Asymptotics of Wiener Functionals and Applications to Mathematical Finance

by

Sean Paolo Nicholas Violante

Department of Mathematics
Imperial College London
London SW7 2AZ
United Kingdom

Submitted to Imperial College London
for the degree of
Doctor of Philosophy

2012
Declaration

I, the undersigned, hereby declare that the work presented in this thesis is my own. When material from other authors has been used, these have been duly acknowledged. This thesis has not previously been presented for this or any other PhD examinations.

Sean Violante
Abstract

In this thesis we study asymptotic expansions for option pricing with emphasis on small noise “singular perturbations” which are, as we shall see, better suited than the more popular small time asymptotics to approximate typical stochastic volatility models. In particular, we argue that analytic solutions are unlikely for more advanced models, and therefore numerical methods of calculation are required. The following are the main results of the thesis. We show that zeroth order implied volatility is given by the large deviation rate function under minimal assumptions. We then show a small noise sample path large deviations principle for a class of two dimensional positive diffusions of relevance to finance. We numerically calculate the large deviations rate function for an example process, Gatheral’s Double CEV model, and highlight the speed and accuracy of the approximation. We then investigate Yoshida-Watanabe asymptotic expansions and develop a Mathematica program to derive them automatically. Lastly, we develop a small noise asymptotic expansion for marginal densities of solutions of SDEs (joint work). Using this we determine the large strike implied volatility for the Stein-Stein model and the Schöbel and Zhu model by rescaling into a small noise problem.
To my parents
Acknowledgments

Thanks to my supervisor, Professor Dan Crisan, for inspiration and thoughtful guidance.

Thanks to Professor Peter Friz and all the staff at T.U. Berlin for their support and insights.

Thanks to my parents and my wife for their constant support and encouragement and finally to my children for putting up with it all!
# Contents

**Common Notation**

1 **Introduction**
   1.1 Derivatives Models ................................................... 17
   1.2 Small Noise and Small Time Asymptotic Expansions ............... 19
   1.3 Regular and Singular Perturbations ................................ 23
      1.3.1 Singular Perturbations ........................................... 25
         1.3.1.1 Calculation of energy function ...................... 26
         1.3.1.2 Asymptotic expansions of the transition density ... 29
   1.4 Our Contribution ...................................................... 32

2 **Preliminaries**
   2.1 Large deviations ..................................................... 35
   2.2 Malliavin Calculus ................................................... 37
      2.2.1 Application to SDEs ............................................. 41

3 **Small Noise Implied Volatility Asymptotics and Large Deviations**
   3.1 Implied Volatility asymptotics from the Large Deviations Principle .... 46
   3.2 Small Noise Asymptotics of Implied Volatility and Gatheral and Wang’s Variational most likely path .............................................. 51

4 **Wentzell-Freidlin Large Deviations for a class of two-dimensional positive diffusions**
   4.1 Gatheral’s Double CEV model ........................................ 56
   4.2 Background Results .................................................. 57
   4.3 Application to DCEV process ....................................... 62
      4.3.1 Monotonicity of Rate Function ............................... 67
   4.4 Numerical Results ................................................... 69
      4.4.1 Numerical method ................................................. 72
   4.A Comparison Principle .............................................. 73
CONTENTS

5 Yoshida Watanabe expansions using Mathematica 75
  5.1 Asymptotic Expansion 75
  5.2 Iterated Stochastic Integrals with Mathematica 80
  5.3 Application to DCEV Model 83
  5.4 Results 87

6 Marginal density expansions and Large Strike Expansions of the Stein 91
  6.1 The main result and its corollaries 93
    6.1.1 Corollary on tail expansions 96
    6.1.2 Corollary on short time expansions 97
  6.2 Energy Minimising Paths and Hamiltonian Flows 98
  6.3 Large Strike expansions of Stein-Stein and Schöbel-Zhu Models 106
    6.3.1 Check of Theorem Assumptions 121
  6.4 Comments on Heston [39] and Lions–Musiela [54] 124

Bibliography 127

A Appendix 135
  A.1 Feller’s test for explosions 135
  A.2 Wentzell-Freidlin Sample Path Large Deviations by Azencott’s method 137
  A.3 Implicit Function Theorem on a Hilbert Space 139
  A.4 Non-focality and infinite-dimensional non-degeneracy 140
    A.4.1 Jacobi variation 145
  A.5 Asymptotic expansion of the marginal density 148
  A.6 C++ Code for Chapter 4 152
  A.7 Mathematica Code for Chapter 5 166
List of Figures

1.2.1 Small Time vs Small Noise asymptotics of DCEV implied volatility . . . 23
1.3.1 SABR Implied Volatility at Zero Order . . . . . . . . . . . . . . . . . . . . 28

4.3.1 Region of integration . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
4.4.1 Numerical investigation of small time and small noise asymptotics of
DCEV implied volatility . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

5.4.1 Yoshida-Watanabe small time price convergence . . . . . . . . . . . . . . 88
5.4.2 Yoshida-Watanabe small noise price convergence . . . . . . . . . . . . . . 88
5.4.3 Implied Volatility Smile for DCEV VIX Option: Watanabe Small Time,
Small Noise and Large deviations. Circles are 5%,10%,...,95% probab-
ility points. Cross is the forward or mean value. . . . . . . . . . . . . . . 90
LIST OF FIGURES
List of Tables

4.1 DCEV parameter settings ........................................ 70
4.2 Implied Volatility from Monte Carlo Simulation ................ 70
4.3 Small Noise - Monte Carlo Implied Volatility Error ............ 70
Common Notation

\{F_t\} \quad \text{Asset Process}

\bar{F}_T \quad \text{Expected value of asset price at time } T

K \quad \text{Vanilla Option Strike}

C(m, s; k) \quad \text{Lognormal call option formula for option strike } k, \text{ mean } m \text{ and lognormal variance } s.

C(m, s; k) := \int_{\bar{k}}^{\infty} \frac{x - k}{\sqrt{2\pi s x}} \exp\left(-\frac{(\log(x/m) + 1/2s)^2}{2s}\right)dx.

\sigma_I(K, T) \quad \text{Implied (lognormal) volatility for call/put option on Asset price at maturity } T \text{ years and strike } K.

\mathbb{E}[(F_T - K)^+] = C[\bar{F}_T, \sigma_I(K, T)^2T; K]

\sigma_{I,\epsilon}(K, T) \quad \text{small noise implied (lognormal) volatility.}

\mathbb{E}[(F^\epsilon_T - K)^+] = C[\bar{F}^\epsilon_T, \epsilon^2\sigma_{I,\epsilon}(K, T)^2T; K]

\sigma_{I,0}(K, T) \quad \lim_{\epsilon\to0} \sigma_{I,\epsilon}(K, T)

\partial_{x^i} V(x) \quad \frac{\partial}{\partial x^i} V(x) \text{ for } x = (x^1, \ldots, x^d)

\partial_x V(x) \quad (\frac{\partial}{\partial x^x} V(x), \ldots, \frac{\partial}{\partial x^d} V(x))

\partial_x^m \partial_y^n V(x, y) \quad \frac{\partial^{m+n}}{\partial x^m \partial y^n} V(x, y) \text{ with } x, y \text{ scalar}

W = (W^1, W^2, \ldots, W^m), \text{ } m\text{-dimensional Brownian motion. Components uncorrelated unless explicitly stated.}
\( X_t \) \( \mathbb{R}^d \)-valued process depending on small parameter \( \epsilon > 0 \)

\( X_t^{\epsilon,i} \) \( i \)-th component of \( \mathbb{R}^d \)-valued process depending on small parameter \( \epsilon > 0 \)

**Small noise SDE** An SDE where the drift term is \( O(\epsilon^0) \) and dispersion \( O(\epsilon) \) for small parameter \( \epsilon \downarrow 0 \), eg.

\[
\begin{align*}
  dx_t & = b(X_t) \, dt + \epsilon \sigma(X_t) \, dW \\
  x_0 & = x_0
\end{align*}
\]

**Small time SDE** An SDE where the drift term is \( O(\epsilon^2) \) and dispersion \( O(\epsilon) \) for small parameter \( \epsilon \downarrow 0 \), eg.

\[
\begin{align*}
  dx_t & = \epsilon^2 b(X_t) \, dt + \epsilon \sigma(X_t) \, dW \\
  x_0 & = x_0
\end{align*}
\]

**Local Volatility model:** A model of an asset price or discounted asset price where the dispersion coefficient is a function of the asset and time only eg

\[
\begin{align*}
  dF_t & = a(F_t, t) \, dW_t^{1} \\
  F_0 & = f_0 > 0
\end{align*}
\]

**SABR model**

\[
\begin{align*}
  dF_t & = \alpha_t F_t^\beta \, dW_t^1 \\
  F_0 & = f_0 > 0, 0 \leq \beta \leq 1 \\
  d\alpha_t & = \nu \alpha_t \, dW_t^2 \\
  \alpha_0 & > 0, \nu \geq 0 \\
  d\langle W^1, W^2 \rangle_t & = \rho \, dt \\
  \rho \in (-1, 1)
\end{align*}
\]

**DCEV model**

\[
\begin{align*}
  dS_t & = \sqrt{v_t} S_t \, dW_t^3 \\
  dv_t & = \nu (v_t - v_t^0) \, dt + \xi_1 v_t^\alpha (\sqrt{1 - \rho^2} \, dW_t^1 + \rho dW_t^2) \\
  dv_t' & = c (z_3 - v_t') \, dt + \xi_2 v_t'^\beta dW_t^2 \\
  \text{Joint model of SPX stock index and VIX volatility index. In this thesis we only consider options on VIX, so } S_t \text{ is not modelled.}
\end{align*}
\]

Calibrated parameters:

\begin{align*}
  S_0, v_0, v_0' & > 0, \\
  \kappa & > c > 0 \text{ mean reversion,} \\
  \xi_1, \xi_2 & \geq 0 \text{ volatility of volatility,} \\
  \rho & \in (-1, 1), \text{ Correlation} \\
  \alpha, \beta & \in [1/2, 1] \text{ CEV power}
\end{align*}

**VIX**

VIX asset - in DCEV model given by \( VIX_T = \sqrt{a_1 v_T^0 + a_2 v_T' + a_3 z_3} \) with \( a_1, a_2, a_3 > 0 \)

\( p_t(x, y) \) transition density from \( x \) to \( y \) in time \( t \)
\begin{itemize}
\item \{\mu_\epsilon\} family of probability measures indexed by small parameter \(\epsilon > 0\)
\item \(\mu_\epsilon[g(F_T)] = \int g(x)\mu_\epsilon(dx)\). Used as a generalisation of \(\mathbb{E}[g(F_T^\epsilon)]\) for a family of probability measures, \(\mu_\epsilon\), defined on the Borel sets of \(\mathbb{R}\).
\item \(\mathcal{C}^d = \mathcal{C}([0, T], \mathbb{R}^d)\), space of continuous functions from \([0, T]\) to \(\mathbb{R}^d\) endowed with uniform norm, \(\|\cdot\|_T\)
\item \(\mathcal{C}^d_{x_0}\) closed subset of \(\mathcal{C}^d\) starting at \(x_0 \in \mathbb{R}^d\)
\item \(\|f\|_T = \sup_{t \in [0, T]}|f(t)|\)
\item \(\mathcal{H}^m\) Hilbert space \{\(h \in \mathcal{C}^0_0 : \|h\|_H < \infty\)\} with norm \(\|\cdot\|_H\)
\item \(\|h\|^2_H = \int_0^T \|\dot{h}(s)\|^2 ds\)
\item \(I : \mathcal{H}^m \to [0, \infty] = \frac{1}{2} \|h\|^2_H\) - large deviation rate function for Brownian motion,
\item \(\phi^h(x_0)\) In Chapter 4, solution of ODE \(\dot{g}(t) = b(g(t)) + \sigma(g(t))\dot{h}(t), \quad g(0) = x_0\)
\item \(J(g) = \inf \{I(h) : h \in \mathcal{H}^m, \phi^h(x_0) = g\}\), for \(g \in \mathcal{C}^d_{x_0}\)
\item \(\mathcal{J}(g) = \frac{1}{2} \int_0^T (\dot{g}_t - b(g_t))^T \Sigma^{-1} (\dot{g}_t - b(g_t)) dt, \quad \Sigma = \sigma\sigma^T\), for \(\sigma\) invertible and \(g \in \mathcal{C}^d_{x_0}\)
\item \(K_a\) \(\{h \in \mathcal{H}^m : \|h\|_H \leq a\}\)
\item \(X^\epsilon\) solution of \(dX_t^\epsilon = b(\epsilon, X_t^\epsilon) dt + \epsilon \sigma(X_t^\epsilon) dW_t, \quad \text{with } X_0^\epsilon = x_0^\epsilon \in \mathbb{R}^d\)
\item \((Y_l) = : \Pi_l X_l := (X_{l1}^\epsilon, \ldots, X_{ll}^\epsilon)\), \(l \in \{1, 2, \ldots, d\}\) projected diffusion process
\item \(\Pi_l : \mathbb{R}^d \to \mathbb{R}^l\) projection onto first \(l\) components
\item \(f(y, t)\) density of projected process \(Y\) at time \(t\)
\item \(f^\epsilon(y, T)\) density of process \(Y^\epsilon\) at time \(T\) and small parameter \(\epsilon\).
\item \(V\) \(V_0 = b(0, \cdot), V_k = \sigma_k, k = 1, \ldots, m, V = (V_1, \ldots, V_m)\)
\item \(\phi^h_T(x_0)\) solution at time \(T\) of controlled ODE \(d\phi^h_T(x_0) = V_0(\phi^h_T(x_0)) dt + \sum_{k=1}^m V_k(\phi^h_T(x_0)) dh^k_t, \quad \phi^h_0 = x_0 \in \mathbb{R}^d\)
\item \(\phi^h_{T-\epsilon}(x_t) = \phi^h_T \circ (\phi^h_T)^{-1}(x_t)\)
\end{itemize}
$\Phi^h_{T\leftarrow t}$ differential of flow from $t$ to $T$:

$$\Phi^h_{T\leftarrow t} : y = \frac{d}{ds} \phi^h_{T\leftarrow t}(\phi^h_t + sy) \bigg|_{s=0}$$

$g : \mathbb{R}^d \rightarrow \mathbb{R}$ In Chapter 6, $C^\infty$ function of $x_T$. Except in section 6.2, we restrict $g$ to the projection $\Pi_t$.

$N_a \{ x \in \mathbb{R}^d : g(x) = a \}$.

$\mathcal{K}_a \{ h \in \mathcal{H}^m : \phi^h_T(x_0) \in N_a \}$

$\mathcal{K}_a^{\min} \{ h \in \mathcal{K}_a : I(h) = \Lambda (a) \}$

$D$ Malliavin derivative of random variable.

$D$ deterministic Malliavin derivative. Fréchet derivative of functional of control, $h$. $DF(h)[k] = \lim_{\tau \to 0} \frac{F(h+\tau k) - F(h)}{\tau}$ $k \in \mathcal{H}^m$

$C^h_{\tau}$ deterministic Malliavin covariance of $\phi^h_t(x_0)$

$H_0$ Nullspace of $Dg \left( \phi^h_{T\leftarrow 0} \right)$, for fixed $h_0 \in \mathcal{K}_a^{\min}$.

$I''(h_0) : H_0 \times H_0$ Second derivative, evaluated at $h_0$, of $I|_{\mathcal{K}_a}$, restriction of energy $I$ to manifold $\mathcal{K}_a$, $h_0 \in \mathcal{K}_a^{\min}$.

$\mathcal{N}(h_0)$ Nullspace of $I''(h_0)$: $\{ h \in H_0 : I''(h_0)[h,h] = 0 \}$

$\mathcal{H}(x,p)$ Hamiltonian associated to ODE $dx_t = V_0(x_t)dt + V(x_t)\dot{h}(t)dt$

$H_{t\leftarrow 0} (x_0, p_0)$ Hamiltonian flow from $(x_0, p_0)$ to $(x_t, p_t)$
Chapter 1

Introduction

1.1 Derivatives Models

In the following we will explore the use of asymptotic expansions of probability densities to problems in mathematical finance in particular for the development of fast approximations to option prices. Models of asset price processes are often defined as solutions of stochastic differential equations (SDE’s) with respect to a pricing measure. Then pricing a derivative on an asset corresponds to calculating the expectation of some functional of the asset path. The need for fast approximations arises from the need of fast and accurate calibration of the asset model: the coefficients of the SDEs must be inferred from asset and option prices in the market, rather than being estimated from historical data. We will therefore be concerned with approximation to the vanilla options that are typically used to calibrate SDEs.

Given some random process for an asset price, \( F = \{ F_t, t \geq 0 \} \), in an appropriate pricing measure, the value of a vanilla call option on \( F \) with expiry/maturity \( T \), and strike \( K \) at time 0 is given by the expectation \( E[(F_T - K)^+] \). In particular, assuming \( F_T \) is lognormally distributed, with mean \( E[F_T] = \bar{F}_T \), and variance \( E[(F_T - \bar{F}_T)^2] = (\bar{F}_T)^2 (\exp(s) - 1) \), we have the lognormal call option price formula given by

\[
C(\bar{F}_T, s; K) := \int_K^\infty \frac{x - K}{\sqrt{2\pi sx}} \exp(-\frac{(\log(x/\bar{F}_T) + 1/2s)^2}{2s}) dx.
\]

Given a call option price (whether given by the market or derived from some model), we define the implied volatility \( \sigma_I(K, T) \) for the call option price at strike \( K \) and maturity \( T \) for an arbitrary positive asset price (not necessarily log-normal) \( F \)

\footnote{We will always assume interest rates to be zero in this thesis}
with $\bar{F}_T = \mathbb{E}[F_T]$ as the unique nonnegative solution of

$$
\mathbb{E}[(F_T - K)^+] = C[\bar{F}_T, \sigma_I(K, T)^2 T; K].
$$

A unique solution exists, since for any model

$$(\bar{F}_T - K)^+ \leq \mathbb{E}[(F_T - K)^+] < \bar{F}_T,$$

whilst $C[\bar{F}_T, 0; K] = (\bar{F}_T - K)^+$, $\lim_{s \to \infty} C[\bar{F}_T, s; K] = \bar{F}_T$ and $\partial_s C[\bar{F}_T, s; K] > 0$ (on $s \geq 0$). Option prices are typically quoted in terms of implied volatility, since this allows one to reduce the expected variation in option prices due to stock price movements, maturity, strike etc. Since around 1987, when a major crash occurred in the US stock market, implied volatilities for different markets have taken on a persistent shape when plotted against strike, $K$, and maturity, $T$ ([21]). The following stylized properties have been identified in studies of market implied volatilities.

1. Smile - for a given maturity, implied volatility tends to grow higher as the strike moves away from the forward, $\bar{F}_T$.

2. Skew - the smile may be asymmetric around the forward.

3. Term structure - for a given strike, implied volatility varies much less with maturity.

These “facts” have led to the development of stochastic volatility models of the following form:

$$
\begin{align*}
\text{d}F_t &= \sigma_t F_t \text{d}W_1^t \\
\text{d}\sigma_t &= \kappa (\sigma_\infty - \sigma_t) \text{d}t + \nu g(\sigma_t) \text{d}W_2^t \\
\text{d}\langle W_1, W_2 \rangle_t &= \rho \text{d}t.
\end{align*}
$$

Here the asset is labelled $F$, and its instantaneous volatility $\sigma$ is driven by another stochastic factor. The stochastic volatility, $\sigma$, generates the smile, the correlation, $\rho$, affects the skew and the term structure is determined by the mean reversion parameters, $\kappa$, and $\sigma_\infty$. Given that there may be hundreds of points on a market implied volatility surface, moving relatively independently (according to supply and demand), we can see that this is a drastic simplification - we cannot expect to accurately fit more than a handful of the implied volatilities, and the same parameters determine both current implied volatilities and their dynamics. In fact one of the key research areas in quantitative finance is the development of more realistic models of the implied volatility surface ([32, 16, 19]). This however is severely hampered by the lack
of fast methods to price vanilla options in order to calibrate the model parameters. Our work is an attempt to use asymptotic expansions to produce such fast approximations. As a test case, we study the DCEV model of Gatheral([32]), in which the stochastic volatility, \( \sigma_t = \sqrt{v_t} \), is driven by a two dimensional SDE

\[
\begin{align*}
  dv_t &= \kappa (v_t' - v_t) dt + \xi_1 v_t^\alpha (\sqrt{1 - \rho^2} dW_1^1 + \rho dW_1^2) \\
  dv'_t &= c (z_3 - v'_t) dt + \xi_2 v'_t^\beta dW_2 \\
  v_0, v'_0 > 0,
\end{align*}
\]

where \( \kappa > c > 0, z_3 > 0, |\rho| \leq 1, \alpha, \beta \in [1/2, 1] \). Although one typically starts by finding asymptotic expansions for the price of vanilla options, one then extrapolates to non vanishing values of the parameter by developing an asymptotic expansion for the implied volatility or similar model parameters ([35, 43, 4, 65, 37]). This is because of the exponentially increasing behaviour of option prices with the asymptotic small parameter (see Chapter 3) making extrapolation error prone; implied volatility is much more linear, and therefore a perturbation power series will have a larger range of validity. In much of the thesis we will be interested in the asymptotic limit of implied volatility as a parameter \( \epsilon \downarrow 0 \). Given a positive asset price process, \( \{F_\epsilon t\} \) dependent on the small parameter \( \epsilon \), then we define the \( (\epsilon-) \) implied volatility, \( \sigma_{I,\epsilon}(K,T) \), for the call option at strike \( K > 0 \) and maturity \( T \) for fixed \( \epsilon > 0 \) as the unique solution\(^2\) of

\[
\mathbb{E}[(F_\epsilon - K)^+] = C[F_\epsilon, \epsilon^2 \sigma_{I,\epsilon}^2(K,T); K], \quad F_\epsilon^\epsilon := \mathbb{E}[F_\epsilon] .
\]  

The insertion of an \( \epsilon^2 \) factor in the variance term is a notational convenience allowing us to unify the small noise and small time implied volatility limit definitions ( \( t = \epsilon^2 T \downarrow 0 \)).

### 1.2 Small Noise and Small Time Asymptotic Expansions

We investigate the limiting behaviour of solutions of SDEs dependent on a parameter \( \epsilon \) and in particular consider sample path small noise asymptotics, rather than the more commonly investigated small time asymptotics. Consider a generic SDE, with time homogeneous coefficients and elliptic diffusion coefficients,

\[
dX_t = b(X_t) dt + \sigma(X_t) dW_t \quad X_0 = x_0 .
\]

\(^2\)In Chapter 3, our definition is more generalised and uses the family of measures on \( \mathcal{B}(\mathbb{R}) \), \( \{\mu_\epsilon\} \), because our analysis there does not require the processes to be diffusion processes defined on the same space, so there we use the notation \( \mu_\epsilon((F - K)^+) \).
Then we can consider the limiting behaviour of the transition density, say, of \( X_T \) as \( T \to 0 \). This is a small time asymptotic. Alternatively, we can consider the family of scaled diffusion processes

\[
dX^\epsilon_t = \epsilon^2 b(X^\epsilon_t)dt + \epsilon \sigma(X^\epsilon_t)dW_t \quad X^\epsilon_0 = x_0,
\]

and Brownian scaling implies that \( X_t \overset{D}{=} X^{\sqrt{T}}_1 \), so we can also study the small time behaviour by considering \( \epsilon \to 0 \) in the above. In this thesis, we instead concentrate on small noise case where we consider the limiting behaviour as \( \epsilon \downarrow 0 \) of SDEs of the form

\[
dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad X_0 = x_0,
\]

which have been less frequently studied. We study these because market implied volatilities tend to have significant maturity dependence, which in turn implies that the model SDE’s calibrated to the market have a significant drift component. As we will see in Chapter 4 and 5, this implies that asymptotic expansions in time provide worse approximations than small noise approximations for models calibrated to market data. Similarly, if the drift or diffusion coefficients are time dependent,

\[
dX_t = b(X_t, t)dt + \epsilon \sigma(X_t, t)dW_t \quad X_0 = x_0,
\]

then the small noise asymptotic (at order 0 in \( \epsilon \)) will capture this dependence, whereas the small time asymptotic (at order 0 in \( T \)) will not: the zero order in \( T \) behaviour typically depends only on \( b(\cdot, 0) \) and \( \sigma(\cdot, 0) \). To illustrate the time-dependent case we consider the following lognormal model

\[
\frac{dF_t}{F_t} = \sigma(t)dW^1_t \quad F_0 = f_0,
\]

and \( \sigma : [0, \infty) \to \mathbb{R} \). We wish to establish an approximate formula for the implied volatility. Then it is clear that the implied volatility \( \sigma_I(K, T) \) for a \( T \)-maturity option in such a time-varying lognormal model is

\[
\sigma_I(K, T) = \sqrt{\frac{1}{T} \int_0^T \sigma^2(t)dt}
\]

If we were to look for an asymptotic expansion in powers of \( T \) as \( T \to 0 \), then the solution is a Taylor expansion in \( T \) around \( T = 0 \), and could well require a large number of terms depending on the particular nature of \( \sigma(\cdot) \). Therefore a small time asymptotic expansion of the transition density or some related quantity will require a
similarly high order for accuracy. The small noise problem for this model is specified as
\[
\frac{dF_t}{F_t} = \epsilon \sigma(t) dW_t^1 \quad F_0 = f_0.
\]
Then \(\sigma_{1,\epsilon}(K,T) \equiv \sigma_{1,1}(K,T)\), so any small noise expansion will be exact at order zero. The case of mean reversion is similar, consider the degenerate model
\[
\begin{align*}
\frac{dX_t^1}{X_t^1} &= \sqrt{X_t^2} dW_t^1 \\
\frac{dX_t^2}{X_t^2} &= \kappa(\alpha_\infty - X_t^2) dt \\
X_0^1 &= f_0 \quad X_0^2 = \alpha_0.
\end{align*}
\]
The implied volatility is then
\[
\sigma_{I}(K,T) = \sigma_{I,1}(K,T) = \sqrt{(\alpha_0 - \alpha_\infty) \left(1 - \exp(-\kappa T)\right) / \kappa T + \alpha_\infty}.
\]
Again we see that the asymptotic expansion to order 0 in \(\epsilon\) will be exact, whereas we will require significant number of terms to get good accuracy in the small time expansion. This decaying exponential behaviour with time carries over to the non zero vol-of-vol case and cannot be easily approximated by an asymptotic expansion in powers of \(T\). We should point out that this maturity dependence is precisely the reason mean-reverting models are typically chosen: they are a natural way to capture the observed maturity dependence of market data (in a time homogeneous model).

In using small noise asymptotic expansion as an approximation for the \(\epsilon = 1\) case, our working assumption is that the original diffusion coefficients (calibrated to market data) and/or \(T\) are small enough that \(\sigma_{1,1}(K,T) \approx \sigma_{1,0}(K,T)\). In the small time case we require that the drifts too are small enough for the asymptotic limit to be a good approximation. Assume we are given a model specified by
\[
\begin{align*}
dX_t &= b(X_t) dt + \sigma(X_t) dW_t \\
X_0 = x_0,
\end{align*}
\]
where the drift, \(b\), and dispersion coefficient, \(\sigma\), have been calibrated to market data up to maturity \(T\), and we are interested in the distribution of \(X_T\). To develop a small noise asymptotic approximation, we normalise time to 1, and introduce a normalised dispersion coefficient \(\tilde{\sigma}(X) = \frac{\sigma(X)}{\sigma(x_0)}\). Then introducing \(\epsilon_c = \sigma(x_0) \sqrt{T}\), we have \(X_T \overset{D}{=} \)
$X_t^{\epsilon_c}$, where

$$dX_t^\epsilon = T.b(X_t^\epsilon)dt + \epsilon\tilde{\sigma}(X_t^\epsilon)dW_t,$$

$$X_0^\epsilon = x_0$$

and we study the limiting distribution of $X_t^\epsilon$ as $\epsilon \to 0$. Assuming $\epsilon_c \ll 1$, then we expect that the limiting behaviour of $X_t^\epsilon$ will be a good approximation of the original $X_T$. We note that the small noise expansion makes no requirements on the drift term.

Turning now to a small time expansion (but using $\epsilon$-notation), we have

$$dX_t^\epsilon = \epsilon^{2}\tilde{b}(X_t^\epsilon)dt + \epsilon\tilde{\sigma}(X_t^\epsilon)dW_t,$$

$$X_0^\epsilon = x_0$$

Here again we want the normalised coefficients, $|\tilde{b}|, |\tilde{\sigma}| \leq 1$, using the same normalisation for $\sigma$, implies that $b(x) = O(\sigma^2(x))$. In other words, for the small time expansion we require both $\epsilon_c = \sigma(x_0)\sqrt{T} \ll 1$ and $b(x) = O(\sigma^2(x))$. Conversely, if $|b/\sigma^2|$ is small, we would expect small time and small noise expansions to work equally well. Applying this normalisation to the two-dimensional DCEV model calibrated to market data with $T = 0.5$ years, we find $\epsilon_c := \max_{i \in \{1,2\}} \sqrt{T}\sigma^i = 0.025$, but $b/\sigma^2 = \{31, 68\}$. In other words, one would expect that we can get a good approximation to the 6 months problem with a small noise asymptotic, but not small time.

Whilst this is the standard argument for asymptotic expansions, we are still left with the problem of identifying how large $\epsilon_c$ can be and still permit a good approximation to implied volatility. If we knew $\partial_s \sigma_{I,\epsilon}$, then this would clearly provide the natural scale; however, the whole reason for using asymptotic expansions is to tackle SDEs that do not have a closed form solution. Therefore in the rest of this thesis, we do not make these normalisations, and simply consider $\epsilon \in (0, 1]$ as scaling down the original SDE and determine how well the approximation works by numerical simulation. Getting slightly ahead of ourselves, we plot in figure 1.2.1 the behaviour of implied volatility of VIX options in Gatheral’s Double CEV model (which we describe in Chapter 4) for both the small noise and small time asymptotic regime. The x-axis is $\epsilon^2 T$, and we compare $\epsilon^2$ decreasing and $T$ fixed to $\epsilon = 1$ and $T$ decreasing. Our goal is to develop approximations for the right endpoints of the small noise curves (i.e. a given $T$, $\epsilon = 1$). As can be seen the small noise behaviour is much more linear and slowly varying for all times, and we can hope that a low order expansion will be effective. The plot was generated using Gatheral’s parameters calibrated to market data; full details of the model and parameter are provided in Chapter 4.

Since we consider sample path large deviations and Wiener Functional asymptot-
1.3 Regular and Singular Perturbations

In this thesis we also distinguish between two classes of asymptotic expansions, singular and regular perturbations. Although these categories are common in applied mathematics they do not seem to have been applied in mathematical finance. Let us explain briefly the difference between the two perturbation approaches. Consider a set of SDEs

\[ dX_t^\epsilon = b(X_t^\epsilon)dt + \epsilon\sigma(X_t^\epsilon)dW_t \]
\[ X_0^\epsilon = x_0 \]

with suitably smooth coefficients and depending on a small parameter \( \epsilon \in (0, 1] \). Whereas the solution of the SDE, for each \( \epsilon > 0 \) is a stochastic process, setting \( \epsilon = 0 \) in the above SDE results in an ODE, and a deterministic process. We consider
investigation of $X^\epsilon$ as $\epsilon$ tends to 0 a singular perturbation problem, since the $\epsilon = 0$ case is qualitatively different. Now take $b \equiv 0$ for simplicity and consider a change of variables,

$$Y^\epsilon_t = \frac{X^\epsilon_t - x_0}{\epsilon}.$$ 

Then the SDE for $Y^\epsilon$ is given by

$$dY^\epsilon_t = \sigma(x_0 + \epsilon Y^\epsilon_t) dW_t$$

Now clearly, $Y^\epsilon$ stays stochastic also for $\epsilon = 0$, hence we consider the investigation of the limiting process $Y^\epsilon$ as $\epsilon \to 0$ a regular perturbation problem. This rescaling causes the ‘typical’ coefficient range to narrow around $\sigma(x_0)$, so our leading order term is a Wiener integral, and this is quite standard. Then densities or expectations are found in terms of Gaussian random variables. This same categorisation applies to PDE methods. In the class of regular perturbations, we would place the investigation based on the PDE based approach by Hagan and Woodward ([37]) showing an expansion for the option price and then for the implied volatility associated to the system of SDEs

$$dF^\epsilon_t = \epsilon A(F^\epsilon_t) dW_t$$

$$F^\epsilon_0 = f_0.$$ 

This represents a local volatility model, where the volatility $A$ of the asset $F$ is a deterministic function of the asset price alone. Here they showed that by performing the above rescaling, the corresponding backward Fokker-Planck equation for the option price can be solved by assuming that the option price,

$$V^\epsilon(T, f_0) := E[(F^\epsilon_T - K^\epsilon)^+], \quad K^\epsilon = f_0 + \epsilon x, \quad x \in \mathbb{R}$$

can be written as a power series in $\epsilon$, leading to the solution of a hierarchy of PDE problems, where the lowest order solution is simply the Gaussian and higher order terms are derivatives of the Gaussian case:

$$V^\epsilon(T, f) = \epsilon G(T, x) + \epsilon^2 \nu_1 T x \partial_T G(T, x) + \cdots$$

where $G(s, x)$ is the Gaussian option price for a call option with strike $x$ on a normal asset with mean 0 and variance $s$. The predominant SDE applications of regular

\[ \text{Note the authors say they use singular perturbation techniques, which is justified in the sense that one is rescaling the singular perturbation problem to convert it into a regular perturbation problem (see eg. [13] Section 7.2, example 1).} \]
perturbations are the Yoshida-Watanabe expansions advocated by Takahashi and co-workers \([74, 71, 47]\), and developed using Malliavin calculus. This approach (as opposed to PDE methods) allows non Markovian processes to be investigated analysed so that marginals and path dependent quantities can be directly analysed. Again, solutions for option prices are found as expansions around a Gaussian model by employing the same rescaling as above. Although the method involves straightforward computations, the exponential growth in the number of terms with expansion order makes manual calculation forbidding. In Chapter 5 we develop a method to compute the expansion using Mathematica – it is to our knowledge, the first time such a procedure has been introduced. The Mathematica program can then generate C++ code for the specified model.

### 1.3.1 Singular Perturbations

Considering now singular perturbations, a fundamental paper was that of Berestycki, Busca and Florent \([14]\), often abbreviated to BBF, that showed that for a local volatility model,

\[
\frac{dF_t}{dt} = a(F_t, t) dW_t \\
F_0 = f_0
\]

satisfying suitable regularity and ellipticity conditions, the implied volatility, \(\sigma_I(K, T)\), satisfies a quasilinear degenerate PDE, and, using viscosity solution methods, showed that the small time limit of the PDE is well defined and given by

\[
\lim_{T \to 0} \frac{\log(K/f_0)}{\sigma_I(K, T)} = \int_{f_0}^{K} \frac{1}{a(x, 0)} dx. \tag{1.3.1}
\]

The key issue to note that the limit is for a fixed \(K\), whereas the regular perturbation approaches are valid only in the joint limit as \((K - f_0) \to 0\) and \(T \to 0\). We note, however, that this singular perturbation only provides the zero order limit, whereas the regular perturbation approaches above can go to arbitrarily high order. BBF went on to show \([15]\) that a similar result holds for stochastic volatility models

\[
\frac{dX_t}{dt} = b(X_t) + \sigma(X_t) dW_t \\
X_0 = x_0,
\]

where we take \(X^1\) to correspond to the asset price. Now the right hand side of (1.3.1) is replaced by the Riemannian distance, \(d(x, y)\), between the initial condition and the
hyperplane \((K, \cdot)\) induced by the metric
\[
ds^2 = \sum_{i,j} g_{ij} dx^i dx^j
\]
where \((g_{ij})\) is the inverse of the diffusion matrix \(\Sigma = \sigma \sigma^T\). Avellaneda, Boyer-Olson, Busca, and Friz ([7]) showed that this small time asymptotic limit for implied volatility could also be derived from Varadhan’s ([78, 79]) result for the small time large deviations rate function of the probability transition function, \(p_t(x, y)\), of an elliptic diffusion process
\[
\lim_{t \to 0} t \log p_t(x, y) = \frac{1}{2} d^2(x, y).
\]
In Chapter 3, we show that the BBF small time result and its generalisations follow naturally with minimal assumptions other than that the underlying asset price random variable satisfies a large deviations principle with respect to the given parameter (e.g. small time, small noise). Although the small time behaviour has been extensively studied, the small noise sample path behaviour has not. So we are led to investigate the required conditions on the rate function (such as monotonicity and continuity) in a more general setting. In Chapter 4 subsection 4.3.1, we show monotonicity of the rate function of a (possible path dependent) asset price from general principles. Similarly in proposition 6.2.5, we generalise the Hamiltonian optimality conditions from the transversality condition for the shortest path to a hyperplane, to a more general condition for the minimal energy path to a surface \(\{g(x_T) = K\}\), for some specified \(g\). Lastly we study existence of minimising paths and the continuity of the rate function in Theorem 6.2.4 under a hypoelliptic setting which is of interest for continuously monitored Asian options, since the running integral and the asset are not jointly elliptic, but are hypoelliptic. Furthermore we note that many other asymptotic problems, eg large strike and/or large time asymptotics can be turned into small time/small noise problems by suitable rescaling as we do in Chapter 6.

