\Pi(0) = \Pi_0$:

$$\frac{d\Pi}{(\Pi - \mathcal{G})} = \frac{1}{(e^{\lambda(t-T)} - 1)}(\lambda dt + \sqrt{\lambda} dW)$$

The mean of the portfolio $\mathbb{E}[\Pi]$ satisfies the ODE

$$\frac{d\mathbb{E}[\Pi]}{dt} = \lambda \frac{\mathbb{E}[\Pi] - \mathcal{G}}{e^{\lambda(t-T)} - 1}$$

The solution is

$$\mathbb{E}[\Pi] - \mathcal{G} = (\Pi_0 - \mathcal{G}) \frac{(1 - e^{-\lambda(t-T)})}{(1 - e^{\lambda T})}$$

One can then verify that the condition $\mathbb{E}[\Pi_T] = \mathcal{G}$ is well satisfied. Finally, the optimal control ϕ_0^* at order zero satisfies the following SDE:

$$\frac{d[\phi_0^* S]}{\phi_0^* S} = -\lambda t + \frac{\sqrt{\lambda}}{(e^{\lambda(t-T)} - 1)} dW_t$$

and the solution is

$$\phi_0^* = \frac{\lambda}{\mu_0} S^{-1} \frac{(\mathcal{G} - \Pi_0)}{(1 - e^{-\lambda T})^{\frac{3}{2}}} (1 - e^{-\lambda(t-T)})^{\frac{1}{2}}$$
$$e^{-\lambda t + \frac{1}{2(e^{\lambda(t-T)}-1)} - \frac{1}{2(e^{-\lambda T}-1)} + \int_0^t \frac{\sqrt{\lambda}}{(e^{\lambda(t'-T)}-1)} dW'_t}$$

As explained in the first section, the hedging strategy depends explicitly on the history of the Brownian motion. To obtain a return and a volatility that only depend on S and t, we will take $\bar{\phi_0}^*(t, S) = \phi^*(t, S, \mathbb{E}^{\mathbb{P}}[\Pi])$ as the demand of the large trader:

$$\bar{\phi}_0^*(t, S) = \frac{\lambda}{\mu_0} S^{-1} (\mathcal{G} - \Pi_0) \frac{e^{-\lambda t}}{(1 - e^{-\lambda T})} \tag{12.30}$$

We should note that the only models considered by Merton are models of dynamical portfolio optimization with no age effects [32].
Let us derive now the one-order correction to this hedging function. First, we give from (12.22)-(12.23) $\frac{\mu^2}{\sigma^2}$ up to the first order in ϵ (using (12.30)):

$$\frac{\mu^2}{\sigma^2} = \frac{1}{\sigma_0^2}(\mu_0^2 + 2\epsilon(\partial_t \bar{\phi}_0^* + \frac{1}{2}\sigma_0^2 S^2 \partial_S^2 \bar{\phi}_0^*)) \tag{12.31}$$
$$= \frac{1}{\sigma_0^2}(\mu_0^2 + 2\epsilon \bar{\phi}_0^*(-\lambda + \sigma_0^2)) \tag{12.32}$$

Then the equation (12.28) at order one is

$$\partial_t X_1 - (\mu_0 + \frac{1}{2}\sigma_0^2)\partial_u X_1 - \frac{1}{2}\sigma_0^2 \partial_u^2 X_1 = \frac{2\bar{\phi}_0^*}{\sigma_0^2}(-\lambda + \sigma_0^2) \tag{12.33}$$

with boundary condition $X_1(0, u) = 0$. The solution is then given by

$$X_1(t, S) = 2S^{-1} \frac{(\mathcal{G} - \Pi_0)}{(e^{\lambda T} - 1)} \frac{\lambda}{\mu_0} \frac{(\sigma_0^2 - \lambda)}{(\mu_0 - \lambda)\sigma_0^2} (e^{\lambda(T-t)} - e^{\mu_0(T-t)}) \tag{12.34}$$

and after a sightly long computation we find the correction to ϕ_0^* which depends on S, t and Π:

$$\phi^*(t, S, \Pi) = \frac{(\Pi - \mathcal{G})S^{-1}\lambda}{\mu_0(e^{\lambda(t-T)} - 1)} \left(1 + \epsilon \bar{\phi}_0^* \frac{(-\lambda + \sigma_0^2 + \mu_0)}{\mu_0} \right.$$
$$\left. - \epsilon \frac{\mu_0}{\lambda} X_1 - \epsilon \frac{X_1}{(1 - e^{-\lambda(t-T)})} \right)$$

To obtain the next corrections for the implied volatility and return, one should take $\phi^*(t, S, \mathbb{E}[\Pi])$ as the demand of the large trader. By plugging this expression into the volatility (12.22), we found that the large trader strategy induces a local volatility model (and therefore an implied volatility) which is given at the first-order by:

$$\sigma(t, S) = \sigma_0(1 - \epsilon\alpha e^{-\lambda t} S^{-1}) + o(\epsilon^2) \tag{12.35}$$

with α a constant depending on the characteristic of the portfolio (Π_0, $\mathcal{G}$, T, μ_0, σ_0).

Conclusion

In this chapter, we have explained how to incorporate easily the effect of the hedging strategy in the market prices. The derivative hedging gives a negative effect and the portfolio optimization hedging a positive one. The main difficulty with our optimization scheme was that the hedging function depends on age effects, which was simply solved by taking the mean over the portfolio value for S fixed. As explained, the portfolio optimization and the option pricing lead to relatively simple non-linear PDEs (12.25)-(12.17) that can be numerically solved.

Appendix A

Saddle-Point Method

Classical references for the saddle-point methods are [5] and [15].
Let Ω be a bounded domain on $\mathbb{R}^n$, $S : \Omega \to \mathbb{R}$, $f : \Omega \to \mathbb{R}$ and $\lambda > 0$ be a large positive parameter. The *Laplace method* consists in studying the asymptotics as $\lambda \to \infty$ of the multi-dimensional Laplace integrals

$$F(\lambda) = \int_\Omega f(x) e^{-\lambda S(x)} dx$$

Let S and f be smooth functions and we assume that the function S has a minimum only at one interior non-degenerate critical point $x_0 \in \Omega$:

$$\nabla_x S(x_0) = 0 \ , \ \nabla_x^2 S(x_0) > 0 \ , \ x_0 \in \Omega$$

x_0 is called the *saddle-point*. Then in the neighborhood of x_0 the function S has the following Taylor expansion

$$S(x) = S(x_0) + \frac{1}{2}(x - x_0)^\dagger \nabla_x^2 S(x_0)(x - x_0) + o((x - x_0)^3)$$

As $\lambda \to \infty$, the main contribution of the integral comes from the neighborhood of x_0. Replacing the function f by its value at x_0, we obtain a Gaussian integral where the integration over x can be performed

$$\begin{aligned} F(\lambda) &\approx f(x_0) e^{-\lambda S(x_0)} \int_\Omega e^{-\frac{\lambda}{2}(x-x_0)^\dagger \nabla_x^2 S(x_0)(x-x_0)} dx \\ &\approx f(x_0) e^{-\lambda S(x_0)} \int_{\mathbb{R}^n} e^{-\frac{\lambda}{2}(x-x_0)^\dagger \nabla_x^2 S(x_0)(x-x_0)} dx \end{aligned}$$

One gets the leading asymptotics of the integral as $\lambda \to \infty$

$$F(\lambda) \sim f(x_0) e^{-\lambda S(x_0)} \left(\frac{2\pi}{\lambda}\right)^{\frac{n}{2}} [\det(\nabla_x^2 S(x_0))]^{-\frac{1}{2}}$$

More generally, doing a Taylor expansion at the n-th order for S (resp. $n-2$ order for f) around $x = x_0$, we obtain

$$F(\lambda) \sim e^{-\lambda S(x_0)} \left(\frac{2\pi}{\lambda}\right)^{\frac{n}{2}} [\det(\nabla_x^2 S(x_0))]^{-\frac{1}{2}} \sum_{k=0}^{\infty} a_k \lambda^{-k}$$

with the coefficients a_k expressed in terms of the derivatives of the functions f and S at the point x_0. For example, at the first-order (in one dimension), we find

$$F(\lambda) \sim \sqrt{\frac{2\pi}{\lambda S''(c)}} e^{-\lambda S(x_0)} \left(f(x_0) \right.$$

$$\left. - \frac{1}{\lambda} \left(- \frac{f''(x_0)}{2S''(x_0)} + \frac{f(x_0) S^{(4)}(x_0)}{8 S''(x_0)^2} + \frac{f'(x_0) S^{(3)}(x_0)}{2S''(x_0)^2} - \frac{5 \left(S'''(x_0) \right)^2 f(x_0)}{24 S''(x_0)^3} \right) \right)$$

Appendix B

Monte-Carlo Methods and Hopf Algebra

B.1 Introduction

Since the introduction of the Black-Scholes paradigm, several alternative models which allow to better capture the risk of exotic options have emerged : local volatility models, stochastic volatility models, jump-diffusion models, mixed stochastic volatility-jump diffusion models, etc. With the growing complication of Exotic and Hybrid options that can involve many underlyings (equity assets, foreign currencies, interest rates), the Black-Scholes PDE, which suffers from the curse of dimensionality (the dimension should be strictly less than four in practice), cannot be solved by finite difference methods. We must rely on Monte-Carlo methods.