1.3.1.1 Calculation of energy function

The results of BBF highlight the key importance of the minimum energy path to determining small time asymptotics of the implied volatility. For elliptic diffusion coefficients, this can be viewed as a problem of calculating the geodesics on a Riemannian manifold. Unfortunately, there are only a few closed form solutions even for 2 dimensional surfaces. The majority are determined by isometric mappings of known manifolds. For instance, the BBF result (for a time homogeneous model) corresponds
to the Euclidean distance under the change of variables

\[ Y(F) = \int_{f_0}^{F} \frac{1}{a(x)} \, dx, \]

and the SABR model (35)

\[
\begin{align*}
    dF_t &= \alpha_t F_t^\beta \, dW^1_t, \quad F_0 = f_0 > 0, \ 0 \leq \beta \leq 1 \\
    d\alpha_t &= \nu \alpha_t \, dW^2_t, \quad \alpha_0 > 0, \ \nu \geq 0 \\
    d\langle W^1, W^2 \rangle_t &= \rho \, dt, \quad |\rho| \leq 1,
\end{align*}
\]

corresponds to the distance in the hyperbolic plane under the change of variables

\[
\Phi(F, \alpha) = \left( \frac{1}{\sqrt{1 - \rho^2}} \left( \int_{f_0}^{F} \frac{1}{x^\beta} \, dx - \frac{\rho \alpha}{\nu} \right), \frac{\alpha}{\nu} \right).
\]

Bourgade and Croissant (18) went on to investigate a range of other possible stochastic volatility models that could be reduced by isometric mappings, but found few classes that were financially relevant. Osajima (63, 64) investigated deriving the large deviations rate function (or energy) of the 1-dimensional marginal transition density of the solution of an SDE by an asymptotic expansion of the Hamiltonian equations governing the minimum energy path to achieve a state \(x_T\) with \(x_1^T = y\), starting from state \(x_0\) (with \(x_1^0 = \bar{F}_T\)) at time 0. He proved the following result:

**Theorem 1.3.1.** There is a constant \(r_0 > 0\) such that the energy function \(e\) is twice continuously differentiable on \([\bar{F}_T - r_0, \bar{F}_T + r_0]\) and a constant \(C_0 > 0\) such that the asymptotic expansion of the energy function satisfies

\[
|e(K) - \left[ \frac{1}{2b_1} (K - \bar{F}_T)^2 - \frac{b_2}{3b_1^2} (K - \bar{F}_T)^3 + \left( -\frac{b_3}{4b_1^4} + \frac{b_2^2}{2b_1^3} \right) (K - \bar{F}_T)^4 \right] | \leq C_0 |K - \bar{F}_T|^5
\]

Here \(b_1, b_2, b_3\) are constants derived from the SDE coefficients. This is an expansion for small \((\bar{F}_T - K)\), and cannot be expected to be accurate for a large range of strikes, which will be a particular issue for longer times. The small noise zero order implied volatility expansion is given by

\[
\sigma_{I,0}(K, T) = \frac{\left| \log(K/\bar{F}_T) \right|}{\sqrt{2e(K)T}}
\]

(see Chapter 3 for derivation). Therefore Osajima’s expansion for the energy gives

\footnote{We analyse the same energy and Hamiltonian equations for our marginal density expansions (see Chapter 6, 6.2).}
a short time implied volatility asymptotic with relative error \( O(K - \bar{F}_T)^3 \) for \( K - \bar{F}_T \ll 1 \). We can investigate the problems encountered by considering the SABR model for which the closed form for the energy is known. We plot an example below using the SABR parameters in [65], \((f_0 = 4, \alpha_0 = 0.3, \beta = 0.7, \nu = 0.4, \rho = -0.5)\). The blue curve is the exact short time implied volatility limit for SABR and the red curve is the implied volatility using Osajima’s asymptotic expansion of energy. We have assessed the effect of using a higher order expansion by calculating a 6th order Taylor series expansion of the (exact) energy around \( x_0^1 \), which is plotted in green (the spike is where the asymptotic approximation of the energy goes to zero, and then becomes negative. As can be seen, the asymptotic expansion breaks down away from \( x_0^1 \), and using a higher order expansion does not increase the range (as is typical for asymptotic expansions).

We have pursued small noise asymptotics in this thesis, but given the small class of closed form geodesics even for just two dimensional surfaces, we did not expect minimum energies for small noise problems to be in closed form. We therefore also investigated the numerical computation of minimum energy paths. We are unaware of other published results using numerical computation. We point out that our calculation took typically approximately 1ms on a two dimensional problem, whereas our
finite difference calculation of a single option price took 15 seconds. We note however, that large deviations by itself only provides the zero order expansion of the implied volatility.

1.3.1.2 Asymptotic expansions of the transition density

We now consider singular perturbation expansions that go beyond the zero order limit. These form the subject of Chapter 6. Perhaps the most famous and widely used such expansion is that of the SABR model expansion by Hagan et al. (35)

\[
\begin{align*}
\frac{dF_t^e}{\epsilon} &= \epsilon \alpha_t^e (F_t^e)^\beta dW_t^1 \\
\frac{d\alpha_t^e}{\epsilon} &= \epsilon \nu \alpha_t^e dt \\
\langle W^1, W^2 \rangle_t &= \rho t, \quad \alpha_0^e = \alpha_0 > 0, \quad F_0^e = f_0 > 0,
\end{align*}
\]

\(\beta \in [0, 1], |\rho| < 1, \nu \in \mathbb{R}^+\). Hagan et al. derived a short time asymptotic expansion of the transition density by the WKB method ([40]) and after integrating out derived an asymptotic expansion for the implied volatility up to order\(^5 O(T^2)\). It was soon realised ([36, 38]) that the expansion could be derived more directly using the small time asymptotics of the heat kernel \(p_t(x, y)\) on a Riemannian manifold \((M, g)\), namely the hyperbolic plane. Here \(M\) is a manifold endowed with inner product \(g\) and the \((g_{ij})\) is the inverse of the (elliptic) diffusion coefficient matrix. We define \(d(x, y)\) as the distance between two points on \(M\) induced by the inner product \(g\) on \(M\). A point \(y \in M\) is in the cut locus of \(x\), \(\text{Cut}(x)\), if there is no unique minimising geodesic connecting \(x\) and \(y\), or the points \(x\) and \(y\) are conjugate. Let \(C_M \subset M \times M\) be the set of pairs of points \((x, y)\) such that \(y \in \text{Cut}(x)\), then Minakshisundaram and Pleijel [58] showed the following theorem\(^6\).

**Theorem 1.3.2.** Let \((M, g)\) be a smooth, complete Riemannian manifold of dimension \(n\). Then there are smooth functions \(H_i(x, y)\) defined on \((M \times M) \setminus C_M\) such that the asymptotic expansion

\[
p_t(x, y) \sim \left(\frac{1}{2\pi t}\right)^{n/2} e^{-d^2(x, y)/2t} \sum_{i=0}^{\infty} H_i(x, y)t^i
\]

holds uniformly as \(t \searrow 0\) on compact subsets of \((M \times M) \setminus C_M\). Further, if \(y = \exp_x(Y)\), then \(H_0(x, y)\) is given by the reciprocal of the square root of the Jacobian of \(\exp_x\) at \(Y\).

---

\(^5\)Various approximations made it also an expansion in \(|K - f_0|\). These were subsequently removed by Paulot, [65].

\(^6\)Although some work has been done on establishing asymptotic expansions also in the cut-locus [59], the variety of singularities precludes a general theory.
This theorem has been further investigated by numerous theoretical physicists and mathematicians, and many coefficients of the expansion have been derived for specific cases ([57]). The relevance of geometric concepts of geodesics, conjugate points and cut locus to determining the small time transition density of the solution of an SDE becomes clearer from the Laplace method on Wiener Space. We recall that the Laplace method in $\mathbb{R}^n$ is used to develop an asymptotic expansion of

$$
\lim_{\epsilon \to 0} \int_{\mathbb{R}^n} g(w) \exp\left( -\frac{1}{\epsilon^2} e(w) \right) dw
$$

for smooth functions $e, g$. Assuming a single nondegenerate (i.e., strictly positive second derivative) minimum of $e(\cdot)$ at $w_0$, and using the change of variables, $z = \frac{w - w_0}{\epsilon}$,

$$
\int_{\mathbb{R}^n} g(w) \exp\left( \frac{1}{\epsilon^2} e(w) \right) dw \\
= |\epsilon|^n \int_{\mathbb{R}^n} g(w_0 + \epsilon z) \exp\left( -\frac{1}{\epsilon^2} e(w_0 + \epsilon z) \right) dz \\
\sim |\epsilon|^n \exp\left( -\frac{1}{\epsilon^2} e(w_0) \right) \int_{\mathbb{R}^n} g(w_0 + \epsilon z) \exp\left( -\frac{1}{2} \sum_{i,j=1}^{n} \partial_{z_i} \partial_{z_j} e(w_0 + \epsilon z) \bigg|_{z=0} z_i z_j + O(\epsilon) \right) dz
$$

Gaussian integral, therefore

$$
= |\epsilon|^n g(w_0) \exp\left( -\frac{1}{\epsilon^2} e(w_0) \right) (2\pi)^{n/2} (\det \partial_{z_i} \partial_{z_j} e(w_0 + \epsilon z) \bigg|_{z=0})^{-1/2} + O(\epsilon^{n+1})
$$

We have ignored discussion of control of the tails to concentrate on the key requirements for a single\textsuperscript{7} non degenerate minimum i.e.

$$
\left( \frac{\partial}{\partial z} e(w_0 + \epsilon z) \right) \bigg|_{z=0} = 0,
$$

$$
\left( \frac{\partial^2}{\partial z_i \partial z_j} e(w_0 + \epsilon z) \right) \bigg|_{z=0} > 0.
$$

In determining asymptotics of the small time transition density, the rate function plays the role of $e(w)$. Geodesics are the minimising paths between a given starting point, $x$, and the destination $y$, and non degeneracy of the minimum corresponds to non conjugacy of the initial and final points along the minimising geodesic. In Chapter \textsuperscript{6} we\textsuperscript{8} extend the work of Ben Arous [11, 6] and Bismut [17] that used the Laplace method on Wiener Space and Malliavin calculus to derive asymptotic ex-

\textsuperscript{7}a finite number of minima is also handled straightforwardly
\textsuperscript{8}This is joint work with Prof P. Friz, Prof. J-D Deuschel and Dr A. Jacquier, [23]
expansions of the transition density in the hypo-elliptic case, which corresponds to a sub-Riemannian geometry. This is one of the areas where Malliavin Calculus has proved to be most successful. In the elliptic case, it was well known (e.g.\cite{59},\cite{78} ) that the small-time asymptotics depended crucially on the Riemannian geometry induced by the diffusion coefficients of the SDE. In the hypo-elliptic case, the associated geometry had not been studied before the interest shown from probabilists. In fact one of the key achievements of Bismut’s analysis (\cite{17}) was to show under what condition geodesics were given as solutions of the corresponding Hamiltonian system ( in the Riemannian case always). Furthermore, Bismut showed that the non-degeneracy of the minimum path was determined by the rank of the Jacobian associated with the (finite-dimensional ) Hamiltonian system, generalising the notion of conjugate points to the sub-Riemannian case. Our extensions involve two main changes. Firstly we perform a small noise expansion rather than small time, so the drift enters also at the zeroth order expansion. Secondly, we directly work on the marginal density: let \(X^\epsilon_T\) be the solution of a small noise SDE, then we define \(Y^\epsilon_T = \Pi_l X^\epsilon_T\) as the projection of \(X^\epsilon_T\) onto its first \(l\) components. Working with the projection, \(Y^\epsilon_T\), and therefore the minimal energy to a hyperplane rather than a point, requires us to generalise the concept of (sub-Riemannian) conjugate points to the corresponding notion of (sub-Riemannian) focal points. In Chapter 6 we therefore consider focal points, and develop a Hamiltonian test for focal points, generalising Bismut’s test for conjugate points.

Using this small noise expansion, we are able to consider the large strike asymptotics of certain stochastic volatility models. Consider the Stein and Stein model of stochastic volatility given by

\[
\begin{align*}
\frac{dY_t}{Y_t} &= -\frac{1}{2}Z_t^2 dt + Z_t dW^1_t, \quad Y_0 = y_0 = 0 \\
\frac{dZ_t}{Z_t} &= (a + bZ_t)dt + c dW^2_t, \quad Z_0 = \sigma_0 > 0,
\end{align*}
\]

where \(Y_t\) is the log of the asset price at time \(t\), and \(Z_t\) is its associated volatility. Recently, Gulisashvili and Stein (\cite{34}) showed that the density of the log of the stock price \(x = Y_t\) had the following asymptotic expansion

\[
B_1 e^{-B_3 x} e^{-B_2 \sqrt{x} x^{-\frac{1}{2}}} \left(1 + O\left(x^{-\frac{1}{2}}\right)\right) \text{ as } x \to \infty,
\]

where \(B_1, B_2, B_3\) are constants determined from the coefficients of the SDE. The following simple argument show how large strike asymptotics can be studied through small noise asymptotic results by employing a suitable scaling. Consider \(Y^\epsilon = \epsilon^k Y\), for some fixed \(k \geq 1\), then clearly \(\mathbb{P}(Y_T \geq 1/\epsilon^k) = \mathbb{P}(Y^\epsilon_T \geq 1)\), so we can study the
large strike asymptotics of $Y$ by considering a rescaled process $Y^\epsilon$. In the case of the Stein-Stein model, we take the scaling $Y^\epsilon := \epsilon^2 Y$, $Z^\epsilon := \epsilon Z$ to get

$$
\begin{align*}
    dY^\epsilon_t &= -\frac{1}{2} (Z^\epsilon_t)^2 dt + \epsilon Z_t^\epsilon dW_1^\epsilon, \quad Y^\epsilon_0 = 0 \\
    dZ^\epsilon_t &= (\epsilon a + bZ^\epsilon_t) dt + \epsilon c dW^2_2, \quad Z^\epsilon_0 = \epsilon \sigma_0, \quad \epsilon > 0.
\end{align*}
$$

This leads to an interesting small noise problem ($\epsilon \downarrow 0$). Note in particular, that the drift of $Z^\epsilon$ has a term of order $O(\epsilon)$, unlike the more common small time expansion where the drift is $O(\epsilon^2)$, or the more commonly studied small noise perturbations where the drift has no $\epsilon$-dependence. Since the SDE coefficients are not smooth functions of $s = \epsilon^2$ in a neighbourhood of zero, Kusuoka and Stroock’s asymptotic expansion ([50]) is not applicable. Furthermore, the initial condition depends on $\epsilon$ and causes the diffusion to degenerate at $\epsilon = 0$. Our expansion based on Ben Arous’ work is applicable and allows us to extend Gulisashvili’s result for the Stein-Stein model to a correlated generalisation, the Schöbel and Zhu model ([69]):

$$
\begin{align*}
    dY_t &= \frac{1}{2} Z^2_t dt + Z_t dW_1^1, \quad Y_0 = y_0 = 0 \\
    dZ_t &= (a + bZ_t) dt + cdW^2_t, \quad Z_0 = \sigma_0 > 0 \\
    d\langle W^1, W^2 \rangle_t &= \rho dt.
\end{align*}
$$

1.4 Our Contribution

Here we summarise our contribution to research on Wiener functional asymptotics and applications to Mathematical Finance.

We are (to our knowledge) the first to analyse small noise sample path asymptotics of implied volatility. Essentially all previous work has focussed on small time asymptotics. In Chapter 3 we develop the LDP argument in a general enough framework to encompass small noise sample path asymptotics. In particular, the large deviations rate function is not a Riemannian distance function because now the drift term also determines the rate function. Secondly, considering (small noise) sample path asymptotics allows us to analyse functionals of the diffusion path, such as the running maximum or time integral, rather than simply the final value. We are led to provide (Lemma 4.3.5) a generic proof that the rate function for the asset induced by the sample path rate function is monotonic. In the small time case, where the rate function is a distance (to a point or a hyperplane), this is more straightforward.

---

9 or rather $1/2$ the squared distance
We show by numerical example that small noise asymptotics provide much better approximations than small time asymptotics for stochastic volatility models with mean reversion (using parameters calibrated to the market). We argue that small noise asymptotics will also provide a better approximation for models with time varying parameters. In particular we show that small noise sample path asymptotics give a rigorous justification of Gatheral and Wang’s “Heat Kernel most likely path approximation” see section 3.2. Future work will therefore consider the analysis of path dependent options using the sample path large deviations results.

In Chapter 4, we show small noise sample path asymptotics for a class of two dimensional positive diffusions, which covers the Double CEV model (DCEV), one of a new generation of two dimensional stochastic volatility models. The key issue here is that the dispersion coefficients of the DCEV model are only Hölder continuous rather than Lipschitz continuous. We develop a localisation argument for a two dimensional process to allow us to use Baldi and Caramellino’s [9] one dimensional results for positive diffusions. This also requires an extension of the comparison theorem presented in [42], Proposition 5.2.18 for one dimensional diffusions with different drifts to a two dimensional case 4.A.

In Chapter 5, we develop a Mathematica program to automatically produce Yoshida-Watanabe asymptotic expansions of diffusion processes and apply it to developing a small noise expansion for a call option on the VIX in the DCEV model. Although such expansions have been derived before for particular models, no automated method has been presented; here we present a Mathematica program to perform the calculation. Mathematica is then used to generate C++ code to calculate the small noise expansion up to order $\epsilon^3$. The resulting program is over 100,000 lines long, and could clearly not have been produced without computer support. Nevertheless, the option (approximate) valuation takes only 10 milliseconds. In that chapter we also compare to a small time expansion of order $\epsilon^4$ by Chenxu Li ([52]) which is seen to perform significantly worse. As we mentioned above, PDE approaches can also be used and will lead to the same expansions, so the same computational problem must be faced. There is certainly room for further developments here in analysing the structure of the expansion and reduce the exponential growth of terms.

Lastly in Chapter 6, we extend Ben Arous’ small time expansion of the transition density of an SDE, to a small noise expansion of the marginal of an SDE. In doing so we derive a new result for the large strike expansion of the Schöbel and Zhu model. I was responsible for identifying the key requirement for a focal point condition, beyond the conjugate point condition (in particular if a path is a non-degenerate minimiser to a plane, then it is necessarily a non-degenerate minimiser to that minimising point on the plane, but not vice versa). I extended the Hamiltonian analysis to functions
of the endpoints, and not just the orthogonal projection onto the first $i$ coordinates of the endpoints. The goal of this extension is to develop a Hamiltonian analysis of the minimising paths that define the energy function for more general assets, such as VIX (a smooth function of the endpoints) and certain path dependent quantities such as the time integral of a state variable (as required for a continuously monitored Asian option). We note that a hypo-elliptic framework is then necessary because adding the running time integral as an additional state variable turns an elliptic problem into a hypo-elliptic problem. In the analysis of the Stein and Stein model worked on the Hamiltonian analysis of the Stein and Stein model, in particular the focal point condition, and generalised the analysis to the correlated case of the Schöbel and Zhu model. Converting the large strike limit into a small noise limit provides an elegant analysis when applicable. One key issue raised by our analysis of the large strike problem is the need to prove an LDP also in the case that the initial state moves towards the origin. This is problematic for many stochastic volatility models, since the boundary behaviour (at zero volatility) often leads to the diffusion coefficient not being Lipschitz continuous there. There are few results on proving an LDP in these situations, and the localisations we apply in Chapter 4 are not directly applicable, since the initial state is moving towards the origin. A further area of research is therefore investigating whether an LDP can be shown in these conditions.
Chapter 2

Preliminaries

2.1 Large deviations

Large deviations studies the exponential rate of decay of a family of probability measures \( \{\mu_\epsilon\} \) on a measurable space \((\chi, \mathcal{B})\) as \( \epsilon \to 0 \) using a rate function which provides upper and lower bounds for subsets of \( \chi \). We follow the definitions in ([20]).

**Definition 2.1.1.** A rate function \( I \) is a lower semicontinuous mapping: the level sets \( \{x \in \chi : I(x) \leq a\} \) are closed for \( a \in [0, \infty) \). A good rate function has compact level sets.

We note that this implies that the infimum of a good rate function on a non-empty closed set in \( \chi \) is achieved.

**Definition 2.1.2.** The family of probability measures \( \{\mu_\epsilon\} \) satisfies the large deviation principle with a rate function \( I \), if for all \( \Gamma \in \mathcal{B} \),

\[
- \inf_{x \in \Gamma^o} I(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x)
\]

We will primarily be interested in large deviations on the sample paths of Brownian motion and solutions of SDE. For \( T > 0 \), let \( \mathcal{C}^d = \mathcal{C}([0, T], \mathbb{R}^d) \) denote the Banach space of continuous paths \([0, T] \to \mathbb{R}^d\) endowed with the topology of uniform convergence. Then for \( f \in \mathcal{C}^d \) we define

\[
\|f\|_T = \sup_{t \in [0, T]} |f_t|.
\]

We set \( \mathcal{C}^d_x \) as the closed set of paths starting at \( x \in \mathbb{R}^d \). Let \( \mathcal{H}^m = \mathcal{H}([0, T], \mathbb{R}^m) \) be the subspace of \( \mathcal{C}^m_0 \) consisting of paths that are absolutely continuous and whose derivative is square integrable on \([0, T]\) and endowed with the Hilbert norm \( \|\cdot\|_H \) and
inner product \(<\cdot,\cdot>_H\), that is
\[\|h\|_H^2 = \int_0^T \|h(s)\|^2 \, ds,\]
\[\langle h_1, h_2 \rangle_H = \int_0^T \langle h_1(s), h_2(s) \rangle \, ds.\]

For \(h \in \mathcal{C}_0^m\), we set
\[I(h) = \begin{cases} \frac{1}{2} \|h\|_H^2 & \text{if } h \in \mathcal{H}_m \\ +\infty & \text{otherwise.} \end{cases} \]

(2.1.1)

Let \(w_t\) be standard Brownian motion on \(\mathbb{R}^m\) and consider the scaled process \(w_{\epsilon t} = \epsilon w_t\). Let \(\nu_\epsilon\) denote the measures induced by \(w_{\epsilon t}\) on \(\mathcal{C}_0^m\).

**Theorem 2.1.3.** (Schilder [68]). \(\nu_\epsilon\) satisfies an LDP in \(\mathcal{C}_0^m\) with good rate function \(I(h)\).

We note that \(I(h)\) is weakly lower semicontinuous in \(\mathcal{H}_m\), which implies in particular that its infimum is achieved on closed balls in \(\mathcal{H}_m\).

**Theorem 2.1.4** (Contraction principle, [20] Theorem 4.2.1). Let \(\mathcal{X}\) and \(\mathcal{Y}\) be Hausdorff topological spaces and \(f: \mathcal{X} \to \mathcal{Y}\) a continuous function. Consider a good rate function \(I: \mathcal{X} \to [0, \infty)\).

1. For each \(y \in \mathcal{Y}\), define \(I'(y) := \inf\{I(x) : x \in \mathcal{X}, \quad y = g(x)\}\). Then \(I'\) is a good rate function on \(\mathcal{Y}\).

2. If \(I\) controls the LDP associated with a family of probability measures \(\{\mu_\epsilon\}\) on \(\mathcal{X}\), then \(I'\) controls the family of probability measures \(\{\mu_\epsilon \circ g^{-1}\}\) on \(\mathcal{Y}\).

If solutions of SDEs were continuous functions of Brownian motion, then the contraction principle would allow us to conclude that the family of probability measures induced by solutions of the SDE
\[dX^\epsilon_t = b(X^\epsilon_t)dt + \epsilon \sigma(X^\epsilon_t)dW_t,\]
\[X^\epsilon_0 = x_0\]
satisfy a Large Deviations Principle governed by the rate function
\[\mathcal{J}(g) = \inf\{I(h) : g = \phi^h(x_0)\} \quad g \in \mathcal{C}_{x_0}^d\]
where \( \phi^h(x_0) \) is the solution of the ODE

\[
d\phi^h_t(x_0) = b(\phi^h_t(x_0))dt + \sigma(\phi^h_t(x_0))\dot{h}(t)dt
\]

\[
\phi^h_0(x_0) = x_0 \in \mathbb{R}^d.
\]

Although solutions of SDEs fail to be continuous (in the supremum norm) in general, they nevertheless satisfy the above LDP. Wentzell and Freidlin proved this under restricted conditions on the coefficients by approximating the SDE solution with a continuous functional, namely the Euler-approximation, with an error that had exponentially decaying probability of exceeding any fixed bound. Azencott \[8\] subsequently introduced the notion of quasi-continuity which allowed him to weaken the conditions on the SDE coefficients and allowed him to show an LDP even for Hypo-elliptic SDEs. In Chapter 4, we will use Azencott’s method to show an LDP for a class of two-dimensional positive diffusions. We then use the contraction principle to show an LDP for a continuous functional of the SDE solution see lemma 4.3.5.

Since SDE solutions satisfy an LDP, a Laplace method can be employed on Wiener space \([17, 6]\) to determine the small noise asymptotic probability density of solutions of SDEs. We apply this method to develop an asymptotic expansion of marginal densities of SDE solutions in Chapter 6.

### 2.2 Malliavin Calculus

In this section we review the basics of Malliavin calculus as applied to the determination of asymptotic expansions of probability densities for solutions of SDEs, following the survey in \[80\] from which the below theorems are taken. Malliavin Calculus is used in Chapters 5 and 6 to develop asymptotic expansions. We provide more in depth references from \[70\], and note that \[41\] Section 5.8-5.10 contains an expanded version of \[80\]. Let \((\mathcal{C}_0^m, \mathbb{P})\) be the \(m\)-dimensional Wiener space, with \(\mathbb{P}\), the standard Wiener measure on the \(\mathbb{P}\)-completion of the Borel field over \(\mathcal{C}_0^m\). To simplify notation, we will use \(W = \mathcal{C}_0^m\) and \(H = \mathcal{H}^m\). Given a real separable Hilbert space \(E\), with norm \(\| \cdot \|_E\), we denote by \(L_p(E)\) the \(L_p\)-space, \(1 \leq p < \infty\), of \(E\)-valued Wiener functionals. A function \(f : W \mapsto E\) is called an \(E\)-valued polynomial if there exist \(n \in \mathbb{Z}^+, h_1, h_2, \ldots, h_n \in H\) and a polynomial with \(E\)-valued coefficients \(p : \mathbb{R}^n \mapsto E\) such that

\[
f(w) = p([h_1](w), [h_2](w), \ldots, [h_n](w)),
\]

for \(h_i \in H\) and

\[
[h_i](w) = \sum_{j=1}^m \int_0^T \dot{h}_j^i(t) dW_j^i.
\]
are defined in the sense of Ito’s stochastic integrals. We denote the space of \( E \)-valued polynomials on \( W \), by \( \mathcal{P}(E) \). Using the \( L_p(E) \) norm \( \|F\|_p = (\int_W |F|^p_E \mathbb{P}(dw))^{1/p} \), we define a sequence of norms on \( \mathcal{P}(E) \) (see [70] Definition 4.8)

\[
\|F\|_{p,s} = \|(I - \mathcal{L})^{s/2} F\|_p \quad s \in \mathbb{R}, \quad p \in (1, \infty), \quad F \in \mathcal{P}(E).
\]

Here \( \mathcal{L} \) is the Ornstein-Uhlenbeck operator and

\[
(I - \mathcal{L})^{s/2} F = \sum_{n=0}^{\infty} (1 + n)^{s/2} J_n F, \quad F \in \mathcal{P}(E),
\]

where \( J_n \) are the projection operators in the Wiener chaos decomposition in \( L^2(E) \). We define \( D^s_p(E) \) as the Banach space formed by the completion of \( \mathcal{P}(E) \) with respect to \( \|\cdot\|_{p,s} \). Then \( D^0_p = L_p(E) \) and we have the following relationship (see [70] Proposition 4.9)

\[
D^{s'}_{p'}(E) \hookrightarrow D^s_p(E) \quad \text{if} \quad p \leq p' \quad \text{and} \quad s \leq s'
\]

where \( \hookrightarrow \) denotes continuous inclusion. The dual space of \( D^s_p(E) \) is \( D^{-s}_{q^*}(E) \), for \( s \in \mathbb{R}, \quad p > 1, \quad 1/p + 1/q = 1. \) Then \( D^\infty(E) \) is defined as

\[
D^\infty(E) = \cap_{s>0} \cap_{1<p<\infty} D^s_p(E),
\]

and its dual by

\[
D^{-\infty}(E) = \cup_{s>0} \cup_{1<p<\infty} D^{-s}_p(E).
\]

Elements in \( D^s_p \) are standard Wiener functionals for \( s \geq 0 \) (since \( D^0_p(E) = L_p(E) \)), but if \( s < 0 \), some elements of \( D^s_p(E) \) are not, and are named generalised Wiener functionals. We also introduce the spaces

\[
\tilde{D}^\infty(E) = \cap_{s>0} \cup_{1<p<\infty} D^s_p(E)
\]

and

\[
\tilde{D}^{-\infty}(E) = \cup_{s>0} \cap_{1<p<\infty} D^{-s}_p(E)
\]

We will abbreviate the corresponding \( \mathbb{R} \) valued space by \( \mathcal{P}, \, D^s_p \) etc. We define the H-derivative \( D : \mathcal{P}(E) \to \mathcal{P}(H \otimes E) \) by

\[
DF(w)[h,e] = \left. \frac{d}{dt} \langle F(w + th), e \rangle_E \right|_{t=0}, \quad h \in H, \quad e \in E, \quad F(w) \in \mathcal{P}(E)
\]

and

\[
D^k : \mathcal{P}(E) \to \mathcal{P}(H \otimes H \otimes \cdots \otimes H \otimes E)
\]
successively by $D^n = D(D^{n-1})$. Then we have the following result -

**Theorem 2.2.1** (cf. [70], Theorem 4.4 Meyer’s Equivalence of Norms). The operator $D : \mathcal{P}(E) \mapsto \mathcal{P}(H \otimes E)$ is uniquely extended to a linear operator $D^{-\infty}(E) \mapsto D^{-\infty}(H \otimes E)$ which is continuous in the sense that its restriction $D_{p+1}^s(E) \mapsto D_p^s(H \otimes E)$ is continuous for every $p \in (1, \infty)$ and $s \in \mathbb{R}$.

An $m$-dimensional Wiener functional $F : C_0^m \mapsto \mathbb{R}^d$ is smooth (in the sense of Malliavin), if $F \in D_{-\infty}(\mathbb{R}^d)$. In this case, $\Xi^{ij}(w) = \langle DF_i(w), DF_j(w) \rangle_H \in D_{-\infty}^\infty$, $i, j = 1, 2, \ldots, d$, and the Wiener functional $\Xi = (\Xi^{ij})$ with values in non-negative definite symmetric $d \times d$ matrices is called the *Malliavin covariance* of $F$. $F$ is said to be nondegenerate if $\det \Xi(w) > 0 \text{ P.a.s}$ and $1/\det \Xi(w) \in L^p \forall p > 1$. (2.2.1)

Let $\mathcal{S}(\mathbb{R}^d)$ be the real Schwartz space of rapidly decreasing $C^\infty$-functions on $\mathbb{R}^d$ and set $\|\phi\|_{2n} = \| (1 + |x|^2 - \Delta)^n \phi \|_\infty$, $\phi \in \mathcal{S}(\mathbb{R}^d)$, $n \in \mathbb{Z}$ where $\|\|_\infty$ is the supremum norm and $\Delta = \sum_{i=1}^{d} (\partial_{x_i})^2$. Let $\mathcal{S}_{2k}$ be the completion of $\mathcal{S}(\mathbb{R}^d)$ by the norm $\|\|_{2k}$. Then we have $\mathcal{S}(\mathbb{R}^d) \hookrightarrow \cdots \hookrightarrow \mathcal{S}_2 \hookrightarrow \mathcal{S}_0 = \mathcal{C}(\mathbb{R}^d) \hookrightarrow \mathcal{S}_{-2}(\mathbb{R}^d) \hookrightarrow \cdots \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$

where $\mathcal{C}(\mathbb{R}^d)$ is the Banach space of all real continuous functions on $\mathbb{R}^d$ tending to 0 at infinity and endowed with the supremum norm, and $\mathcal{S}'(\mathbb{R}^d)$ is the Schwartz space of real tempered distributions on $\mathbb{R}^d$. Furthermore, $\mathcal{S}(\mathbb{R}^d) = \bigcap_{n=1}^{\infty} \mathcal{S}_{2n}$ and $\mathcal{S}'(\mathbb{R}^d) = \cup_{k=1}^{\infty} \mathcal{S}_{-2k}$. Introduction of the Schwartz spaces allows us to represent the Dirac measure on $\mathbb{R}^d$ and other functionals as distributions that can be differentiated by way of an integration by parts formula (see e.g. [53], Chapter 6). In particular the Dirac function $\delta_y \in \mathcal{S}'(\mathbb{R}^d)$ is in $\mathcal{S}_{-2p}(\mathbb{R}^d)$ for $p > d/2$ (see [41] Lemma 5.9.1).

**Theorem 2.2.2** ([41] Theorem 5.9.1). Let $F \in D^\infty(\mathbb{R}^d)$ be given and satisfy the nondegeneracy condition (2.2.1). Then for every $p \in (1, \infty)$ and $n \in \mathbb{N}$, there exists a positive constant $C = C_{p,n}$ such that $\| \phi \circ F \|_{p,-2n} \leq C \| \phi \|_{-2n}$ $\forall \phi \in \mathcal{S}(\mathbb{R}^d)$. 
Therefore the map \( \phi \in \mathcal{S}(\mathbb{R}^d) \mapsto \phi \circ F \in D^\infty \) can be extended uniquely to a linear map

\[
T \in \mathcal{S}'(\mathbb{R}^d) \mapsto T \circ F \in D^{-\infty},
\]
such that its restriction \( T \in \mathcal{S}_{-2n} \mapsto T \circ F \in D^{-2n}_p \) is continuous for every \( p \in (1, \infty) \) and \( n \in \mathbb{N} \). In particular, \( T \circ F \in \tilde{D}^{-\infty} = \bigcup_{n=1}^{\infty} \cap_{1 < p < \infty} D^{-n}_p \) for every \( T \in \mathcal{S}'(\mathbb{R}^d) \).

This then allows the properties (such as smoothness) of the density of the Wiener functional to be analysed.

We now come to asymptotic expansions of Wiener functionals.

**Definition 2.2.3.** We say that \( F(\epsilon, w) = O(\epsilon^k) \) as \( \epsilon \downarrow 0 \) in \( D^s_p \) if \( F(\epsilon, w) \in D^s_p \) for all \( \epsilon \in (0, 1] \) and

\[
\limsup_{\epsilon \downarrow 0} \frac{\|F(\epsilon, w)\|_{s,p}}{\epsilon^k} < \infty \quad k \in \mathbb{R},
\]

Let \( F(\epsilon, w) \in D^\infty \) for all \( \epsilon \in (0, 1] \), and \( f_0, f_1, \cdots \in D^\infty \), then we say that

\[
F(\epsilon, w) \sim f_0 + \epsilon f_1 + \cdots \quad \text{in } D^\infty(E) \text{ as } \epsilon \downarrow 0
\]

if for every \( p \in (1, \infty), \, s > 0, \, k \in \mathbb{Z}, \)

\[
F(\epsilon, w) - \sum_{i=0}^{k-1} \epsilon^i f_i = O(\epsilon^k) \quad \text{in } D^s_p \text{ as } \epsilon \downarrow 0.
\]

Similar definitions apply for \( \tilde{D}^\infty, D^{-\infty}, \) and \( \tilde{D}^{-\infty} \). In particular if

\[
\Phi(\epsilon, w) \sim \Phi_0 + \epsilon \Phi_1 + \cdots \quad \text{in } D^{-\infty}(E) \text{ as } \epsilon \downarrow 0,
\]

then the asymptotic expansion of its expectation is given by

\[
E[\Phi(\epsilon, w)] \sim E[\Phi_0] + \epsilon E[\Phi_1] + \cdots \quad \text{as } \epsilon \downarrow 0.
\]

**Definition 2.2.4.** Given a family \( F(\epsilon, w), \, \epsilon \in (0, 1], \) of elements in \( D^\infty(\mathbb{R}^d) \), we say it is *uniformly nondegenerate* if for each \( \epsilon \in (0, 1] \), \( F(\epsilon, w) \) is nondegenerate (2.2.1), and

\[
\limsup_{\epsilon \downarrow 0} \| \det \Xi(\epsilon, w)^{-1} \|_p < \infty \quad \forall p \in (1, \infty),
\]

where \( \Xi(\epsilon, w) \) is the Malliavin covariance of \( F(\epsilon, w) \).

**Theorem 2.2.5** ([80] Theorem 2.3). Let \( F(\epsilon, w) \in D^\infty(\mathbb{R}^d), \, \epsilon \in (0, 1], \) be uniformly
nondegenerate and possess the asymptotic expansion

\[ F(\epsilon, w) \sim F_0 + \epsilon F_1 + \cdots \text{ in } D^\infty \text{ as } \epsilon \downarrow 0. \]

Then for every \( T \in \mathcal{S}'(\mathbb{R}^d) \), \( T(F(\epsilon, w)) \in \tilde{D}^{-\infty} \) has the asymptotic expansion in \( \tilde{D}^{-\infty} \) as \( \epsilon \downarrow 0 \):

\[ T(F(\epsilon, w)) \sim \Phi_0 + \epsilon \Phi_1 + \cdots \text{ in } \tilde{D}^{-\infty} \text{ as } \epsilon \downarrow 0, \]

and \( \Phi_i \in \tilde{D}^{-\infty} \) are determined by the formal Taylor expansion

\[ T(F_0 + [\epsilon F_1 + \epsilon^2 F_2 + \cdots]) = \sum_n \frac{1}{n!} D^n T(F_0) [\epsilon F_1 + \epsilon^2 F_2 + \cdots]^n \]

where \( n = (n_1, \ldots, n_d) \) is a multi-index, \( n! = n_1! \cdots n_d! \), \( a^n = a_1^{n_1} \cdots a_d^{n_d} \), for \( a \in \mathbb{R}^d \) and \( D^n = \partial_x^{n_1} \cdots \partial_x^{n_d} \).

The above theorem then justifies the asymptotic expansion formulas for densities of Wiener functionals and the expectations of other generalised Wiener Functionals that we use in Chapter 5.

### 2.2.1 Application to SDEs

Let \( V_j(x) = (V_j^1(x), V_j^2(x), \ldots, V_j^d), j = 0, 1, \ldots, m \) be a system of \( \mathbb{R}^d \)-valued functions defined on \( \mathbb{R}^d \). We suppose \( V_j^j(x) \) are \( C^\infty \)-functions with bounded derivatives of all orders, and consider the Stratonovich SDE’s (denoted by \( \circ dW \)) on \( \mathbb{R}^d \)

\[ dX_t = \sum_{j=1}^m V_j(X_t) \circ dW_t^j + \epsilon^2 V_0(X_t) dt \quad X_0 = x_0 \in \mathbb{R}^d. \]

Then a unique solution to the above SDE exists for every \( x_0 \in \mathbb{R}^d \) and

1. \( t \to X(t, x_0, w) \) is a sample path of A-diffusion process starting at \( x_0 \),

2. with probability one, \( (t, x_0) \to X(t, x_0, w) \) is continuous.

Let \( \Omega_t = (\Omega^j_k(t, x_0, w)) \) be defined as the unique solution of the following matrix SDE equation,

\[ d\Omega_t = \sum_{k=1}^m \partial_x V_k(X_t) \Omega_t \circ dW_t^k + \partial V_0(X_t) \Omega_t dt \quad \Omega_0 = 1 : \text{ the identity matrix.} \]

or in component notation
\[
d\Omega_{i,j} = \sum_{k=1}^{m} \sum_{p=1}^{d} \partial_{x^p} V_k^i(X_t) \Omega_{i,j}^{p} \circ dW_t^k + \sum_{p=1}^{d} \partial_{x^p} V_0^i(X_t) \Omega_{i,j}^{0} dt
\]
\[
\Omega_{0,j} = \delta_{ij} : \text{ the identity matrix.}
\]

We then have the following theorem

**Theorem 2.2.6** ([41] Proposition 5.10.1). Let \( t > 0 \) and \( x_0 \in \mathbb{R}^d \) be fixed. Then \( X(t, x_0, w) \) is smooth in the sense that

\[
X(t, x_0, w) \in D^\infty(\mathbb{R}^d)
\]

Furthermore the Malliavin covariance \( \Xi_{ij}^t(w) = \langle DX^i(t, x_0, w), DX^j(t, x_0, w) \rangle_H \) is given by

\[
\Xi_{ij}^t = \sum_{k=1}^{m} \int_0^t \left( \Omega_{4t}^{-1} \Omega_k^{-1} \right)^i (\Omega_{4t}^{-1} \Omega_k^{-1})^j ds \quad (2.2.2)
\]

**Definition 2.2.7.** We define the Hörmander condition

\[
\dim \mathcal{L} \{ V_{\alpha_n}, \ldots, V_{\alpha_1}, V_{\alpha_0} \ldots \} \cap \mathbb{R}^d : \quad 0 \leq n \leq n_0, \text{ where } \alpha_0 \in \{1, 2, \ldots, m\} \text{ and }\alpha_i \in \{0, 1, \ldots, m\} \text{ for } i \in \{1, 2, \ldots, n_0\} = d, \quad (2.2.3)
\]

for some \( n_0 \). Here

\[
[V, U]^i(x) = \sum_{j=1}^{d} \left\{ V^j(x) \partial_{x^j} (U^i(x)) - U^j(x) \partial_{x^j} (V^i(x)) \right\},
\]

and \( \mathcal{L} \) denotes the linear hull in \( \mathbb{R}^d \).

**Theorem 2.2.8** (Kusuoka and Stroock [48]). If \( (2.2.3) \) is satisfied at \( x_0 \in \mathbb{R}^d \) then for every \( t > 0 \), \( X(t, x_0, w) \in D^\infty(\mathbb{R}^d) \) satisfies the nondegeneracy condition \( (2.2.1) \). More precisely, there exists a positive integer \( d \) depending only on \( n_0 \) in \( (2.2.3) \) and, for each \( 1 < p < \infty \), a positive constant \( c = c(p, x_0) \) such that

\[
\|\| (\det \Xi_t)^{-1} \|_p \leq ct^{-d} \text{ for all } t > 0.
\]

(2.2.4)

If \( (2.2.3) \) is satisfied everywhere in a domain \( \mathcal{D} \) of \( \mathbb{R}^d \), then the estimate \( (2.2.4) \) holds uniformly in \( x_0 \in K \) for any bounded set \( K \subset \mathcal{D} \).

Therefore assuming \( (2.2.3) \) is satisfied we can consider Schwartz distributions of \( X(t, x_0, w) \), and therefore analyse the density of \( X(t, x_0, w) \). Now consider the fam-
2.2. Malliavin Calculus

\[ dX_t^\epsilon = \epsilon \sum_{j=1}^{m} V_j(X_t^\epsilon) dW_t^j + \epsilon^2 V_0(X_t^\epsilon) dt \quad \epsilon \in (0, 1] \]

then each SDE has a unique, continuous solution and \( X^\epsilon(t, x_0, w) \in D^\infty \). Furthermore, \( X^\epsilon(1, x_0, w) \in D^\infty(\mathbb{R}^d) \) has the asymptotic expansion,

\[ X^\epsilon(1, x_0, w) \sim f_0 + \epsilon f_1 + \cdots \text{ in } D^\infty(\mathbb{R}^d) \text{ as } \epsilon \downarrow 0, \]

derived by repeatedly applying Ito’s formula to the integrands of the SDE. We illustrate with the expansion up to the first order:

\[ X_t^\epsilon = x_0 + \epsilon \int_0^t V_\alpha(X_s^\epsilon) dW_s^\alpha + \epsilon^2 \int_0^t V_0(X_s^\epsilon) ds \]

Clearly \( f_0 = x_0 \in D^\infty(\mathbb{R}^d) \), and similarly \( f_1 = \int_0^t V_j(x_0) dW_s^j \in D^\infty(\mathbb{R}^d) \). By considering the quadratic variation and using the Burkholder-Davis-Gundy inequality one can see that the iterated Ito integrals are in \( L^p \). Similarly, by considering the SDE solved by the Malliavin derivative \( DX^\epsilon \), one can show that the elements are in \( D^1_p \), \( p > 1 \), and so continuing in this way, one can show that the expansion and remainder terms are in \( D^\infty \).

We cannot consider the asymptotic expansion of the density of \( X^\epsilon \) directly, because \( X^\epsilon \) is not uniformly nondegenerate, \( f_0 \) is after all deterministic. We therefore consider the change of variables \( F(\epsilon, w) = \frac{(X^\epsilon(1, x_0, w) - x_0)}{\epsilon} \). Watanabe is then able to study the small time asymptotics of the heat kernel on the diagonal, \( p(t, x_0, x_0) \), (i.e. the probability density that the SDE solution started at \( x \) returns to \( x \) at time \( t \)) and we see that this is a natural rescaling for this problem. On the other hand, Watanabe employs an alternative approach for determining the off diagonal elements of the heat kernel, \( p(t, x_0, y), y \neq x_0 \). The essential idea is the following. A small time/small noise large deviations principle holds for strong solutions of SDEs so that the probability of
reaching a given point $y$ is dominated by those Brownian paths that stay arbitrarily close to the paths in $\mathcal{H}^m$ with minimal $|| \cdot ||_H$ norm such that the corresponding ODE solution reaches $y$ (see Chapter 4). Then by applying a Girsanov Transform using such a minimal energy path we are returned to the original problem. In Chapter 6 we apply a similar approach to recover the asymptotic expansion of the marginal density of the solution of an SDE. The Yoshida-Watanabe expansion approach instead considers a different asymptotic limit, namely to consider $\lim_{\epsilon \to 0} p(\epsilon^2 t, x_0, x_0 + \epsilon z)$. As discussed in the introduction, we consider this a transformation to a regular perturbation problem, and naturally leads to an expansion around a Gaussian variable, since our rescaling causes us to approach the origin, $x_0$, so that the corresponding diffusion coefficients approach a constant. This is the transformation which we will use in Chapter 5. Our main interest in asymptotic limits is to approximate option prices under typical maturities and strike ranges. Therefore we are interested in which limits converge fastest, ie those limiting sequences that vary little from the typical market parameters to the asymptotic limit.
Chapter 3

Small Noise Implied Volatility
Asymptotics and Large Deviations

In this chapter we show how the implied volatility of a vanilla option price is determined by the large deviations rate function as the associated probability measure decays to the Dirac measure. Although there have been specific proofs for specific asymptotic regimes, we are unaware of a general proof. Secondly we introduce small noise asymptotic expansions of the implied volatility. Although the majority of research has focussed on small-time asymptotics ([14, 15]), we argue that small noise is a more natural asymptotic regime to investigate for time varying models and models with significant drift terms such as the majority of stochastic volatility models (e.g. Heston) which have mean reverting behaviour. In fact it is well known that although small time-approximations of models without mean reversion or time dependence such as the SABR model, and the Berestycki et al. local volatility approximation perform well even up to several years ([35, 14]), mean reverting models have equivalent accuracy on the order of weeks or months ([65, 26]). In particular we consider Wentzell-Freidlin sample path large deviations. We point out that Gatheral and Wang’s ([83]) variational most likely path approximation for implied volatility in a time-varying local (i.e state-dependent) volatility model is more clearly justified by Wentzell-Freidlin small noise asymptotics rather than their proposed small-time argument.
3.1 Implied Volatility asymptotics from the Large Deviations Principle

We first show how call or put option asymptotics follow from the Large deviations principle. Then we show how this then allows us to infer the asymptotics of the associated implied volatility. We consider a family of measures \( \{ \mu_\varepsilon, \varepsilon \in (0, 1] \} \) defined on \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \), and the associated real random variable \( F_T \), corresponding to the financial asset or index at observation time \( T \). We use the notation \( \mu_\varepsilon[g(F_T)] := \int g(x)\mu_\varepsilon(dx) \), and will be interested in vanilla call option \( \mu_\varepsilon[F_T - K]^+ \). We use this notation in this chapter rather than \( \mathbb{E}[(F_T^\varepsilon - K)^+] \) of the other chapters, because our proof is at the level of general large deviations, rather than assuming say a family of diffusion processes all defined on the same space. We make the following assumption.

Assumption 3.1.1.

a) The family of measures \( \{ \mu_\varepsilon \} \) converges weakly to the Dirac measure \( \delta_{F_0^T} \), for some \( F_0^T \in \mathbb{R} \).

b) The family of measures \( \{ \mu_\varepsilon \} \) satisfies a large deviations principle in \( \mathbb{R} \) with speed \( \varepsilon^2 \) and good rate function \( I : \mathbb{R} \rightarrow [0, \infty] \).

c) The rate function \( I \) decreases monotonically away from \( F_0^T \) and \( I(K) > I(F_0^T) = 0 \) for \( K \neq F_0^T \).

d) The family of measures, \( \{ \mu_\varepsilon \} \), satisfies the integrability condition

\[
\lim_{\varepsilon \to 0} \mu_\varepsilon[|F_T|^p] < \infty \text{ for some } p > 1.
\]

We define the mean or forward \( \bar{F}_T^\varepsilon := \mu_\varepsilon[F_T] \), then our assumption ensures \( \lim_{\varepsilon \to 0} \bar{F}_T^\varepsilon = F_0^T \).

Theorem 3.1.2. Let the family of measures, \( \{ \mu_\varepsilon \} \), satisfy Assumption 3.1.1. Then at a point of continuity \( K \) of the rate function, \( I \),

\[
\lim_{\varepsilon \to 0^+} -\varepsilon^2 \log \mu_\varepsilon[(F_T - K)^+] = \begin{cases} 
I(K), & \text{for } K \geq F_0^T; \\
0, & \text{otherwise.}
\end{cases} \tag{3.1.1}
\]

Remark 3.1.3. We note similar results have been demonstrated in several papers ([62, 27]) when looking at various different large deviations implied volatility asymptotics (small time, large time etc.). Note that we do not require \( K \) or \( F_T \) to be positive, which is important because one can have options on indices that can become negative such as inflation.
Proof. The in-the-money call case $F_T^0 > K$ follows from the bounds
\[
\mu_\epsilon[(F_T^0]^p]^{1/p} + |K| \geq \mu_\epsilon((F_T - K)^+) \geq (F_T^0 - K)^+.
\]

We now consider the out-of-the-money call ($K \geq F_T^0$). Since we have assumed the rate function monotonically increases away from $F_T^0$ and $K \geq F_T^0$,
\[
\inf_{x \geq K} I(x) = I(K).
\]

We rewrite the call option as
\[
\mu_\epsilon[(F_T - K)^+] = C_\epsilon + R_\epsilon
\]
where
\[
C_\epsilon := \int_{(K, \tilde{K})} \mu_\epsilon[F_T \geq K] dK_1
\]
\[
R_\epsilon := \mu_\epsilon[(F_T - \tilde{K})^+]
\]
and the cut-off strike $\tilde{K} > K$ to be fixed. We make use of the following standard estimates, (eg. [20], Lemma 1.2.15) where $A_i^s \geq 0$,
\[
\lim_{s \to 0} s \log \sum_{i=1}^n A_i^s \geq \max \lim_{s \to 0} s \log(A_i^s)
\]
\[
\lim_{s \to 0} s \log \sum_{i=1}^n A_i^s \leq \max \lim_{s \to 0} s \log(nA_i^s) = \max \lim_{s \to 0} s \log(A_i^s)
\]

In words, the 'log asymptotics' of a finite sum of nonnegative terms is determined by the log asymptotic of the maximum term. We claim $\lim_{\epsilon \to 0} \epsilon^2 \log C_\epsilon = -I(K)$. Then this implies
\[
\lim_{\epsilon \to 0} \epsilon^2 \log \mu_\epsilon[(F_T - K)^+] \geq -I(K)
\]
by the above bounds. Then the result will follow if we can show that there exists some $\tilde{K}$ such that $\lim_{s \to 0} \epsilon^2 \log R_\epsilon < -I(K)$. By assumption, there exist $\epsilon_c, B \in \mathbb{R}^+$ such that $\mu_\epsilon[(F_T^0[^p]^{1/p} < B$ for all $\epsilon \leq \epsilon_c$ for some $p > 1$.
\[
R_\epsilon = \mu_\epsilon[(F_T - \tilde{K})^+] \leq (\mu_\epsilon[(F_T - \tilde{K})^p])^{1/p} \mu_\epsilon[F_T \geq \tilde{K}]
\]
\[
\leq 4(B + |\tilde{K}|) \mu_\epsilon[F_T \geq \tilde{K}]
\]
so

\[ \limsup_{\epsilon \to 0} \epsilon^2 \log R_\epsilon \leq I(\tilde{K}). \]

We note that since we have a good rate function on \( \mathbb{R}^+ \), the level sets of \( I(x) \) are compact, so we can always find a \( \tilde{K} \) sufficiently large that \( I(\tilde{K}) > I(K) \). We now consider the upper bound for \( C_\epsilon \),

\[ \int_K^{\tilde{K}} \mu_\epsilon[F_T \geq K_1]dK_1 \leq (\tilde{K} - K)\mu_\epsilon[F_T \geq K] \]

hence

\[ \lim_{\epsilon \to 0^+} \epsilon^2 \log C_\epsilon \leq -I(K). \]

Now we consider the lower bound. By continuity and monotonicity of \( I(\cdot) \) at \( K \), for all \( \eta > 0 \) we can find \( \delta : K < K + \delta \leq \tilde{K} \) such that \( I(K + \delta) < I(K) + \eta \),

\[ \int_K^{\tilde{K}} \mu_\epsilon(F_T > K_1)dK_1 \geq \delta \mu_\epsilon(F_T > K + \delta) \]

\[ \lim_{\epsilon \to 0^+} \epsilon^2 \log C_\epsilon \geq -(I(K) + \eta). \]

Since \( \eta \) is arbitrary, we conclude that

\[ \lim_{\epsilon \to 0^+} \epsilon^2 \log C_\epsilon = -I(K). \]

**Corollary 3.1.4.** *Under the Assumption [3.1.1] above, at a point of continuity \( K \) of the rate function, \( I \),

\[ \lim_{\epsilon \to 0^+} -\epsilon^2 \log \mu_\epsilon[(K - F_T)^+] = \begin{cases} I(K), & \text{for } K \leq F_T^0; \\ 0, & \text{otherwise}. \end{cases} \]

**Proof:** Apply Theorem [3.1.2] to \( G_T = -F_T, G_T^0 = -F_T^0, L = -K, J(x) = I(-x) \).

**Remark 3.1.5.** We remark that our integrability condition in Assumption [3.1.1] (for bounded \( T \)) follows naturally from the assumption of locally Lipschitz coefficients and (sub)linear growth at infinity of coefficients sufficient to prove the existence and uniqueness of a strong solution of an SDE ([67], Chapter IX, Ex 2.10), and that the solution is square integrable. Modifying the standard proof so that the diffusion is
3.1. Implied Volatility asymptotics from the Large Deviations Principle

scaled by $\epsilon$, we see that the SDE solution $\{X_t, t \geq 0\}$ (with initial condition $\xi$) has

$$E[\|X_t\|^2] \leq C_1 (1 + \epsilon^2)(1 + E[\|\xi\|^2]) e^{C_1(1+\epsilon^2)t}; \quad 0 \leq t \leq T.$$ 

We now show how the implied volatility asymptotics are related to the corresponding large deviations rate function. define the ($\epsilon$-) implied volatility, $\sigma_{I,\epsilon}(K,T)$, for the call option price at strike $K$ and maturity $T$ for the positive asset price with $K > 0$ and fixed $\epsilon > 0$ as the unique solution of

$$\mu_{\epsilon}[(F_T - K)^+] = C[F^2_T, \epsilon^2 \sigma^2_{I,\epsilon}(K,T) T; K].$$

(3.1.2)

**Theorem 3.1.6.** Let $\{\mu_{\epsilon}\}$ satisfy Assumption 3.1.1 above and in addition let $F_T$ be $\mu_{\epsilon}$ almost surely positive with $F^0_T > 0$ then for fixed $K$

$$\lim_{\epsilon \to 0} \sigma_{I,\epsilon}(K,T) = \frac{|\log(F^0_T/K)|}{\sqrt{2TI(K)}} \text{ for } F^0_T \neq K$$

**Remark 3.1.7.** Note that we cannot get the implied volatility asymptotic for $F_T = K$, since $I_{\text{LN}}(x)$ is $\beta > 0$. One natural approach would be to show continuity (in $K$) of the smile limit at $K = F^0_T$, and twice differentiability of the energy (see e.g. Osajima’s result Theorem 1.3.1 in Chapter 1), then by l’Hôpital’s rule $\lim_{\epsilon \to 0} \sigma_{I,\epsilon}(F^0_T, T) = \frac{1}{F^0_T \sqrt{TI'(F^0_T)}}$

**Proof** We show the result for out of the money calls ($K > F^0_T$). The same approach applies for out of the money puts ($K < F^0_T$). Since by put-call parity i.e. $F_T - K = (F_T - K)^+ - (K - F_T)^+$, the implied volatilities for call and puts with same parameters are the same, the full result follows. We first note that $\epsilon \sigma_{I,\epsilon} \to 0$. Otherwise $\lim_{\epsilon \to 0} \epsilon \sigma_{I,\epsilon}(K,T) = \beta > 0$, which implies that the call price does not decrease to zero exponentially according to the rate function. Now consider the lognormal families $\text{LN}(\mu, s)$, with density

$$p(x; \mu, s, T) = \frac{1}{\sqrt{2\pi s T x}} \exp(- \frac{(\log(x/\mu) + 1/2sT)^2}{2sT}).$$

They satisfy the large deviation principle with rate function $I_{\text{LN}}(x) = \log(x/\mu)^2$ as $s \to 0^+$. This can be seen by use of the contraction principle to the normal families

$$\{\mu - 1/2sT + \sqrt{sT}x; \ x \sim N(0,1)\}.$$ 

We first consider the case $\bar{F}^0_T = F^0_T$. Applying Theorem 3.1.2 to our lognormal family

---

1 The assumption of positivity is required to work with the lognormal implied volatility, but similar results exist for the normal implied volatility limit.
we have

$$\lim_{s \to 0^+} -s \log C(F_T^0, sT; K) = \frac{(\log(x/F_T^0))^2}{2T}$$

Making the substitutions

$$s \mapsto \epsilon^2 \sigma_{I,\epsilon}^2, \ C(F_T^0, \epsilon^2 \sigma_{I,\epsilon}^2 T; K) \mapsto \mu_\epsilon[(F_T - K)^+]$$

we see

$$\lim_{\epsilon \to 0^+} - (\epsilon \sigma_{I,\epsilon})^2 \log \mu_\epsilon[(F_T - K)^+] = \frac{\log(F_T^0/K)^2}{2T}.$$ 

Since

$$\lim_{\epsilon \to 0^+} -\epsilon^2 \log \mu_\epsilon[(F_T - K)^+] = I(K),$$

we have

$$\sigma_{I,0}^2(K, T) := \lim_{\epsilon \to 0^+} \sigma_{I,\epsilon}^2(K, T) = \frac{\log(F_T^0/K)^2}{2TI(K)}.$$ 

Although for standard stochastic volatility models the underlying asset is a martingale in the appropriate measure (so $F_T \equiv F_T^0$), we would like to use this method also for nonlinear functions of the asset, so we need the result when the mean $F_T$ depends on $\epsilon$. Under a lognormal model, the call option price is a monotonically increasing function of the forward or mean $F_T$, so by choosing $\delta, \epsilon_c$ such that $0 < \delta < K - F_T^0$ and $|F_T - F_T^0| < \delta$ for all $\epsilon \leq \epsilon_c$ we have

$$C(F_T^0 - \delta, \epsilon^2 \sigma_{I,\epsilon}^2 T; K) \leq \mu_\epsilon[(F_T - K)^+] \leq C(F_T^0 + \delta, \epsilon^2 \sigma_{I,\epsilon}^2 T; K)$$

This implies that $C(F_T^0 - \delta, \epsilon^2 \sigma_{I,\epsilon}^2 T; K)$ decays exponentially to zero as $\epsilon \to 0$, which implies that $\epsilon \sigma_{I,\epsilon}$ converges to zero. Therefore

$$\limsup_{\epsilon \to 0} (\epsilon \sigma_{I,\epsilon})^2 \log \mu_\epsilon[(F_T - K)^+] \leq \lim_{\epsilon \to 0} (\epsilon \sigma_{I,\epsilon})^2 \log C(F_T^0 + \delta, \epsilon^2 \sigma_{I,\epsilon}^2 T; K)$$

$$= \frac{\log((F_T^0 + \delta)/K)^2}{2T}$$

and similarly for the lower bound. But since $\delta$ can be made arbitrarily small we have

$$\lim_{\epsilon \to 0} (\epsilon \sigma_{I,\epsilon})^2 \log \mu_\epsilon[(F_T - K)^+] = \frac{\log((F_T^0/K)^2)}{2T}$$

3.2 Small Noise Asymptotics of Implied Volatility and Gatheral and Wang’s Variational most likely path

which together with the original result

$$\lim_{\epsilon \to 0^+} -\epsilon^2 \log \mu_{\epsilon}[(F_T - K)^+] = I(K)$$

yields the result.

Note that we have proved this result for a relatively general positive random variable $F_T$, $F_T$ could be for example the arithmetic average or maximum of a stock price. In fact in the next chapter we will consider the option on the VIX asset which is defined (within the model we study there) as the square root of the weighted sum of the state variables.

### 3.2 Small Noise Asymptotics of Implied Volatility and Gatheral and Wang’s Variational most likely path

Consider a local volatility model (so $S$ and $W$ are 1-dimensional),

$$dS_t = a(S_t, t)dW_t, S_0 = s_0, \quad (3.2.1)$$

Berestycki, Busca, and Florent ([14]) showed that under the assumption that $a$ is globally bounded, uniformly continuous and uniformly elliptic, the short time asymptotic of implied volatility $\sigma_I(K, T)$, is the harmonic mean of the time zero local volatility $a(s_0, 0)$.

$$\lim_{T \to 0} \frac{|\log(K/s_0)|}{\sigma_I(K, T)} = \int_{s_0}^{K} \frac{1}{a(x, 0)} dx = \min_{S: S_0 = s_0, S_1 = K} \int_0^1 \left[ \frac{\dot{S}_t}{a(S_t, 0)} \right]^2 dt =: d(s_0, K)$$

This corresponds to the distance in the Riemannian metric associated with the inverse of the diffusion coefficient (at time zero). To relate this to general Wentzell-Freidlin large deviations, it is useful to recall the length $l(\phi)$ and energy $e(\phi)$ of a

---

2The minimum in this one dimensional case is easily found by minimising the energy functional, (see below), $\frac{1}{2} \int_0^1 \left[ \frac{\dot{S}_t}{a(S_t, 0)} \right]^2 dt$ and using the change of variables $x(S) = \int_{S_0}^S \frac{1}{a(S, 0)} ds$. So we have $\min(\frac{1}{2} \int_0^1 \dot{x}^2 dt; \int_0^1 \dot{x} dt = x(K) - x(S_0))$, then the Cauchy-Schwarz inequality (with $\dot{f} = \dot{x}, \dot{g} = 1$) tells us that the energy is bounded below by $\frac{1}{2}(x(K) - x(S_0))^2$ with the minimum energy achieved when $\dot{x}$ is constant, namely $\dot{x}_t \equiv x(K) - x(s_0)$. In general the existence of the minimum of the rate function can be shown by continuity and compactness arguments, see Assumption 4.2.1b) and Theorem 6.2.4.
curve $\phi$. Typically one works with the energy functional, rather than the length,

$$
e(\phi) := \frac{1}{2} \int_0^1 \left( \frac{\dot{\phi}(s)}{a(\phi(s), 0)} \right)^2 ds
$$

and $2e(\phi) \geq l(\phi)^2$ by the Cauchy-Schwarz inequality applied to $\| \frac{\dot{\phi}(s)}{a(\phi(s), 0)} \|$ and the unit integrand. Whilst the length functional is independent of the curve parametrisation, the energy is not. In fact we have strict equality in the above when $\phi$ is parametrised by a multiple of its own length (so $\| \frac{\dot{\phi}(s)}{a(\phi(s), 0)} \| = k$ for some $k > 0$). This implies that the minimal energy path is the minimum distance path (up to reparameterisation).

Although the Berestycki et al. result that a time varying model has a limiting small time implied volatility is of mathematical interest, as an approximation the BBF limit is typically used only for time homogeneous models, where it provides sufficient accuracy on the order of years. Gatheral and Wang ([33]) proposed the following theorem to develop an approximation for implied volatility in a time-dependent local volatility model.

Define the variational most likely path as the unique solution of the Euler-Lagrange equation corresponding to

$$
\min_{S: S_0 = S_0, S_T = K} \frac{1}{2} \int_0^T \left[ \frac{\dot{S}_t}{a(S_t, t)} \right]^2 dt
$$

Then we have the following

**Theorem 3.2.1** ([33]). Assume the dynamic of the underlying asset $S$ under risk neutral probability measure is governed by the local volatility model (3.2.1) and the local volatility function $a(S_t, t)$ is smooth and bounded with bounded derivatives in $S, t$. Then the implied volatility $\sigma_I$ has the following expansion as $T \to 0^+$,

$$
\sigma_I(K, T) = \sigma_{I, 0}(1 + O(T)),
$$

with

$$
\sigma_{I, 0} = \left( \frac{\sqrt{T}}{\log(s_0/K)} \sqrt{\int_0^T \left[ \frac{\dot{s}(t)}{a(s(t), t)} \right]^2 dt} \right)^{-1},
$$

Gatheral and Wang do not provide justification for this statement. A minimum exists under the assumptions of Theorem 3.2.1 assuming a path to $S_T = K$ exists see Theorem 6.2.4 but it need not be unique. Our Large deviations justification of their theorem does not require uniqueness.
where $s(t)$ is the variational most-likely-path (MLP).

Given that $\sigma_{I,0}$ depends on $T$ in this definition, precisely because of the time dependence of $a(\cdot, \cdot)$, it is unclear why the zero order (in T) implied volatility should not simply converge to the original BBF result (since it is an integral over $T$). Nevertheless, Gatheral and Wang show with numerical experiments that using their implied volatility approximate formula (for fixed $T$) gives good results. We would suggest an alternative explanation and proof. Consider the family of processes solving

$$dS^\epsilon_t = \epsilon a(S^\epsilon_t, t) dW_t, \epsilon \to 0.$$ 

Then assuming some conditions on the coefficients (see the next Chapter) we can show that the sample paths of $S^\epsilon$ satisfy a Wentzell-Freidlin large deviations principle on $C^1$ with good rate function

$$J(\phi) = \begin{cases} \frac{1}{2} \int_0^T \left[ \frac{\dot{\phi}(t)}{a(\phi(t), t)} \right]^2 dt & \text{if } \phi \text{ is absolutely continuous and } \dot{\phi} \in L^2([0, T]) \\ \infty & \text{otherwise} \end{cases}$$

Then by the contraction principle, if $s(\cdot)$ is the variational most likely path [to $S_T = K$], then $\frac{1}{2} \int_0^T \left[ \frac{s(t)}{a(s(t), t)} \right]^2$ is the rate function for $S^\epsilon_T$ evaluated at K. Then providing we can show $S^\epsilon_T$ satisfies the assumptions of 3.1.6, we will have shown that $\sigma_{I,0}$ (with T now fixed) is precisely the small noise asymptotic for the implied volatility $\sigma_{I,\epsilon}(K, T)$.

We are left with showing that $S^\epsilon_T$ has a bounded moment of order $p > 1$, which we see can be proved using Remark 3.1.5. In the next chapter we show that the rate function of a smooth 1 dimensional functional of the path is monotonic so that all the assumptions required for Theorem 3.1.6 are satisfied. There we carry out the full proof for a 2 dimensional positive diffusion. We first describe the Wentzell-Freidlin result as generalised by Azencott and others ([8, 66, 10, 9]) to locally Lipschitz coefficients and (sub)linear growth with no ellipticity requirements. We then show that a particular 2 dimensional positive diffusion, the DCEV model, satisfies a Wentzell-Freidlin LDP, and determine the small noise implied volatility asymptotic of a function of the endpoints.
3.2. Small Noise Asymptotics of Implied Volatility and Gatheral and Wang’s Variational most likely path
Wentzell-Freidlin Large Deviations for a class of two-dimensional positive diffusions

In this chapter we show a Wentzell-Freidlin large deviation principle for a class of 2-dimensional positive diffusions solving an SDE with coefficients which are only Hölder (and not Lipschitz) continuous around their axes. In particular our results apply to Gatheral’s Double CEV model, (31). We base our method on a recent paper by Baldi and Caramellino (9) proving Wentzell-Freidlin sample path large deviations for one dimensional positive diffusions whose coefficients are only Hölder continuous around zero. We then show that sample path large deviations can be used to determine asymptotics for implied volatility of (one dimensional) continuous functionals of the SDE solution path. This applies then for a terminal value, continuous functions of the terminal value(s), and even the maximum of the path in a time interval. In particular we show that we have a rate function for the functional and it is monotonically decreasing away from the mean level. This then allows us to demonstrate that the VIX index, modelled as a continuous function of the state variables at some observation time, satisfies an LDP and to determine the implied volatility asymptotics for the VIX call option. Finally by discretising the optimal path rate function we compute the VIX rate function numerically and compare to Monte Carlo simulations. In the next Chapter we also compare to small time and small noise Yoshida-Watanabe expansions.
4.1 Gatheral’s Double CEV model

A small noise expansion is of most interest for models with drift and models with time-varying coefficients. Here we will consider a two factor mean reverting stochastic volatility model, Gatheral’s double CEV model (DCEV):

\[\begin{align*}
    dS_t &= \sqrt{v_t} S_t dZ_t \\
    dv_t &= \kappa(v'_t - v_t) dt + \xi_1 v_t^{\alpha} (\sqrt{1 - \rho^2} dW^1_t + \rho dW^2_t) \\
    dv'_t &= c(z_3 - v'_t) dt + \xi_2 v_t^{\beta} dW^2_t \\
    S_0, v_0, v'_0 > 0
\end{align*}\]  

(4.1.1)

where \(\kappa > c > 0\), \(z_3 > 0\), \(|\rho| \leq 1\), \(\alpha, \beta \in [1/2, 1]\), where \(W = (W^1, W^2)\) is standard 2-dimensional Brownian motion and \(Z\) is Brownian motion correlated with \((W^1, W^2)\). In our work we will not consider the stock process, \(S\), so to minimise notation we have also not specified the deterministic correlation structure assumed between \(Z\), and \(W^1, W^2\). Nevertheless our results can be straightforwardly extended to the full 3 dimensional process. We also restrict the parameters slightly, \(|\rho| < 1\), and \(\alpha, \beta \in (1/2, 1]\). The restriction on \(\alpha, \beta\) is to ensure a.s positivity of the \(v', v\) state variable, see section 4.3. One could also use \(\beta = \frac{1}{2}\), providing \(cz_3 \geq \frac{1}{2} \xi_2^2\) by the Feller condition we use there. This model has been proposed to capture simultaneously options on the S&P500 index, SPX, and the VIX ([2]) index, a measure of the SPX’s 30 day implied volatility. In SDE (4.1.1) \(S\) corresponds to the SPX index, \(v\) corresponds to its instantaneous variance, and \(v'\) is a (stochastic) mean reversion level for the instantaneous variance of the SPX. The VIX index is defined as the square root of the par variance swap rate over the next 30 days, derived from a prescribed set of SPX
option prices. Therefore in the DCEV model, the VIX has the following formula:

\[
VIX^2(t; t + t_{30}) := \frac{1}{t_{30}} \mathbb{E}[\int_0^{t_{30}} v_{t+s} ds | v_t, v'_t]
\]

\[
= z_3 + (v_t - z_3) \frac{1 - e^{-\kappa t_{30}}}{\kappa t_{30}} + (v'_t - z_3) \frac{\kappa}{\kappa - c} \left\{ \frac{1 - e^{-ct_{30}}}{ct_{30}} - 1 - e^{-\kappa t_{30}} \right\}
\]

\[
= (a_1v_t + a_2v'_t + a_3z_3)
\]

where

\[
ap_1 = \frac{1 - e^{-\kappa t_{30}}}{\kappa t_{30}}
\]

\[
ap_2 = \frac{\kappa}{\kappa - c} \left\{ \frac{1 - e^{-ct_{30}}}{ct_{30}} - 1 - e^{-\kappa t_{30}} \right\}
\]

\[
ap_3 = 1 - \frac{1 - e^{-\kappa t_{30}}}{\kappa t_{30}} - \frac{\kappa}{\kappa - c} \left\{ \frac{1 - e^{-ct_{30}}}{ct_{30}} - 1 - e^{-\kappa t_{30}} \right\}
\]

and \(t_{30} = 30/365\). The value of a European Call option on VIX with expiry \(T\) and strike \(K\) is given by

\[
\mathbb{E}\left[\sqrt{a_1v_T + a_2v'_T + a_3z_3 - K}\right].
\]

As one would expect given \(VIX^2 \geq 0\), the coefficients \(a_1, a_2, a_3\) are all non-negative. We can check by noting that \(a_1\) is the integral of \(e^{-\kappa t}\) wrt \(t\). Similarly, using \(e^{-ct} > e^{-\kappa t}\), and integrating with respect to \(t\) we see that \(a_2\) is positive. Lastly, by using the convexity relation \(e^{-ct} < \frac{c-\kappa}{\kappa} + \frac{\kappa}{c} e^{-\kappa t}\) and again integrating with respect to \(t\) we find that \(a_3\) is positive. We note that the DCEV process may or may not have a stationary distribution. For instance if \(\beta = 1\) it is easy to calculate that the variance of \(v'_t\) is bounded for \(cz_3 > \frac{\xi^2}{2}\).

Since we wish to study the implied volatility for VIX options, we will first consider the LDP for the sample paths of the DCEV model, and then show that the VIX asset, as a continuous functional of the sample paths, also satisfies an LDP.

### 4.2 Background Results

In this section we review the basic framework of sample path large deviations as developed by Azencott and others ([8, 66, 10]) following Baldi and Caramellino’s paper ([9]). For \(\epsilon > 0\), let \(b_\epsilon : \mathbb{R}^d \to \mathbb{R}^d\) and \(\sigma_\epsilon : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^m\) families of vector and matrix fields respectively. Let \(W\) be a \(m\)-dimensional Brownian motion on the canonical Wiener space and \(Y^\epsilon\) the solution of the SDE (4.2.1).
\[ dY^\varepsilon_t = b_t(Y^\varepsilon_t)dt + \varepsilon\sigma_t(Y^\varepsilon_t)dW_t \quad (4.2.1) \]

\[ Y^\varepsilon_0 = x_0 \]

Now fix \( x_0 \in \mathbb{R}^d \) and consider the following assumption.