This appendix is organized as follows:

In the first section, we review basic features in Monte-Carlo simulation from a modern (algebraic) point of view: generation of random numbers and discretization of SDEs. A precise mathematical formulation involves the use of the Taylor-Stratonovich expansion (TSE) that we define carefully. Details can be found in the classical references [20], [28].

In the second section, we show that the TSE can be framed in the setting of Hopf algebras.[1] In particular, TSEs define group-like elements and solutions of SDEs can be written as exponentials of primitive elements (i.e., elements of the universal Lie algebra associated to the Hopf algebra). This section is quite technical and can be skipped by the reader.

This Hopf algebra structure allows to prove easily the Yamato theorem that we explained in the last section. As an application, we classify local volatility models that can be written as a functional of a Brownian motion and therefore can be simulated exactly. The use of Yamato's theorem allows us to reproduce and extend the results found by P. Carr and D. Madan in [69].

[1]The author would like to thank his colleague Mr. C. Denuelle for fruitful collaborations on this subject and for help in the writing of this appendix.

B.1.1 Monte Carlo and Quasi Monte Carlo

Let X_t be an n-dimensional Itô process. According to (2.31), the pricing of a derivative product requires the evaluation of

$$\mathbb{E}[f(X_T)] \tag{B.1}$$

Through a discretization scheme discussed in the next paragraph, we assume that X_T can be approached by a function g of a multidimensional Gaussian variable $G \sim N(0, \Sigma)$. To figure out an approximation of (B.1), one simulates M independent random variables $(G_m \sim N(0, \Sigma))_{1 \le m \le M}$ and computes the associated values $(X_T^m)_{1 \le m \le M}$. Under reasonable hypothesis on the payoff f and the function g, the strong version of the law of large numbers [26] entails that the empirical average $\frac{1}{M}\sum_{m=1}^{M} f(X_T^m)$ provides a good proxy for $\mathbb{E}[f(X_T)]$. Better yet the central limit theorem [26] states that the error

$$\varepsilon_M = \frac{1}{M}\sum_{m=1}^{M} f(X_T^m) - \mathbb{E}[f(X_T)]$$

is asymptotically normally distributed, with mean 0 and covariance $\frac{{}^t\Sigma\Sigma}{M}$. Thus there exists a constant C such that the $\mathbb{L}^2$ Monte Carlo error is

$$||\varepsilon_M||_{\mathbb{L}^2} \le \frac{C}{\sqrt{M}}$$

Even though the $O(\frac{1}{\sqrt{M}})$ bound is steady, variance reduction techniques such as control variates can considerably reduce the constant C. We refer to [20] for further discussions on these methods and the way they speed up the convergence of MC estimates.
As liquidity on the market expands, option pricing requires increasing precision, and the $O(\frac{1}{\sqrt{M}})$ MC bound is simply not good enough. To gain a factor 10 in accuracy the number of simulations must be increased by a factor 100. To elude this pitfall, Quasi Monte-Carlo (QMC) methods have been developed [20]. They consist in low-discrepancy sequences (Van Der Corput, Faure, Sobol...) filling up $(0,1)^d$ uniformly. Contrary to what is often assumed these sequences are not at all random. QMC trajectories are much too uniform to be random. Standard transformations convert the $(0,1)^d$-valued sequences into d-dimensional Gaussian sequences, and lead to well chosen solutions $(X_T^m)_{1 \le m \le M}$. The error coming from the QMC estimator $\frac{1}{M}\sum_{m=1}^{M} f(X_T^m)$ can be as good as $O(\frac{(\ln M)^d}{M})$, which is *nearly* $O(\frac{1}{M})$.

B.1.2 Discretization schemes

For the sake of simplicity, we assume that we are trying to simulate on the time interval $[0, T]$ a stochastic process driven by the one dimensional SDE :

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

Let $(t_n)_{0\leq n\leq N}$ be a subdivision of $[0,T]$

$$(\Delta^n t = t_{n+1} - t_n)_{0\leq n\leq N-1}$$

and

$$\Delta W = (\Delta^n W = W_{t_{n+1}} - W_{t_n})_{0\leq n\leq N-1}$$

From the basic properties of the Brownian motion it is clear that $\Delta W \sim N(0,\Delta t)$ where Δt is the diagonal matrix whose entries are the $(\Delta^n t)$. MC simulation or QMC can therefore be employed to draw M paths for ΔW. As stated in the previous paragraph, discretization schemes will permit the transformation of these paths into paths of the underlying process. The SDE above yields

$$X_{t_{n+1}} = X_{t_n} + \int_{t_n}^{t_{n+1}} \mu(X_t)dt + \int_{t_n}^{t_{n+1}} \sigma(X_t)dW_t$$

Hence a very intuitive discretization is provided by the well-known *Euler scheme* ([28],[20]) which generates the following path $\overline{X}$ conditionally on ΔW:

$$\begin{aligned}\overline{X}_{t_0} &= X_0\\ \overline{X}_{t_{n+1}} &= \overline{X}_{t_n} + \mu(\overline{X}_{t_n})\Delta^n t + \sigma(\overline{X}_{t_n})\Delta^n W\end{aligned}$$

However the Euler scheme does have an inconvenient: the two approximations it consists in are not of the same order

$$\int_{t_n}^{t_{n+1}} \mu(X_t)dt = \mu(\overline{X}_{t_n})\Delta^n t + o(\Delta^n t)$$

whereas

$$\int_{t_n}^{t_{n+1}} \sigma(X_t)dW_t = \sigma(\overline{X}_{t_n})\Delta^n W + o(\sqrt{\Delta^n t})$$

It would therefore make sense to develop :

$$\begin{aligned}\int_{t_n}^{t_{n+1}} \sigma(X_t)dW_t &= \int_{t_n}^{t_{n+1}} \sigma\left(X_{t_n} + \int_{t_n}^{t} \mu(X_s)ds + \int_{t_n}^{t} \sigma(X_s)dW_s\right)dW_t\\ &= \sigma(X_{t_n})\Delta^n W + \sigma'(X_{t_n})\sigma(X_{t_n})\int_{t_n}^{t_{n+1}}(\int_{t_n}^{t} dW_s)dW_t + o(\Delta^n t)\\ &= \sigma(X_{t_n})\Delta^n W + \frac{\sigma'(X_{t_n})\sigma(X_{t_n})}{2}\left((\Delta^n W)^2 - \Delta^n t\right) + o(\Delta^n t)\end{aligned}$$

This refinement is at the origin of the one-dimensional *Milstein scheme*:

$$\begin{aligned}\widehat{X}_{t_0} &= X_0\\ \widehat{X}_{t_{n+1}} &= \widehat{X}_{t_n} + \mu(\widehat{X}_{t_n})\Delta^n t + \sigma(\widehat{X}_{t_n})\Delta^n W + \frac{\sigma'(\widehat{X}_{t_n})\sigma(\widehat{X}_{t_n})}{2}\left((\Delta^n W)^2 - \Delta^n t\right)\end{aligned}$$

To study the performance of discretization schemes the literature distinguishes between *strong and weak order of convergence* ([28], [20]):

DEFINITION B.1 Strong/Weak weak order of convergence *Noting* $h = \max_{n=0,..,N-1} (\Delta^n t)$, *a scheme* $\bar{X}$ *is said to be of strong order* $\beta > 0$ *if there exists a constant* c *such that*

$$\mathbb{E}[\| \bar{X}(t_N) - X(T) \|] \leq ch^{\beta}$$

for some vector norm $\| \cdot \|$. *It is said to be of weak order* $\beta > 0$ *if there exists a constant* c *such that*

$$| \mathbb{E}[f(\bar{X}(t_N)] - \mathbb{E}[f(X(T))] | \leq ch^{\beta}$$

for all f *polynomially bounded in* $C^{2\beta+2}$.

Under Lipschitz-type conditions on μ and σ the Euler scheme can be shown to be of strong order $\frac{1}{2}$ and of weak order 1, whereas the Milstein scheme is of strong and weak order 1.
Writing a multi-dimensional generalization of the Milstein scheme would involve adding antisymmetric Wiener iterated integrals, called *Lévy areas*

$$\mathcal{A}^{ij} = \int_0^1 W_t^i dW_t^j - \int_0^1 W_t^j dW_t^i \tag{B.2}$$

The standard approximation for this term requires several additional random numbers [28]. There are however approaches to avoid the drawing of many extra random numbers by using the relation of this integral to the Lévy area formula [124]:

$$\mathbb{E}[e^{i\lambda\mathcal{A}^{ij}} | W_1^i + \imath W_1^j = z] = \frac{\lambda}{\sinh\lambda} e^{-\frac{|z|^2}{2}(\lambda \coth\lambda - 1)}$$

From this characteristic function, Lévy areas and Brownian motion W_t^i can be simulated exactly. However, this can become quite costly in computational expense.