**Assumption 4.2.1.** The SDE \[ \text{(4.2.1)} \] has a unique strong solution \( \{Y^\varepsilon_t, t \geq 0\} \) for every \( \varepsilon > 0 \) and there exists a vector field \( b : \mathbb{R}^d \to \mathbb{R}^d \) and a matrix field \( \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m) : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^m \) such that

\begin{enumerate}
  \item for every \( h \in \mathcal{H}^m \) the ordinary differential equation
  \[ \dot{g}_t = b(g_t) + \sigma(g_t)h_t, \quad g_0 = x_0 \quad (4.2.2) \]
  \text{has a unique solution on } [0, T].
  \item Let \( \phi^h(x_0) \) denote the solution of \[ \text{(4.2.2)} \]. Therefore \( \phi^h(x_0) : \mathcal{H}^m \to \mathcal{C}^d_{x_0} \). For any \( a > 0 \), the restriction of \( \phi^h(x_0) \) to the compact set \( K_a = \{\|h\|_H \leq a\} \) is continuous with respect to the uniform norm, for any \( \{h_n\}_n \subset K_a \) such that \[ \lim_{n \to \infty} \|h_n - h\|_T = 0 \] then
  \[ \lim_{n \to \infty} \|\phi^{h_n}(x_0) - \phi^h(x_0)\|_T = 0 \]
  \item (The quasi-continuity property) For every \( h \in \mathcal{H}^m \), \( R > 0 \), \( r > 0 \) there exist \( \epsilon_0 > 0 \), \( \alpha > 0 \) such that, if \( \varepsilon < \epsilon_0 \),
  \[ P(\|Y^\varepsilon - \phi^h(x_0)\|_T > r, \|\epsilon W - h\|_T \leq \alpha) \leq e^{-R/\varepsilon^2}. \quad (4.2.3) \]
\end{enumerate}

**Theorem 4.2.2.** (c.f. [9], Theorem 2.4) If Assumption \[ \text{4.2.1} \] holds, the family \( \{Y^\varepsilon\}_\varepsilon \) satisfies a Large Deviations Principle on \( \mathcal{C}^d_{x_0} \) with inverse speed \( \varepsilon^2 \) and (good) rate function

\[ J(g) = \inf \{I(h); \phi^h(x_0) = g\}, \quad (4.2.4) \]

with the understanding \( J(g) = +\infty \) if \( \{h : \phi^h(x_0) = g\} \) is empty. In other words

\[ \lim_{\varepsilon \to 0} \varepsilon^2 \log P(Y^\varepsilon \in F) \leq -\inf_{\psi \in F} J(\psi) \]

\[ \lim_{\varepsilon \to 0} \inf \varepsilon^2 \log P(Y^\varepsilon \in G) \geq -\inf_{\psi \in G} J(\psi) \]

\(^1\) i.e \( K_a \) is compact in the uniform norm topology

\(^2\) Baldi and Caramellino’s paper - specifies quasi continuity uniformly in \( \|h\|_H < a \) and \( |x| < c \), whereas the theorem stated is only for fixed \( x \), ie \( \mathcal{C}^d_{x_0} \) and the proof explicitly uses only a finite number of \( h \), indexing \( \alpha \) by the corresponding \( h \), ie \( \alpha_{h_i} \).
for every closed set $F \subset \mathcal{C}_{x_0}$ and open set $G \subset \mathcal{C}_{x_0}$ and that the level sets of $J$ are compact.

**Corollary 4.2.3.** If in addition $d = m$ and the matrices $\{\sigma^j(\cdot)\}_{1 \leq i, j \leq d}$ are invertible on $\mathbb{R}^d$ then $J(g)$ simplifies to

$$J(g) := \frac{1}{2} \int_0^T (\dot{g}_t - b(g_t))^T \Sigma^{-1}(g_t)(\dot{g}_t - b(g_t)) dt = J(g),$$

(4.2.5)

if $g$ is absolutely continuous and $+\infty$ otherwise, where $\Sigma = \sigma \sigma^T$.

**Remark 4.2.4.** In this chapter we are primarily concerned with the non singular case covered by corollary 4.2.3 (restricting the domain to the positive half line or quadrant), so our rate function is of the form $J(g)$. In Chapter 6 we work with the more general case as we deal with hypo-elliptic diffusions.

**Remark 4.2.5.** We note that the theorem above allows for time dependent coefficients by introducing an extra deterministic state variable representing time. This then allows the time varying BBF result and Gatheral and Wang’s variational most likely path to be derived (as discussed in Section 3.2).

We report the proof of the theorem in the appendix, section A.2. The following conditions on the coefficients enable Assumption 4.2.1 to be satisfied.

**Lemma 4.2.6.** If $b$ and $\sigma$ are locally Lipschitz continuous and have sub-linear growth at infinity,

$$\limsup_{\|x\| \to \infty} \frac{\|b\|^2 + \|\sigma \sigma^T\|}{1 + \|x\|^2} < \infty$$

then 4.2.1 a) and 4.2.1 b) hold. Moreover for every compact set $K \subset \mathbb{R}^d$ and $a > 0$

$$\sup_{x \in K} \sup_{\|h\|_{H} \leq a} \|\phi^b(x)\|_T < \infty.$$ (4.2.6)

**Assumption 4.2.7.**

**a)** The coefficients $b$ and $\sigma$ are locally Lipschitz continuous, have a sub-linear growth at infinity and satisfy

$$\lim_{\epsilon \to 0^+} |b_\epsilon(y) - b(y)| = 0 \quad \lim_{\epsilon \to 0^+} |\sigma_\epsilon(y) - \sigma(y)| = 0$$ (4.2.7)

uniformly on compact sets.

**b)** The coefficients $b_\epsilon$ and $\sigma_\epsilon$ are locally Lipschitz continuous.

---

3This condition is not explicitly stated (compare the assumptions of the more general Theorem 1.1 in [10] where they remark it is (only) required to show existence of a unique strong solution for every $\epsilon > 0$).
Theorem 4.2.8. Under Assumption 4.2.7 for every $R > 0$, $r > 0$, $a > 0$, and compact set $K \subset \mathbb{R}^d$ there exist $\epsilon_0 > 0$, $\alpha > 0$ such that,

\[ P(\| Y^\epsilon - \phi^h(x) \|_T > r, \| \epsilon W - h \|_T \leq \alpha ) \leq e^{-R/\epsilon^2} \text{ for } \epsilon < \epsilon_0, \tag{4.2.8} \]

uniformly for \{ $h \in \mathcal{H}_m : \| h \|_H \leq a$ \} and $x \in K$. Moreover if $b$ and $\sigma$ are bounded and the convergence in (4.2.7) is uniform in $y$, then (4.2.3) is uniform in (the starting point) $x$.

These results are specified for coefficients $b, \sigma, b, \sigma$ that are locally Lipschitz on the whole real space. Baldi and Caramellino consider a positive diffusion, $X^\epsilon$, given as the solution of the SDE with values in $\mathbb{R}^+$

\[ dX^\epsilon_t = b(X^\epsilon_t)dt + \epsilon \sigma(X^\epsilon_t)dW_t, \quad X^\epsilon_0 = x_0 > 0. \tag{4.2.9} \]

Assumption 4.2.9.

a) $\sigma : [0, +\infty) \to \mathbb{R}^+$ vanishes\textsuperscript{4} at 0, is Hölder continuous with exponent $\gamma \geq \frac{1}{2}$, locally Lipschitz continuous on $(0, +\infty)$, and has a sub-linear growth at $\infty$.

b) $b : [0, +\infty) \to \mathbb{R}$ is locally Lipschitz continuous, has a sub-linear growth at $\infty$, and $b(0) > 0$.

Using Feller’s Test for explosions (see Appendix, section A.1), we develop a simple test to check that the origin is unattainable given our standing assumptions of local Lipschitz continuity and sublinear growth.

Lemma 4.2.10. Consider the one-dimensional SDE

\[ dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0 > 0. \]

Let $b$ and $\sigma$ be locally Lipschitz continuous on $(0, \infty)$ and satisfy the sublinear growth condition [A.3.2] and let $\sigma$ be positive on $(0, \infty)$. If there exists $\delta > 0$ such that $b(y) \geq \frac{\sigma^2(y)}{2y}$ for all $y \in (0, \delta)$ then the origin is unattainable for the process $X$.

Proof: Local Lipschitz continuity and the sublinear growth condition ensures that we have a strong solution on every interval $(\frac{1}{n}, n)$ for $n \geq 1$, and no explosions to infinity. Continuity of the coefficients $b, \sigma$ and positivity of $\sigma$ on $\mathbb{R}^+$ ensure that conditions (A.1.2) and (A.1.3) are satisfied. The scale function $p(x)$ (see section A.1) has value $-\infty$ at $0^+$ for the CIR process given by $b(x) \equiv \frac{1}{2}, \sigma(x) = \sqrt{x}$. Since the scale function $p(x)$ is monotonic in the integrand $\frac{b(x)}{\sigma^2(x)}$, $p(0^+) = -\infty$ for any process satisfying the

\textsuperscript{4}ie $\sigma$ is strictly positive except at the origin
above condition. Then we are in either of conditions a) or c) together with \( \mathbb{P}(S = \infty) = 1 \), so the Feller test \( A.1 \) implies that the origin is unattainable.

**Theorem 4.2.11** (\cite{[9]}, Theorem 1.2). Let \( X^\epsilon \) be the solution of (4.2.9) in the time interval \([0, T]\) with \( x_0 > 0 \). Then under Assumption 4.2.9

\[
\limsup_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(X^\epsilon \in F) \leq - \inf_{\psi \in F} J(\psi)
\]

\[
\liminf_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(X^\epsilon \in G) \geq - \inf_{\psi \in G} J(\psi)
\]

for every closed set \( F \subset C_{x_0}([0, T], \mathbb{R}^+) \) and open set \( G \subset C_{x_0}([0, T], \mathbb{R}^+) \), with \( J(\psi) \) as defined in Corollary 4.2.3.

**Proof.** The assumptions ensure the existence and uniqueness of a strong solution of the SDE (\cite{[67]}, Chapter IX, Theorem 3.5). In fact, by Lemma 4.2.10 we see that the diffusion stays strictly positive for sufficiently small \( \epsilon > 0 \): denote \( L \) the Lipschitz constant for \( b \) in a neighbourhood of 0 and \( H \) the corresponding Hölder constant for \( \sigma \), then \( b(x) \geq b(0) - L x \) and \( 0 < \epsilon^2 \sigma(x)^2 \leq \epsilon^2 H^2 x \) for all sufficiently small \( x \in \mathbb{R}^+ \). By applying a localisation argument Baldi and Caramellino show that assumption 4.2.1 is satisfied for all \( x \in \mathbb{R}^+ \). Assume that any solution of the ODE associated with the coefficients of \( X^\epsilon \) stays bounded and bounded away from zero in \( t \in [0, T] \) for all \( \|h\|_H \leq a \) for every \( a < \infty \) and every compact set \( K \in \mathbb{R}^+ \), and denote the region by \( \tilde{K}_{K,a} \). We can find a set of coefficients, \( \tilde{b}, \tilde{\sigma} \) that are locally Lipschitz continuous on the whole of \( \mathbb{R}^+ \), have sub-linear growth at infinity and match the coefficients \( b, \sigma \) within the \( \tilde{K}_{K,a} \) on which the solution exists. So 4.2.1a) and b) hold. Similarly, by matching coefficients also on a tube \( \mathbb{B}(\phi^h(x_0), 2r) \subset [0, T] \times \mathbb{R}^+ \), then since the corresponding SDE solution \( \tilde{X}^\epsilon \) satisfies the quasi-continuity property (4.2.1c), so does \( X^\epsilon \). Continuity on \([0, \infty)\) and sub-linear growth at infinity ensure that we have \( |\sigma(x)| \leq C(1 + |x|), \ x \geq 0 \). Similarly, we have \( |b(x)| \leq C(1 + |x|), \ x \geq 0 \). So it is clear that any solution stays bounded by Gronwall’s lemma. Baldi and Caramellino prove that solutions of the ODE stay bounded away from zero using the following lemma.

**Lemma 4.2.12.** Define the free time cost function,

\[
J_\infty(x) = \inf_{T>0} \inf_{\psi \in C_{x_0}^{1}} J(\psi)
\]

\[
\psi_T = x
\]

\( \text{The quasi-continuity property only has to be shown for arbitrarily small } r \)
where \( x_0 > 0 \) is a fixed starting point, with \( J(\psi) \) as defined in 4.2.5 and \( b \) and \( \sigma \) strictly positive continuous scalar functions on \( \mathbb{R}^+ \). Then for \( 0 < x < x_0 \)

\[
J_\infty(x) = -2 \int_{x_0}^x \frac{b(z)}{\sigma(z)^2} dz.
\]

Using Lemma 4.2.12 with \( b(z) = 1/2, \sigma(z) = \sqrt{z} \), we see that \( J_\infty(x) = \log \frac{x}{x_0} \), which can be made arbitrarily large by taking \( x \) small enough. By the Hölder continuity assumption on \( \sigma \) (and \( \sigma(0) = 0 \)), we see that there exists a \( C < \infty \): \( \sigma(z) \leq C \sqrt{z} \) for \( z \) small enough. Similarly there exists \( \gamma : 0 < \gamma < b(z) \) for \( z \) small enough. So by comparing our integrand to the above case and considering only sections of paths sufficiently close to zero, we can show that ODEs satisfying Assumption 4.2.9 stay bounded away from zero.

### 4.3 Application to DCEV process

We plan to show that the DCEV process satisfies a large deviations principle by extending Baldi and Caramellino’s result to a class of two-dimensional positive diffusions. We first show that the DCEV SDE has a unique strong solution for all \( \epsilon > 0 \).

Let \( X_\epsilon = (X_\epsilon^1, X_\epsilon^2) \) be the solution of the SDE (where \( |\rho| < 1 \))

\[
\begin{align*}
    dX_\epsilon^1 &= b^{(1)}(X_\epsilon^1, X_\epsilon^2) dt + \epsilon \sigma^{(1)}(X_\epsilon^1)(\sqrt{1 - \rho^2} dW_1^1 + \rho dW_2^1), \\
    dX_\epsilon^2 &= b^{(2)}(X_\epsilon^2) dt + \epsilon \sigma^{(2)}(X_\epsilon^2) dW_2^2, \\
    X_0^1 &= x_0^1 > 0, \quad X_0^2 = x_0^2 > 0.
\end{align*}
\]

In our standard form,

\[
\begin{align*}
    b(x) &= (b^1(x), b^2(x)) = (b^{(1)}(x^1, x^2), b^{(2)}(x^2)) \\
    \sigma^1(x) &= (\sigma^{(1)}(x^1) \sqrt{1 - \rho^2}, \sigma^{(1)}(x^1) \rho) \\
    \sigma^2(x) &= (0, \sigma^{(2)}(x^2)).
\end{align*}
\]

**Assumption 4.3.1.**

a) \( \sigma^{(1)} : [0, +\infty) \to \mathbb{R}^+ \) vanishes at 0, is Hölder continuous with exponent \( \gamma > 1/2 \), is locally Lipschitz continuous on \( (0, +\infty) \), and has sub-linear growth at \( \infty \).

b) \( \sigma^{(2)} : [0, +\infty) \to \mathbb{R}^+ \) vanishes at 0, is Hölder continuous with exponent \( \gamma > 1/2 \), is locally Lipschitz continuous on \( (0, +\infty) \), and has sub-linear growth at \( \infty \).

\( ^6 \text{i.e. } \sigma^{(1)}, \sigma^{(2)} > 0 \text{ except at zero} \)
4.3. Application to DCEV process

4.3.1. Assumption

b) \( b^{(1)} : [0, +\infty) \times [0, +\infty) \rightarrow \mathbb{R} \) is locally Lipschitz, has sub-linear growth at \( \infty \), \( b^{(1)}(x, y) \) is monotonically increasing in \( y \), and \( b^{(1)}(0, y) > 0 \) for \( y > 0 \).

c) \( b^{(2)} : [0, +\infty) \rightarrow \mathbb{R} \) is locally Lipschitz continuous, has a sub-linear growth at \( \infty \), and \( b^{(2)}(0) > 0 \).

d) \( b^{(2)} : [0, +\infty) \rightarrow \mathbb{R} \) is locally Lipschitz, has sub-linear growth at \( \infty \), and \( b^{(2)}(0) > 0 \).

Lemma 4.3.2. Under Assumption 4.3.1, the equation (4.3.1) has a unique strong solution for all \( \epsilon > 0 \). Moreover, the solution stays in the positive quadrant for all \( t > 0 \).

Proof. Our assumptions on \( b, \sigma \) ensure local Lipschitz continuity on \( \mathbb{R}^+ \times \mathbb{R}^+ \). Continuity on \( [0, \infty) \times [0, \infty) \) and sub-linear growth at infinity ensure we have sub-linear growth

\[
\|b^{(1)}(x, y)\|^2 + \|b^{(2)}(y)\|^2 + \|\sigma^{(1)}(x)\|^2 + \|\sigma^{(2)}(y)\|^2 \leq K^2(1 + \|x\|^2 + \|y\|^2),
\]

for some \( K \), for the whole quadrant, \( \{x \geq 0, y \geq 0\} \). Therefore by standard existence and uniqueness results \([67], \text{Chap IX, ex. 2.10}\) we can show that a unique solution exists up to the first time \( X^{\epsilon,2} \) or \( X^{\epsilon,1} \) touches \( 1/n \). In fact by results on Hölder continuous coefficients in the one dimensional case \([67], \text{Chap IX, Theorem 3.5}\), we know we have a unique strong solution for the second component, \( X^{\epsilon,2} \). Furthermore, by Feller’s test for explosions \( (4.2.10) \), \( X^{\epsilon,2} \) stays positive for all \( \epsilon > 0 \) by using the same calculations as \( 4.2.11 \) but noting that we have assumed that Hölder continuity with exponent strictly greater than \( 1/2 \). We now show that under the above assumptions \( X^{\epsilon,1} \) also stays positive. Fix the stopping time \( \tau_n := \inf\{t : X^{\epsilon,2}_t \leq 1/n\} \). Define \( Y^{\epsilon,(n)} \) as the solution of

\[
dY^{\epsilon,(n)}_t = b^{(1)}(Y^{\epsilon,(n)}_t, 1/n)dt + \epsilon\sigma^{(1)}(Y^{\epsilon,(n)}_t)(\sqrt{1 - \rho^2}dW^1_t + \rho dW^2_t),
\]

\[
Y^{\epsilon,(n)}_t = x_0 > 0
\]

for \( |\rho| < 1 \). By our assumptions, \( b^{(1)}(x, 1/n) \) is bounded above zero in a sufficiently small neighbourhood of \( x = 0 \). Therefore by Feller’s test for explosions \( 4.2.10 \), \( Y^{\epsilon,(n)} \) stays positive. Clearly,

\[
\min_{0 \leq \ell \leq T} Y^{\epsilon,(n)}_{t \wedge \tau_n} \geq \min_{0 \leq \ell \leq T} Y^{\epsilon,(n)}_t,
\]

and \( b^{(1)}(x, y) \geq b^{(1)}(x, 1/n) \) for \( y \geq 1/n \), so we can use a one dimensional comparison principle extended to include a random stopping time \( (\text{Appendix 4.A}) \) to conclude that

\[
\mathbb{P}[X^{\epsilon,1}_{t \wedge \tau_n} \geq Y^{\epsilon,(n)}_{t \wedge \tau_n} \ 0 \leq t < \infty] = 1
\]
Therefore up to every stopping time \( \tau_n \), the probability of \( X_t^{\epsilon,1} \) hitting 0 is zero. But since \( \lim_{n \to \infty} \tau_n = \infty \) a.s., we have the result that neither \( X_t^{\epsilon,1} \) nor \( X_t^{\epsilon,2} \) reach the origin within any finite time interval so we have a unique strong solution for SDE (4.3.1) such that \( X_t^{\epsilon,1} > 0 \).

We now show that the corresponding deterministic ODE stays in a compact set bounded away from zero, for any bounded subset of \( \mathcal{X}^m \), and all \( x \) in any compact set \( K \) strictly inside \( \mathbb{R}^+ \times \mathbb{R}^+ \). We can then apply the same localisation argument to show that our SDE satisfies Assumptions 4.2.1. As previously noted, local uniformity in \( x \) is only required to show a locally uniform in \( x \) LDP result.

**Proposition 4.3.3.** Under Assumption 4.3.1 the equation

\[
\begin{align*}
\dot{w}_t^1 &= b^{(1)}(w_t^1, w_t^2) + \sigma^{(1)}(w_t^1)(\sqrt{1 - \rho^2}h_t^1 + \rho h_t^2) \\
\dot{w}_t^2 &= b^{(2)}(w_t^2) + \sigma^{(2)}(w_t^2)h_t^2 \\
w_0^1 &= x_0^1, \quad w_0^2 = x_0^2
\end{align*}
\]  

(4.3.3)

for \( h = (h^1, h^2) \in \mathcal{K}^2 \), \( x_0 \in \mathbb{R}^+ \times \mathbb{R}^+ \), and \( |\rho| < 1 \) has a unique solution for \( t \in [0, T] \) for every \( T > 0 \). Moreover for every compact set \( K \subset \mathbb{R}^+ \times \mathbb{R}^+ \) and \( a > 0 \) there exists \( \eta \in \mathbb{R}^+ \times \mathbb{R}^+ \) such that \( w_t^i \geq \eta_i, i \in \{1, 2\} \) for every \( x_0 \in K \), and \( \|h\|_H \leq a \).

**Proof.** Fix a compact set \( K \subset \mathbb{R}^+ \times \mathbb{R}^+ \), and ball \( \|h\|_H \leq a \). Since we have local Lipschitz continuity on \( (0, \infty) \times (0, \infty) \) together with sublinear growth at infinity, we have existence and uniqueness of a solution on \( [0, T] \) up to exit from any given positive subdomain. Therefore we only need to prove the existence of the bound \( \eta \) to show existence and uniqueness of the ODE solution on \( [0, T] \). Consider any solution \( \phi^h(x_0) = (\phi^{h,1}(x_0), \phi^{h,2}(x_0)) \) of (4.3.3) with \( \|h\|_H \leq a \), and \( x_0 \in K \). The energy of the path is bounded by \( \frac{1}{2}a^2 \):

\[
\frac{1}{2}a^2 \geq \frac{1}{2}\|h\|_H^2 = J(\phi^h(x_0)) = \frac{1}{2} \int_0^T l(\phi^h(x_0), t)^T C^{-1} l(\phi^h(x_0), t) dt,
\]

where

\[
l(\phi, t) := \begin{pmatrix} l^1(\phi, t) \\ l^2(\phi, t) \end{pmatrix} = \begin{pmatrix} |\phi^1_t - b^{(1)}(\phi^1_t, \phi^2_t)| \\ |\phi^2_t - b^{(2)}(\phi^2_t)| \end{pmatrix}
\]

(4.3.5)

\[C := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.
\]
Define $\lambda_{\text{max}}$ as the maximum eigenvalue of $C$. Since 
\[ \frac{1}{\lambda_{\text{max}}} x^T x \leq x^T C^{-1} x, \]
\[ \mathcal{L}^i(\phi) := \int_0^T I^i(\phi, t)^2 dt \leq a^2 \lambda_{\text{max}} \quad i \in \{1, 2\}. \]

Note that the same bound is applicable for arbitrary correlation matrices, so the process $(S, \nu, \nu')$ required to determine the stock price large deviations can be handled in exactly the same way. Applying Proposition 3.11 in [9] to $\mathcal{L}^2$ shows there exists $\eta(2) > 0 : \phi^h(x_0) \geq \eta(2)$, $0 \leq t \leq T$ for all $x_0 \in K$, $\|h\|_H \leq a$. This implies we only need to show that there exists a lower bound for $\phi^{h,1}(x_0), \eta(1)$, in the case that the trajectory $\phi^{h,2}(x_0) > \eta(2)$. We now consider a lower bound for the integral $\mathcal{L}^1$. We first assume that $\phi^h(x_0)$ decreases monotonically to some $x$, and consider only the segment of the path with $x < \phi^h_0(x_0) \leq \bar{x}$. We choose $\bar{x}$ sufficiently small that $h(1)(w^1, w^2) \geq \gamma$ and 
\[ \sigma(1)(\phi^h(x_0)) \leq \delta(w^1)^2 \] (by the H"{o}lder continuity assumption) for some $\gamma, \delta > 0$ and $0 \leq w^1 \leq \bar{x}, w^2 > \eta(2)$ (we also take $\bar{x}$ small enough that $w^1 > \bar{x}$ for all $(w^1, w^2) \in K$). Therefore

\[ \mathcal{L}^1 = \int_0^T (I^1(\phi^h(x_0), t)^2 dt, \]
\[ \leq \int_\tau T \frac{(\phi^{h,1}(x_0) - \gamma)^2}{\delta^2 \phi^{h,1}(x_0)} dt \]
\[ \geq -2 \frac{\gamma}{\delta^2} \log \frac{x}{\bar{x}} \text{ by Lemma 4.2.12} \]

However, since this is unbounded as $x \to 0^+$, whereas we have assumed $\mathcal{L}^1 \leq \lambda_{\text{max}} a^2$, there must be some $\eta(1) : 0 < \eta(1) \leq \phi^h_0(x_0), 0 \leq t \leq T$. Then $\eta = (\eta(1), \eta(2))$. In the case that $\phi^{h,1}(x_0)$ is not monotonically decreasing, we can simply consider a new path $\psi = (\psi^1, \psi^2)$ formed from $\phi^h(x_0)$ by removing those sections where $\phi^{h,1}(x_0)$ is not monotonically decreasing. By assumption,
\[ \phi^{h,i}_0 = x_0^i + \int_0^t v^i dt, \quad i \in \{1, 2\}, \]
for some $v^i \in L^2[0, T]$. So we define $\psi$ as follows:
\[ \tilde{\psi}_t^i = x_0^i + \int_0^t v^i_1 \mathbb{I} |\phi^{h,1}_t(x_0) - \min_{0 \leq t \leq t_1} \phi^{h,1}_t(x_0)| dt \]
\[ s(t) = \int_0^t \mathbb{I} (\tilde{\psi}_t^i = \phi^{h,1}_t(x_0)) dt \]
\[ \psi_s^i(t) = (\psi^i, \phi_t^h), \quad \text{for } 0 \leq s \leq s(T). \]

Then the path, $\psi^1$, is absolutely continuous and has square integrable derivative.
Since the integrand is positive and we have only removed segments from the integration, the path, $\psi^1$, cannot have a greater energy than the original path, $\phi^{h,1}(x_0)$, and Baldi and Caramellino’s lower bound is still applicable. Note that we restrict to monotonically decreasing segments in order that $\left(\dot{w}^1 - b^{(1)}(w^1, w^2)\right)^2 \geq (\dot{w}^1 - \gamma)^2$, for $b^{(1)}(w^1, w^2) \geq \gamma > 0$.

**Proposition 4.3.4.** Assume 4.3.1 and let $\{X^\epsilon\}$ be the solutions of (4.3.1) in the time interval $[0, T]$ with $x_0 \in \mathbb{R}^+ \times \mathbb{R}^+$. Then

$$
\limsup_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(X^\epsilon \in F) \leq - \inf_{\psi \in F} \mathcal{J}(\psi)
$$

$$
\liminf_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(X^\epsilon \in G) \geq - \inf_{\psi \in G} \mathcal{J}(\psi)
$$

for every closed set $F \subset C_{x_0}([0, T], \mathbb{R}^+ \times \mathbb{R}^+)$ and open set $G \subset C_{x_0}([0, T], \mathbb{R}^+ \times \mathbb{R}^+)$,
with \( \mathcal{J}(\psi) \) as defined in \ref{4.2.5} with \( \mathcal{J}(\psi) = \infty \) if \( \psi \) is not absolutely continuous.

**Proof.** The proof is by the same localisation argument as for Baldi and Caramellino’s one dimensional result \ref{4.2.11}. \qed

### 4.3.1 Monotonicity of Rate Function

We now show how the rate function for the VIX random variable can be derived from the sample path rate function for the DCEV process. In particular we need to show that the rate function is monotonically decreasing away from the mean VIX value in order to determine the small noise implied volatility asymptotic for the VIX.

**Lemma 4.3.5.** Assume \ref{4.3.1} and let \( \{X^t\} \), be the solutions of \ref{4.3.1} in the time interval \( [0, T] \) with \( x_0 \in \mathbb{R}^+ \times \mathbb{R}^+ \). Denote the rate function for \( \{X^t\} \) by \( \mathcal{J}(\psi) \) as defined in \ref{4.2.5} Let \( Z \) be a continuous function from \( C^2 \) to \( \mathbb{R} \). Then \( Y^t = Z(X^t) \) has good rate function \( \tilde{\mathcal{J}} : \mathbb{R} \to [0, \infty] \)

\[ \tilde{\mathcal{J}}(y) := \inf \{ \mathcal{J}(\phi) : \phi \in C^2_{x_0}; \ Z(\phi) = y \}. \]

The set \( \{ y : \tilde{\mathcal{J}}(y) < \infty \} \) is an interval containing \( \bar{y} = Z(\bar{y}), \) where \( \bar{y} = \phi^0(x_0) \). \( \tilde{\mathcal{J}}(\bar{y}) = 0 \) and \( \tilde{\mathcal{J}}(y) \) is strictly monotonically increasing in \( |y - \bar{y}| \) on this interval.

**Proof.** The first part is by the contraction principle, (c.f. \cite{20}, Theorem 4.2.1). The second part follows from the intermediate value theorem. We will first show that if there exists a path \( \bar{\psi} \in C^2_{x_0} : \mathcal{J}(\psi) < \infty \) and \( Z(\psi) = y > \bar{y} \) then we can construct a path \( \hat{\psi} \) that achieves \( Z(\hat{\psi}) = \bar{y} \) with \( y > \bar{y} > \bar{y} \) with \( \mathcal{J}(\hat{\psi}) < \mathcal{J}(\psi) \). Now if \( \psi \) is a minimising path for \( y \) so that \( \mathcal{J}(\psi) = \tilde{\mathcal{J}}(y) \) then this implies that \( \tilde{\mathcal{J}}(\bar{y}) < \tilde{\mathcal{J}}(y) \). For \( \psi \in C^2_{x_0} \), we define the family of curves \( \{ \psi^\tau \} \subset C^2_{x_0} \) by

\[
\psi^\tau_t = \begin{cases} 
\psi_t & 0 \leq t \leq \tau \\
\phi_{2 \tau}^0(\psi_T) & \tau < t \leq T 
\end{cases}
\]

for \( \tau \in [0, T] \). Then \( \tau \to \psi^\tau \) is a continuous mapping from \( [0, T] \) to \( C^2_{x_0} \). Therefore the composition \( \tau \to Z \circ \psi^\tau \) is also continuous. Assume that there is a \( \psi \in C^2_{x_0} \) such that \( Z(\psi) > \bar{y} \) and \( \mathcal{J}(\psi) < \infty \). Then \( \psi^0 \equiv \bar{y} \) and \( \psi^T \equiv \psi \), so by the intermediate value theorem, there is a \( \tau_1 \in (0, T) \) such that \( Z(\psi^{\tau_1}) \) achieves any value in \( (y, \bar{y}) \). Furthermore \( \psi^{\tau_1} \) is in \( \mathcal{H}^d_{x_0} \). For such a \( \tau_1 \), since \( Z(\psi^{\tau_1}) < y, |\xi - b(\xi)| > 0 \) on a set of positive measure in \( (\tau_1, T] \). Furthermore, since \( \mathcal{J}(\psi) < \infty \), we know that \( \psi \) stays in a bounded domain strictly inside the positive quadrant, so that the integrand of \( \mathcal{J}(\psi) \) is strictly positive on that set. Therefore \( \mathcal{J}(\psi^{\tau_1}) < \mathcal{J}(\psi) \). This shows that \( \{ y : \tilde{\mathcal{J}}(y) < \infty \} \) is an interval containing \( \bar{y} = Z(\bar{y}) \). Since \( \tilde{\mathcal{J}}(\cdot) \) is a good rate function,
the minimum path is achieved if the energy is finite. Therefore by choosing \( \psi \) as a minimising path (we allow multiple minima) that achieves a given level \( y > \bar{Z}(\bar{g}) \), so that \( \mathcal{J}(\psi) = \tilde{\mathcal{J}}(y) \) and pursuing the same argument as above we see that the rate function, \( \tilde{\mathcal{J}}(\cdot) \), is also strictly monotonically increasing away from \( \bar{y} \) on the region it is finite (and \( \tilde{\mathcal{J}}(\bar{y}) \) is clearly zero).

\[\square\]

**Corollary 4.3.6.** In the DCEV model the family of VIX random variables \( \{VIX^\epsilon_T\}_\epsilon \) given by

\[ VIX^\epsilon_T = \sqrt{a_1 v_T^\epsilon + v_T'^\epsilon + a_3 z_3} \]

where

\[
dv^\epsilon_T = \kappa(v_T^\epsilon - v_T'^\epsilon)dt + \epsilon \xi_1 (v_T^\epsilon)^\alpha (\sqrt{1 - \rho^2}dW^1_t + \rho dW^2_t) \\
dv'^\epsilon_T = c(z_3 - v_T'^\epsilon)dt + \epsilon \xi_2 (v_T'^\epsilon)^\beta dW^2_t \\
v_0 = v_0 > 0, \ v_0'^\epsilon = v_0'^\epsilon > 0
\]

with \( a_1, a_2, a_3 > 0 \) satisfy an LDP with good rate function

\[ \tilde{\mathcal{J}}(y) : = \inf\{\mathcal{J}(\psi): \psi \in \mathcal{C}_2(v_0, v_0'); \sqrt{a_1 \psi_1^T + a_2 \psi_2^T + a_3 z_3} = y\}, \]

\[ \mathcal{J}(\psi) = \frac{1}{2} \int_0^T l(\psi, t) \psi C^{-1} l(\psi, t) dt \]

\[ l((v, v'), t) = \begin{pmatrix} l^1((v, v'), t) \\ l^2((v, v'), t) \end{pmatrix} = \begin{pmatrix} \frac{\dot{v}_t - \kappa(v_t' - v_1)}{\xi_1 v_t^\epsilon} \\ \frac{\dot{v}_t' - c(z_3 - v_t')}{\xi_2 v_t'^\epsilon} \end{pmatrix} \]

\[ C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \]

\( \tilde{\mathcal{J}}(y) \) is finite on \((0, \infty)\) and monotonically increasing away from \( F^0_0 = \sqrt{a_1 \bar{v}_T + a_2 \bar{v}'_T + a_3 z_3} \) where

\[
\bar{v}'_T = e^{-\kappa T} v_0'^\epsilon + (1 - e^{-\kappa T}) z_3 \\
\bar{v}_T = e^{-\kappa T} v_0 + \kappa \left( \frac{v_0'^\epsilon - z_3}{c - \kappa} \right) (e^{-\kappa T} - e^{-cT}) + (1 - e^{-\kappa T}) z_3
\]

and the small noise implied volatility limit for a VIX (vanilla) option with strike \( K \) and maturity \( T \) (at a continuity point of the energy \( \tilde{\mathcal{J}} \)) is

\[ \lim_{\epsilon \to 0} \sigma_{T, \epsilon}(T, K) = \frac{|\log F^\epsilon_0(K)|}{\sqrt{2T \tilde{\mathcal{J}}(K)}} \text{ for } K \neq F^0_0 \]
Proof. Since the VIX random variable, given by $\sqrt{a_1 v_T + a_2 v_T' + a_3 z_3}$, is a continuous functional of the path $(v, v')$, we see that this result applies. Furthermore for any (positive) value of the VIX, we can find a finite energy path that achieves it, namely a straight line to any candidate $(v_T, v_T')$, therefore the energy is finite for all positive values of the VIX, and strictly monotonically increasing. We now show that the VIX has a bounded quadratic moment (for bounded $\epsilon$ and $T$).

Setting a stopping time, $s_n := \inf\{t : v_t \wedge v_t' \leq 1/n\}$, the stopped process $X_{t \wedge s_n}^\epsilon$ has locally bounded quadratic moment. Then $E\|X_{t \wedge s_n}^\epsilon\|^2$ is uniformly bounded for all $n$ and bounded $\epsilon$ (and $t$). We denote this bound for $\epsilon, t \leq 1$, by $M$. Since $\{v = 0\} \cup \{v' = 0\}$ is unattainable in finite time, then $\lim_{n \to \infty} s_n = \infty$ so $E[\|X_t^\epsilon\|^2] \leq \lim_{n \to \infty} E[\|X_{t \wedge s_n}^\epsilon\|^2] \leq M$ by Fatou’s lemma. Since $E[(VIX_T^\epsilon]^2) = E[a_1 v_T + a_2 v_T' + a_3 z_3] \leq a_3 z_3 + \|(a_1, a_2)\| \sqrt{E[(v_T, v_T')]^2}$, we see that this also implies that $E[(VIX_T^\epsilon]^2]$ is bounded for bounded $\epsilon, T$. Then the VIX random variable (derived from the DCEV model) satisfies a small noise LDP with rate function monotonically increasing away from the mean value, and the VIX random variable is square integrable, so we can use Theorem 3.1.6 to derive the small noise implied volatility asymptotic for VIX in Gatheral’s DCEV model. 