B.1.3 Taylor-Stratonovich expansion

Using the tools described in the previous sections, we explain the *Taylor-Stratonovich* expansion (TSE) equivalent to a Taylor expansion in the deterministic case. Note that a similar expansion, called *Taylor-Itô* expansion, that uses the Itô calculus exists. In order to be simulated, the discretization scheme eventually needs to be set in Itô form. However, we prefer to present TSE as its algebraic structure is simpler.

Let X_t be an n-dimensional Itô process following the Stratonovich SDE

$$dX_t = V_0 dt + \sum_{i=1}^{m} V_i \diamond dW_t^i \ , \ X_{t=0} = X_0 \in \mathbb{R}^n$$

By introducing the notation $dW_t^0 \equiv dt$, we have

$$dX_t = \sum_{i=0}^{m} V_i \diamond dW_t^i \tag{B.3}$$

Without loss of generality, we assume that we have a time-homogeneous SDE. A time-inhomogeneous SDE can be written in this normal form (B.3) by including an additional state X_t^{n+1}

$$dX_t^{n+1} = dt$$

The Stratonovich calculus entails that for f in $C^1(\mathbb{R}^m, \mathbb{R})$

$$f(X_T) = \sum_{i=0}^{m} \int_0^T \ V_i f(X_t) \diamond dW_t^i$$

By iterating this equation, we obtain

$$f(X_T) = f(X_0) + \sum_{i=0}^{m} \left(V_i f(X_0) \int_0^T \ \diamond dW_t^i \right.$$
$$\left. + \sum_{j=0}^{m} \int_0^T \left(\int_0^t V_i V_j f(X_s) \diamond dW_s^j \right) \diamond dW_t^i \right)$$

With a repeated application of the Stratonovich formula, the iterated integral defined as

$$\int_{0 \leq t_1 < \ldots < t_k \leq T} \ \diamond dW_{t_1}^{i_1} \ldots \diamond dW_{t_k}^{i_k}$$

appears naturally. From a $\mathbb{L}^2$-norm perspective, $\diamond dW_t^0$ counts as $[dt]$ whereas $\diamond dW_t^i$ $(i = 1, ..., m)$ counts as $[\sqrt{dt}]$. As a consequence, following [28], we introduce a graduation on multi-indexes $(i_1, ..., i_k)$ by

$$\deg(i_1, \cdots, i_k) = k + \sharp\{j \text{ such as } i_j = 0\}$$

and take

$$\mathcal{A}_r = \{(i_1, ..., i_k) \text{ such as } \deg(i_1, ..., i_k) \leq r\}$$

Finally, continuing the calculation sketched above, we obtain [28]:

THEOREM B.1 Taylor-Stratonovich expansion

$$f(X_T) = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_r} V_{i_1}\ldots V_{i_k} f(X_0) \int_{0\leq t_1<\ldots<t_k\leq T} \diamond dW_{t_1}^{i_1}\ldots\diamond dW_{t_k}^{i_k} + R_r(T,X_0,f)$$

where $\sup_{X_0\in\mathbb{R}^n} ||R_r(T,X_0,f)||_2 \leq CT^{\frac{r+1}{2}} \sup_{(i_1,\ldots,i_k)\in\mathcal{A}_{r+2}\backslash\mathcal{A}_r} ||V_{i_1}\ldots V_{i_k} f||_\infty$.

REMARK B.1 Actually, the TSE remains correct if we replace $\diamond dW_t$ by any continuous semi-martingale. □

TSE provides a link with the discretization schemes mentioned earlier. Taking for instance $r=1$ and $f(X_t)=X_t$ yields the Euler scheme (if writing in Itô form)

$$X_{t_{n+1}} = X_{t_n} + V_0\left(X_{t_n}\right)\Delta^n t + \sum_{i=1}^m V_i\left(X_{t_n}\right)\diamond\Delta^n W^i$$

For $r=2$, we have

$$\begin{aligned} X_{t_{n+1}} &= X_{t_n} + V_0\left(X_{t_n}\right)\Delta^n t + \sum_{i=1}^m V_i\left(X_{t_n}\right)\diamond\Delta^n W^i \\ &+\frac{1}{2}\sum_{1\leq i\leq j\leq m}\left(\frac{1}{2}[V_i,V_j]\left(X_{t_n}\right)\mathcal{A}_\diamond^{ij} + V_iV_j\left(X_{t_n}\right)\Delta^n W^i\diamond\Delta^n W^j\right)\end{aligned}$$

which happens to be strictly equivalent to the multidimensional Milstein scheme. Here $[\cdot,\cdot]$ is the Lie bracket (see exercise 4.3) defined by

$$[V_i,V_j] = V_iV_j - V_jV_i$$

Note that when the Lie algebra generated by the $\{V_i\}_{i=1,\cdots,m}$ is *Abelian*

$$[V_i,V_j]=0\ ,\ \forall\, i,j=1,\cdots,m$$

the Milstein scheme simplifies as it does not require the simulation of Lévy areas.

B.2 Algebraic Setting

As seen in the previous section, the MC methods rely on discretization of SDEs. Convergence issues require in certain cases to select a very small discretization time-step. The simulation can become quite time-consuming for multi-asset hybrid options. A possible resolution of this problem is to represent asset prices as functionals of simple processes that can be simulated exactly (Brownian motion, Ornstein-Uhlenbeck, ...). This is the starting point of Markov functional LMMs (see exercise 8.1) and forward variance swap models [62], [68]. In the last section, we state the Yamato theorem giving a necessary and sufficient condition for representing solutions of SDEs as functionals of Brownian motions. This theorem originates from the Hopf algebra structure of (Chen) iterated integrals which appear in TSEs.

B.2.1 Hopf algebra

To study algebraic relations between iterated integrals we replace the vector fields $V_0, V_1, ..., V_m$ by letters $\varepsilon_0, \varepsilon_1, ..., \varepsilon_m$.

B.2.1.1 Algebra

We endow the real vector space $\mathcal{H}$ generated by $\{\varepsilon_i\}_{i=0,\cdots,m}$ with an unital graded algebra structure $(\mathcal{H}, +, ., \varepsilon)$. The noncommutative product corresponds to the *concatenation product*

$$. : (\varepsilon_i, \varepsilon_j) \longmapsto \varepsilon_i.\varepsilon_j = \varepsilon_i\varepsilon_j$$

The *words* are obtained by sums of concatenations of generators. The neutral element . is the empty word ε. It satisfies for any word w:

$$w.\varepsilon = \varepsilon.w = w$$

Consistently with TSEs, we introduce a graduation on words equivalent to the graduation introduced for multi-indices

$$\deg(\varepsilon_{i_1}...\varepsilon_{i_k}) = \deg(i_1, ..., i_k)$$

We have

$$\mathcal{H} = \bigoplus_{r=0}^{\infty} \mathcal{H}_r$$

We draw the reader's attention to the fact that in $\mathcal{H}_r$ all words of graduation greater than r have a coefficient set to 0. For instance in $\mathcal{H}_3$ we get

$$(2\varepsilon_1 + \varepsilon_0\varepsilon_d - \varepsilon_1\varepsilon_0\varepsilon_1).\varepsilon_1 = 2\varepsilon_1\varepsilon_1 + \varepsilon_0\varepsilon_d\varepsilon_1$$

$(\mathcal{H}, +, ., \varepsilon)$ then has the structure of a graded associative algebra with unit ε. The product . is associative and commutes with the identity Id on $\mathcal{H}$, which we will summarize by the diagram B.1.

REMARK B.2 Note that the tensor algebra $(\mathcal{H} \otimes \mathcal{H}, +, .)$ is a graded algebra (its graduation is induced by $\deg(x \otimes y) = \deg(x) + \deg(y)$) with operations $+$ and . deriving from those of $\mathcal{H}$. □

TABLE B.1: Associativity diagram.

$$
\begin{array}{cccccc}
 & & . \otimes Id & & \\
 & \mathcal{H}_r \otimes \mathcal{H}_r \otimes \mathcal{H}_r & \longrightarrow & \mathcal{H}_r \otimes \mathcal{H}_r & \\
Id \otimes . & \downarrow & & \downarrow & . \\
 & \mathcal{H}_r \otimes \mathcal{H}_r & \longrightarrow & \mathcal{H}_r & \\
 & & . & &
\end{array}
$$

TABLE B.2: Co-associativity diagram.

$$
\begin{array}{cccccc}
 & & \Delta \otimes Id & & \\
 & \mathcal{H}_r \otimes \mathcal{H}_r \otimes \mathcal{H}_r & \longleftarrow & \mathcal{H}_r \otimes \mathcal{H}_r & \\
Id \otimes \Delta & \uparrow & & \uparrow & \Delta \\
 & \mathcal{H}_r \otimes \mathcal{H}_r & \longleftarrow & \mathcal{H}_r & \\
 & & \Delta & &
\end{array}
$$

B.2.1.2 Co-product

Next we define the *co-product* $\Delta : \mathcal{H} \otimes \mathcal{H} \to \mathcal{H}$ on the generators

$$\Delta(\varepsilon_i) = \varepsilon_i \otimes \varepsilon + \varepsilon \otimes \varepsilon_i$$

Then we extend the action of Δ on words by imposing that Δ is a morphism for the concatenation product, meaning that for two noncommutative polynomials x and y

$$\Delta(x.y) = \Delta(x).\Delta(y)$$

The action of Δ can be seen (see diagram B.2) as reversing the arrows in the diagram B.1

B.2.1.3 Co-unit

We also introduce a *co-unit* $\eta : \mathcal{H}_r \to \mathbb{R}$ by

$$\eta(x) = 1 \ , \ x \in \mathcal{H}_r$$

B.2.1.4 Antipode

To complete the setting we bring up a final operation $a : \mathcal{H}_r \to \mathcal{H}_r$ called an *antipode.* a is a linear application characterized by its action on words:

$$a(\varepsilon_{i_1}\varepsilon_{i_2}...\varepsilon_{i_k}) = (-1)^k \varepsilon_{i_k}...\varepsilon_{i_2}\varepsilon_{i_1}$$

It follows from this definition that for any two noncommutative polynomials x, y in $\mathcal{H}_r$ we have an anti-homomorphism

$$a(xy) = a(y)a(x)$$

Finally $(\mathcal{H}, +, ., \Delta, a, \varepsilon)$ endowed with these five operations is a Hopf algebra (see [17] for details).[2]

DEFINITION B.2 Hopf algebra *A Hopf algebra $\mathcal{H}$ is a vector space endowed with five operations*

$$\begin{aligned}
\cdot &: \mathcal{H} \otimes \mathcal{H} \to \mathcal{H} \text{ multiplication} \\
\varepsilon &: \mathbb{R} \to \mathcal{H} \text{ unit map} \\
\Delta &: \mathcal{H} \to \mathcal{H} \otimes \mathcal{H} \text{ coproduct} \\
\eta &: \mathcal{H} \to \mathbb{R} \text{ counit} \\
a &: \mathcal{H} \to \mathcal{H} \text{ antipode}
\end{aligned}$$

which possess the following properties

$$\begin{aligned}
& m \circ (Id \times m) = m \circ (m \times Id) \text{ (associativity)} \\
& m \circ (Id \times \varepsilon) = id = m \circ (\varepsilon \times Id) \text{ (existence of unit)} \\
& (Id \times \Delta) \circ \Delta = (\Delta \times Id) \circ \Delta \text{ (coassociativity)} \\
& (\eta \times Id) \circ \Delta = Id = (Id \times \eta) \circ \Delta \text{ (existence of counit)} \\
& m \circ (Id \times a) \circ \Delta = \varepsilon \circ \eta = m \circ (a \times Id) \circ \Delta \\
& \Delta \circ m = (m \times m) \circ (\Delta \times \Delta)
\end{aligned}$$

There happens to be very particular elements of $\mathcal{H}$ called *primitive elements*

DEFINITION B.3 Primitive element *An element $\mathcal{L}$ satisfying*

$$\Delta(\mathcal{L}) = \mathcal{L} \otimes \varepsilon + \varepsilon \otimes \mathcal{L}$$

is called a primitive element.

By equipping the set $\mathcal{G}$ of primitive elements with the bracket defined by

$$[\mathcal{L}, \mathcal{L}'] \equiv \mathcal{L}.\mathcal{L}' - \mathcal{L}'.\mathcal{L}$$

it follows that $(\mathcal{G}, [\cdot, \cdot])$ forms a Lie algebra.

PROOF Indeed if $\mathcal{L}, \mathcal{L}' \in \mathcal{G}$,

$$\begin{aligned}
\Delta(\mathcal{L}.\mathcal{L}') &= \Delta(\mathcal{L}).\Delta(\mathcal{L}') \\
&= (\mathcal{L} \otimes 1 + 1 \otimes \mathcal{L}).(\mathcal{L}' \otimes 1 + 1 \otimes \mathcal{L}') \\
&= \mathcal{L}\mathcal{L}' \otimes 1 + 1 \otimes \mathcal{L}\mathcal{L}' + \mathcal{L} \otimes \mathcal{L}' + \mathcal{L}' \otimes \mathcal{L}
\end{aligned}$$

[2]These operations are so rich and beautiful that some physicists conjecture that God should surely be a Hopf algebra.

Hence

$$\begin{aligned}\Delta([\mathcal{L},\mathcal{L}']) &= \Delta(\mathcal{L}.\mathcal{L}') - \Delta(\mathcal{L}'.\mathcal{L})\\ &= \mathcal{L}\mathcal{L}'\otimes 1 + 1\otimes \mathcal{L}\mathcal{L}' - \mathcal{L}'\mathcal{L}\otimes 1 - 1\otimes \mathcal{L}'\mathcal{L}\\ \Delta([\mathcal{L},\mathcal{L}']) &= [\mathcal{L},\mathcal{L}']\otimes 1 + 1\otimes [\mathcal{L},\mathcal{L}']\end{aligned}$$

□

DEFINITION B.4 Grouplike elements *Elements g of $\mathcal{H}$ satisfying*

$$\Delta(g) = g\otimes g$$

are called grouplike elements.

The set G of grouplike elements forms a group for the concatenation product.

PROOF Indeed if $g, h \in G$

$$\begin{aligned}\Delta(g.h) &= \Delta(g).\Delta(h)\\ &= (g\otimes g).(h\otimes h)\\ \Delta(g.h) &= gh\otimes gh\end{aligned}$$

□

DEFINITION B.5 Exponential map *We define an exponential map on $\mathcal{H}_r$ by the usual series expansion*

$$\exp(\mathcal{L}) = \sum_{k=0}^{\infty}\frac{1}{k!}\mathcal{L}^k$$

The expression above is a finite sum because $\deg(\mathcal{L}^k) = k\deg(\mathcal{L})$ and $\mathcal{L}^k$ is equal to 0 as soon as $k\deg(\mathcal{L}) > r$.

This definition takes its sense from the fact that the exponential of a primitive element in $\mathcal{G}_r$ is a group-like element in G_r:

PROPOSITION B.1

$$G_r = \exp(\mathcal{G}_r)$$

PROOF We do the proof in one way: If $\mathcal{L} \in \mathcal{G}_r$ then

$$
\begin{aligned}
\Delta(\exp(\mathcal{L})) &= \Delta(\sum_{k=0}^{\infty} \frac{\mathcal{L}^k}{k!}) \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \Delta(\mathcal{L})^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} (\mathcal{L} \otimes \varepsilon + \varepsilon \otimes \mathcal{L})^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} (\mathcal{L}^j \otimes \mathcal{L}^{k-j}) \\
&= \exp(\mathcal{L}) \otimes \exp(\mathcal{L})
\end{aligned}
$$

∎

B.2.2 Chen series

Now that this setting has been introduced we can almost reach our goal and see how it appears in TSE. The idea is to use a morphism

$$\Gamma_r : \varepsilon_{i_1} ... \varepsilon_{i_k} . \varepsilon \in \mathcal{H}_r \longmapsto V_{i_1} ... V_{i_k}(X_0)$$

For a semi-martingale ω we define its *Chen series* in $\mathcal{H}_r$ as the pre-image of the Taylor-Stratonovich expansion:

$$X_{0,1}(\omega) = \sum_{(i_1,...,i_k) \in \mathcal{A}_r} \varepsilon_{i_1} ... \varepsilon_{i_k} \int_{0 \leq t_1 < ... < t_k \leq 1} d\omega^{i_1}(t_1) ... d\omega^{i_k}(t_k)$$

The introduction of the Hopf algebra setting allows us to provide a simple proof for the major result that follows

THEOREM B.2 Chen [74]
$X_{0,1}(\omega)$ is a group-like element of $\mathcal{H}_r$.