In the next section, we report results from calculating the rate function numerically.

4.4 Numerical Results

The success of the small time asymptotic expansion for SABR is in large part due to the model not being time dependent or having mean reversion etc, so that the implied volatility varies slowly with time. When the stochastic model is time varying, a small noise expansion can make use of the same properties. We illustrate below, in the case of Gatheral’s double CEV model. We compare convergence as $\epsilon \to 0$ to $T \to 0$, by plotting the implied volatility $\sigma_{T, \epsilon}(K, \epsilon^2 T; F^\epsilon)$ against $\epsilon^2 T$ for Monte Carlo simulations (10 million samples) of the model with varying $\epsilon$ and $T$. The model parameters were taken from [52]. They are based on Gatheral’s ([31]) calibration to real market data.

The plots below in red show the implied volatility profile as time varies from 2 weeks to 6 months, keeping $\epsilon = 1$; the plots in blue, fix $T$ and reduce $\epsilon$—the steps are 0.25, 0.5, 1. Finally, the zero point is our large deviations calculation for the small noise implied volatility asymptotic. The spikes are due to Monte Carlo noise, as the probability of being in the money becomes exponentially smaller as $\epsilon \to 0$. Nevertheless,
4.4. Numerical Results

<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_0$</td>
<td>0.0137</td>
</tr>
<tr>
<td>$v'_0$</td>
<td>0.0208</td>
</tr>
<tr>
<td>$c$</td>
<td>0.1</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>5.5</td>
</tr>
<tr>
<td>$z_3$</td>
<td>0.078</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>2.6</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.44</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Table 4.1: DCEV parameter settings

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Strike</th>
<th>14%</th>
<th>16%</th>
<th>18%</th>
<th>20%</th>
<th>22%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 months</td>
<td>89.20%</td>
<td>92.32%</td>
<td>94.70%</td>
<td>96.58%</td>
<td>98.14%</td>
<td></td>
</tr>
<tr>
<td>2 months</td>
<td>66.59%</td>
<td>69.97%</td>
<td>72.75%</td>
<td>75.08%</td>
<td>77.06%</td>
<td></td>
</tr>
<tr>
<td>4 months</td>
<td>52.59%</td>
<td>55.29%</td>
<td>57.62%</td>
<td>59.64%</td>
<td>61.42%</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>45.04%</td>
<td>47.22%</td>
<td>49.13%</td>
<td>50.82%</td>
<td>52.33%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Implied Volatility from Monte Carlo Simulation

it can be seen that the implied volatility varies much less with noise than with time: consider approximating a two month option (70% in the graphs below) with the corresponding zero order asymptotic, small noise is out by a couple of percent, whilst extrapolating the red graphs to zero, implies a small time estimate well over 100%! Furthermore one can see that the implied volatility is converging to our theoretical calculation.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Strike</th>
<th>14%</th>
<th>16%</th>
<th>18%</th>
<th>20%</th>
<th>22%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 months</td>
<td>1.24%</td>
<td>1.49%</td>
<td>1.73%</td>
<td>1.95%</td>
<td>2.13%</td>
<td></td>
</tr>
<tr>
<td>2 months</td>
<td>3.52%</td>
<td>3.67%</td>
<td>3.77%</td>
<td>3.86%</td>
<td>3.93%</td>
<td></td>
</tr>
<tr>
<td>4 months</td>
<td>3.83%</td>
<td>4.16%</td>
<td>4.40%</td>
<td>4.59%</td>
<td>4.75%</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>3.55%</td>
<td>3.90%</td>
<td>4.18%</td>
<td>4.42%</td>
<td>4.62%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Small Noise - Monte Carlo Implied Volatility Error
4.4. Numerical Results

Figure 4.4.1: Numerical investigation of small time and small noise asymptotics of DCEV implied volatility
4.4.1 Numerical method

To numerically determine the rate function, \( \tilde{J}(K) \), for the VIX asset at time \( T \),

\[
\tilde{J}(K) = \min \{ J(\psi) : \psi \in C^2_{(v_0, v_0')} ; \sqrt{a_1 \psi_1^2 + a_2 \psi_2^2 + a_3 z_3} = \bar{K} \},
\]

\[
J(\psi) = \frac{1}{2} \int_0^T l(\psi, t)^T C^{-1} l(\psi, t) \, dt
\]

\[
l((v, v'), t) = \begin{pmatrix} l_1((v, v'), t) \\ l_2((v, v'), t) \end{pmatrix} = \begin{pmatrix} \frac{v_1 - \kappa(v_1' - v_1)}{\xi_1 v_1' - \xi_3 z_3 - v_1} \\ \xi_2 v_1'^2 \end{pmatrix}
\]

\[
C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},
\]

we discretise the integral to be minimised as a Riemann sum, and parameterise the curves \( \psi \) as piecewise linear. We use the Cholesky Decomposition of the inverse of the correlation matrix \( C^{-1} = LL^T \),

\[
h(t) = L^T l(\psi, t)
\]

\[
J(\psi) \approx \sum_{i=1}^{N} l(i \Delta T)^T C^{-1} l(i \Delta T) \Delta T
\]

\[
= \sum_{i=1}^{N} (h^1(i \Delta T)^T h^1(i \Delta T) + h^2(i \Delta T)^T h^2(i \Delta T) \Delta T).
\]

We seek the lowest energy path, such that \( VIX = K \). This corresponds to \( (a_1, a_2) \cdot \psi_T = K^2 - a_3 z_3 \). We now have a standard constrained nonlinear least squares optimisation problem, so by fixing the (first coefficient of the) gradient at the last point so that \( VIX = K \), we turn the problem into an unconstrained nonlinear least squares problem to which we apply the Levenberg-Marquardt algorithm ([44]). In our results we used 32 time-points, which seemed to give good convergence - approx 10% relative error on the SABR exact asymptotic. Our C++ implementation took on average 1 millisecond to converge. In comparison our finite difference implementation of the PDE equation (using NAG’s D03RAF [1] routine) for a VIX call option price took approximately 15 seconds to provide an accurate result.

Future areas of investigation would be to investigate alternative parameterisations of the curves to increase speed and accuracy (such as using piecewise polynomial functions). Furthermore one could try to numerically solve the Hamiltonian ODEs for the minimising paths which we describe in Chapter 6 section 6.2.
4.A Comparison Principle

We modify Proposition 5.2.18 in [42], to allow for a random drift and stopping time.

Proposition 4.A.1. Suppose that on a certain probability space \((\Omega, \mathcal{F}, P)\) equipped with a filtration \(\{\mathcal{F}_t\}\) which satisfies the usual conditions, we have a standard, one-dimensional Brownian motion \(\{W_t, \mathcal{F}_t; 0 \leq t < \infty\}\), a stopping time \(\tau\), a continuous adapted process \(Y\) and two continuous, adapted processes \(X^{(1)}, X^{(2)}\) such that

\[
X_t^{(j)} = X_0^{(j)} + \int_0^{\tau \wedge t} b_j(s, X_s^{(j)}, Y_s) ds + \int_0^{\tau \wedge t} \sigma(s, X_s^{(j)}) dW_s^1; \quad 0 \leq t < \infty
\]

holds a.s. for \(j=1,2\). We assume that

1. the coefficients \(\sigma(t, x), b_j(t, x, z)\) are continuous, real-valued functions on \([0, \infty) \times \mathbb{R}\) and \([0, \infty) \times \mathbb{R}^2\) respectively,

2. the dispersion matrix satisfies \(|\sigma(t, x) - \sigma(t, y)| \leq h(|x - y|)\) for every \(0 \leq t \leq \infty\) and \(x, y \in \mathbb{R}\) and \(h : [0, \infty) \to [0, \infty)\) is a strictly increasing function with \(h(0) = 0\) and

\[
\int_{(0, \delta)} h^{-2}(u) du = \infty; \quad \forall \delta > 0
\]

3. \(X_0^{(1)} \leq X_0^{(2)}\) a.s.,

4. \(b_1(t, x, Y_t) \leq b_2(t, x, Y_t), \forall 0 \leq t \leq \tau, x \in \mathbb{R},\) and

5. either \(b_1(t, x, z)\) or \(b_2(t, x, z)\) satisfies

\[
|b(t, x, Y_t) - b(t, y, Y_t)| \leq K|x - y|,
\]

for every \(0 \leq t \leq \tau\) and \(x, y \in \mathbb{R}\), where \(K\) is a positive constant

Then

\[
\mathbb{P}[X_t^{(1)} \leq X_t^{(2)}, \forall 0 \leq t < \infty] = 1
\]

Proof: Define the process \(\Delta_t = X_t^{(1)} - X_t^{(2)}\). Let \(\{\phi_n\}_n\) be a nondecreasing sequence of \(\mathcal{C}^2(\mathbb{R})\) functions converging pointwise to \(x^+\), with

\[
0 \leq \phi_n'(x) \leq 1_{(0, \infty)}(x)
\]

and

\[
0 \leq \phi_n''(x) \leq \frac{2}{n h^2(x)} 1_{(0, \infty)}(x)
\]
for $x \in \mathbb{R}$ (see Proposition 5.2.13 and 5.2.18 in [42] for precise definition). Then following Karatzas and Shreve’s proof, we have

\[
\phi_n(\Delta t \wedge \tau) = \int_0^{t \wedge \tau} \phi_n' (\Delta_s) [b_1(s, X^{(1)}_s, Y_s) - b_2(s, X^{(2)}_s, Y_s)] ds \\
+ \frac{1}{2} \int_0^{t \wedge \tau} \phi_n'' (\Delta_s) [\sigma(s, X^{(1)}_s) - \sigma(s, X^{(2)}_s)]^2 ds \\
+ \int_0^{t \wedge \tau} \phi_n'(\Delta_s) [\sigma(s, X^{(1)}_s) - \sigma(s, X^{(2)}_s)] dW_s.
\]

Then

\[
\frac{1}{2} \mathbb{E} \left[ \int_0^{t \wedge \tau} \phi_n'' (\Delta_s) [\sigma(s, X^{(1)}_s) - \sigma(s, X^{(2)}_s)]^2 ds \right] \leq \frac{1}{n} \mathbb{E} \int_0^{t \wedge \tau} 1_{(0,\infty)}(\Delta_s) ds \leq \frac{t}{n},
\]

so

\[
\mathbb{E}[\phi_n(\Delta(t \wedge \tau))] - \frac{t}{n} \leq \mathbb{E} \int_0^{t \wedge \tau} \phi_n'(\Delta_s) [b_1(s, X^{(1)}_s, Y_s) - b_2(s, X^{(2)}_s, Y_s)] ds \\
+ \mathbb{E} \int_0^{t \wedge \tau} \phi_n'(\Delta_s) [b_1(s, X^{(2)}_s, Y_s) - b_2(s, X^{(2)}_s, Y_s)] ds \\
\leq \mathbb{E} \int_0^{t \wedge \tau} \phi_n(\Delta_s) [b_1(s, X^{(1)}_s, Y_s) - b_1(s, X^{(2)}_s, Y_s)] ds \\
\leq K \mathbb{E} \int_0^{t \wedge \tau} 1_{(0,\infty)}(\Delta_s) |\Delta_s| ds.
\]

Now by using the stopped process, $\Delta(t \wedge \tau)$, on the right hand side, we can extend the integral up to $t$, so we finally have

\[
\mathbb{E}[\phi_n(\Delta(t \wedge \tau))] - \frac{t}{n} \leq K \int_0^{t} \mathbb{E}[\Delta^+_{t \wedge \tau}] ds.
\]

Letting $n \to \infty$, we obtain $\mathbb{E}[\Delta^+_{t \wedge \tau}] \leq K \int_0^{t} \mathbb{E}[\Delta^+_{t \wedge \tau}] ds; 0 \leq t < \infty$. So by the Gronwall inequality, we have $\mathbb{E}[\Delta^+_{t \wedge \tau}] = 0$, which implies, by sample path continuity that

\[
\mathbb{P}[X^{(1)}_{t \wedge \tau} \leq X^{(2)}_{t \wedge \tau}, \forall 0 \leq t < \infty] = 1.
\]

\[\square\]
Chapter 5

Yoshida Watanabe expansions using Mathematica

In this chapter we give an outline Watanabe-Yoshida of asymptotic expansions of Generalised Wiener functionals and provide a set of Mathematica routines to derive them automatically. To our knowledge, this is the first time such a program is published. These expansions can be computed through repeated formal differentiation and Taylor expansions of the Generalised Wiener functionals and associated random variables. They have been derived for a number of different models in finance ([47, 43, 75, 4]). Although each step is straightforward, the number of terms grows exponentially with the order required, so manual calculations soon becomes impractical for all but the most simple one dimensional expansion. An automated procedure allows even high dimensional SDEs to be attempted. Here we show that by extending Tocino’s work on stochastic Taylor expansions in Mathematica ([77]) to SDEs and conditional expectations, Watanabe-Yoshida expansions can be derived straightforwardly. Providing this methodology will allow a better investigation of the application of Yoshida-Watanabe expansions. Although one could investigate very many variations such as small time or small noise expansions, expansions of the log of the stock process versus the stock itself etc., the number of computations involved discourage a thorough evaluation. In a recent development [75], an alternative method of deriving the expansion coefficients has been suggested which involves solving nested series of ODEs. However no implementation scheme was provided.

5.1 Asymptotic Expansion

We follow the presentation of the method in [75]. Let $W = (W_1, \ldots, W_m)$ be a $m$-dimensional Wiener process, we assume that the individual Wiener components have
a deterministic correlation \( \rho_{i,j}(t), 1 \leq i, j \leq m \). We depart from [75] and use correlated Brownian drivers since this reduces the number of terms in the asymptotic expansions of SDEs used in finance. We consider a \( d \)-dimensional diffusion process \( X^\epsilon_t = (X^\epsilon_{t,1}, X^\epsilon_{t,2}, \cdots, X^\epsilon_{t,d}) \) which is the solution of the following stochastic differential equation:

\[
dX^\epsilon_t = V_0(X^\epsilon_t, \epsilon)dt + \epsilon V(X^\epsilon_t) dW_t
\]

\( X^\epsilon_0 = x_0 \in \mathbb{R}^d \) and \( \epsilon \in (0, 1] \) is a known parameter. Next suppose that a function \( g : \mathbb{R}^d \to \mathbb{R} \) is smooth and all derivatives have polynomial growth. Then, for \( \epsilon \downarrow 0 \), \( g(X^\epsilon_t) \) has an asymptotic expansion (see Chapter 2):

\[
g(X^\epsilon_T) = \sum_{n=0}^{\infty} \epsilon^n g_{nT} = g(X^0_T) + \epsilon \partial_x g(X^0_T) X^\epsilon_T + \epsilon^2 \left( \frac{1}{2} \partial_{x,i} \partial_{x,j} g(X^0_T) X^\epsilon_{T,i} X^\epsilon_{T,j} + \partial_{x,i} g(X^0_T) \partial^2 \epsilon X^\epsilon_{T,i} \right) + O(\epsilon^3),
\]

where we have implicitly used Einstein summation and all derivatives with respect to \( \epsilon \) are evaluated at \( \epsilon = 0 \). A similar expansion can be derived for (\( \epsilon \)-families of ) Generalised Wiener functionals, in particular the application of a Schwartz distribution \( T \in \mathcal{S}'(\mathbb{R}) \) to a smooth uniformly nondegenerate family of random variables \( \{F^\epsilon\}_{\epsilon \in D} \) (Theorem 2.2.5). Consider \( T(x) = (x)^+ \). The restriction to uniformly nondegenerate random variables means we cannot compute \( \lim_{\epsilon \to 0} \mathbb{E}[(g(X^\epsilon_T) - K)^+] \) by this method. Instead we are led to considering \( \mathbb{E}[(Y^\epsilon - k)^+] \) where

\[
Y^\epsilon = \frac{g(X^\epsilon_T) - g(X^0_T)}{\epsilon},
\]

which in terms of the original variables can be written

\[
\frac{1}{\epsilon} \mathbb{E}[(g(X^\epsilon_T) - (g(X^0_T) + \epsilon k))^+]
\]

Taking \( \epsilon \mapsto \epsilon \sigma \), with \( \sigma \) denoting the standard deviation of the Gaussian random variable \( g_{1T} \), we see that the Yoshida-Watanabe method considers the asymptotics of a European call option whose strike is always \( k \) standard deviations from the mean, \( g(X^0_T) \), as the standard deviation \( \epsilon \sigma \downarrow 0 \), rather than a fixed strike. This is what we
call a regular perturbation in Chapter 1 and we clearly expect that the method will work better the less the coefficients of the SDE vary over the domain.

The coefficients in the expansion can be obtained by a Taylor expansion around $\epsilon = 0$. Let

$$A_{n,t} = \partial^n \frac{X^\epsilon_t}{\epsilon} \big|_{\epsilon = 0}$$

and $A_i^\epsilon(t), i = 1, \ldots, d$ denote the i-th element of $A_n(t)$. Then $A_n(t)$ can be represented as iterated Wiener-Itô integrals. We illustrate this with an example with $d = m = 1$,

$$dX_t^\epsilon = \epsilon \sigma(t, X_t^\epsilon) dW_t \quad X_0^\epsilon = x_0$$

$$A_{n,T} = \partial^n \frac{X^\epsilon_T}{\epsilon} \big|_{\epsilon = 0}$$

$$A_{0,t} \equiv x_0$$

$$dA_{1,t} = \sigma(t, A_{0,t}) dW_t$$

$$dA_{2,t} = 2 \partial_x \sigma(t, A_{0,t}) A_{1,t} dW_t$$

$$dA_{3,t} = 3 \partial_x^2 \sigma(t, A_{0,t}) (A_{1,t})^2 + 3 \partial_x \sigma(t, A_{0,t}) A_{2,t} dW_t$$

Applying Itô’s formula to $(A_{1,t})^2$, allows us to express $A_3$ also as an iterated Wiener integral. As can be seen, the $\{A_n\}$ variables can be calculated iteratively, since each new variable, $A_n$, depends only on the preceding variables. Given a representation for iterated Wiener integrals, we see that the key computational issue is deriving products (and integer powers) of iterated Itô and Lebesgue integrals. The two additional steps are adding an extra layer of integration and calculating expectations of iterated Wiener integrals. In the next section we will show how Mathematica can be used to derive these products and expectations. Itô integration with respect to the component $W^j_t$ will be denoted by $dW^j_t$, and $dW^j_0$ for Lebesgue integrals. Then we will write the iterated Itô/Lebesgue integral of $f_1 \cdot f_2 \cdots f_r$ with respect to the integration variable indices $(j_1, j_2, \ldots, j_r)$ by

$$I_{(j_1, j_2, \ldots, j_r)}[f_1, f_2, \ldots, f_r](t) = \int_0^t \left( \int_0^{s_r} \cdots \left( \int_0^{s_2} f_1(s_1) dW^{j_1}_{s_1} \right) \cdots f_{r-1}(s_{r-1}) dW^{j_{r-1}}_{s_{r-1}} \right) f_r(s_r) dW^{j_r}_{s_r}.$$

We draw attention to the fact that the outermost integration is the last in the list. This can be confusing - in fact, in the list of conditional expectations in [75], the notation (25) defines the wrong ordering (for their stated formulas) and should be reversed to
\[ F_n(f_1, \cdots, f_n) := \int_0^T \int_0^{t_{n-1}} \cdots \int_0^{t_2} f_1(t_1) \cdots f_n(t_n) dt_1 \cdots dt_n, \quad n \geq 1. \]

So in our example,

\[ A_{1,t} = I_{(1)}[\sigma(\cdot, A_0, \cdot)] = \int_0^t \sigma(t_1, x_0) dW_t^1 \]

\[ A_{2,t} = I_{(1,1)}[2\sigma(\cdot, x_0) \partial_x \sigma(\cdot, x_0)](t) = \int_0^t 2 \partial_x \sigma(t_2, x_0) \int_0^{t_2} \sigma(t_1, x_0) dW_t^1 dW_{t_2}^1 \]

Our goal is to compute asymptotic expansions of the expectation of the generalised Wiener Functional \( T(Y^x_T) \), where \( Y^x_T \) is uniformly non degenerate and has expansion \( Y^x_T \sim Y_0 + \epsilon Y_1 + \cdots \) in \( D^\infty \). We have shown that the asymptotic expansion of \( T(Y^x_T) \) can be written as a sum of iterated Wiener-Itô integrals. We can compute the expectation of these integrals by conditioning on the Gaussian random variable \( Y_0 \) using the following lemma.

**Lemma 5.1.1.** Let \( \xi \) be a random variable on a probability space \( L^2(\Omega, \mathcal{F}, P) \), and \( G \) be a Gaussian random variable, with zero mean and variance \( \Sigma > 0 \), on the same space. Then a representation of the conditional expectation of \( \mathbb{E}[\xi | G] \) is given by

\[
\mathbb{E}[\xi | G] = \sum_{n=0}^{\infty} \frac{\mathbb{E}[H_n(G; \Sigma)]}{n! \Sigma^n} H_n(G; \Sigma),
\]

where \( H_n(x, \Sigma) \) is the Hermite polynomial of degree \( n \) which is defined as

\[
H_n(x; \Sigma) := (-\Sigma)^n e^{x^2/2\Sigma} \partial_x^n e^{-x^2/2\Sigma} \quad n \in \mathbb{N}. \quad (5.1.1)
\]

In particular, for the Wiener Integral

\[
\int_0^T q(t) \cdot dW_t, \quad q \in L^2([0, T]; \mathbb{R}^m),
\]

the conditional expectation is given by

\[
\mathbb{E}[\xi | G] = \sum_{i=0}^{\infty} \frac{1}{\|q\|_{L^2}^{2n}} \mathbb{E}[\xi J_n[q](T)] H_n(J_1(q); \|q\|_{L^2}^2)
\]

where

\[
\|q\|_{L^2}^2 := \int_0^T \sum_{i,j=1}^m \rho_{ij}(t) q_i(t) q_j(t) dt
\]

\[ ^1 \text{Nualart’s definition includes a factor } 1/n!, \text{ so his highest order term is } x^n/n! \]
and \( J_n[q](T) \) refers to the \( n \)-times iterated Wiener integral of the \( m \) dimensional function \( q : \mathbb{R}^+ \to \mathbb{R}^m \):

\[
J_n[q](T) := \int_0^T \ldots \left( \int_0^{t_2} q(t_1) \cdot dW_{t_1} \right) \ldots q(t_n) \cdot dW_{t_n}, \ n \in \mathbb{Z}^+
\]

\( J_0[q](T) := 1. \)

**Proof.** Since the Hermite polynomials of \( G \) form a complete orthogonal basis for the closed subspace of \( L^2(\Omega, \mathcal{F}, P) \) generated by \( \sigma(G) \), then the conditional expectation of any random variable in \( L^2(\Omega, \mathcal{F}, P) \) on \( \sigma(G) \) can be represented as a Hermite polynomial expansion in \( G \). Then we use the result that the iterated Itô integral \( J_n[q](T) \) is equal to

\[
J_n[q](T) = \frac{1}{n!} H_n(J_1[q](T), \|q\|_{L^2([0,T])}^2)
\]

(e.g. [61], Proposition 1.1.4), and

\[
\mathbb{E}[J_n[q](T)J_r[q](T)] = \delta_{nr} \frac{\|q\|_{L^2([0,T])}^{2n}}{n!}.
\]

Since our asymptotic expansions are represented as iterated integrals, one could calculate the projection onto a given Hermite polynomial order \( k \) by first computing the product with \( J_k[q](T) \) as a linear combination of iterated integrals and then computing the integral of only the purely deterministic iterated integrals. However, it is more efficient to compute the projection directly as a projection on iterated integrals because of the orthogonality between different orders of iterated Wiener integrals.

**Proposition 5.1.2.** Let the \( r \) times Wiener/Lebesgue iterated integral

\[
I_{(k_1, \ldots, k_r)}[f_1, f_2, \ldots, f_r](t)
\]

have \( n \leq r \) Wiener integrals, then the expectation of its product with \( J_{\bar{n}}[q](T) \) (\( \bar{n} \in \mathbb{N} \))

---

\(^2\)here we explicitly exclude Lebesgue integrals
is given by

\[
E[I_{(k_1, \ldots, k_m)}[f_1, f_2, \ldots, f_r](t)J_n[q](T)] =
\begin{cases}
\int_0^{t_{\bar{n}}} \left( \int_0^{t_{\bar{n}}-1} g_{\bar{n}}(t_{\bar{n}}) \, dt_{\bar{n}} \right) g_{\bar{n}}(t_{\bar{n}}) \, dt_{\bar{n}} & \bar{n} = n \\
0 & \bar{n} \neq n
\end{cases}
\]

and \( g_i = \begin{cases} f_i & \text{if } k_i = 0 \\
\sum_{j=1}^m f_i \rho_{k_i,j} q_j & \text{otherwise.} \end{cases} \)

**Proof.** By induction on \( r \). For \( r = 1 \), by applying Itô’s formula to the product, we see that

\[
E[I_{(k_1, \ldots, k_r)}[f_1](t)J_n[q](T)] =
\begin{cases}
\int_0^{t_{\bar{n}}} g_1(t_1) \, dt_1 & \bar{n} = n \\
0 & \text{otherwise.}
\end{cases}
\]

Now assume the relationship is true up to a given \( r \in \mathbb{Z}^+ \). Then

\[
E[I_{(k_1, \ldots, k_r, k_{r+1})}[f_1, \ldots, f_r, f_{r+1}](t)J_n[q](T)]
\begin{cases}
\int_0^{t_{\bar{n}}} E[I_{(k_1, \ldots, k_r)}[f_1, \ldots, f_r](t_{r+1})J_n[q](t_{r+1})]g_{r+1}(t) \, dt_{r+1} & \text{if } k_{r+1} = 0 \\
\int_0^{t_{\bar{n}}} E[I_{(k_1, \ldots, k_r)}[f_1, \ldots, f_r](t_{r+1})J_{n-1}[q](t_{r+1})]g_{r+1}(t) \, dt_{r+1} & \text{otherwise.}
\end{cases}
\]

If \( I_{(k_1, \ldots, k_{r+1})} \) has \( n \) Wiener integrals, then \( I_{(k_1, \ldots, k_r)} \) has either \( n \) or \( n-1 \), according to whether \( k_{r+1} \) is zero or not, so substituting the result for \( r \), we see that the result holds also for \( r + 1 \). \( \square \)

### 5.2 Iterated Stochastic Integrals with Mathematica

We now set about expressing the above theorems in Mathematica. We first note that the key objects are iterated Wiener integrals, \( I_n[f_1, \ldots, f_r](t) \). In fact, since we are considering small noise expansions, we find it simpler to introduce a prefactor function, \( g(t) \). Consider the SDE with drift

\[
dY_t^\epsilon = (bY_t^\epsilon + c) \, dt + \epsilon \sigma(t, Y_t^\epsilon) \cdot dW_t
\]

Since the drift term does not have a factor \( \epsilon^n \), \( n \geq 1 \), the asymptotic expansion of \( Y^\epsilon \) of order \( n \), does not follow from inserting the expansion to order \( n - 1 \) into the right hand side. We therefore remove the drift by using the integrating factor, \( \exp(-bt) \), so
we obtain

\[ Y_t^\epsilon = Y_0^\epsilon e^{bt} + ce^{bt} \int_0^t e^{-bs} dt + ce^{bt} \int_0^t e^{-bs} \sigma(s,Y_s^\epsilon) \cdot dW_t \]

The same approach works for drifts that are not simply a linear function of \( Y_t^\epsilon \), but in this case we have to first determine the SDE for the asymptotic expansion, so that the drift becomes linear in the required (highest order) term. In any case, we use the prefactor function to enable some time dependent function to be represented as in the case above. So we consider an iterated Itô integral \( \text{ito}[g; (j_1, j_2, \ldots, j_r); f_1, f_2, \ldots, f_r](t) := g(t)I_{(j_1, j_2, \ldots, j_r)}[f_1, \ldots, f_r](t) \). In Mathematica we choose to represent each integrand and prefactor as a pure function, and multi-indices, \( \alpha \), (and integrands, \( f \)) as lists. The use of pure functions (i.e. functions without an argument specified, \([f(\cdot)]\)), seems natural, given that the time variables are arbitrary integration variables and would have to be changed as one calculated products of Itô integrals. However, it does lead to complications, in that natural simplifications of integrands e.g. \( \exp[-t] \exp[t] \Rightarrow 1 \) are not automatically performed by Mathematica in the pure function representation. We write \( \text{ito}[f_0; (0, 1, 0); (f_1, f_2, \ldots, f_r)] \) in Mathematica as

\[ \text{ito}[f_0[\#]&, \{0, 1, 0\}, \{f_1[\#]&, f_2[\#]&, f_3[\#]&\}], \]

so

\[ \text{ito}[\sin; (2, 1); (\exp, \cos)] = \sin(\cdot) \int_0^s \cos(s) \left( \int_0^{s_2} \exp(s_1) dW_{s_1}^2 \right) dW_{s_2}^1 \]

becomes \( \text{ito}[\sin[#]&, \{2, 1\}, ](\exp[#]&, \cos[#]&) \). As seen above, expectations and conditional expectations of iterated Wiener/Lebesgue integrands are straightforward computationally, since they only depend on a formal inspection of the integration variables and pairwise multiplication of integrands. The key mathematical computation is products of iterated Wiener/Lebesgue integrals. We first introduce some notation. Given \( l \in \mathbb{N} \), a vector of the form \( \alpha = (j_1, \ldots, j_l) \) with \( (j_1, \ldots, j_l) \in \{0, 1, \ldots, m\} \) is called a multi-index of length \( l(\alpha) = l \). For \( \alpha = (j_1, \ldots, j_l) \), \( \alpha^- \) denotes the multi-index \((j_1, \ldots, j_{l-1})\), with the understanding that \((j_1) = v\) is a multi-index of length equal to 0. The following two lemmas allow products of arbitrary iterated integrals with unit integrands to be calculated.

Lemma 5.2.1. ([45]) If \( j, j_1, \ldots, j_l \in \{0, 1, \ldots, m\} \) then

\[ I_{(j)}I_{(j_1, \ldots, j_l)} = \sum_{i=0}^{l} I_{(j_1, \ldots, j_i, j_{i+1}, \ldots, j_l)} + \sum_{i=1}^{l} \mathbb{1}\{j_i = j \neq 0\}I_{(j_1, \ldots, j_i, 0, j_{i+1}, \ldots, j_l)} \]

Products of integrals of any length are given recursively by applying Itô’s Lemma

Lemma 5.2.2. ([46]) If \( \alpha = (j_1, \ldots, j_l) \) and \( \beta = (j'_1, \ldots, j'_r) \) where \( j_i, j'_k \in \{0, 1, \ldots, m\} \)
then

\[ I_\alpha(t)I_\beta(t) = \int_0^t I_\alpha(s)I_\beta(s)dW^j_s + \int_0^t I_\alpha(s)I_\beta(s)dW^j_s + \int_0^t I_\alpha(s)I_\beta(s)\mathbb{1}_{\{j_t = j'_t \neq 0\}}ds \]

for \( t \geq 0 \).

Tocino shows how the above two lemmas can be implemented in Mathematica, so that an arbitrary product of iterated integrals can be calculated. We extend his work to allow non unit integrands, together with correlated Brownian increments and a time dependent prefactor. Then the above two lemmas become

\[
I[g_1, (j), (f)]I[g_2, (j_1, \ldots, j_l), (f_1, \ldots, f_l)] = \\
\sum_{i=0}^{l} I[g_1g_2, (j_1, \ldots, j_i, f), (f_1, \ldots, f, f_{i+1}, \ldots, f_l)] \\
+ \sum_{i=1}^{l} I[g_1g_2, (j_1, \ldots, j_{i-1}, 0, j_{i+1}, \ldots, j_l), (f_1, \ldots, f_{i-1}, f \cdot f_i \rho_{j, j_i}, f_{i+1}, \ldots, f_l)]
\]

where we also extend \( \rho \) to zero indices so that \( \rho_{j,0} = \rho_{0,j} \equiv 0 \), \( j \in \{0, 1, \ldots, m\} \).

\[
ito[g_1; (j_1, \ldots, j_l); (f_1, \ldots, f_l)](t)ito[g_2; (\tilde{j}_1, \ldots, \tilde{j}_r); (\tilde{f}_1, \ldots, \tilde{f}_r)](t) = \\
g_1(t)g_2(t) \int_0^t ito[1; (j_1, \ldots, j_l); (f_1, \ldots, f_l)](s)ito[1; (\tilde{j}_1, \ldots, \tilde{j}_r-1); (\tilde{f}_1, \ldots, \tilde{f}_{r-1})](s)\tilde{f}_r(s)dW^j_s \\
+ g_1(t)g_2(t) \int_0^t ito[1; (j_1, \ldots, j_{l-1}); (f_1, \ldots, f_{l-1})](s)ito[1; (\tilde{j}_1, \ldots, \tilde{j}_r); (\tilde{f}_1, \ldots, \tilde{f}_r)](s)f_l(s)dW^j_s \\
+ g_1(t)g_2(t) \int_0^t ito[1; (j_1, \ldots, j_{l-1}); (f_1, \ldots, f_{l-1})](s)\rho_{ji, j_i}(s)f_{j_i}(s)f_{j_r}(s)ds
\]

These formulas allow us to take any products of iterated Itô integrals and convert them into sums of iterated Itô integrals. Calculating the conditional expectations of these iterated integrals given \( J_{[q]} \) is then straightforward. Using the above representation for an iterated Itô integral, we replace each Itô integral, \( \int_0^t f_i (\cdots) dW^j_s \) with the Lebesgue integral \( \int_0^t f_i \sum_{l=1}^{m_i} \rho_{j_i, q_i} q_i (\cdots) dW^0 \) as specified by Proposition \[5.1.2\].
5.3 Application to DCEV Model

\begin{align*}
    dv_t &= \kappa (v_t' - v_t) dt + \epsilon \xi_1 v_t'^2 dW_t^1 \quad v_0 > 0 \\
    dv_t' &= c(z_3 - v_t') dt + \epsilon \xi_2 v_t'^\beta dW_t^2 \quad v_0' > 0 \\
    d\langle W^1, W^2 \rangle_t &= \rho dt
\end{align*}

We note we can rewrite this as

\begin{align*}
    v_t &= \exp(-\kappa t) \left( v'_0 + \kappa \int_0^t \exp(\kappa u) v_u' du + \epsilon \xi_1 \int_0^t \exp(\kappa u) v_u'^2 dW_u^1 \right) \\
    v_t' &= \exp(-c t) \left( v'_0 + cz_3 \int_0^t \exp(c u) du + \epsilon \xi_2 \int_0^t \exp(c u) v_u'^\beta dW_u^2 \right) \\
    v_0, v_0' &> 0
\end{align*}

In Mathematica we write this as follows:

\begin{verbatim}
Vp[e_, t_] := Module[{s, W1s, W2s},
    Exp[-c t]vp0 + c Exp[-c t]Integrate[Exp[c s]z3, {s, 0, t}] + \epsilon \xi_2 \exp[-c t] Integrate[Exp[c W2s]Vpt[e, W2s]^\beta, {W2s, 0, t}]
]
V[e_, t_] := Module[{s, W1s, W2s},
    Exp[-\kappa \times t]v0 + \kappa \exp[-\kappa t]Integrate[Exp[\kappa s] Vpt[e, s], {s, 0, t}]
    + \epsilon \xi_1 \exp[-\kappa t] Integrate[ Exp[\kappa Wls]Vt[e, Wls]^\alpha, {Wls, 0, t}]
]
\end{verbatim}

Although Mathematica has no representation for Itô integrals, we can use its representation for Lebesgue integrals providing the integrand is an undefined function of the integration variable in order to develop the asymptotic expansions by formally differentiating these integrals in powers of \( \epsilon \). We define functions \( \text{dnVde} \) and \( \text{dnVpde} \), which return \( \partial_{\epsilon=0}^2 v \), \( \text{dnVde}[2, t] \) and \( \partial_{\epsilon=0}^1 v' \), \( \text{dnVpde}[1, t] \).

\begin{verbatim}
In[46]:= dnVde[i_Integer, t_] := Module[{e}, D[V[e, t], {e, i}] /. e -> 0];
In[47]:= dnVpde[i_Integer, t_] := Module[{e}, D[Vp[e, t], {e, i}] /. e -> 0];
\end{verbatim}

So below we show \( \partial_{\epsilon=0}^2 v \), \( \text{dnVde}[2, t] \) and \( \partial_{\epsilon=0}^1 v' \), \( \text{dnVpde}[1, t] \).
Having got Mathematica to derive the asymptotic expansion of each \( A_k \), we define them in our iterated integral representation.

\[
\text{In[51]:= dnVpdeIto[1]=ito[\xi^2 \exp[-c#1]&,\{2\},\{\exp[c#1]Vpt[0,\#1]\}]}
\]

The second derivative of \( v_t' \) is an integral of the first

\[
\text{In[52]:= dnVpde[2, t]=2e^{-ct}\left(\int_0^t e^{\xi W_2 s^16915}Vpt[0,W2s^16915]dW2s^16915\right)\xi^2}
\]

Continuing in this way, we can derive Mathematica representations of all required asymptotic expansion of the state variables, \( A_{k,t} \).