PROOF (Sketch) Assuming $m = 2$ for simplicity, we prove the result for $r = 2$. In that case

$$
\begin{aligned}
X_{0,1}(\omega) = \varepsilon &+ \varepsilon_1 \int_{0 \leq t \leq 1} d\omega^1(t) + \varepsilon_2 \int_{0 \leq t \leq 1} d\omega^2(t) + \varepsilon_0 \\
&+ \varepsilon_1 \varepsilon_2 \int_{0 \leq t \leq 1} \omega^1(t) d\omega^2(t) + \varepsilon_2 \varepsilon_1 \int_{0 \leq t \leq 1} \omega^2(t) d\omega^1(t)
\end{aligned}
$$

Remembering the graduation introduced on $\mathcal{H}_r \otimes \mathcal{H}_r$ we compute

$$\begin{aligned}
X_{0,1}(\omega) \otimes X_{0,1}(\omega) = \varepsilon \otimes \varepsilon &+ (\varepsilon_0 \otimes \varepsilon + \varepsilon \otimes \varepsilon_0) \\
&+ (\varepsilon_1 \otimes \varepsilon + \varepsilon \otimes \varepsilon_1) \int_{0 \le t \le 1} d\omega^1(t) \\
&+ (\varepsilon_2 \otimes \varepsilon + \varepsilon \otimes \varepsilon_2) \int_{0 \le t \le 1} d\omega^2(t) \\
&+ (\varepsilon_1\varepsilon_2 \otimes \varepsilon + \varepsilon \otimes \varepsilon_1\varepsilon_2) \int_{0 \le t \le 1} \omega^1(t) d\omega^2(t) \\
&+ (\varepsilon_2\varepsilon_1 \otimes \varepsilon + \varepsilon \otimes \varepsilon_2\varepsilon_1) \int_{0 \le t \le 1} \omega^2(t) d\omega^1(t) \\
&+ (\varepsilon_1 \otimes \varepsilon_2 + \varepsilon_2 \otimes \varepsilon_1) \int_{0 \le t \le 1} d\omega^1(t) \int_{0 \le t \le 1} d\omega^2(t)
\end{aligned}$$

On the other hand we have

$$\begin{aligned}
\Delta(X_{0,1}(\omega)) = \varepsilon \otimes \varepsilon &+ (\varepsilon_1 \otimes \varepsilon + \varepsilon \otimes \varepsilon_1) \int_{0 \le t \le 1} d\omega^1(t) \\
&+ (\varepsilon_2 \otimes \varepsilon + \varepsilon \otimes \varepsilon_2) \int_{0 \le t \le 1} d\omega^2(t) \\
&+ (\varepsilon_0 \otimes \varepsilon + \varepsilon \otimes \varepsilon_0) + \Delta(\varepsilon_1\varepsilon_2) \int_{0 \le t \le 1} \omega^1(t) d\omega^2(t) \\
&+ \Delta(\varepsilon_2\varepsilon_1) \int_{0 \le t \le 1} \omega^2(t) d\omega^1(t)
\end{aligned}$$

Noting that

$$\Delta(\varepsilon_1\varepsilon_2) = \Delta(\varepsilon_1)\Delta(\varepsilon_2) = (\varepsilon_1\varepsilon_2 \otimes \varepsilon + \varepsilon \otimes \varepsilon_1\varepsilon_2 + \varepsilon_1 \otimes \varepsilon_2 + \varepsilon_2 \otimes \varepsilon_1)$$

we see that

$$\Delta(X_{0,1}(\omega)) = X_{0,1}(\omega) \otimes X_{0,1}(\omega)$$

is equivalent to

$$\begin{aligned}
\int_{0 \le t \le 1} d\omega^1(t) \int_{0 \le t \le 1} d\omega^2(t) &= \int_{0 \le t \le 1} \omega^1(t) d\omega^2(t) + \int_{0 \le t \le 1} \omega^2(t) d\omega^1(t) \\
&= \int_{0 < s \le t < 1} d\omega^1(s) d\omega^2(t) + \int_{0 < s \le t < 1} d\omega^2(s) d\omega^1(t)
\end{aligned}$$

This last is a consequence of Fubini's theorem. The same proof holds for a general r, using the fact that Δ is a morphism for words. ∎

As a consequence, if ω is a semi-martingale, we can write

$$X_{0,1}(\omega) = \exp(\mathcal{L})$$

for a primitive element (Lie polynomial) $\mathcal{L}$ in $\mathcal{G}$.

B.3 Yamato's theorem

By applying Γ_r on the Chen series $X_{0,1}(\diamond W)$, we obtain that the TSE up to degree r on the interval $[0,1]$ can be written as

$$X_{0,1}(\diamond W) = \exp(\mathcal{L})(X_0)$$

$X_{0,1}(\diamond W)$ is the solution of a flow at time $t = 1$ driven by the (random) vector field $\Gamma_r(\mathcal{L})$:

$$\frac{dX_{0,1}}{dt} = \Gamma_r(\mathcal{L})$$

For completeness, we recall the definition of a flow:

DEFINITION B.6 Flow *A flow on a manifold M generated by a vector field Z at time t, denoted by e^{tZ}, is a one parameter group of transformations $\phi_t(x_0) \equiv e^{tZ} x_0$ which satisfies:*

- *For each $t \in [0, \infty)$, ϕ_t is a diffeomorphism of M.*
- *$\phi_t \circ \phi_s = \phi_{t+s}$ for any $t, s \in [0, \infty)$.*
- *$\lim_{t \to 0} \phi_t(x_0) = x_0$.*
- *Solution to the ODE $\frac{d\phi_t(x_0)}{dt} = Z(\phi_t(x_0))$.*

REMARK B.3 Beware, $\phi_t(x_0) \equiv e^{tZ} x_0$ is only a notation, and the solution at time t to the ODE $d\phi_t(x_0) = Z(\phi_t(x_0))dt$ is not necessarily $e^{\int_0^t Z(\phi_u(x_0))du} x_0$. This will only be the case if the vector fields $\{Z(\phi_u(x_0))\}_{0 \leq u \leq t}$ commute. □

Before stating the Yamato theorem, we introduce a last object, the free Lie algebra. (Note that this free Lie algebra is slightly different (resp. identical) from the one which appears in the Hörmander (resp. Frobenius) theorem.)

DEFINITION B.7 Free Lie algebra *The free Lie algebra spanned by the generators $(V_0, V_1, ..., V_m)$ is the Lie algebra consisting of linear combinations of elements of the form*

$$[V_{i_1}, ..., [V_{i_{k-1}}, V_{i_k}]]$$

with $(i_1, ..., i_k) \in \{0, ..., m\}$.

Finally, from the Chen theorem B.2, we have

THEOREM B.3 Yamato [143]
Let X_t be a solution of

$$dX_t = V_0 dt + \sum_{j=1}^{m} V_j \diamond dW_t^j \tag{B.4}$$

with $X_{t=0} = X_0$. Then it is represented as

$$X_t = \exp\left(\mathcal{L}_t\right) X_0$$

where

$$\mathcal{L}_t = tV_0 + \cdots + W_t^m V_m + \sum_{r=2} \sum_{J \in \mathcal{A}_r} c_J W_t^J V^J$$

with the iterated Brownian integrals

$$W_t^J = \int_{0 \leq t_{j_1} < \cdots < t_{j_m} \leq t} dW_{u_1}^{j_1} \cdots \diamond dW_{u_m}^{j_m}$$

and

$$V^J = [\cdots [V_{j_1}, V_{j_2}] \cdots V_{j_m}] \ , \ J = (j_1, \cdots, j_m)$$

c_J are some constants.

COROLLARY B.1
Let X_t be a solution of (B.4). If $V_0, V_1, \cdots, V_m$ forms an Abelian Lie algebra, i.e., $[V_i, V_j] = 0$ for each i and j, then the solution of the SDE above starting at X_0 is represented as

$$X_t = \exp\left(tV_0 + W_t^1 V_1 + \cdots W_t^m V_m\right) X_0 \tag{B.5}$$

This means that X_t can be represented as a functional of the Brownian motions $\{W_t^i\}_{i=1,\cdots,m}$. The functional is obtained by solving the flow (B.5). As a consequence, the SDE can be simulated exactly. In the following example, we classify the local volatility function for which asset prices can be written as a functional of a Brownian motion.

Example B.1 Exact discretization of (Abelian) LV models
Let us consider the local volatility model

$$df_t = \sigma(t, f_t) dW_t \tag{B.6}$$

Using the Stratonovich calculus, this gives the time-homogeneous SDE

$$\begin{aligned} df_t &= -\frac{1}{2}\sigma(u, f_t)'\sigma(u, f_t)dt + \sigma(u, f_t) \diamond dW_t \\ du &= dt \end{aligned}$$

for which we read the vector fields

$$\begin{aligned} V_0 &= -\frac{1}{2}\sigma(u, f)'\sigma(u, f)\partial_f + \partial_u \\ V_1 &= \sigma(u, f)\partial_f \end{aligned}$$

The prime $'$ indicates a derivative with respect to f. From corollary B.1, f_t is a functional $f_t \equiv \Phi(t, W_t)$ of the time and the Brownian W_t if and only

$$\begin{aligned} [V_0, V_1] &= \left(\partial_u \sigma(u, f) + \frac{1}{2}\sigma(u, f)^2 \sigma(u, f)''\right)\partial_f \\ &= 0 \end{aligned} \tag{B.7}$$

Setting $F(u, f) \equiv \sigma(u, f)^2$, the constraint above gives

$$\partial_u F + \frac{FF''}{2} - \frac{F'^2}{4} = 0$$

A solution is the hyperbolic model

$$\sigma(t, f) = \sqrt{a(t)f^2 + b(t)f + c(t)}$$

with

$$\begin{aligned} a(t) &= a_0 \\ b(t) &= b_0 \\ c(t) &= c_0 e^{-a_0 t} + \frac{b_0^2}{4a_0}\left(1 - e^{-a_0 t}\right) \end{aligned}$$

∎

Example B.2 Nilpotent step 1 LV models

According to the Yamato theorem, the forward f_t driven by the SDE (B.6) is a functional $f_t = \Phi\left(W_t, W_t^{[01]}\right)$ of the Brownian motion W_t and the stochastic iterated integral

$$\begin{aligned} W_t^{[01]} &\equiv \frac{1}{2}\int_{0<s<u<t} (ds dW_u - dW_s du) \\ &= \int_0^t u dW_u - \frac{t}{2}W_t \end{aligned}$$

if and only if the vectors V_0, V_1 generate a step 1 nilpotent Lie algebra:

$$[V_1, [V_0, V_1]] = 0 \ , \ [V_0, [V_0, V_1]] = 0$$

From (B.7), we get

$$[V_1, [V_0, V_1]] = \left(\sigma \partial_u \sigma' - \sigma' \partial_u \sigma + \frac{1}{2} \sigma^3 \sigma''' + \frac{1}{2} \sigma^2 \sigma' \sigma'' \right) \partial_f = 0$$

$$\begin{aligned} [V_1, [V_0, V_1]] = & \left(-\frac{1}{2} \sigma \sigma' \partial_u \sigma' + \partial_u^2 \sigma + \frac{1}{2} \partial_u \left(\sigma^2 \sigma'' \right) + \frac{1}{2} \partial_u \sigma \left(\sigma \sigma' \right)' \right) \partial_f \\ & + \left(-\frac{1}{4} \sigma \sigma' \left(\sigma^2 \sigma'' \right)' + \frac{1}{4} \sigma^2 \sigma'' \left(\sigma \sigma' \right)' \right) \partial_f = 0 \end{aligned}$$

Note that the processes W_t and $\int_0^t u dW_u$ are normal r.v. with zero mean and variance (using Itô's isometry)

$$\begin{pmatrix} t & \frac{t^2}{2} \\ \frac{t^2}{2} & \frac{t^3}{3} \end{pmatrix}$$

The forward can therefore be simulated exactly and efficiently in a nilpotent step 1 LV model.
When $\sigma = \sigma(f)$ is time-homogeneous, the constraints above reduce to

$$\sigma \sigma''' + \sigma' \sigma'' = 0 \tag{B.8}$$

$$-\left(\sigma' \right)^2 \sigma'' - \sigma \sigma' \sigma''' + \sigma \left(\sigma'' \right)^2 = 0 \tag{B.9}$$

The first equation (B.8) gives

$$\sigma \sigma'' = A \tag{B.10}$$

with A an integration constant. The equation (B.9) reduces

$$\frac{A^2}{\sigma} = 0$$

for which we deduce that $A = 0$ and $\sigma(f) = \alpha f + \beta$ with α and β two constants. Therefore a time-homogeneous nilpotent step 1 LV model is a displaced diffusion model which is already an Abelian LV model.
Finally, we exhibit an example which is a nilpotent step 1 (but not Abelian) LV model:

$$\sigma(t, f) = (\alpha t + \beta)(f + \lambda)$$

with α, β and λ three constants.

$\square$

References

Books, monographs

[1] Abramowitz, M., Stegun, I. A. : *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, Dover.

[2] Arnold, V. : *Chapitres supplémentaires de la théorie des équations différentielles ordinaires*, p58-59. Edition MIR, 1980.

[3] Berline N., Getzler, E., Vergne M. : *Dirac Operators and Heat Kernel*, Springer (2004).

[4] Birrell, N.D., Davies, P.C.W. : *Quantum Fields in Curved Space*, Cambridge University Press (1982).

[5] Bleistein, N., Handelsman, R.A. : *Asymptotic Expansions of Integrals*, Chapter 8, Dover (1986).

[6] Bouchaud, J-P, Potters, M. : *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, Cambridge University Press, 2nd edition (2003).

[7] Brigo, D., Mercurio, F. : *Interest Rate Models - Theory and Practice With Smile, Inflation and Credit*, Springer Finance. 2nd ed. (2006).

[8] Cheeger, J., Ebin, D.G. : *Comparison Theorems in Differential Geometry*, North-Holland/Kodansha (1975).

[9] Cont, R., Tankov, P. : *Financial Modelling with Jump Processes*, Chapman & Hall / CRC Press, (2003).

[10] Cooper, F., Khare, A., Sukhatme, U. : *Supersymmetry in Quantum Mechanics*, World Scientific (2001).

[11] Delbaen, F. , Schachermayer, W. : *The Mathematics of Arbitrage*, Springer Finance (2005).

[12] Duffie, D. : *Security Markets, Stochastic Models*, Academic Press (1988).

[13] DeWitt, B.S. : *Quantum Field Theory in Curved Spacetime*, Physics Report, Volume 19C, No 6 (1975).

[14] Eguchi, T., Gilkey, P.B., Hanson, A.J. : *Gravitation, Gauge Theories and Differential Geometry*, Physics Report, Vol. 66, No 6, December (1980).

[15] Erdély, A. : *Asymptotic Expansions*, Dover (1956).

[16] Fouque, J-P, Papanicolaou, G., Sircar, R. : *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, Jul. (2000).

[17] Fuchs, J. : *Affine Lie Algebras and Quantum Groups*, Cambridge University Press (1992).

[18] Gatheral, J. : *The Volatility Surface: A Practitioner's Guide*, Wiley Finance (2006).

[19] Gilkey, P.B. : *Invariance Theory, The Heat Equation, and the Atiyah-Singer Index Theorem*, CRC Press (1995).

[20] Glasserman, P. : *Monte Carlo Methods in Financial Engineering*, Stochastic Modelling and Applied Probability, Springer (2003).

[21] Gutzwiller, M.C. : *Chaos in Classical and Quantum Mechanics*, Springer (1991).

[22] Hebey, E. : *Introduction à l'Analyse Non linéaire sur les Variétés*, Diderot Editeur (1997).

[23] Hsu, E.P. : *Stochastic Analysis on Manifolds*, Graduate Studies in Mathematics, Vol. 38, AMS (2001).

[24] Ikeda, N., Watanabe, S. : *Stochastic Differential Equations and Diffusion Processes*, 2nd edition, North-Holland/Kodansha (1989).

[25] Itô, K., McKean, H. : *Diffusion Processes and their Sample Paths*, Springer (1996).

[26] Jacod, J., Protter, P. : *Probability Essentials*, Springer, Universitext, 2nd ed. (2003).

[27] Karatzas, I., Shreve, S.E. : *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, 2nd edition, Springer (1991).

[28] Kloeden, P. , Platen, E. : *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, (1995).

[29] Lewis, A. : *Option Valuation under Stochastic Volatility*, Finance Press (2000), CA.

[30] Lipton, A. : *Mathematical Methods for Foreign Exchange*, World Scientific (2001).

[31] Malliavin, P., Thalmaier, A. : *Stochastic Calculus of Variations in Mathematical Finance*, Springer (2006).

[32] Merton, R.C. : *Continuous Time Finance*, Blackwell, London (1992).

[33] Nualart, D. : *The Malliavin Calculus and Related Topics*, 2nd Edition, Springer (2006).

[34] Oksendal, B. : *Stochastic Differential Equations: An Introduction with Applications*, Springer, 5th ed. (1998).

[35] Rudin, W. : *Functional Analysis*, McGraw-Hill Science/Engineering/ Math, 2nd edition (1991).

[36] Reed, M., Simon, B. : *Methods of Mathematical Physics*, vol. 2, Academic Press, New York (1975).

[37] Sznitman, A-S : *Brownian Motion, Obstacles and Random Media*, Monographs in Mathematics, Springer (1998).

[38] Stroock, D. W., Varadhan, S. R. S. : *Multidimensional Diffusion Processes*, Die Grundlehren der mathematischen Wissenschaften, vol. 233, Springer-Verlag, Berlin and New York (1979).

[39] Terras, A. : *Harmonic Analysis on Symmetric Spaces and Applications*, Vols. I, II, Springer-Verlag, N.Y., 1985, 1988.

Articles

[40] Albanese, C., Campolieti, G., Carr, P., Lipton, A. : *Black-Scholes Goes Hypergeometric*, Risk Magazine, December (2001).

[41] Albanese, C., Kuznetsov, A. : *Transformations of Markov Processes and Classification Scheme for Solvable Driftless Diffusions*, http://arxiv.org/abs/0710.1596.

[42] Albanese, C., Kuznetsov, A. : *Unifying the Three Volatility Models*, Risk Magazine, March (2003).

[43] Andersen, L. : *Efficient Simulation of the Heston Stochastic Volatility Model*, Available at SSRN: http://ssrn.com/abstract=946405, January 23 (2007).

[44] Andersen, L., Andreasen, J. : *Volatility Skews and Extensions of the Libor Market Model*, Applied Mathematical Finance 7(1), pp 1-32 (2000).

[45] Andersen, L., Andreasen, J. : *Volatile Volatilities*, Risk 15(12), December 2002.

[46] Andersen, L., Brotherton-Ratcliffe, R. : *Extended Libor Market Models with Stochastic Volatility*, Journal of Computational Finance, Vol. 9, No. 1, Fall 2005.