We now consider the expansion of our smooth function,

\[
\text{In[39]:= X[0]=Coefficient[SquareRootF[e], e, 0]}
\]

We identify the terms of the rescaled variable \( Y^c = (F^c - F_0)/\epsilon \xi = Y_0 + \epsilon Y_1 + \cdots \), where \( \xi \) is the standard deviation of the zero mean Gaussian variable \( F_1 \).
We now calculate Taylor series for the call option payoff, and specify the derivatives (ie the Heaviside and Dirac Delta functions).

\[
\text{In[42]} := \text{Series}[f[y-e]/.y->Xa[1]+e \, Xa[2]+e^2 \, Xa[3],\{e,0,2\}]
\]

\[
\text{Out[42]} = f[-y+Xa[1]]+Xa[2]f'[-y+Xa[1]]e
\]

\[
+\left(\frac{1}{2}\text{Xa}[2]^2 f''[-y+Xa[1]]\right) e^2 + O[e]^3
\]

\[
\text{In[42]} :=%/.\{\text{Derivative}[1][f][0] -> (\Theta[\#1])\},
\text{Derivative}[n_][f] -> (\text{Derivative}[n-2][f][0])
\]

\[
\text{Out[42]} = f[-y+Xa[1]]+Xa[2]\Theta[-y+Xa[1]]e
\]

\[
+\left(\frac{1}{2}\text{Xa}[2]^2 \Theta[-y+Xa[1]]\right) e^2 + O[e]^3
\]

We substitute in the value for \(X\), and expand the derivatives of \(F\) of order higher than 1 into the derivatives of the state variables \(\nu'\) and \(\nu\).

\[
\text{In[56]} = \%/.\{Xa[1] -> X1, Xa -> Xm,}
\text{Derivative}[n_][F] -> \text{Derivative}[n][\{a_1 \text{VS}[\#1] + a_2 \text{VpS}[\#1] + a_3 \text{z}_3\}]
\]

\[
\text{Out[56]} = f[X1 - y] + \frac{\Theta[X1 - y](-2X1^2 + a_2 \text{VS}'[0]+a_1 \text{VS}''[0])e}{4X1 \sqrt{F[0]}}
\]

\[
+\left(\frac{\delta[X1 - y](-2X1^2 + a_2 \text{VS}'[0]+a_1 \text{VS}''[0])^2}{4X1 \sqrt{F[0]}}
\right.
\]

\[
+\frac{\Theta[X1 - y]((6X1^2 - 3X1 \delta(a_2 \text{VS}'[0]+a_1 \text{VS}''[0]) + \sqrt{F[0]}(a_2 \text{VS}'[0]+a_1 \text{VS}''[0]))}{12X1 \sqrt{F[0]}}
\]}

\)

We now select terms of order \(e^1\), and identify the particular products of \(A_k\) required.

\[
\text{In[60]} := \text{Expand}[\text{Coefficient}[\%56,e,1]]
\]

\[
\text{Out[60]} = -\frac{X1^2 x_1 \Theta[X1 - y]}{2 \sqrt{F[0]}} + \frac{a_2 X1 \text{VS}'[0] + a_1 \Theta[X1 - y] \text{VS}''[0]}{4X1 \sqrt{F[0]}}
\]

\[
\text{In[61]} := \text{Order2CoefficientList} =
\text{CoefficientRules}[\%60,\{\text{Derivative}[1][\text{VpS}][0],\text{Derivative}[1][\text{VS}][0],\text{Derivative}[2][\text{VpS}][0],\text{Derivative}[2][\text{VS}][0],\text{Derivative}[3][\text{VpS}][0],\text{Derivative}[3][\text{VS}][0]\}][[\text{All},1]]
\]

\[
\text{Out[61]} = \{\{0,0,1,0,0,0\},\{0,0,0,1,0,0\},\{0,0,0,0,0,0\}\}
\]

Similarly for order \(e^2\) we get

\[
\text{In[61]} := \text{Expand}[\text{Coefficient}[\%56,e,2]]
\]

\[
\text{Out[61]} = \frac{X1^2 x_1^2 \Theta[X1 - y]}{8F[0]} + \frac{X1^4 x_1^2 \Theta[X1 - y]}{2F[0]} - \frac{X1^2 a_2 \text{VS}'[0] + a_1 \Theta[X1 - y] \text{VS}''[0]}{4F[0]}
\]

\[
-\frac{a_2 \text{VS}'[0] + a_1 \Theta[X1 - y] \text{VS}''[0]}{32x_1^2 F[0]}
\]

\[
-\frac{a_2 \text{VS}'[0] + a_1 \Theta[X1 - y] \text{VS}''[0]}{8F[0]}
\]

\[
-\frac{a_2 \text{VS}'[0] + a_1 \Theta[X1 - y] \text{VS}''[0]}{8F[0]}
\]

\[
\text{Similarly for order } e^2 \text{ we get}
\]
After conditioning on $X_1$, we will end up with a series of expectations of the form $\mathbb{E}[\theta(X_1 - y)H_n(X_1; \Sigma)]$. But these are straightforward to evaluate using the Hermite polynomial definition \ref{eq:hermite_def} for $n \in \mathbb{Z}^+$, where $\Sigma$ is the variance of $X_1$ (in fact 1).

$$\mathbb{E}[\theta(X_1 - y)H_n(X_1; \Sigma)] = \int_{-\infty}^{\infty} (-\Sigma)^n e^{x^2/2\Sigma} \left( \frac{\partial^n}{\partial x^n} e^{-x^2/2\Sigma} \right) \frac{1}{\sqrt{2\pi \Sigma}} e^{-x^2/2\Sigma} dx = (-\Sigma)^n \left( \frac{\partial^{n-1}}{\partial x^{n-1}} e^{-y^2/2\Sigma} \right) \frac{1}{\sqrt{2\pi \Sigma}} e^{-y^2/2\Sigma} = \Sigma H_{n-1}(y; \Sigma) \frac{1}{\sqrt{2\pi \Sigma}} e^{-y^2/2\Sigma}$$

We now set about calculating these monomials of $A_k$. This is straightforward using Tocino's approach, as each $A_k$ is represented as an iterated Itô integral. We are left with computing the conditional expectations. Rather than displaying some actual results for the DCEV model, we select a few calculations from \cite{75} section 3.3 and show how these are computed in Mathematica.

$$E \left[ \int_0^T q_2(t_2) dW_{t_2} \int_0^T q_1(t_1) dW_{t_1} = x \right] = F_3(q_2 \cdot q_1, q_3 \cdot q_1, q_4 \cdot q_1) \frac{H_3(x; \Sigma)}{\Sigma^3} + F_2(q_2 \cdot q_3, q_4 \cdot q_1) \frac{H_1(x; \Sigma)}{\Sigma}$$

$$E \left[ \int_0^T \left( \int_0^T q_2(u) dW_u \right) \left( \int_0^T q_4(s) dW_s \right) q_4(t) dW_t \right] = F_3(q_2 \cdot q_1, q_3 \cdot q_2 \cdot q_1, q_4 \cdot q_1) \frac{H_3(x; \Sigma)}{\Sigma^3} + F_2(q_2 \cdot q_3, q_4 \cdot q_1) \frac{H_1(x; \Sigma)}{\Sigma}$$
5.4 Results

In the figures below we investigate the performance of the Yoshida-Watanabe expansions for calculating the VIX option price and implied volatility. Chenxu Li \cite{52} has investigated a small time Yoshida-Watanabe expansion. Written as a small noise problem this corresponds to

\begin{align*}
\frac{dv_t}{t} &= \epsilon^2 \kappa (v'_t - v_t) dt + \epsilon \xi_1 v_t W_t^1 \quad v_0 > 0 \\
\frac{dv'_t}{t} &= \epsilon^2 c (z_3 - v'_t) dt + \epsilon \xi_2 v'_t W_t^2 \\
d \langle W^1, W^2 \rangle_t &= \rho dt,
\end{align*}

so in other words the drift terms are scaled down by \( \epsilon^2 \) with respect to our small noise SDE family. Although his calculation methodology is different from the standard calculations suggested by Takahashi, the resulting expansion is the same. We plot convergence of the option price as \( t \to 0 \) in Figure 5.4.1. As discussed at the beginning of the chapter, since our limit is not for a fixed strike, so much as a fixed standard deviation, that is what we plot against. We have chosen 0.25 standard deviations of the Gaussian random variable underlying the \( O(\epsilon^1) \) result. The call price to expansion order \( O(\epsilon^n) \) is labelled \( O_n \) in the graph. The strike started at around 12.5% at the smallest maturity and reached 14.4% at 6 months. We saw similar results for other strikes. We show the results of his expansion (using his code) below for orders 1 (Gaussian) to 4. As can be seen, the \( O(\epsilon^2) \) term causes the expansion to significantly depart from our Monte Carlo simulation. Nevertheless the \( O(\epsilon^3) \) and \( O(\epsilon^4) \) results improve performance again. This suggests that the expansion has been calculated correctly. We used our own methodology to recompute the \( O(\epsilon^1) \) and \( O(\epsilon^2) \) results and confirmed his calculations. Given that the terms are not getting smaller, one would be wary of using the higher order terms and the simple Gaussian approximation seems best.

We next plot the Yoshida-Watanabe small noise expansion we calculated using our Mathematica routines (Figure 5.4.2). The two graphs are not directly comparable. Firstly we have fixed maturity to six months, and it is \( \epsilon \) that is increasing. Given that we start with a different Gaussian expansion, our strike is different, it is however close (going from 15% at \( \epsilon^2 = 0.001 \) to 15.5% at \( \epsilon^2 = 1 \)). We have plotted against \( \epsilon^2 \) to make the results comparable to the small time results. As can be seen, the expansion is much better behaved – the \( O(\epsilon^2) \) term actually improves convergence, and not coincidentally, the expansion terms are much smaller in the small noise case.

Finally we consider the full smile for maturities from half a month to 6 months.
5.4. Results

Figure 5.4.1: Yoshida-Watanabe small time price convergence

Figure 5.4.2: Yoshida-Watanabe small noise price convergence
We computed the VIX implied volatility (i.e., by Newton-Raphson root search) from our Monte Carlo call option price simulation, as well as the various Yoshida Watanabe expansions. The circles indicate the 5% probability percentiles calculated by numerically differentiating the Monte Carlo call option prices. These provide a range for the analysis - small maturities clearly have a smaller spread than larger maturities, correspondingly the relevant range of liquidly traded option strikes varies too. Practitioners use the concept of “Delta” to refer to a similar concept. The forward or mean value is also plotted with a cross. The Watanabe expansions are closed form approximations, and here the points where implied volatility is not plotted correspond to the option price being non positive or off the scale. The Large Deviations result corresponds to our numerical minimisation of the energy as described in Chapter 4, Subsection 4.4.1 - here the implied volatility is directly calculated from the rate function value. We note the peculiar ‘peak’ in the Large Deviations result for 6 months maturity. This is explained by the energy going to zero as the strike approaches the forward as mentioned in Remark 3.1.7, so care must be taken in determining the smile directly around the forward. In general, the $O(\epsilon^3)$ small noise gives the best results in the 30-80% probability range, whereas the small noise large deviations results are (much) better outside this range. The small time $O(\epsilon^4)$ is not competitive with either the large deviations or Yoshida-Watanabe $O(\epsilon^3)$ small noise expansions and becomes negative for larger maturities. Despite requiring $> 100,000$ lines of C++ code, our small noise $O(\epsilon^3)$ valuation took 10ms to calculate a single option price, the much smaller large deviations energy minimisation program took 1ms to calculate the implied volatility. To provide a comparison, a 2D PDE numerical solver using the NAG routine D03RAF ([1]) to value a single European option on the VIX accurately took around 15 seconds. We see that both the small noise expansions provide significant accuracy for a fraction of the time required to evaluate a full finite difference numerical method. One direction for future research would be to see how these two expansions could best be combined. We note that which method performs best depends on the particular application: if one is developing a smile model valid for all strikes then the large deviations approach seems more suitable, whereas if one needs the approximation to calibrate to a few liquidly traded strikes in the 30% to 80% probability range, then the small noise Watanabe expansion should be more appropriate.

---

3The relevant forward here is the $\epsilon = 0$ forward, rather than the $\epsilon = 1$ forward marked with a cross
Figure 5.4.3: Implied Volatility Smile for DCEV VIX Option: Watanabe Small Time, Small Noise and Large deviations. Circles are 5%, 10%, ..., 95% probability points. Cross is the forward or mean value.
Chapter 6

Marginal density expansions and Large Strike Expansions of the Stein Stein model

Given a multi-dimensional hypoelliptic diffusion process \( X_t = (X^1_t, \ldots, X^d_t : t \geq 0) \), started at \( X_0 = x_0 \), we are interested in the behaviour of the probability density function \( f = f(y, t) \) of the projected (in general non-Markovian) process

\[
Y_t := \Pi_l X_t := \left( X^1_t, \ldots, X^l_t \right), \quad l \in \{1, 2, \ldots, d\}.
\]

Both short time asymptotics and tail asymptotics, in presence of some scaling, can be derived from the small noise problem

\[
dX^l_t = b(\epsilon, X^l_t) \, dt + \epsilon \sigma(X^l_t) \, dW_t, \quad \text{with } X^l_0 = x^l_0 \in \mathbb{R}^d
\]

where we note that the initial condition \( x^l_0 \) also depends on \( \epsilon \). Using the Laplace method on Wiener space as developed by Ben Arous [11, 6], allows us to derive a density expansion for \( Y^l_t := \Pi_l \circ X^l_t \) of the form

\[
f^\epsilon(y, T) = e^{-c_1/\epsilon^2} e^{c_2/\epsilon} e^{-l(c_0 + O(\epsilon))} \text{ as } \epsilon \downarrow 0 \text{ for } y, T \text{ fixed and } x^l_0 \rightarrow x_0. \quad (6.0.2)
\]

Following the original idea of Bismut [17], Ben Arous’ method can be viewed as an extension of the Laplace method in finite dimensions (see eg. [25]). Consider the Laplace method applied to the limit of the following integral

\[
I(\epsilon) = \left(2\pi \epsilon^2 \right)^{-n/2} \int_{\{z \in \mathbb{R}^n : F(z) = a\}} e^{-\frac{1}{2\epsilon} \|z\|^2} \, dz \text{ as } \epsilon \downarrow 0.
\]
Then the following conditions are typical assumptions at least for the most straightforward applications of the method.

1. There must be finitely many minima of \( \|z\|^2 \) in the constraint set \( \mathcal{K} = \{ z \in \mathbb{R}^n : F(z) = a \} \).

2. The constraint set \( \mathcal{K} \) must have a differentiable manifold structure around each minimum.

3. Each minimum must be nondegenerate, in the sense that at each minimum the second derivative of \( \|z\|^2 \) restricted to \( \mathcal{K} \) is positive definite.

As Bismut [17] noted, the same conditions apply to the infinite dimensional case of solutions of SDEs viewed as functionals of Brownian motion. Bismut considered a small time point to point problem under strictly hypoelliptic conditions. He analysed the asymptotics of the probability density of solutions of an SDE at a fixed final time \( T \) with drift coefficients of \( O(\epsilon^2) \) and dispersion coefficients of order \( O(\epsilon) \) as \( \epsilon \downarrow 0 \). The large deviations theory of Wentzell-Freidlin applies (as in Chapter 4) and we are led to work on the Cameron-Martin space \( \mathcal{H}^m \), and the constraint subset consisting of those elements of \( \mathcal{H}^m \) solving the corresponding controlled ODE and reaching a given endpoint at time \( T \). We will see that as for the finite dimensional case, the existence of a constraint surface (locally around a minimum) can be shown using the implicit function theorem, assuming the continuous differentiability of \( F \) (in \( \mathcal{H}^m \)) and that its differential \( DF \) at the minimum is of maximum rank ([17]). Bismut showed that the minimising paths in \( \mathcal{H}^m \) could be identified using the (finite dimensional) Hamilton’s equations ([5]) describing the minimal energy paths under this same assumption. He further identified a condition for the minimum to be nondegenerate in terms of a rank condition on the Jacobian associated to the Hamiltonian system.

In this chapter we consider the asymptotics of the marginal probability density of solutions of an SDE at a fixed time \( T \) with drift coefficients of order \( O(1) \) and dispersion coefficients of order \( O(\epsilon) \), where also the initial condition depends on \( \epsilon \). We are therefore led to adapt his analysis and that of Ben Arous ([11, 6]) under these different assumptions. Using this expansion we are able to determine the large strike asymptotics for the Stein-Stein model [73] stochastic volatility model as well as a correlated extension, the Schöbel and Zhu model [69]. Since the expansion depends on the minimal energy (and its derivative) to reach a given point, we are led to solve Hamilton’s equations for the model. This is joint work; those sections in which I was not significantly involved have been moved to the appendix.
Consider a $d$-dimensional diffusion $(X_t^\epsilon)_{t \geq 0}$ given by the stochastic differential equation

\[ dX_t^\epsilon = b(\epsilon, X_t^\epsilon) \, dt + \epsilon \sigma(X_t^\epsilon) \, dW_t, \quad \text{with} \quad X_0^\epsilon = x_0^\epsilon \in \mathbb{R}^d \]  

(6.1.1)

and where $W = (W^1, \ldots, W^m)$ is an $m$-dimensional Brownian motion. Unless otherwise stated, we assume $\sigma = (\sigma_1, \ldots, \sigma_m) : \mathbb{R}^d \to L(\mathbb{R}^m, \mathbb{R}^d)$ to be smooth and bounded with bounded derivatives of all orders, similarly we assume the same for $b(\epsilon, \cdot)$ for $\epsilon \in [0,1)$ with $b : [0,1) \times \mathbb{R}^d \to \mathbb{R}^d$. Set $V_0 = b(0, \cdot)$ and $V = (V_1, \ldots, V_m)$ with $V_k = \sigma_k$, $k \in \{1,2,\ldots,m\}$, and define $\Sigma$, the $d \times d$ positive semidefinite diffusion matrix with elements

\[ \Sigma^{i,j} = \sum_{k=1}^m V_k^i V_k^j. \]  

(6.1.2)

Assume that, for every multi-index $\alpha$, the drift vector field $b(\epsilon, \cdot)$ converges to $V_0$ in the following sense

\[ \partial^\alpha_x b(\epsilon, \cdot) \to \partial^\alpha_x b(0, \cdot) = \partial^\alpha_x V_0(\cdot) \]  

uniformly on compacts as $\epsilon \downarrow 0$.  

(6.1.3)

We shall also assume that

\[ \partial_\epsilon b(\epsilon, \cdot) \to \partial_\epsilon b(0, \cdot) \]  

uniformly on compacts as $\epsilon \downarrow 0$.  

(6.1.4)

We draw attention to the fact that the initial condition $x_0^\epsilon$ is a function of $\epsilon$, $x_0 : [0,1) \to \mathbb{R}^d$. We assume that $x_0^\epsilon$ has an asymptotic expansion

\[ x_0^\epsilon = x_0 + \epsilon x_0' + o(\epsilon) \]  

as $\epsilon \downarrow 0$.  

(6.1.5)

To ensure the existence of a density for $X^\epsilon$ and therefore also $Y^\epsilon$, we assume that the Hörmander condition \[2.2.3\] holds at $x_0$, that is the linear span of $V_1, \ldots, V_m$ and all Lie brackets of $V_0, V_1, \ldots, V_m$ is full. Since this condition is “open” it also holds, thanks to (6.1.3), for $\epsilon > 0$ small enough, with $V_0$ and $x_0$ replaced by $b(\epsilon, \cdot)$ (or $\tilde{b}(\epsilon, \cdot)$, cf. previous footnote) and $x_0^\epsilon$, respectively. It then is a classical result (due to Hörmander, Malliavin) that the $\mathbb{R}^d$-valued random variable $X_T^\epsilon$ admits a (smooth) density for all times $T > 0$ (see Theorem \[2.2.8\]) and so does its $\mathbb{R}^l$-valued projection $Y_T^\epsilon$. We denote

---

1 If (6.1.1) is understood in Stratonovich sense, so that $dW$ is replaced by $\circ dW$, the drift vector field $b(\epsilon, \cdot)$ is changed to $\tilde{b}(\epsilon, \cdot) = b(\epsilon, \cdot) - (\epsilon^2/2) \sum_{i=1}^d V_i \cdot \partial_\epsilon V_i$. In particular, $V_0$ is also the limit of $\tilde{b}(\epsilon, \cdot)$ in the sense of (6.1.3).
the probability density of \( Y_T^\varepsilon \) by

\[
f^\varepsilon (\cdot, T) \equiv f^\varepsilon (y, T) \quad \text{with} \quad y \in \mathbb{R}^d.
\]

We also will make use of the strong Hörmander condition at every \( x \in \mathbb{R}^d \)

\[
\forall x \in \mathbb{R}^d : \text{Lie} [V_1, \ldots, V_m] |_x = \mathbb{R}^d. \tag{H1}
\]

see in particular Theorem 6.2.4.

For \( h \in \mathcal{H}^m \) let \( \phi^h_T \) denote the time-\( T \) solution to the controlled ordinary differential equation

\[
d\phi^h_t = V_0 (\phi^h_t) \, dt + \sum_{j=1}^m V_j (\phi^h_t) \, dH^j_t, \quad \phi^h_0 = x_0 \in \mathbb{R}^d. \tag{6.1.6}
\]

We will call elements of \( \mathcal{H}^m \), controls, and the solutions of the controlled ODE paths. Given a \( \mathbb{R}^n \) valued functional on \( \mathcal{H}^m \), \( F : \mathcal{H}^m \to \mathbb{R}^n \), we call its (\( \mathcal{H}^m \)-valued) Fréchet derivative the deterministic Malliavin derivative, which satisfies

\[
DF(h)[k] = \lim_{\tau \to 0} \frac{F(h + \tau k) - F(h)}{\tau} \quad k \in \mathcal{H}^m.
\]

Denote the subset of \( \mathcal{H}^m \) that satisfies the constraint \( \Pi_l \phi^h_T = a \) by

\[
\mathcal{K}_{x_0, T; a} = \mathcal{K}_a := \{ h \in \mathcal{H}^m : \Pi_l \phi^h_T = a \} \tag{6.1.7}
\]

We recall (2.1.1) the energy

\[
I(h) = \begin{cases} 
\frac{1}{2} \| h \|_H^2 & \text{if} \ h \in \mathcal{H}^m \\
+\infty & \text{otherwise},
\end{cases}
\]

defined for \( h \in \mathcal{C}^m \), the large deviations rate function for Brownian motion. We define

\[
\Lambda (a) = \inf \left\{ \frac{1}{2} \| h \|_H^2 : h \in \mathcal{K}_a \right\}, \tag{6.1.8}
\]

which is the rate function for \( \{ Y_T^\varepsilon \} \). We will always use the convention that the infimum of an empty set is \( +\infty \). We also define the set

\[
\mathcal{K}_a^{\text{min}} := \{ h \in \mathcal{K}_a : I(h) = \Lambda (a) \},
\]

**Assumption 6.1.1.** Assume
6.1. The main result and its corollaries

a) \( b(\epsilon, \cdot) \rightarrow V_0(\cdot) \) in the sense of (6.1.3), (6.1.4).

b) \( X_0^\epsilon \equiv x_0^\epsilon \rightarrow x_0 \) as \( \epsilon \rightarrow 0 \) in the sense of (6.1.5).

c) The weak Hörmander condition (2.2.3) at \( x_0 \in \mathbb{R}^d \).

Assumption 6.1.2. Fix \( y \in \mathbb{R}^l \), then we assume

a) The number of minimising controls, \( \#K_{y}^{\min} \), is finite and non-zero.

b) At each \( h_0 \in K_{y}^{\min} \), the deterministic Malliavin derivative of \( h \rightarrow \phi^h_T \), \( D\phi^h_T \) has maximal rank \( d \).

c) Each \( h_0 \in K_{y}^{\min} \) is a non-degenerate minimum of the energy \( I := \frac{1}{2} \| \cdot \|_H^2 \) restricted to the Hilbert manifold \( K_y \); i.e.

\[
I''(h_0)[k,k] > 0 \quad \forall 0 \neq k \in T_{h_0}K_y,
\]

where \( T_{h_0}K_y \) is the tangent space of the constraint manifold \( K_y \) at \( h_0 \).

d) The energy to reach \( y \), \( \Lambda(y) \), is smooth in a neighbourhood of \( y \).

Theorem 6.1.3. (Small noise) Fix \( x_0, y \) and \( T > 0 \) and assume 6.1.1 and 6.1.2 hold. Then there exists \( c_0 = c_0(x_0,y,T) > 0 \) such that

\[
Y_\epsilon = \Pi_l X_\epsilon = \left( X_\epsilon^{\epsilon,1}, \ldots, X_\epsilon^{\epsilon,l} \right), \quad 1 \leq l \leq d
\]

admits a density with expansion

\[
f^\epsilon(y,T) = e^{-\frac{\Lambda(y)}{\epsilon^2}} e^{\max\left\{ \frac{\Lambda'(y) \cdot Y_T(h_0) - \epsilon h_0 \in K_y^{\min}}{\epsilon} \right\}} \epsilon^{l}(c_0 + O(\epsilon)) \quad \text{as} \quad \epsilon \downarrow 0.
\]

Here \( \hat{Y} = \hat{Y}(h_0) = \left( \hat{Y}_1, \ldots, \hat{Y}_l \right) \) is the projection, \( \hat{Y} = \Pi_l \hat{X} \), of the solution to the following (ordinary) differential equation

\[
\begin{align*}
d\hat{X}_t &= \left( \partial_x b \left( 0, \phi^h_t \right) + \partial_x V(\phi^h_t) \hat{h}_0(t) \right) \hat{X}_t dt \\ + \partial_x b \left( 0, \phi^h_t \right) dt,
\end{align*}
\]

(6.1.9)

\[
\hat{X}_0 = \partial_x x_0^\epsilon|_{\epsilon=0}.
\]

Remark 6.1.4. We outline the proof in the Appendix, Section A.5. Essentially it is an adaptation of Ben Arous’s proof for the point-to-point problem (\( l = d \)). The key insight is that since Ben Arous identifies the transition probability density using the Fourier
inverse of its characteristic function, the marginal density can be derived directly using

\[ E \left[ \exp \left( i \xi \cdot Y_T^\epsilon \right) \right] = E \left[ \exp \left( i (\xi, 0) \cdot X_T^\epsilon \right) \right] \]

where we write \((\xi, 0) = (\xi^1, \ldots, \xi^l, 0, \ldots, 0) \in \mathbb{R}^d\). In other words, it suffices to restrict the characteristic function of \(X_T^\epsilon\), the full (Markovian) process evaluated at time \(T\), to obtain the characteristic function of \(Y_T^\epsilon\). The density is then obtained by Fourier-inversion (as for the full process). The differences to the setting of [11], aside from (i) allowing for \(l < d\), is that (ii) our drift-term does not vanish of order \(\epsilon^2\) and (iii) that the starting point is allowed to depend on \(\epsilon\). In fact, (ii), (iii) are responsible for the additional exponential \(\exp\{\ldots\}/\epsilon\) factor in our expansion. The non-vanishing drift plays little influence and is covered by Ben Arous’ more abstract paper ([6]). In [11] the asymptotic expansion of the transition density \(p_t(x, y)\) is uniform on compacts in \(\mathbb{R}^d \times \mathbb{R}^d\) satisfying the assumptions, so (iii) is also covered. Therefore we see that perhaps the key issue is to identify when the assumptions 6.1.2 are satisfied for our \(l < d\) case, and identify the energy and minimising controls. We will address these issues in Sections 6.2 and A.4. In particular we develop a generalisation to the \(l < d\) case of Bismut’s finite dimensional test for non-degeneracy of the minimisers in terms of the Hamilton’s equations.

When applied to small time expansions, the weak Hörmander condition in the above theorem automatically becomes the strong Hörmander condition at \(x_0\); indeed, the “drift” vector field in the weak condition will be the limit of \(\epsilon^2\) times the original drift vector fields; plainly this is zero and therefore does not figure in the span.

### 6.1.1 Corollary on tail expansions

We have the following application to tail behaviour of, say, the first component (i.e. \(l = 1\) here) of a diffusion processes at a fixed time \(T\). The scaling assumption below is met in a number of stochastic volatility models.

**Corollary 6.1.5. (Tail behaviour)** Assume \(x_0^\epsilon \to 0 \in \mathbb{R}^d\) as \(\epsilon \to 0\) and some diffusion process \(X^\epsilon\), started at \(x_0^\epsilon\), satisfies the assumptions of Theorem [6.1.3] with \(x_0 = 0\) and \(N_1 = (1, \cdot) \subset \mathbb{R} \times \mathbb{R}^{d-1}\); in particular, \(\{0\} \times (1, \cdot)\) is assumed to satisfy condition (ND). Assume also \(\theta\)-scaling by which we mean the scaling relation

\[ Y_T^\epsilon \overset{\text{law}}{=} \epsilon^\theta Y_T \text{ where } Y \equiv \Pi_1 X \]

for some \(\theta \geq 1\). Then the probability density function of \(Y_T\) has the expansion

\[ f(y) = e^{-c_1 y^\theta} e^{c_2 y^{\frac{1}{\theta}} - 1} \left( \alpha_0 + O \left( 1/y^{1/\theta} \right) \right) \text{ as } y \to \infty \]

(6.1.10)
where \( \alpha_0 \) is a function that does not depend on \( y \) and

\[
\begin{align*}
  c_1 & = \Lambda(1) \\
  c_2 & = \hat{Y}_T \Lambda'(1) = \frac{2\hat{Y}_T}{\theta} \Lambda(1)
\end{align*}
\]

In particular, when \( \theta = 1 \) we have a Gaussian tail behaviour of the precise form

\[
f(y) = e^{-\Lambda(1)y^2} e^{2\hat{Y}_T \Lambda(1)y} (c_0 + O(1/y));
\]

while \( \theta = 2 \) leads to the exponential tail of the precise form

\[
f(y) = e^{-\Lambda(1)y} e^{\hat{Y}_T \Lambda(1)\sqrt{y}y^{-1/2}} (c_0 + O(1/\sqrt{y})).
\]

**Proof.** Let \( f^\epsilon \) denote the density of \( Y^\epsilon_T \). Since \( f(y/\epsilon^{\theta}) = \epsilon^{\theta} f(y) \) we can take \( y = 1 \) and \( \epsilon^{\theta} = y^{-1} \) in the theorem above. Another observation is that the assumed scaling implies

\[
\Lambda(y) = y^{2/\theta} \Lambda(1)
\]

and hence \( \Lambda'(1) = \frac{2}{\theta} \Lambda(1) \).

### 6.1.2 Corollary on short time expansions

Finally, we have the following application to short time asymptotics. Note that for \( l < d \), the projection of \( X \) is non-Markovian and there is no Fokker-Planck equation that describes the evolution of \( f \). In particular, there is no direct PDE approach that leads to the expansion below.

**Corollary 6.1.6. (Short time)** Consider \( dX_t = b(X_t) dt + V(X_t) dW \), started at \( X_0 = x_0 \in \mathbb{R}^d \), with \( C^\infty \)-bounded vector fields such that the strong Hörmander condition holds for all \( x \in \mathbb{R}^d \):

\[
\forall x \in \mathbb{R}^d : \text{Lie} [V_1, \ldots, V_m] | x = \mathcal{T}_x \mathbb{R}^d.
\]

(H1)

For fixed \( l \in \{1, \ldots, d\} \) assume \( \{x_0\} \times N_y \), where \( N_y := (y, \cdot) \) for some \( y \in \mathbb{R}^l \), satisfies condition (ND). Let \( f(t, \cdot) = f(t, y) \) be the density of \( Y_t = (X_{t_1}, \ldots, X_{t_l}) \). Then

\[
f(t, y) \sim (\text{const}) \frac{1}{t^{d/2}} \exp \left( -\frac{d^2(x_0, y)}{2t} \right) \text{ as } t \downarrow 0
\]

where \( d(x_0, y) \) is the sub-Riemannian distance\(^2\) based on \( (V_1, \ldots, V_m) \), from the point \( x_0 \) to the affine subspace \( N_y \).

\(^2\)see e.g. [60] for its definition
Proof. After Brownian scaling, we apply the theorem with $T = 1$, $\epsilon^2 = t$ so that

$$b(\epsilon, \cdot) = \epsilon^2 b(\cdot) \to V_0(\cdot) \equiv 0;$$

which explains why there is no drift vector field in the present Hörmander condition $H1$. Also $x'_0 = x_0$ here. The identification of the energy with $1/2$ times the square of the sub-Riemannian (or: control -, Carnot-Caratheodory -) distance from $x$ to $N_y$ is classical. At last, the unique ODE solution to (6.1.9) is then given by $\hat{Y} \equiv 0$ and there is no $\exp \{ (...) / \epsilon \}$ factor.

6.2 Energy Minimising Paths and Hamiltonian Flows

We now analyse the minimal energy paths following Bismut’s pioneering work [17]. As noted in the introduction, Bismut considered the small time (ie drift of order $O(\epsilon^2)$) point-to-point problem; Takanobu and Watanabe [76] considered the small time point-to-subspace problem; we are interested in the small noise point to subspace problem, in particular that the first $l$ coordinates achieve a given $a \in \mathbb{R}^l$. We first work with a rather more general problem, before restricting to this when considering nondegeneracy of the minimal energy paths.

For $h \in \mathcal{H}^m$ let $\phi^h_T$ denote the unique time-$T$ solution to the controlled ordinary differential equation

$$d \phi^h_t = V_0(\phi^h_t) dt + \sum_{i=1}^m V_i(\phi^h_t) \dot{h}^i(t), \quad \phi^h_0 = x_0 \in \mathbb{R}^d. \quad (6.2.1)$$

We note that $\phi^h_T$ is a diffeomorphism as a function of $x_0 \in \mathbb{R}^d$. We define $\phi^h_{T-t} := \phi^h_T \circ (\phi^h_t)^{-1}$. We follow the flow notation of Malliavin ([56], Section 9.1). We denote its differential with respect to $x_t = \phi_t^h$ by $\Phi^h_{T-t}$, which is the $d \times d$ matrix that satisfies

$$\Phi^h_{T-t} \cdot y = \frac{d}{ds} \phi^h_{T-t}(\phi_t^h + sy) \bigg|_{s=0} \forall y \in \mathbb{R}^d.$$

It solves the following matrix ODE

$$d \Phi^h_{T-t} = \partial_x V_0(\phi^h_t) \Phi^h_{T-t} dt + \partial_x V_\alpha(\phi^h_t) \Phi^h_{T-t} \dot{h}^\alpha(t) dt \quad \text{ (6.2.2)}$$

$$\Phi^h_{T-t=0} = I.$$
and its inverse solves
\[
\begin{align*}
d\Phi_{0\leftarrow t}^h &= -\Phi_{0\leftarrow t}^h \partial_x V_0(\phi_t^h)dt - \Phi_{0\leftarrow t}^h \partial_x V_\alpha(\phi_t^h)\dot{\alpha}(t)dt, \\ \Phi_{0\leftarrow 0}^h &= I.
\end{align*}
\] (6.2.3)

Our assumptions ensure that \(\phi_t^h\) has continuous Fréchet derivatives of all orders from \(\mathcal{H}^m\) into \(\mathbb{R}^d\). In fact similarly to the Malliavin derivative (2.2.2, [70] Proposition 6.6), the Fréchet derivative of \(\phi_t^h\) in direction \(k \in \mathcal{H}^m\), \(D\) (which we term the deterministic Malliavin derivative) can be expressed succinctly using \(\Phi_{t\leftarrow s}^h\).

**Theorem 6.2.1.** (Bismut, [17] Theorem 1.1) If \(h_n\) converges weakly to \(h \in \mathcal{H}^m\) then \(\phi_{tn}^h\) converges to \(\phi_t^h\) uniformly on \([0, T]\) for the topology \(C_\infty\), of uniform convergence of function and derivatives on compact subsets of \(\mathbb{R}^d\). Moreover for any \(x_0 \in \mathbb{R}^d\), \(t \in [0, T]\), the mapping \(h \to \phi_t^h\) is a \(C_\infty\) mapping from \(\mathcal{H}^m\) in \(\mathbb{R}^d\). In particular, for \(h, k \in \mathcal{H}^m\) then
\[
\langle D\phi_t^h, k \rangle_H = \int_0^t \sum_{r=1}^m \Phi_{t\leftarrow s}^h V_r(\phi_t^h)\dot{k}_r ds,
\] (6.2.4)

The representation for the derivative is just the same as for the Malliavin derivative. We note also that the representation for the Fréchet derivative does not (explicitly) depend on the drift, so that many of the subsequent derivations performed by Bismut [17] and Takanobu and Watanabe [76] carry over exactly to our case with drift.

Then the deterministic Malliavin covariance [17] \(C_t^h\) is defined (cf the [stochastic] Malliavin covariance in Theorem 2.2.6) as the \(d \times d\) positive semi-definite matrix (for fixed \(x_0 \in \mathbb{R}^d\), \(h \in \mathcal{H}^m\) and \(0 \leq t \leq T\)) with elements
\[
C_t^{hi,j} = \int_0^t \sum_{r=1}^m \left[ \Phi_{t\leftarrow s}^h V_r(\phi_t^h) \right]^i \left[ \Phi_{t\leftarrow s}^h V_r(\phi_t^h) \right]^j ds \quad i, j \in \{1, 2, \ldots, d\}.
\]

Since \(\Phi_t^h\) is invertible, \(C_t^h\) is invertible if and only if \(D\phi_t^h\) has maximum rank. We prefer to work with the maximum rank condition (following [50] [76]), as this is a more standard condition for the implicit function theorem that we require. Given a smooth function of the endpoint, \(g \in C_\infty(\mathbb{R}^d; \mathbb{R}^l)\), where \(1 \leq l \leq d\), and \(a \in \mathbb{R}^l\) we consider the constraint set
\[
K_{x_0, T;a} = K_a := \left\{ h \in \mathcal{H}^m : g(\phi_T^h) = a \right\}
\] (6.2.5)

We note that for our full expansion we restrict to \(g = \Pi_1\) ie the projection onto the
first \( l \) components to enable us to consider the Fourier transform of the density. Nevertheless, the Hamiltonian result we develop here is useful for applications that only require a large deviations result (as we did in Chapter 4) as well as for alternative expansions such as [50, 76]. Furthermore one could still consider other linear functions of the endpoints using the Fourier transform, as would be necessary for Basket options which depend on the arithmetic average of the asset endpoints.

As discussed in the introduction, in order to apply the Laplace method to a constraint set, it has to have a manifold structure around each minimising \( h \in \mathcal{K}_a \). By the implicit function theorem (see Appendix, Theorem \[A.3.1\] from [50], Theorem A.2), \( \mathcal{K}_a \) does have a smooth manifold structure around \( h \in \mathcal{K}_a \) providing \( g(\phi_T) \in C^\infty(H_m; \mathbb{R}^l) \) and \( Dg(\phi_T^h) \) has maximal rank \( l \). We then identify the tangent space of \( \mathcal{K}_a \) at such an \( h \):

\[
T_h \mathcal{K}_a \cong \ker Dg(\phi_T^h) =: H_0.
\]

We will subsequently see that the maximal rank condition on \( Dg(\phi_T^h) \) also allows us to use Hamilton’s equations to identify the minimising \( h \in \mathcal{K}_a \). The following theorem provides us with a simple criterion that \( Dg(\phi_T^h) \) has maximal rank \( l \).

**Proposition 6.2.2.** Assume \( h \in \mathcal{H}^m \), \( \partial_x g(x_T) \) has rank \( l \) and

\[
\exists t \in [0, T] : \text{span}\ [V_1, \ldots, V_m] |_{x_t} = \mathbb{R}^d
\]

where \( x_t := \phi_t^h \). Then \( Dg(\phi_T^h) \) also has rank \( l \).

**Proof.** Clearly we just need to show that \( D\phi_T^h \) has rank \( d \), so we follow the standard proof for nondegeneracy of the Malliavin covariance and consider the quadratic form \( Q(p) : \mathbb{R}^d \to \mathbb{R} \) given by \( Q(p) = \| p D\phi_T^h \|^2_H \), then

\[
Q(p) = \int_0^T \sum_{j=1}^m \left( p \Phi_T^{h_s} V_j(x_s) \right)^2 ds = \int_0^T \left< p \Phi_T^{h_s} V(x_s) V^T(x_s) , p \Phi_T^{h_s} \right> ds.
\]

Assume \( \exists p \neq 0 : Q(p) = 0 \). By assumption \( \text{span}[V_1, \ldots, V_m] |_{x_t} = \mathbb{R}^d \) for some \( t \in [0, T] \), and this clearly remains valid in a small enough open interval containing \( t \) which is enough to conclude \( p\Phi_T^{h_s} \equiv 0 \). By the invertibility of \( \Phi_T^{h_s} \), this implies \( p = 0 \) and so \( D\phi_T^h \) has rank \( d \), as claimed. \( \square \)

\(^4\)A sufficient condition for the invertibility of \( C_T^h \) for every \( h \neq 0 \) in a strictly sub-elliptic setting is given as condition (H2) by [17], although much stronger than Hörmander’s condition, it does apply to examples such as the 3-dimensional Heisenberg group.
We recall (2.1.1)
\[ I(h) = \begin{cases} \frac{1}{2} \|h\|_H^2 & \text{if } h \in \mathcal{H}^m \\ +\infty & \text{otherwise.} \end{cases} \]
defined for \( h \in \mathcal{C}_0^m \). We will define the infimum over an empty set to be \(+\infty\).

**Definition 6.2.3.** The function \( \Lambda(a) \) on \( \mathbb{R}^l \) is defined by
\[ \Lambda(a) := \inf \{ I(h), h \in K_a \} \tag{6.2.6} \]
In words, \( \Lambda(a) \) is the minimal energy required to reach the target submanifold \( N_a := \{ x \in \mathbb{R}^d : g(x) = a \} \).

at time \( T \) starting from \( x_0 \in \mathbb{R}^d \) at time 0. We also define the (possibly empty set) \( K_{a}^{\text{min}} \)
\[ K_{a}^{\text{min}} := \{ h \in K_a : I(h) = \Lambda(a) \}, \]
the set of minimizers or minimizing controls.

**Theorem 6.2.4** (cf. [17] Theorem 1.14). \( \Lambda : \mathbb{R}^l \to [0, \infty] \) is a good rate function: its level sets are compact. For every \( a \in \mathbb{R}^l \) such that \( K_a \neq \emptyset \), there exists \( h \in K_a \) such that \( \Lambda(a) = I(h) \).

If \( V_1 \ldots V_m \) verify the strong Hörmander condition \( [H] \) at every \( x \in \mathbb{R}^d \), then for all \( a \) in the image of \( g \), \( K_a \) is non empty. If this holds and \( g(x) \) has maximal rank \( l \) for all \( x \in \mathbb{R}^d \), then the function \( \Lambda \) is finite and continuous on \( \mathbb{R}^l \).

**Proof:** Since the level sets of \( I(h) \) are weakly compact, and \( h \to g(\phi_T^h) \) is weakly continuous, then Lemma 1.3 in [8], Chapter III, shows that \( \Lambda \) is a good rate function. Similarly, since \( I \) is weakly lower semi continuous, \( g(\phi_T^h) = a \) is weakly closed in \( \mathcal{H}^m \), and \( \{ h : I(h) \leq \Lambda(a) + 1 \} \) is weakly compact, there exists an \( h \in \mathcal{H}^m \) such that \( \Lambda(a) = I(h) \). In the driftless case, Bismut shows that \( \tilde{K}_x := \{ h \in \mathcal{H}^m : \phi_T^h = x \} \) is nonempty for all \( x \in \mathbb{R}^d \) under the strong Hörmander condition. Since we assume the drift is bounded, we can always find a control to overcome it. So our result is a natural consequence\(^5\). Let \( E(x) = \inf \{ I(h) : h \in \mathcal{H}^m, \phi_T^h = x \} \). Theorem 1.14 in [17] shows the continuity of \( E(\cdot) \), assuming the strong Hörmander condition at every \( x \in \mathbb{R}^d \). So using Bismut’s theorem for the continuity of \( E(x) \) we can show the continuity of \( \Lambda(a) \). For any \( a \in \mathbb{R}^l \), choose \( x \) such that \( a = g(x) \) and \( \Lambda(a) = E(x) \), and

\(^5\)We will henceforth always assume \( a \) is in the image of \( f \).
consider a sequence $a^{(n)} \to a$. Then by the inverse function theorem, there exists an $x^{(n)} : g(x^{(n)}) = a^{(n)}$, and $x^{(n)} \to x$. We have $\Lambda(a^{(n)}) \leq E(x^{(n)})$, then by the continuity of $E(\cdot)$, we have $\limsup_{n \to \infty} \Lambda(a^{(n)}) \leq E(x) = \Lambda(a)$. We have shown $\Lambda$ is upper and lower semicontinuous, therefore it is continuous. \hfill \Box

We would like to identify these minimum energy paths for the given functional. If the functional’s value can be written as a smooth function of the endpoints (possibly in an augmented state space), then we can hope to solve the problem using Hamilton’s equations (see [5] for definitions associated with Hamiltonian and Lagrangian mechanics). This includes the problem of finding the minimum energy path such that $\Pi_t x_T = y$ but we can also consider other problems such as where the functional involves a time integral of the path by introducing an extra state variable corresponding to the running integral. If our diffusion coefficients were elliptic, then we could justify the use of Hamilton’s equations by the Legendre transformation of the corresponding Lagrangian function $L : \mathcal{C}^d \times [0, T] \to \mathbb{R}$

$$L(\phi, t) = (\dot{\phi}(t) - V_0(\phi(t)))^T \Sigma(\phi(t))^{-1}(\dot{\phi} - V_0(\phi(t))).$$

where $\Sigma$ is the diffusion matrix corresponding to $\{V_k\}_{k=1}^m$. Bismut showed a sufficient condition to use Hamilton’s equation also in the non-elliptic case. We introduce the Hamiltonian

$$\mathcal{H}(x, p) := \frac{1}{2} \sum_{k=1}^m \langle p, V_k(x) \rangle^2 + \langle p, V_0(x) \rangle$$

and $H_{t=0}(x_0, p_0)$ as the flow associated to the vector field $(\partial_p \mathcal{H}, -\partial_x \mathcal{H})$ on the cotangent bundle $T^*\mathbb{R}^d$. Then $(x_t, p_t) := H_{t=0}(x_0, p_0)$ solves the Hamiltonian ODEs in $T^*\mathbb{R}^d \simeq \mathbb{R}^d \oplus \mathbb{R}^d$

$$\begin{pmatrix} \dot{x}_t \\ \dot{p}_t \end{pmatrix} = \begin{pmatrix} \partial_p \mathcal{H}(x_t, p_t) \\ -\partial_x \mathcal{H}(x_t, p_t) \end{pmatrix}, \quad (6.2.7)$$

$$= \begin{pmatrix} \sum_{j=1}^m \langle p_t, V_j(x_t) \rangle V_j(x_t) + V_0(x_t) \\ -\sum_{j=1}^m \langle p_t, V_j(x_t) \rangle \partial_x V_j(x_t)^* p_t - \partial_x V_0(x_t)^* p_t \end{pmatrix}, \quad (6.2.8)$$

where we remember that $\partial_x V_r(x) = (\partial_{x_1} V_r(x), \ldots, \partial_{x_d} V_r(x))$, $r \in \{0, 1, \ldots, m\}$. Using Einstein summation over repeated indices, we have the component form

$$\begin{pmatrix} \dot{x}_t^i \\ \dot{p}_t^i \end{pmatrix} = \begin{pmatrix} p_t^i V_j^i(x_t) V_j^i(x_t) + V_0^i(x_t) \\ -p_t^i V_j^i(x_t) p_t^i \partial_x V_j^i(x_t) - p_t^i \partial_x V_0^i(x_t) \end{pmatrix}.$$
We define \( h_j(x,p) = \langle p, V_j(x) \rangle \), \( j = 0, 1, \ldots, m \) then taking
\[
\dot{h}_j = h_j(H_t \leftarrow 0 (x_0, p_0)), \quad j = 1, 2, \ldots, m
\]
we see (using (6.2.3)) that
\[
(\phi^h_t, \Phi^h_{0 \leftarrow t} p_0) = H_{t \leftarrow 0} (x_0, p_0), \quad 0 \leq t \leq T.
\]
(6.2.9)
(where \( \Phi^h_{0 \leftarrow t} \) is the adjoint of \( \Phi^h_{0 \leftarrow t} \). So (keeping \( x_0 \) fixed) for every \( p_0 \) we have a well defined \( h(p_0) \). We are left with the question, if for every minimising \( h_0 \in K_{a}^{\min} \) there exists a \( p_0 \), such that \( h_0 = h(p_0) \).

**Proposition 6.2.5.** If

a) \( h_0 \in K_{a}^{\min} \) is a minimizing control

b) The Fréchet derivative \( D\phi^h_T \) has maximal rank \( d \) and the Jacobian \( \partial_x g(x_T) \) has maximal rank \( l \), where \( x_T = \phi^h_{T} \),

then there exists a unique \( p_0 = p_0(h_0) \in T_{x_0}^{\ast} \mathbb{R}^d \), such that
\[
h_0 = D\phi^h_T \left[ \Phi^h_{0 \leftarrow T} p_0 \right].
\]

(6.2.10)

Furthermore
\[
(\phi^h_t, \Phi^h_{0 \leftarrow t} p_0) = H_{t \leftarrow 0} (x_0, p_0), \quad 0 \leq t \leq T
\]
\[
h_0(p_0) = h_0
\]

(6.2.11)

(6.2.12)

(\( \pi \) denotes the projection from \( T_{*}^{\ast} \mathbb{R}^d \) onto \( \mathbb{R}^d \); in coordinates \( \pi(x,p) = x \).

the minimizing control \( h_0 = h_0(\cdot) \) is recovered by
\[
\dot{h}_0 = \begin{pmatrix}
    h_1^t(x, p) \\
    \vdots \\
    h^m(x, p)
\end{pmatrix}
\]

(6.2.13)

and with \( C := \mathcal{H}(x_t, p_t) \) independent of \( t \in [0, T] \),
\[
\Lambda(a) = \frac{1}{2} \| h_0 \|^2_{\mathcal{H}} = TC - \int_0^T h^0(x_t, p_t) dt.
\]

(6.2.14)

At last, crucial for actual computations, \( (x_t, p_t) = H_{t \leftarrow 0} (x_0, p_0) \) satisfies the Hamiltonian ODEs (6.2.7) as boundary value problem, subject to the following initial -,
Energy Minimising Paths and Hamiltonian Flows

**minal - and transversality conditions,**

\[ x_0 \in \mathbb{R}^d \text{ is given} \]

\[ g(x_T) = a \in \mathbb{R}^l \]

\[ p_T = \lambda \partial_x g(x_T) \in \mathbb{R}^d. \]  \hspace{1cm} (6.2.15)

*In particular, if \( g = \Pi_l \) then*

\[ x_T = (a, \cdot) \in \mathbb{R}^l \oplus \mathbb{R}^{d-l} \]

\[ p_T = (\cdot, 0) \in \mathbb{R}^l \oplus \mathbb{R}^{d-l}. \]  \hspace{1cm} (6.2.16)

**Proof.** Recall \( T_h K_a = \ker D (g(\phi^h_t)) =: H_0. \) By our maximal rank assumption, \( h_0 \in \ker D \left( g(\phi^{h_0}_{0\leftarrow t}) \right) \perp , \) i.e. there exists a unique \( \lambda \in \mathbb{R}^l \) such that

\[ h_0 = D\phi^{h_0}_{T\leftarrow 0} [\lambda \partial_x g(\phi^{h_0}_T)]. \]

Since \( D\phi^{h_0}_{T\leftarrow 0} \) has maximal rank \( d \) we see that \( p_T := \lambda \partial_x g(\phi^{h_0}_T) \) is the unique solution to the condition \( h_0 = D\phi^{h_0}_{T\leftarrow 0} p_T, \) and so \( p_0 = \Phi^{h_0}_{0\leftarrow t} p_T \) is the unique solution of 6.2.10.

Also (using 6.2.4)

\[ \dot{h}_0(t) = V(\phi^{h_0}_T)^* \Phi^{h_0}_{0\leftarrow t} p_0 \text{ a.e. in } [0, T]. \]  \hspace{1cm} (6.2.17)

Using 6.2.17 and setting \( \tilde{p}_0 = \Phi^{h_0}_{0\leftarrow t} p_0 \) we see that

\[ \partial_t \phi^{h_0}_t (x_0) = \sum_{k=1}^{m} \left\langle \tilde{p}_t, V_k(\phi^{h_0}_t) \right\rangle V_k(\phi^{h_0}_t) + V_0(\phi^{h_0}_t) \]

\[ \partial_t \tilde{p}_t = - \sum_{k=1}^{m} \left\langle \tilde{p}_t, V_k(\phi^{h_0}_t) \right\rangle \partial_x V_k^*(\phi^{h_0}_t) \tilde{p}_t - \partial_x V_0^*(\phi^{h_0}_t) \tilde{p}_t \]

so we see that \( (\phi^{h_0}_t, \Phi^{h_0}_{0\leftarrow t} p_0) \) solves the Hamiltonian system 6.2.7 with initial conditions \((x_0, p_0) : \)

\[ (\phi^{h_0}_t, \Phi^{h_0}_{0\leftarrow t} p_0) = H_{t\leftarrow 0} (x_0, p_0), \ 0 \leq t \leq T \]

and \( h(p_0) = h_0. \)

**Remark 6.2.6.** Existence and uniqueness for Boundary Value problems for ODEs are much more delicate than the corresponding initial value problems. Nevertheless, the nature of our problems simplifies the analysis considerably. Namely, we know that a minimising control exists as soon as \( K_a \) is nonempty, and by the above this implies that a solution to the Hamiltonian boundary value problem exists. Furthermore, we have uniqueness up to the path taken - there may be multiple paths of minimal
energy that achieve a given target, but each such path has a unique $p_0$ which determines the minimising control, $h$. We will see an instance of this in the case of the Stein-Stein model in proposition [6.3.8](#).

In this section we have shown that the minimal energy paths achieving a given smooth function of the end points are the solutions of Hamiltonian ODEs. This allows us to investigate Hamiltonian methods to identify the zero order small noise or small time implied volatility expansion. In particular we plan in future to consider options on VIX (as investigated in Chapter 4) as well as Asian options which require an additional state variable to represent the time integral of the asset, leading to a hypoelliptic problem in the extended state-space. In all subsequent sections we restrict the function $g$ to a linear projection onto the first $l$ coordinates to allow us to use Ben Arous’s asymptotic expansion of the density.

**Proposition 6.2.7.** Under the assumptions of the proposition [6.2.5](#) in particular $h_0 \in K_{a_{min}}^{0}$ with associated $p_0 = p_0(h_0) \in T_{x_0}^* \mathbb{R}^d$, the following are equivalent:

(iii) $h_0 \in K_a$ is a non-degenerate minimum of the restriction of the energy $I := \frac{1}{2} \| \cdot \|_H^2$ to the Hilbert manifold $K_a$; i.e.

$$I''(h_0) [k, k] > 0 \quad \forall 0 \neq k \in H_0 \cong T_{h_0} K_a$$

(iii') $x_0$ is non-focal for $N_a = (a, \cdot)$ along $h_0$ in the sense that, with

$$(x_T, p_T) := H_{T \leftarrow 0}(x_0, p_0(h_0)) \in T^* \mathbb{R}^d,$$

$$\partial_{(q, z)} \pi_{H_{0 \leftarrow T}}(x_T + (0, z), p_T + (q, 0))|_{(q, z) = (0, 0)}$$

is non-degenerate (as $d \times d$ matrix; here we think of $(q, z) \in \mathbb{R}^l \times \mathbb{R}^{d-l} \cong \mathbb{R}^d$ and recall that $\pi$ denotes the projection from $T^* \mathbb{R}^d$ onto $\mathbb{R}^d$; in coordinates $\pi(x, p) = x$).

**Remark.** I have put the proof in the appendix as it was not carried out by myself and is quite lengthy.

**Remark 6.2.8.** A simple example of a focal point is given by the parabola $x = y^2$. Consider moving a point $(z, 0) \in \mathbb{R}^2$ along the positive x-axis, starting in the neighborhood of the origin. Then the energy, half the square of distance to the parabola, is given by $\frac{1}{2} z^2$, since the closest point is at the origin. The energy between this point, $(z, 0)$, and a point $(w^2, w)$ on the parabola $x = y^2$, is given by $\frac{1}{2}(z - w^2)^2 + \frac{1}{2}w^2$. Taking the second derivative with respect to $w$, at $w = 0$, we see that the energy will be

---

6In other words, given the time reversed Hamiltonian flow from $t = T$ to $t = 0$, we consider the Jacobian of $x_0$ with respect to $(x_T^{l+1}, \ldots, x_T^d, p_T^l, \ldots p_T^l)$. 
degenerate, in that its second derivative is zero, at \( z = \frac{1}{2} \). This is the focal point (in our sense) of the parabola.

**Definition 6.2.9** (Condition (ND) generalized \( \notin \)-cut-locus condition). We say that \( \{x_0\} \times N_a \) where \( N_a := (a, \cdot) := \{x \in \mathbb{R}^d : \Pi_I x = a \in \mathbb{R}^l\} \) satisfies condition (ND) if

(i) \( 1 \leq \#K_{a}^{\min} < \infty \),

(ii) the deterministic Malliavin covariance matrix \( C_T^h \) is invertible, \( \forall h \in K_{a}^{\min} \);

(iii) \( x_0 \) is not focal for \( N_a \) along \( h \), for any \( h \in K_{a}^{\min} \).

**Remark 6.2.10.** When \( V_0 \equiv 0 \) and \( l = d \), i.e. \( N_a = \{a\} \), and \( \#K_{a}^{\min} = 1 \), condition (ND) says precisely that \( (x_0, a) \) is not contained in the sub-Riemannian cut-locus in the sense of Ben Arous [11]; extending the usual Riemannian meaning. In this sense our (global) condition (ND) is effectively a generalization of the well-known “\( \notin \)-cut-locus” condition in the context of heat-kernel expansions. It will not be true in general, when \( \#K_{a}^{\min} > 1 \), that \( \Lambda(a) \) is automatically smooth near \( a \). The sphere provides the standard example of this, in particular the unit circle, where the clockwise and anticlockwise geodesics from the north pole have differentiable (wrt the endpoint) distances, but the minimising path distance is not differentiable at the south pole, precisely because there are multiple minimising paths to the south pole which decrease in different (opposite) directions.

**Remark 6.2.11.** Using this Hamiltonian analysis, we are now able to address particular applications and identify the energy and minimal energy paths and test nonfocality for particular stochastic volatility models, which we do in the next section.

### 6.3 Large Strike expansions of Stein-Stein and Schöbel-Zhu Models

One of our main motivations comes from recent density expansions by Gulisashvili–Stein. In [34, Theorem 2.1] they prove that the stock-price in the *uncorrelated* Stein–Stein stochastic volatility model admits a density with expansion\(^7\)

\[
B_1 s^{-B_3} e^{B_2 \sqrt{\log s} \left( \log s \right)^{-\frac{1}{2}}} \left( 1 + O \left( \log s \right)^{-\frac{1}{2}} \right) \quad \text{as} \; s \uparrow \infty
\]

with explicitly computable constants; asymptotic formulae of the implied volatility in the large (similar: small) strike regime are then obtained as corollaries. When writing this expansion in terms of log-price \( Y = \log S \), it indeed has the form (6.0.2) with \( y = \log s = 1/\epsilon^2 \). More generally, we can show that the tail behaviour of \( Y_T \in \mathbb{R}^1 \)

\(^7\)Strictly speaking, their \( O \)-term is \( \log s \) with power \(-1/4\); the authors have informed us, however, that a closer look at their argument indeed gives power \(-1/2\).
for fixed $T > 0$, subject to a certain scaling with parameter $\theta \in \{1, 2\}$ in the full Markovian specification of the model, has the form

$$f(y, T) = e^{-c_1 y^{2/\theta}} e^{c_2 y^{1/\theta}} y^{1/\theta} \left( c_0 + O \left( \frac{1}{y^{1/\theta}} \right) \right) \text{ as } y \uparrow \infty. \quad (6.3.1)$$

It is worth mentioning that such an expansion leads immediately to call price and then (Black–Scholes) implied volatility expansions in the large strike regime, cf. [34, 29]; in the case $\theta = 2$ typical for stochastic volatility,

$$\sigma_I(k, T)^2 T = (\beta_1 k + \beta_2 + o(1))^2 \text{ as the log-strike } k \to \infty;$$

$$\beta_1 = \sqrt{2} \left( \frac{\sqrt{c_1} - \sqrt{c_1 - 1}}{\sqrt{c_1}} \right),$$

$$\beta_2 = c_2 \sqrt{2} \left( \frac{1}{\sqrt{c_1 - 1}} - 1/\sqrt{c_1} \right).$$

(Small strike asymptotics are similar and will not be discussed here.) The leading order behaviour described by $\beta_1 = \beta_1(c_1)$ is well understood [51, 12]; the second order behaviour is given by $\beta_2 = \beta_2(c_1, c_2)$. Further terms in this expansion are in principle possible [29]; in particular, the next term would involve $c_0$. When applied to the Stein–Stein [72] stochastic volatility model the afore-mentioned scaling indeed leads to a small-noise, hypoelliptic diffusion problem with non-vanishing second order exponential factor, as is handled by our main theorem. We then solve a problem left open in the afore-mentioned work [34, Theorem 2.1] in that we are able to compute the expansion in the correlated case. The importance of allowing for correlation in stochastic volatility models is well-documented, e.g. [30, 54], and evidence from estimation of parametric stochastic volatility models suggests the correlation parameter $\rho \approx -0.7$ or $\rho \approx -0.8$ for S&P 500, for instance; a finding fairly robust across models and time periods [3]. With this in mind, we shall focus on the case $-1 < \rho \leq 0$ in our explicit analysis and derive explicit expressions for $c_1, c_2$. In principle, the Laplace method on which we rely yields an explicit expression for $c_0$, cf. [6] Thm 4, p 135), [49].

For given parameters, $a \geq 0, b < 0, c > 0, \sigma_0 \geq 0$, the Stein–Stein model expresses log of the stock price, $Y$, under the forward measure, via

$$dY_t = -\frac{1}{2} Z^2_t dt + Z_t dW^1_t, \quad Y_0 = y_0 = 0 \quad (6.3.2)$$

$$dZ_t = (a + bz_t) dt + cdW^2_t, \quad Z_0 = \sigma_0 > 0,$$
where the Brownian motions, \( W^1, W^2 \) are mutually independent. We will call the case when the Brownian motions are correlated, with \( d \langle W^1, W^2 \rangle_t = \rho dt, \rho \in [-1, 1] \), the Schöbel-Zhu model, [69]. We will be interested in the behaviour, and in particular the tail-behaviour, of the probability density function of \( Y_T \). In fact, there is no loss of generality to consider \( T = 1 \). Applying Brownian scaling, it is a straight-forward computation to see that the pair \( (\tilde{Y}, \tilde{Z}) \) given by

\[
\tilde{Y}(t) := Y(tT), \quad \tilde{Z}(t) := Z(tT) T^{1/2}
\]

satisfies the same parametric SDE form as Stein-Stein, but with the following parameter substitutions

\[
a \leftarrow \tilde{a} \equiv aT^{3/2}, \quad b \leftarrow \tilde{b} \equiv bT, \quad c \leftarrow \tilde{c} \equiv cT, \quad \sigma_0 \leftarrow \tilde{\sigma}_0 \equiv \sigma_0 T^{1/2}.
\]

In particular then, \( Y_T = Y_T (a, b, c, \sigma_0, \rho) \) has the same law as \( Y_1 \left( \tilde{a}, \tilde{b}, \tilde{c}, \tilde{\sigma}_0, \rho \right) \).

**Theorem 6.3.1.** In the Stein-Stein model, the probability density function of the log of the stock price \( Y_T \) has the expansion

\[
f(y) = e^{-c_1 y} e^{c_2 y^2 \frac{1}{2} y^{-\frac{1}{2}} \left( a_0 + O \left( \frac{1}{y^{1/6}} \right) \right)} \quad \text{as} \quad y \to \infty \quad (6.3.3)
\]

where \( a_0 \) is independent of \( y \),

\[
c_1 = \left( \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{b^2}{c^2} + \frac{r_1^2}{c^2 T^2}} \right)
\]

\[
c_2 = q_0^+ \left( \sigma_0 + a \frac{T}{r_1} \tan \left( \frac{r_1}{2} \right) \right)
\]

\[
q_0^+ (\xi) = \frac{2}{c^2 T^2} \sqrt{\frac{2r_1^3 \xi^2 T}{(2c_1 - 1)(2r_1 - \sin (2r_1))}}
\]

and \( r_1 \) denotes the first positive root to

\[
r \cos (r) - bT \sin (r) = 0.
\]

**Theorem 6.3.2.** In the Schöbel-Zhu model with \( \rho \in (-1, 0] \), the probability density function of the log of the stock price \( Y_T \) has the expansion

\[
f(y) = e^{-c_1 y} e^{c_2 y^2 \frac{1}{2} y^{-\frac{1}{2}} \left( a_0 + O \left( \frac{1}{y^{1/6}} \right) \right)} \quad \text{as} \quad y \to \infty \quad (6.3.4)
\]
where \( \alpha_0 \) is independent of \( y \) and

\[
\begin{align*}
  c_1 &= p(r_1) \\
  c_2 &= q_0^+(1) \left( \frac{T}{r_1} \tan \frac{r_1}{2} \right) \\
  q_0^+(\xi) &= \frac{2}{\epsilon^2 T^2} \sqrt{\frac{2r_1^3 \xi}{r_1}} \left( (2c_1(1 - \rho^2) - (1 + 2\rho/c)) (2r_1 - \sin(2r_1)) + 2\rho r_1/c_T (1 - \cos(2r_1)) \right) \\
  p(r) &= \frac{1}{2} (1 - \rho^2) \left\{ \left( 1 + 2\rho \frac{b}{c} \right) + \sqrt{\left( 1 + 2\rho \frac{b}{c} \right)^2 + 4(1 - \rho^2) \left[ \frac{b^2}{c^2} + \frac{\epsilon^2}{c^2 T^2} \right]} \right\}
\end{align*}
\]

and \( r_1 \) denotes the first positive root to

\[ r \cot(r) = (b + \rho cp(r)) T. \]

**Proof of Theorem 6.3.1** Setting

\[
Y^\epsilon := \epsilon^2 Y, Z^\epsilon := \epsilon Z
\]

yields the small noise problem

\[
\begin{align*}
  dY^\epsilon_t &= -\frac{1}{2} (Z^\epsilon_t)^2 dt + Z^\epsilon_t \epsilon dW^1_t, \ Y^\epsilon_0 = 0 =: y_0 \ \forall \epsilon > 0 \ (6.3.5) \\
  dZ^\epsilon_t &= (a \epsilon + b Z^\epsilon_t) dt + \epsilon c dW^2_T, \ Z^\epsilon_0 = \epsilon \sigma_0 \to 0 =: z_0 \ \text{as} \ \epsilon \downarrow 0. \ (6.3.6)
\end{align*}
\]

We postpone the justification that we may indeed apply corollary 6.1.5 to Subsection 6.3.1 since it depends on an analysis of the Hamiltonian ODEs.

**Solving the Hamiltonian ODEs and computing** \( c_1 \) After replacing \( \epsilon dW \) by a control \( dh \), and taking \( \epsilon \downarrow 0 \) elsewhere in (6.3.5), we have to consider the controlled ordinary differential equation

\[
\begin{align*}
  dy_t &= -\frac{1}{2} z_t^2 dt + zdh^1_t, \ y_0 = 0 \quad \text{subject to} \ y_T = \xi \equiv 1 > 0. \ (6.3.7) \\
  dz_t &= b z_t dt + \epsilon c dh^2_t, \ z_0 = 0,
\end{align*}
\]

minimizing the energy, \( \frac{1}{2} \int_0^T \left| \dot{h}_t \right|^2 dt \) subject to \( y_T = \xi \equiv 1 > 0. \)

**Remark 6.3.3.** One of the attractions of the large deviations approach to determining tail asymptotics is that there are many qualitative statements that can be made without detailed computation. Since \( c_1 = \Lambda(1) \) corresponds to the large deviations
rate function for the small noise problem, we have

\[ c_1 := \Lambda(1) = \inf\left\{ \frac{1}{2} \| h\|_H^2 : \phi_0^h = (0, 0), \phi_T^h \in (1, \cdot) \right\} \]

where \( \phi_0^h = (y_0, z_0) \) is the solution of the controlled ODE [6.3.7] under the control \( h \).

It then follows that \( c_1 \) depends on \( b, c, T \), but not on \( a, \sigma_0 \). The same is true for the optimal control \( h^\star := h_0 \) and its associated path \( \phi^\star := \phi^{h_0} \).

Similarly the \( \Lambda' \) factor in \( c_2 \) also only depends on the parameters \( b, c, T \) (but not on \( a, \sigma_0 \)). It remains to analyze the factor \( \tilde{Y}_T \) where \( \begin{pmatrix} \tilde{Y}_t, \tilde{Z}_t : t \geq 0 \end{pmatrix} \) solves the ODE

\[
\begin{align*}
  d\tilde{Y}_t &= (-\phi_t^{+\star} - h_t^\star t)\tilde{Z}_t dt, \quad \tilde{Y}_0 = 0 \\
  d\tilde{Z}_t &= b\tilde{Z}_t dt + adt, \quad \tilde{Z}_0 = \sigma_0.
\end{align*}
\]

Since \( \tilde{Z}_t = \sigma_0 e^{bt} + a \int_0^t e^{b(t-s)} ds \) it follows that \( \tilde{Z}_T \) is linear in \( \sigma_0, a \) with coefficients depending on \( b \) and \( T \). Furthermore, noting that

\[
\tilde{Y}_T = \int_0^T (-\phi_t^{+\star} - h_t^\star t)\tilde{Z}_t dt
\]

a similar statement is true for \( \tilde{Y}_T \) and then \( c_2 = \Lambda'(1) \times \tilde{Y}_T^1 \). Namely, for constants \( C_1 = C_1(b, c; T) \)

\[
c_2 = C_1(b, c; T) \sigma_0 + C_2(b, c; T) a.
\]

We should note that this general structure is not at all obvious from the detailed analysis of [34]. It is interesting to compare this with the Heston result [28] where the constant \( c_2 \) also depends linearly on spot-vol \( \sigma_0 = \sqrt{v_0} \).

We now write out the Hamiltonian (6.2.7) associated to (6.3.7),

\[
\mathcal{H}\left(\begin{pmatrix} y \\ z \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix}\right) = \left( -\frac{1}{2} z^2 p + b z q \right) \begin{pmatrix} p \\ q \end{pmatrix} + \frac{1}{2} \left| \begin{pmatrix} z \\ 0 \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} \right|^2 + \frac{1}{2} \left| \begin{pmatrix} 0 \\ c \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} \right|^2 \]

\[
= -\frac{1}{2} z^2 p + b z q + \frac{1}{2} \left( z^2 p^2 + c^2 q^2 \right).
\]
The Hamiltonian ODEs then become

\[
\begin{pmatrix}
\dot{y}_t \\
\dot{z}_t \\
\dot{p}_t \\
\dot{q}_t
\end{pmatrix} = \begin{pmatrix}
z_t^2 (p_t - \frac{1}{2}) \\
bz_t + c^2 q_t \\
0 \\
p_t z_t (1 - p_t) - bq_t
\end{pmatrix},
\tag{6.3.9}
\]

Trivially, \( p_t \equiv p_0 \) which we shall denote by \( p \) from here on. As it turns out there is a simple expression for the energy, \( \Lambda(\xi) \). Although we shall ultimately take the target value for \( y_T, \xi \), to be 1, it is convenient to carry out the following analysis for general \( \xi > 0 \).

**Remark 6.3.4.** We first note that the Stein-Stein model satisfies \( \theta \)-scaling with \( \theta = 2 \) in the sense of Corollary 6.1.5 which implies that the rate function \( \Lambda(\xi) \) is linear in \( \xi \).

**Lemma 6.3.5.** For any \( h_0 \) given by (6.2.13), i.e.

\[
\dot{h}_0(t) = \begin{pmatrix} p_{zt} \\ q_{zt} \end{pmatrix},
\tag{6.3.10}
\]

where \( (y, z; p, q) \) satisfies (6.3.9), subject to boundary conditions \( (y_0, z_0) = (0, 0) \) and \( y_T = \xi > 0, q_T = 0 \), we have

\[
\frac{1}{2} \int_0^T \left| \dot{h}_0(t) \right|^2 dt = p \xi.
\]

In particular, we see that \( p \geq 0 \), and \( \Lambda(\xi) \) is given by \( p_1 \xi \), where \( p_1 \) is independent of \( \xi \) and is the minimum nonnegative \( p \) amongst the solutions \( \{ (y, z; p, q) \} \) of the above Hamiltonian system.

**Proof.** We give an argument using the Hamiltonian ODEs. The idea is to express \( \left| \dot{h}_0(t) \right|^2 \) as a time-derivative which then allows for immediate integration over \( t \in [0, T] \). Indeed,

\[
\left| \dot{h}_0(t) \right|^2 = p^2 z_t^2 + c^2 q_t^2 = p^2 z_t^2 + \partial_t (z_t q_t) - z_t^2 (p^2 - p) = 2p z_t^2 (p - 1/2) + \partial_t (z_t q_t) = 2p y_t + \partial_t (z_t q_t)
\]

where we used the ODEs for \( z, q \) as given in (6.3.9). It follows that

\[
\int_0^T \left| \dot{h}_0(t) \right|^2 dt = 2p (y_T - y_0) + (z_T q_T - z_0 q_0)
\]
and we conclude with the initial/terminal/transversality conditions $y_0 = z_0 = 0$, $y_T = \xi$ and $q_T = 0$. The properties of $p_t$ then follow from the fact that $\Lambda(\xi)$ is the minimal energy to reach $\xi$, and is linear in $\xi$.