[47] Andersen, L., Piterbarg, V. : *Moment Explosions in Stochastic Volatility Models*, Finance and Stochastics, Volume 11, Number 1, January 2007, pp. 29-50(22).

[48] Aronsov, D.G. : *Non-negative Solutions of Linear Parabolic Equations*, Ann. Scuola Norm. Sup. Pisa, Cl. Sci. (4) 22 (1968) 607-694, Addendum, 25 (1971) 221-228.

[49] Avellaneda, M., Zhu, Y. : *A Risk-Neutral Stochastic Volatility Model*, Applied Mathematical Finance (1997).

[50] Avellaneda, M., Boyer-Olson, D., Busca, J., Fritz, P. : *Reconstructing the Smile*, Risk Magazine, October (2002).

[51] Avramidi, I.V., Schimming, R. : *Algorithms for the Calculation of the Heat Kernel Coefficients*, Quantum Theory under the Influence of External Conditions, Ed. M. Bordag, Teubner-Texte zur Physik, vol. 30 (Stuttgart: Teubner, 1996), pp. 150-162, http://arxiv.org/abs/hep-th/9510206.

[52] Bally, V. : *An Elementary Introduction to Malliavin Calculus*, No. 4718. Fevrier 2003, INRIA. http://www.inria.fr/rrrt/rr-4718.html.

[53] Benaim, S., Friz, P. : *Regular Variations and Smile Dynamic*, Math. Finance (forthcoming), http://arxiv.org/abs/math/0603146.

[54] Benaim, S., Friz, P. : *Smile Asymptotics II: Models with Known Moment Generating Function*, to appear in Journal of Applied Probability (2008), http://arxiv.org/abs/math.PR/0608619.

[55] Benaim, S., Friz, P., Lee, R. : *On the Black-Scholes Implied Volatility at Extreme Strikes*, to appear in Frontiers in Quantitative Finance, Wiley (2008).

[56] Ben Arous, G. : *Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus*, Ann. sci. Ec. norm. supér, vol. 21, no3, pp. 307-331 (1988).

[57] Benhamou, E. : *Pricing Convexity Adjusment With Wiener Chaos*, working paper. Available at SSRN: http://ssrn.com/abstract=257208.

[58] Benhamou, E. : *Optimal Malliavin Weighting Function for the Computation of the Greeks*, Mathematical Finance, Volume 13, Issue 1, 2003.

[59] Berestycki, H., Busca, J., Florent, I. : *Asymptotics and Calibration of Local Volatility Models*, Quantitative Finance, 2:31-44 (1998).

[60] Berestycki, H., Busca, J., Florent, I. : *Computing the Implied Volatility in Stochastic Volatility Models*, Comm. Pure Appl. Math., 57, num. 10 (2004), p. 1352-1373.

[61] Bergomi, L. : *Smile Dynamics*, Risk Magazine, September (2004).

[62] Bergomi, L. : *Smile Dynamics II*, Risk Magazine, October (2005).

[63] Bermin, H.P., Kohatsu-Higa, A., Montero, M. : *Local Vega Index and Variance Reduction Methods*, Mathematical Finance, 13, 85-97 (2003).

[64] Björk, T, Svensson, L : *On the Existence of Finite-Dimensional Realizations for Nonlinear Forward Rate Models*, Mathematical Finance, Volume 11, Number 2, April 2001, pp. 205-243(39).

[65] Black, F., Scholes, M. : *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy, 1973, vol. 81, issue 3, pages 637-54.

[66] Bonneau, G., Faraut, J., Valent, G. : *Self-adjoint Extensions of Operators and the Teaching of Quantum Mechanics*, Am.J.Phys. 69 (2001) 322, http://arxiv.org/abs/quant-ph/0103153.

[67] Brace, A., Gatarek, D., Musiela, M. : *The Market Model of Interest Rate Dynamics*, Mathematical Finance, 7:127-154, 1996.

[68] Buehler, H. : *Consistent Variance Curve Models*, Finance and Stochastics, Volume 10, Number 2 (2006).

[69] Carr, P., Madan, D. : *Determining Volatility Surfaces and Option Values From an Implied Volatility Smile*, Quantitative Analysis in Financial Markets, Vol II, M. Avellaneda, ed, pp. 163-191, 1998.

[70] Carr, P., Lipton, A., Madan, D. : *Reduction Method for Valuing Derivative Securities*, Working paper, http://www.math.nyu.edu/research/carrp/papers/pdf/symmetry17.pdf.

[71] Carr, P., Lewis, K. : *Corridor variance swaps*, Risk Magazine, February (2004).

[72] Carr, P., Lee, R. : *Realized Volatility and Variance: Options via Swaps*, Risk Magazine, May (2007).

[73] Cartier, P., Morette-DeWitt, C. : *Characterizing Volume Forms*, Fluctuating Paths and Fields, Axel Pelster, ed., World Sci. Publishing, River Edge, N. J., 2001, pp. 139-156.

[74] Chen, K.T : *Iterated Path Integrals*, Bull. Amer. Math. Soc., 83, 1977, p 831-879.

[75] Choy, B., Dun, T., Schlogl, E. : *Correlating Market Models*, Risk Magazine, 17, 9, pp. 124-129, September (2004).

[76] Cox, J.C : *Notes on Option Pricing I: Constant Elasticity of Variance Diffusions*, Journal of Portfolio Management, 22, 15-17, (1996).

[77] Cox, J.C, Ingersoll, J.E, Ross, S.A : *A Theory of the Term Structure of Interest Rates*, Econometrica, 53, 385-407, (1985).

[78] Davies, E.B : *Gaussian Upper Bounds for the Heat Kernel of Some Second-Order Operators on Riemannian Manifolds*, J. Funct. Anal. 80, 16-32 (1988).

[79] Davydov, D. and V. Linetsky : *The Valuation and Hedging of Barrier and Lookback Options under the CEV Process*, Management Science, 47 (2001) 949-965.

[80] Delbaen, F., Shirakawa, H. : *A Note on Option Pricing for the Constant Elasticity of Variance Model*, Financial Engineering and the Japanese Markets, Volume 9, Number 2, 2002 , pp. 85-99(15).

[81] Dupire, B. : *Pricing with a Smile*, Risk Magazine, 7, 18-20 (1994).

[82] Dupire, B. : *A Unified Theory of Volatility*, In Derivatives Pricing: The Classic Collection, edited by Peter Carr, Risk publications.

[83] Durrleman, V. *From implied to spot volatilities*, working paper, http://www.cmap.polytechnique.fr/%7Evaldo/research.html, 2005.

[84] Ewald, C-O : *Local Volatility in the Heston Model: A Malliavin Calculus Approach*, Journal of Applied Mathematics and Stochastic Analysis, 2005:3 (2005) 307-322.

[85] Fournié, E., Lasry, F., Lebuchoux, J., Lions, J., Touzi, N. : *An Application of Malliavin Calculus to Monte Carlo Methods in Finance*, Finance and Stochastics, 4, 1999.

[86] Fournié, E., Lasry, F., Lebuchoux, J., Lions, J. : *Applications of Malliavin Calculus to Monte Carlo Methods in Finance. II*, Finance and Stochastics, 5, 201-236, 2001.

[87] Forde, M. : *Tail Asymptotics for Diffusion Processes with Applications to Local Volatility and CEV-Heston Models*, arXiv:math/0608634.

[88] Fritz, P. : *An Introduction to Malliavin Calculus*, Lecture notes, 2005, http://www.statslab.cam.ac.uk/ peter/.

[89] Gatheral, J. : *A Parsimimonious Arbitrage-Free Implied Volatility Parametrization with Application to the Valuation of Volatility Derivatives*, ICBI conference Paris (2004).

[90] Gaveau, B. : *Principe de moindre action, propagation de la chaleur et estimées sous-elliptiques sur les groupes nilpotents d'ordre deux et pour es opérateurs hypoelliptiques*, Comptes-Rendus Académie des Sciences, 1975, t.280, p.571.

[91] Gentle, D̀ : *Basket Weaving*, Risk Magazine 6(6), 51-52.

[92] Gilkey, P.B., Kirsten, K., Park, J.H., Vassilevich, D. : *Asymptotics of the Heat Equation with "Exotic" Boundary Conditions or with Time Dependent Coefficients*, Nuclear Physics B - Proceedings Supplements, Volume 104, Number 1, January 2002 , pp. 63-70(8), http://arxiv.org/abs/math-ph/0105009.

[93] Gilkey, P.B., Kirsten, K., Park, J.H. : *Heat Trace Asymptotics of a Time Dependent Process*, J. Phys. A:Math. Gen Vol 34 (2001) 1153-1168, http://arxiv.org/abs/hep-th/0010136.

[94] Glasserman, P., Zhao, X. : *Arbitrage-free Discretization of Log-Normal Forward Libor and Swap Rate Models*, Finance and Stochastics 4:35-68.

[95] Grigor'yan, A., Noguchi, M. : *The Heat Kernel on Hyperbolic Space*, Bulletin of LMS, 30 (1998) 643-650.

[96] Grigor'yan, A. *Heat Kernels on Weighted Manifolds and Applications*, Cont. Math. 398 (2006) 93-191.

[97] Grosche, C. : *The General Besselian and Legendrian Path Integrals* : J. Phys. A: Math. Gen. 29 No 8 (1996).

[98] Gyöngy, I. : *Mimicking the One-dimensional Marginal Distributions of Processes Having an Itô Differential*, Probability Theory and Related Fields, Vol. 71, No. 4 (1986), pp. 501-516.

[99] Hagan, P.S., Woodward, D.E. : *Equivalent Black Volatilities*, Applied Mathematical Finance, 6, 147-157 (1999).

[100] Hagan, P.S., Kumar, D., Leniewski, A.S., Woodward, D.E. : *Managing Smile Risk*, Willmott Magazine, pages 84-108 (2002).

[101] Hagan, P.S., Lesniewski, A.S., Woodward, D.E. : *Probability Distribution in the SABR Model of Stochastic Volatility*, March. 2005, unpublished.

[102] Henry-Labordère, P : *A General Asympotic Implied Volatility for Stochastic Volatility Models*, to appear in Frontiers in Quantitative Finance, Wiley (2008).

[103] Henry-Labordère, P. : *Short-time Asymptotics of Stochastic Volatility Models*, to appear in Encyclopedia of Quantitative Finance, Wiley (2009).

[104] Henry-Labordère, P. : *Combining the SABR and LMM models*, Risk Magazine, October (2007).

[105] Henry-Labordère, P. : *Solvable Local and Stochastic Volatility Models: Supersymmetric Methods in Option Pricing*, Quantitative Finance, Volume 7, Issue 5 Oct. (2007), pages 525-535.

[106] Heston, S. : *A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, review of financial studies, 6, 327-343.

[107] Heston, S., Loewenstein, M., Willard, G. : *Options and Bubbles*, Review of Financial Studies, Vol. 20, No. 2. (March 2007), pp. 359-390.

[108] Hull, J., White, A. : *Forward Rate Volatilities, Swap Rate Volatilities, and the Implementation of the Libor Market Model*, Journal of Fixed Income, Vol. 10, No. 3, September (2000), pp. 46-62.

[109] Hull, J., White, A. : *The Pricing of Options on Assets with Stochastic Volatilities*, The Journal of Finance, 42, 281-300.

[110] Hunt, P., Kennedy, J., Pelsser, A. : *Markov-Functional Interest Rate Models*, Finance and Stochastics, 2000, vol. 4, issue 4, pages 391-408.

[111] F. Jamshidian : *Libor and Swap Market Models and Measures*, Finance and Stochastics, 1(4):293-330 (1997).

[112] Jäckel, P., Kahl, C. : *Not-so-complex Logarithms in the Heston Model*, Wilmott, September 2005, pp. 94-103.

[113] Joshi, M.S. : *Rapid Drift Computations in the LIBOR Market Model*, Wilmott, May. (2003).

[114] Jourdain, B. : *Loss of Martingality in Asset Price Models with Log-normal Stochastic Volatility*, CERMICS working paper, http://cermics.enpc.fr/reports/CERMICS-2004/CERMICS-2004-267.pdf.

[115] Junker, G. : *Quantum and Classical Dynamics: Exactly Solvable Models by Supersymmetric Methods*, International Workshop on Classical and Quantum Integrable Systems, L.G. Mardoyan, G.S. Pogosyan and A.N. Sissakian eds., (JINR Publishing, Dubna, 1998) 94-103, http://arxiv.org/abs/quant-ph/9810070.

[116] MCKean, H.P. : *An Upper Bound to the Spectrum of* Δ *on a Manifold of Negative Curvature*, J. Diff. Geom, 4 (1970) 359-366.

[117] Kawai, A. : *A New Approximate Swaption Formula in the Libor Model Market: an Asymptotic Approach*, Applied Mathematical Finance, 2003, vol. 10, issue 1, pages 49-74.

[118] Krylov, G. : *On Solvable Potentials for One Dimensional Schrödinger and Fokker-Planck Equations*, http://arxiv.org/abs/quant-ph/0212046.

[119] Kunita, H. : *Stochastic Differential Equations and Stochastic Flows of Diffeomorphims* (Theorem II 5.2), Ecole d'été de Probabilités de Saint-Flour XII-1982, Lecture Notes in Mathematics, 1097, Springer, 143-3003.

[120] Kuznetsov, A. : *Solvable Markov Processes*, Phd. Thesis (2004), http://www.unbsj.ca/sase/math/faculty/akuznets/publications.html.

[121] Lee, P., Wang, L., Karim, A. : *Index Volatility Surface via Moment-Matching Techniques*, Risk Magazine, December (2003).

[122] Lee, R. : *The Moment Formula for Implied Volatility at Extreme Strikes*, Mathematical Finance, Volume 14, Number 3, July (2004).

[123] Lesniewski, A. : *Swaption Smiles via the WKB Method*, Seminar Math. fin., Courant Institute of Mathematical Sciences, Feb. (2002).

[124] Levy, P. : *Wiener's Random Function, and Other Laplacian Random Functions*, Proc. 2nd Berkeley Symp. Math. Stat. Proba., vol II, 1950, p. 171-186, Univ. California.

[125] Li, P., Yau, S.T. : *On the Parabolic Kernel of the Schrödinger Equation*, Acta Math. 156 (1986), p. 153-201.

[126] Linetsky, V. : *The Spectral Decomposition of the Option Value*, International Journal of Theoretical and Applied Finance (2004), vol 7, no. 3, pages 337-384.

[127] Liskevich, V., Semenov, Yu. : *Estimates for Fundamental Solutions of Second Order Parabolic Equations*, Journal of the London Math. Soc., 62 (2000) no. 2, 521-543.

[128] Milson, R. : *On the Liouville Tranformation and Exactly-Solvable Schrödinger Equations*, International Journal of Theoretical Physics, Vol. 37, No. 6, (1998).

[129] Natanzon, G.A : *Study of One-dimensional Schrödinger Equation Generated from the Hypergeometric Equation*, Vestnik Leningradskogo Universiteta, 10:22, 1971, http://arxiv.org/abs/physics/9907032v1.

[130] Nash, J. : *Continuity of Solutions of Parabolic and Elliptic Equations*, Amer. J. Math., 80 (1958) 931-954.

[131] Norris, J., Stroock, D.W. : *Estimates on the fundamental solution to heat flows with uniformly elliptic coefficients*, Proc. Lond. Math. Soc. 62 (1991) 375-402.

[132] Øksendal, B. : *An Introduction to Malliavin Calculus with Applications to Economics*, Lecture Notes from a course given at the Norwegian School of Economics and Business Administration (NHH), NHH Preprint Series, September (1996).

[133] Ouhabaz, E. M. : *Sharp Gaussian Bounds and L^p-Growth of Semigroups Associated with Elliptic and Schrödinger Operators*, Proc. Amer. Math. Soc. 134 (2006), 3567-3575.

[134] Pelsser, A. : *Mathematical Foundation of Convexity Correction*, Quantitative Finance, Volume 3, Number 1, 2003, pp. 59-65(7).

[135] Pietersz, R., Regenmortel, M. : *Generic Market Models*, Finance and Stochastics, Vol. 10, No 4, December 2006, pp 507-528(22).

[136] Piterbarg, V. : *Time to Smile*, Risk Magazine, May (2005).

[137] Piterbarg, V. : *A Practitioner's Guide to Pricing and Hedging Callable Libor Exotics in Forward Libor Models*, Working paper (2003), Available at SSRN: http://ssrn.com/abstract=427084.

[138] Rebonato, A. : *On the Pricing Implications of the Joint Log-normal Assumption for the Swaption and Cap Markets*, Journal of Computational Finance, Volume 2 / Number 3, Spring (1999).

[139] Simon, B. : *Schrödinger Semigroups*, Bull. Amer. Math. Soc. (N.S) 7 (1982), 447-526.

[140] Sin, C.A. : *Complications with Stochastic Volatility Models*, Adv. in Appl. Probab. Volume 30, Number 1 (1998), 256-268.

[141] Sturm, K-T : *Heat Kernel Bounds on Manifolds*, Math. Annalen, 292, 149-162 (1992), Springer.

[142] Wu, L. : *Fast at-the-money Calibration of Libor Market Model through Lagrange Multipliers*, Journal of Computational Finance, Vol. 6, No. 2, 39-77 (2003).

[143] Yamato, Y. : *Stochastic Differential Equations and Nilpotent Algebras*, Z. Wahrscheinlichkeitstheorie und verw. Gebiete 47, 231-229 (1979).

[144] Zhang, Q.S. : *A Sharp Comparison Result Concerning Schrödinger Heat Kernels*, Bulletin London Math. Society, 35 (2003), no. 4, pp. 461-472.

[145] Zuhlsdorff, C. : *The Pricing of Derivatives on Assets with Quadratic Volatility*, Applied Mathematical Finance, Volume 8, Number 4, December 2001, pp. 235-262(28).

Index