Lemma 6.3.6. [Partial Hamiltonian Flow] Consider (6.3.9) as an initial value problem, with initial data $(y_0, z_0) = (0, 0)$ and $(p, q_0)$. Assume\(^{10}\) \(\chi_p^2 := c^2 p(p-1) - b^2 \geq 0\). Then the explicit solution is given by

$$
y_t = q_0^2 c^4 \frac{2p-1}{8\chi_p^3} \left(2\chi_p t - \sin (2\chi_p t)\right), \tag{6.3.12}
$$

$$
z_t = \frac{q_0 c^2}{\chi_p} \sin (\chi_p t),
$$

$$
p_t \equiv p,
$$

$$
q_t = q_0 \left(\cos (\chi_p t) - \frac{b}{\chi_p} \sin (\chi_p t)\right).
$$

Remark 6.3.7. The given solutions remain valid when $\chi_p^2 < 0$; it suffices to consider $\chi_p$ as purely imaginary; then, if desired, rewrite as $\cos (\chi_p t) = \cosh (|\chi_p| t)$ etc. Below, we shall solve (6.3.9) as boundary value problem, subject to $(y_0, z_0) = (0, 0)$, $y_T = \xi > 0$ and $q_T = 0$; we shall see then that (6.3.11) is always satisfied and in fact $\chi_p^2 > 0$.

Proof. Let us first remark that the path $(p_t)_{t \geq 0}$ is constant, $p_t = p$ for all $t \in [0, T]$. From the Hamiltonian ODEs, the couple $(z_t, q_t)_{t \geq 0}$ solves a linear ODE in $\mathbb{R}^2$, so that the solution must be a linear function of $(z_0, q_0) = (0, q_0)$. Indeed, a simple computation gives

$$
q_t = q_0 \left(\cos (\chi_p t) - \frac{b}{\chi_p} \sin (\chi_p t)\right) \quad \text{and} \quad z_t = \frac{q_0 c^2}{\chi_p} \sin (\chi_p t),
$$

Elementary trigonometric identities then give $(y_t)_{t \geq 0}$ by direct integration; indeed

$$
y_t = \left(p - \frac{1}{2}\right) \int_0^t z_s^2 ds = \frac{q_0^2 c^4 (2p-1)}{8\chi_p^3} \left(2\chi_p t - \sin (2\chi_p t)\right).
$$

This proves the lemma.\(\square\)

For the next proposition we recall the standing assumptions $T > 0$, $b \leq 0$ (which

\(^{10}\)All explicit solutions given in (6.3.12) are even functions of $\chi_p$ and have a removable singularity for $\chi_p = 0$. By convention we shall always assume $\chi_p \geq 0$ although the sign of $\chi_p$ does not matter.
models mean-reversion) and $\xi > 0$.

**Proposition 6.3.8.** The ensemble of solutions to the Hamilton ODEs as boundary value problem

$$(y_0, z_0) = (0, 0) \text{ and } y_T = \xi, q_T = 0$$

with $\xi = 1 > 0$ are characterized by inserting, for any $k \in \{1, 2, \ldots\}$ and any choice of sign in (6.3.14) below,

$$p = p_k = \frac{1}{2} \left( 1 + \sqrt{1 + \frac{4b^2}{c^2} + \frac{4r_k^2}{c^2T^2}} \right), \quad (6.3.13)$$

$$q_{0,k}^\pm = \pm \frac{2}{c^2} \sqrt{\frac{2r_k^3\xi}{(2p_k^* - 1)T^3(2r_k - \sin(2r_k))}} \quad (6.3.14)$$

in (6.3.12). Here $\{r_k : k = 1, 2, \ldots\}$ denotes the set of (increasing) strictly positive roots to

$$r \cos(r) - bT \sin(r) = 0.$$ 

**Remark 6.3.9.** As the proof will show, $p$ as given in (6.3.13) is the unique positive root to

$$c^2p(p - 1) - b^2 = \left( \frac{r_k}{T} \right)^2;$$

in particular, assumption (6.3.11) in the previous lemma is met.

**Proof.** By assumption and (6.3.12),

$$0 = q_T = q_0 \left( \cos(\chi_pT) - \frac{b}{\chi_p} \sin(\chi_pT) \right). \quad (6.3.15)$$

At this stage, $\chi_p$ could be a complex number (when $\chi_p^2 < 0$). Let us note straight away that we must have $q_0 \neq 0$ for otherwise $(y_t)_{t \geq 0}$, which depends linearly on $q_0$ as is seen explicitly in (6.3.12) - would be identically equal to zero in contradiction with $y_T = \xi > 0$. Let us also note that $\chi_p \neq 0$ for otherwise (6.3.15), which has a removable singularity at $\chi_p = 0$, leads to the contradiction $0 = 1 - bT$. (Recall $b \leq 0, T > 0$.) But then $r := \chi_pT$ is a root, i.e. maps to zero, under the map

$$r \in \mathbb{C} \mapsto r \cos r - bT \sin r = r \left( \cos r - \frac{bT}{r} \sin r \right). \quad (6.3.16)$$

A complex analysis lemma [34, Lemma 4] asserts that this map has only real roots, provided

$$-bT \geq 0. \quad (6.3.17)$$
It follows that $\chi_p$ is real and so $\chi_p^2 \geq 0$; actually $\chi_p^2 > 0$, since we already noted that $\chi_p \neq 0$. Note that (6.3.15), and in fact all further expressions involving $\chi_p$, are unchanged upon changing sign of $\chi_p$, we shall agree to take $\chi_p > 0$ as the positive square-root of $\chi_p^2$. In particular, (6.3.15) is equivalent to the existence of $\chi_p > 0$ such that

$$\chi_p T \cos (\chi_p T) - b T \sin (\chi_p T) = 0.$$  

It follows that $\chi_p T \in \{ r_k : k = 0, 1, 2, \ldots \}$, the set of zeros of (6.3.16) written in increasing order. We deduce that, for each $k = 0, 1, 2, \ldots$ there is a choice of $p$ arising from

$$\chi_p^2 = c^2 p (p - 1) - b^2 = \left(\frac{r_k}{T}\right)^2.$$  

For each $k$, there is a negative solution, say $p = p_k^- < 0$ which we may ignore thanks to lemma 6.3.5, and a positive solution, namely

$$p = p_k^+ = \frac{1}{2} \left( 1 + \sqrt{1 + \frac{4b^2}{c^2} + \frac{4r_k^2}{c^2 T^2} + 4} \right) > 1.$$  

We now exploit $y_T = \xi$. From the explicit expression of $y_t$ given in (6.3.12) we get

$$\xi = y_T = \frac{q_0^2 c^4 (2p - 1)}{8\chi_p^2} (2\chi_p T - \sin (2\chi_p T))$$

$$= \frac{q_0^2 c^4 (2p - 1) T^3}{8r_k^3} (2r_k - \sin (2r_k))$$

and thus

$$q_0^2 = \frac{8r_k^3}{c^4 (2p - 1) T^3 (2r_k - \sin (2r_k))} \xi.$$  

It follows that, for each $k \in \{1, 2, \ldots \}$, we can take

$$p = p_k^+ = \frac{1}{2} \left( 1 + \sqrt{1 + \frac{4b^2}{c^2} + 4 \left(\frac{r_k}{cT}\right)^2} \right)$$

$$q_0 = q_{0,k}^\pm = \pm \frac{2}{c^2 T^2} \sqrt{\frac{2r_k^3 T \xi}{(2p_k^+ - 1) (2r_k - \sin (2r_k))}}$$

and any such choice in (6.3.12) leads to a solution of the boundary value problem.  

So far, we have for each $k \in \{1, 2, \ldots \}$ two choices of $(p, q_0)$, depending on the sign in (6.3.14) so that the resulting Hamiltonian ODE solutions, started from $(y_0, z_0) = (0, 0)$ and $(p, q_0)$, describe all possible solutions of the boundary value problem given
by the Hamiltonian ODEs with mixed initial/terminal data

\[(y_0, z_0) = (0, 0) \text{ and } y_T = \xi, q_T = 0.\]

It remains to see which choice (or choices) lead to minimizing controls; i.e. \( h_0 \in K^\text{min}_\xi \). But this is easy since we know from lemma 6.3.5 that, for any \( p \in \{ p_k^+ : k = 1, 2, \ldots \} \),

\[
\frac{1}{2} \int_0^T \left| \dot{h}_0(t) \right|^2 \, dt = p \xi.
\]

Since \( p_k^+ \) is plainly (strictly) increasing in \( k \in \{ 1, 2, \ldots \} \), we see that the energy is minimal if and only if \( p = p_1^+ \). On the other hand, we are left with two choices for \( q_0 \), namely \( q_{0,1}^+ \) and \( q_{0,1}^- \). Using (6.3.10) we then see that there are two minimizing controls,

\[ K^\text{min}_\xi = \{ h_0^+, h_0^- \}, \]

given by

\[ \dot{h}_0^\pm(t) = \left( \begin{array}{c} \frac{p q_0 c^2}{\chi_p} \sin (\chi_p t) \\ c q_0 \left( \cos (\chi_p t) - \frac{b}{\chi_p} \sin (\chi_p t) \right) \end{array} \right) \text{ with } (p, q_0) \leftarrow \left( p_1^+, q_{0,1}^+ \right) \text{ resp. } \left( p_1^+, q_{0,1}^- \right). \]

Of course, \( h_0^+ \) stands for \( h_0^+ \) resp. \( h_0^- \) depending on the chosen substitution above. In \((y, z)\)-coordinates, note that both \( h_0^+ \) and \( h_0^- \) have identical \( y \)-components; their \( z \)-components only differ by a flipped sign due to \( q_{0,1}^- = -q_{0,1}^+ \). (This reflects a fundamental symmetry in our problem which is in fact invariant under \((y, z) \mapsto (y, -z)\)).

We summarize our findings in stating that

\[ \Lambda (\xi) = \frac{1}{2} \| h_0^+ \|^2_H = \frac{1}{2} \| h_0^- \|^2_H = p_1^+ \xi \]  \hspace{1cm} (6.3.18)

and upon taking \( \xi = 1 \) we have computed the leading order constant

\[ c_1 = \Lambda (1) = p_1^+ = \frac{1}{2} \left( 1 + \sqrt{1 + \frac{b^2}{c^2} + 4 \left( \frac{r_1}{cT} \right)^2} \right) \]

where we recall that \( r_1 \) is the first strictly positive root of the equation \( r \cos(r) - bT \sin(r) = 0 \).

**Computing \( c_2 \)** According to Corollary 6.1.5, cf. equation (6.1.9), we need to compute certain ODEs for each minimizer, \( h_0^+ = (h_{0,1}^+, h_{0,2}^+) \) and \( h_0^- = (h_{0,1}^-, h_{0,2}^-) \), determined in the previous section. For ease of notation we shall write \((p, q_{0,1}^+)\) instead of \( (p_1^+, q_{0,1}^+) \)
and \((p^+_1,q^+_0,1)\) in this section. Related to equation (6.3.5) we then have to consider the following ODE along \(h_0^+\) (and then along \(h_0^-\))

\[
\frac{d}{dt} \begin{pmatrix} \hat{Y}_t \\ \hat{Z}_t^2 \end{pmatrix} = \begin{pmatrix} 0 & -z_t^+ \\ 0 & b \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{Y}_{0,t} \\ \hat{Z}_{0,t}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ a \end{pmatrix}
\]

with \(\begin{pmatrix} \hat{Y}_0 \\ \hat{Z}_0^2 \end{pmatrix} = \begin{pmatrix} 0 \\ \sigma_0 \end{pmatrix}\).

Here, we used the fact that \(\dot{h}_0^{+,-} = p z_t^+, z_t^+\) indicates the chosen sign of \(q_{0,1}\) upon which it depends, cf. (6.3.14). The ODE along \(h_0^-\) for \(\hat{Y} = \hat{Y}^-\) is similar, with \(z_t^-, \dot{h}_0^{+,-}\) replaced by \(z_t^-, h_0^{+,-}\) respectively. We can solve these ODEs explicitly.

In a first step (regardless of the chosen sign for \(z, h_0\))

\[
\hat{Z}_t = \begin{cases} \sigma_0 e^{bt} + \frac{a}{b} (e^{bt} - 1) & \text{for } b < 0 \\ \sigma_0 + at & \text{for } b = 0 \end{cases}
\]

and since

\[
\hat{Y}_T^\pm = (p - 1) \int_0^T z_t^T \hat{Z} dt
\]

we see that \(\hat{Y}_T^- = -\hat{Y}_T^+.\) In fact, under the (usual) model parameter assumptions \(a > 0, \sigma_0 > 0\) we see that \(\hat{Z}_t > 0\). We then note that

\[
z_t^+/q_t^+ = \frac{c^2}{\chi_p} \sin (\chi_p t) \geq 0 \text{ for } t \in [0,T];
\]

indeed we saw that \(\chi_p T \in [\pi/2, \pi]\) which implies \(\chi_p t \in [0, \pi]\) and hence \(\sin (\chi_p t) \geq 0\).

In particular, given that \(q_0^+ > 0\) and \(p > 1\) we see that \(\hat{Y}_T^+ > 0\) (and then \(\hat{Y}_T^- < 0\)). It follows that

\[
c_2 := c_2^+ = \Lambda'(1) \times \hat{Y}_T^{+,-1} = p (p - 1) \int_0^T z_t^\pm \hat{Y}_t^2 dt
\]

whereas the contribution from \(c_2^- = \Lambda'(1) \times \hat{Y}_T^{+,-1}\) is exponentially smaller and will not figure in the expansion (cf. remark A.5.1). In fact, given the explicit form of \(t \mapsto z_t^+\) resp. \(\hat{Y}_t^2\) in terms of \(\sin(\cdot)\) and \(\exp(\cdot)\), it is clear that the integration in (6.3.19) can be
6.3. Large Strike expansions of Stein-Stein and Schöbel-Zhu Models

Carried out in closed form. In doing so, one exploits a cancellation due to
\[-\chi_p \cos (\chi_p T) + b \sin (\chi_p T) = 0\]
and also the equality \(\chi_p^2 + b^2 = c^2 (p - 1)\), one is led to
\[c_2 = q_0^+ \left\{ \sigma_0 + a \frac{\tan (\chi_p T / 2)}{\chi_p} \right\} .\]

It is possible, of course, to substitute the explicitly known quantities \(q_0^+, \chi_p\) but this does not yield additional insight.

**Proof of Theorem 6.3.2**

The Schöbel-Zhu model \([69, 55]\) is an extension of the Stein-Stein model to correlated Brownian motions. For fixed parameters, \(a \geq 0, b < 0, c > 0, \sigma_0 \geq 0,\) and \(d \langle W^1, W^2 \rangle = \rho dt\)
\[dY = \frac{1}{2} Z^2 dt + Z dW^1, \ Y(0) = y_0 = 0\]
\[dZ = (a + bZ) dt + c dW^2, \ Z(0) = \sigma_0 > 0.\]

We thus have the diffusion matrix
\[
\begin{pmatrix}
    z^2 & \rho c z \\
    \rho c z & c^2
\end{pmatrix}
\]

In view of financial applications \([30]\) it makes sense to focus on the case \(\rho \in (-1, 0]\). This will also prove convenient in our analysis below, although there is no doubt that the case \(\rho > 0\), less interesting in practice, could also be handled within the present framework. The Hamiltonian becomes
\[
\mathcal{H} \begin{pmatrix}
    y \\
    z
\end{pmatrix} \begin{pmatrix}
    p \\
    q
\end{pmatrix} = -\frac{1}{2} z^2 p + b z q + \frac{1}{2} (z^2 p^2 + c^2 q^2) + \rho c z p q
\]
\[
= -\frac{1}{2} z^2 \tilde{b} + b z q + \frac{1}{2} (z^2 \tilde{b}^2 + c^2 q^2)
\]
with
\[
\tilde{b} := \tilde{b}_p := b + \rho c p
\]

Noting \(\partial_{(y, z)} \tilde{b} = (0, 0)', \partial_{(p, q)} \tilde{b} = (\rho c, 0)'\), we see that the Hamiltonian equations for \(\tilde{z}, \tilde{p}, \tilde{q}\) are thus identical as in the uncorrelated case, one just has to replace \(b\) by \(\tilde{b}\).

In particular, \(p_t\) is again seen to be constant and we denote its value by \(p\). The

---

\[11\] Note that \(Y''\)’s diffusion coefficient is \(Z\) and not \(|Z|\). This makes a difference in the correlated case.
Hamiltonian equation for $\dot{y} = \partial_p \mathcal{H}$ has, in comparison to the uncorrelated case, an additional term, namely \(\partial_p \tilde{b} z_t q_t = \rho c z_t q_t\). In summary, the Hamiltonian ODEs are

\[
\begin{align*}
\begin{pmatrix}
\dot{y}_t \\
\dot{z}_t \\
\dot{p}_t \\
\dot{q}_t
\end{pmatrix}
= \begin{pmatrix}
\left( z_t^2 \left( p_t - \frac{1}{2} \right) + \rho c z_t q_t \\
\tilde{b}_t z_t + c^2 q_t \\
0 \\
p_t z_t (1 - p_t) - \tilde{b}_t q_t
\end{pmatrix}.
\end{align*}
\]

The following lemma is then obvious (only $y$ requires a computation, due to the additional term in the Hamiltonian ODEs).

**Lemma 6.3.10.** [Partial Hamiltonian Flow, correlated case] Consider the above Hamiltonian ODEs as initial value problem, with initial data \((y_0, z_0) = (0, 0)\) and \((p, q_0)\) and assume

\[
\chi_p^2 := c^2 p (p - 1) - \tilde{b}_p^2 \geq 0. \tag{6.3.20}
\]

Then the explicit solution for $z, p, q$ are then identical to the uncorrelated case, one just has to replace $b$ by $\tilde{b}$ throughout. The explicit solution for $y$ is modified to

\[
y_t = \frac{q_0^2 c^2}{8 \chi_p^2} \left[ \left( c^2 (2p - 1) - 2 \rho c \tilde{b}_p \right) (2 \chi_p t - \sin (2 \chi_p t)) + 2 \rho c \chi_p (1 - \cos (2 \chi_p t)) \right]. \tag{6.3.21}
\]

In our explicit analysis of the uncorrelated case (more precisely, in solving the coupled ODEs $\dot{z}_t = b z_t + c^2 q_t, \dot{q}_t = p_t z_t (1 - p_t) - b q_t$) we made use of the (model) assumption $b \leq 0$, cf. \((6.3.17)\). Conveniently, this remains true when $\rho \in (-1, 0]$. Indeed, the following lemma shows we must have $p \geq 0$, so that (with $\rho \leq 0, c > 0$)

\[
\tilde{b} = b + \rho c p \leq 0. \tag{6.3.22}
\]

**Lemma 6.3.11.** Let $\xi > 0$. Then $\Lambda (\xi) = p_1 \xi$ and therefore $p_1 \geq 0$.

**Proof.** We saw in the proof of lemma \textbf{6.3.5} that, in the uncorrelated case, as a direct consequence of the Hamiltonian ODEs,

\[
p^2 z_t^2 + c^2 q_t^2 = 2p \dot{y}_t + \partial_t (z_t q_t).
\]

The correlated case has the identical Hamiltonian ODEs provided we substitute

\[
b \leftarrow \tilde{b} \text{ and } \dot{y} \leftarrow \dot{y} - \rho c z_t q_t.
\]
We therefore have
\[
\left| \dot{h}_0(t) \right|^2 = (p \ q_t) \left( \frac{z^2}{\rho c z} \ \frac{\rho c z}{c^2} \right) \left( \frac{p}{q_t} \right) = p^2 z_t^2 + c^2 q_t^2 + 2\rho c p z_t q_t
\]
\[
= 2p (\dot{y}_t - \rho c z_t q_t) + \partial_t (z_t q_t) + 2\rho c p z_t q_t = 2p \dot{y}_t + \partial_t (z_t q_t)
\]
and then conclude with the boundary data, exactly as in lemma 6.3.5.

As already noted, \( \tilde{b} \leq 0 \) allows to recycle all closed form expressions for \( z, q \) obtained in the uncorrelated case - it suffices to replace \( b \) by \( \tilde{b} \). In particular, for some yet unknown \( p, q_0 \) which may and will depend on \( \rho \),
\[
z_t = \frac{q_0 e^2}{\chi_p} \sin (\chi_p t),
\]
\[
q_t = q_0 \left( \cos (\chi_p t) - \frac{\tilde{b}}{\chi_p} \sin (\chi_p t) \right)
\]
where \( \chi_p^2 := c^2 p (p - 1) - \tilde{b}^2 \) is seen to be positive as in the “uncorrelated” argument. Also, \( q_0 \neq 0 \), seen as in the “uncorrelated” case. Transversality, \( q_T = 0 \), then implies
\[
\chi_p \cos (\chi_p T) - \tilde{b} \sin (\chi_p T) = 0. \quad (6.3.23)
\]

Introducing \( r := \chi_p T \) the gives the equation
\[
r \cot r = (b + \rho cp) T. \quad (6.3.24)
\]

On the other hand, from the very definition of \( \chi_p \), we know
\[
\left( \frac{r}{T} \right)^2 = c^2 p (p - 1) - (b + \rho cp)^2. \quad (6.3.25)
\]

In the uncorrelated case, these two equations were effectively decoupled; in particular, \( r \cot r = bT \) lead to \( r \in \{ r_1^k : k = 1, 2, \ldots \} \subset (0, \infty) \), written in increasing order. Since \( p^+ \) was seen to be monotonically increasing in \( r \), cf. equation (6.3.13), and we were looking for the minimal \( p \), corresponding to the minimal energy (cf. lemma 6.3.11), we were led to seek the first positive root \( r_1^+ \). (In fact, \( r_1^+ \in (\pi/2, \pi) \) as we will also find in the “correlated” discussion below.)

The correlated case is a little more complicated and we start in expressing \( p \) in equation (6.3.24) in terms of \( r \). Indeed, the quadratic equation (6.3.25) shows
\[
p^\pm(r) = \frac{1}{2 (1 - \rho^2)} \left\{ \left( 1 + 2\rho \frac{b}{c} \right) \pm \sqrt{\left( 1 + 2\rho \frac{b}{c} \right)^2 + 4 \left( 1 - \rho^2 \right) \left[ \frac{b^2}{c^2} + \frac{r^2}{c^2 T^2} \right]} \right\}, \quad (6.3.26)
\]
where \( p^- (r) < 0 \) (and hence can be ignored in view of lemma [6.3.11]) and \( p^+ (r) > 0 \).

We now look for \( r \) which satisfies the equation

\[
\frac{r}{\sin r} = (b + \rho c p^+ (r)) T
\]

It is elementary to see that \( \frac{r}{\sin r} \) is non-negative on \([0, \pi/2]\) and then maps \([\pi/2, \pi)\) strictly monotonically to \((-\infty, 0]\). On the other hand, the map \( r \mapsto (b + \rho c p^+ (r)) T \) is \( \leq 0 \) for all \( r \); in particular, there will be a first intersection with the graph of \( r \mapsto \frac{r}{\sin r} \) in \([\pi/2, \pi)\), say at \( r = r^+_1 \). Since \( p^+ (r) \) is plainly strictly increasing in \( r \), the minimal \( p \) must equal to \( p^+_1 := p^+ (r^+_1) \).

We then proceed as in the uncorrelated case, and determine \( q_0 \) from the boundary condition \( y_T = \xi > 0 \) where \( y \) is now given by (6.3.21). This leads to \( q_0 \in \{ q^+_{0,1}, q^-_{0,1} \} \) where

\[
q^+_{0,1} = \pm 2 c \sqrt{\frac{2 r^3 \xi}{T^3 \left( c^2 (2p - 1) - 2 \rho c b \right) + 2 \rho c r / T (1 - \cos (2r))}}
\]

where \( r = r^+_1 \) and \( p = p^+_1 \). Again, we have two minimizing controls, \( \mathcal{K}^\text{min} \) = \( \{ h^+_0, h^-_0 \} \).

We now have

\[
\dot{h}_0 (t) = \begin{pmatrix}
    z_t \sqrt{1 - \rho^2} & 0 \\
    \rho z_t & c
\end{pmatrix}
\begin{pmatrix}
    p \\
    q_t
\end{pmatrix}
\]

(6.3.27)

instead of (6.3.10) and of course lemma [6.3.10] implies that \( z_t \) and \( q_t \) are fully and explicitly determined for each choice of \( (p, q_0) \). In particular for \( (p, q_0) \leftarrow \left( p^+_1, q^+_{0,1} \right) \) resp. \( \left( p^-_1, q^-_{0,1} \right) \) we so obtain \( h^+_0 \) resp. \( h^-_0 \) which can be written explicitly by simple substitution. Moreover, and again as in the uncorrelated case,

\[
\Lambda (\xi) = \frac{1}{2} \| h^+_0 \|^2_H = \frac{1}{2} \| h^-_0 \|^2_H = p^+_1 \xi
\]

(6.3.28)

and upon taking \( \xi = 1 \) we have computed the leading order constant

\[
c_1 = \Lambda (1) = p^+_1 = p^+ (r^+_1)
\]

where we recall that \( r^+_1 \) is the first intersection point of \( r \mapsto \frac{r}{\sin r} \) with \((b + \rho c p^+ (r)) T \) and \( p^+ (\cdot) \) was given in (6.3.26).

At last, we turn to the computation of the second-order exponential constant, \( c_2 \). As in the uncorrelated case, we ease notation by writing \( (p, q_0^\pm) \) instead of \( \left( p^+_1, q^+_{0,1} \right) \).
resp. \((p_{1,1}^+, q_{0,1}^-)\) for the rest of this section. Again, we have to consider ODEs for \((\hat{Y}_t, \hat{Z}_t)\), for each minimizer, \(h_0^+ = (h_{0,1}^+, h_{0,2}^+)\) and \(h_0^- = (h_{0,1}^-, h_{0,2}^-)\). Recall from (6.3.27) that, with \(\tilde{\rho} = \sqrt{1 - \rho^2}\),

\[
\hat{h}_0^+(t) = \begin{pmatrix}
p\tilde{\rho} z_t^+ \\
p\tilde{\rho} z_t^+ + cq_t^+
\end{pmatrix};
\]

where \((\cdot)^\pm\) indicates the chosen sign of \(q_0 \in \{q_0^+, q_0^-\}\) which determines the choice of minimizer. We first determine \(\hat{Y}_T = \hat{Y}_T(h_0^+)\) from the ODE

\[
\frac{d}{dt} \begin{pmatrix}
\hat{Y}_t \\
\hat{Z}_t^2
\end{pmatrix} = \begin{pmatrix}
0 & -z_t^+
0 & b
\end{pmatrix} + \begin{pmatrix}
0 & \tilde{\rho}
0 & 0
\end{pmatrix} \begin{pmatrix}
\hat{h}_{0,t}^+ \\
h_{0,t}^+
\end{pmatrix} + \begin{pmatrix}
0 & \rho
0 & 0
\end{pmatrix} \begin{pmatrix}
\hat{Y}_t \\
\hat{Z}_t
\end{pmatrix} + \begin{pmatrix}
0 \\
an
\end{pmatrix} 
\]

with \(\begin{pmatrix}
\hat{Y}_0 \\
\hat{Z}_0^2
\end{pmatrix} = \begin{pmatrix}
0 \\
s_0
\end{pmatrix}\).

This already shows that we have the identical (closed form) ODE solution for \(\hat{Z}_t\) as in the uncorrelated case. On the other hand, the form of \(\hat{Y}_T\) now exhibits an additional term as is seen in

\[
\hat{Y}_T(h^+) = (p - 1) \int_0^T z_t^+ \hat{Z}_t dt + \rho c \int_0^T q_t^+ \hat{Z}_t dt.
\]

Since \(q_t^+\) is essentially of the same trigonometric form as \(z_t^+\), it is clear that the explicit computations of the uncorrelated case extend. In the end, one finds without too much difficulty

\[
c_2^+ = \Lambda' (1) \times \hat{Y}_T(h_0^+) = \sigma_0 + a \tan \left(\frac{\chi_p T}{2}\right) \frac{\tan (\chi_p T/2)}{\chi_p}.
\]

Now \(\hat{Y}_T(h^-) = -\hat{Y}_T(h^+)\), so \(c_2^- = \Lambda' (1) \times \hat{Y}_T(h_0^-) < c_2^+\).

### 6.3.1 Check of Theorem Assumptions

We have two outstanding issues to address. Firstly, Theorem 6.1.3 assumes that the SDE coefficients are smooth and bounded with bounded derivatives, whereas the Stein-Stein and Schöbel-Zhu models have smooth but unbounded coefficients. The second issue is the check of Assumption 6.1.2. With a view towards the remark on localization in the Appendix, A.5.2, and in particular (A.5.5), we note here that, due
to the particular structure of the SDE, it suffices to localize so as to make $Z$ bounded; e.g. by stopping it upon leaving a big ball of radius $R$. This amounts to, cf. (A.5.5), showing that 
\[
\lim_{R \to \infty} \lim_{\epsilon \to 0} \epsilon^2 \log \Pr \left[ |Z^r|_\infty; [0,T] \geq R \right] = -\infty.
\]

But since \[
\Pr \left[ |Z^r|_\infty; [0,T] \geq R \right] = \Pr \left[ |Z|_\infty; [0,T] \geq R/\epsilon \right]
\]
and $Z$ is a Gaussian process, this is an immediate consequence of Fernique’s estimate.

We now check the non-degeneracy condition (ND), as introduced in definition 6.2.9, which of course is the ultimate justification that an expansion of the form (6.1.10) with the constants computed above holds true. Again, focus is on the case of correlation parameter $\rho \in (-1,0]$. We saw in the previous sections (for $\rho = 0$, then $\rho \leq 0$) that $\# K_{\xi}^{\text{min}} = \# \{ h^+_0, h^-_0 \} = 2$, whenever $\xi > 0$. (In fact, we apply this with $\xi = 1$.)

Secondly, a look at (6.3.7) reveals that the degenerate region is \{(y, z) : z = 0\}, the complement of which is elliptic. Clearly, no controlled path which reaches $y_T = \xi > 0$ can stay in the degenerate region for all times $t \in [0,T]$; after all, this would entail $dy = 0$ and hence $y_T = 0$. We conclude the any ODE solution driven by $h \in K_{\xi}$ must intersect the region of ellipticity; but this already implies non-degeneracy of the corresponding (deterministic) Malliavin covariance matrix.

At last, we check non-focality and focus on $h^+_0$, the other case being similar. We have to check non-degeneracy of the Jacobian of the map $\pi H_{0,-T} (\xi, \cdot ; *, 0)$, evaluated at $\cdot = z_T, * = p_T$ after differentiation, where $z_T, p_T$ are obtained form the Hamiltonian flow at time $T$, cf. lemma 6.3.10 with time 0 initial data $\left( (0,0; p^+_1, q^+_0), 0 \right)$. With some abuse of notation, we write
\[
\begin{pmatrix}
y_0 \\
z_0
\end{pmatrix}
\equiv
\begin{pmatrix}
y_0 (z, p) \\
z_0 (z, p)
\end{pmatrix}
\equiv
\pi H_{0,-T} (\xi, z; p, 0).
\]

Our non-degeneracy condition requires us to show that
\[
\det \begin{pmatrix}
\partial_p y_0 & \partial_z y_0 \\
\partial_p z_0 & \partial_z z_0
\end{pmatrix}_{*,} \neq 0
\] (6.3.29)

where $(\cdot)_{*,}$ indicates evaluation $(\cdot)_{(p,z)=(p^+,z_T)}$ in the sequel. This implies in particular that all expressions which are formulated in terms of the solutions to the Hamiltonian flows, reduced to the corresponding expressions identified in proposition 6.3.8 for $\rho = 0$, resp. in section 6.3 for $\rho \leq 0$. For instance, $(y_0, z_0)_{*,} = (0,0), y_T|_{*,} = \xi, z|_{*,} = z_T \neq 0, \chi_p T|_{*,} \in [\pi/2, \pi)$ and so on. Since $(z, q_*)$ solves a linear ODE, we can
compute

\[
z_0(z, p) = \begin{pmatrix} 1 & 0 \end{pmatrix} e^{-T \begin{pmatrix} \tilde{b}_p & c^2 \\ p(1-p) & -\tilde{b}_p \end{pmatrix}} \begin{pmatrix} z \\ 0 \end{pmatrix} = \frac{z}{\chi_p} \left( \chi_p \cos (\chi_p T) - \tilde{b}_p \sin (\chi_p T) \right).
\]

We first note that \( \partial_z z_0|_* \) is zero; indeed, this follows from (6.3.23). Our next claim is \( \partial_z y_0|_* \neq 0 \). Indeed, from the structure of the Hamilton ODEs,

\[
y_0 - \xi = - \int_0^T y_t dt = z^2 (\ldots)
\]

where \((\ldots)\) does not depend on \(z\). As a result \( \partial_z y_0|_* = 2z (\ldots)|_* = 2 \frac{\chi - \xi}{z}|_* = -2\xi/z_T \neq 0 \).

It remains to check that \( \partial_p z_0|_* \neq 0 \). To this end, recall, as a consequence of the transversality condition, see (6.3.23), that \( \chi_p \cos (\chi_p T) - \tilde{b}_p \sin (\chi_p T) |_* = 0 \). It follows that

\[
\partial_p z_0|_* = \left\{ \frac{z}{\chi_p} \partial_p \left( \chi_p \cos (\chi_p T) - \tilde{b}_p \sin (\chi_p T) \right) \right\}|_*
\]

and since \( z/\chi_p|_* \neq 0 \), it will be enough to show (strict) negativity of \( \partial_p (\ldots)|_* \) above. By scaling, there is no loss of generality in taking \( T = 1 \) and we shall do so from here on. Then

\[
\partial_p \left( \chi_p \cos (\chi_p) - \tilde{b}_p \sin (\chi_p) \right) = \chi_p' \left( 1 - \tilde{b}_p \right) \cos (\chi_p) - \chi_p \sin (\chi_p)\cos (\chi_p) - \rho c \sin (\chi_p).
\]

Since \( \tilde{b}_p|_* \leq 0 \) and \( \chi_p|_* \in [\pi/2, \pi) \) we see that \( [\ldots]|_* < 0 \). Given that \( \chi_p'|_* > 0 \), this already settles the negativity claim in the zero-correlation case. In the case \(-1 < \rho < 0 \), we use (6.3.23) to write

\[
\partial_p \left( \chi_p \cos (\chi_p) - \tilde{b}_p \sin (\chi_p) \right)|_* = \chi_p' \left( 1 - \tilde{b}_p \right) \frac{\tilde{b}_p \sin (\chi_p)}{\chi_p} - \chi_p \sin (\chi_p)|_*.
\]

After division by \( \sin (\chi_p)/\chi_p|_* > 0 \), we have, using \( \tilde{b}_p = b + \rho cp \leq 0 \), \( b \leq 0 \) and again
\[\chi_p' |_* > 0,\]
\[\chi_p' (1 - \tilde{b}_p) \tilde{b}_p - \chi_p^2 | - \rho c |_* \]
\[\leq \chi_p' (1 - \rho cp) \rho cp - \chi_p^2 | - \rho c |_* \]
\[\leq - \rho c (\chi_p - p \chi_p') |_* .\]

With \(-\rho c > 0\), it will then be sufficient to show strict negativity of \(\chi_p - p \chi_p' |_*\). To this end note that the definition, \(\chi_p^2 = c^2 p (p - 1) - \tilde{b}^2\), implies
\[2 \chi_p \chi_p' = c^2 (2p - 1) - 2 \tilde{b} (\rho c)\]
\[\chi_p p \chi_p' = c^2 p (p - 1/2) - \tilde{b} (\rho cp)\]
\[= \chi_p^2 + \frac{c^2 p}{2} + b \tilde{b} > \chi_p^2\]
whenever \(c^2 p/2 + b \tilde{b} > 0\) which is surely the case upon evaluation ... |_. We conclude that \(\partial |_* \neq 0\), and confirm the validity of \((6.3.29)\), for any parameter set \(\rho \in (-1, 0], b \leq 0, c > 0, T > 0\). In other words, we have completed the check of our non-degeneracy condition.

### 6.4 Comments on Heston [39] and Lions–Musiela [54]

We recall from [34, 28] that the density of log-stock price \(Y_T\) in the Heston model,
\[dY_t = \frac{-V_t}{2} + \sqrt{V_t} dW_1^t, \quad Y_0 = y_0 = 0\]
\[dV_t = (a + bV_t) dt + c \sqrt{V_t} dW_2^t, \quad V_0 = v_0 > 0,\]
with \(a \geq 0, b \leq 0, c > 0\) and correlation \(\rho \in (-1, 0]\) has the form
\[f(y) = e^{-c_1 y} e^{c_2 \sqrt{\gamma} y^{-3/4 + a/c^2}} (c_3 + O (1/\sqrt{\gamma})) \text{ as } y \to \infty;\]
with explicitly computable \(c_1 = C_1 (b, c, \rho, T)\) and \(c_2 = \sqrt{v_0} \times C_2 (b, c, \rho, T)\) which do not depend on \(a\). While scaling with \(\theta = 2,\)
\[Y^\epsilon := \epsilon^2 Y, \quad V^\epsilon := \epsilon^2 V\]
indeed yields a small noise problem, namely
\[dY_t^\epsilon = \frac{-V_t^\epsilon}{2} + \sqrt{V_t^\epsilon} \epsilon dW_1^t, \quad Y_0^\epsilon = y_0 = 0\]
\[dV_t^\epsilon = (a \epsilon^2 + bV^\epsilon) dt + c \sqrt{V_t^\epsilon} \epsilon dW_2^t, \quad V_0^\epsilon = v_0 \epsilon^2 > 0.\]
The algebraic factor $y^{-3/4+a/c^2}$ in the above expansion then contradicts the expected factor; cf. (6.1.10)

$$y^{1/2} - 1 = y^{-1/2}.$$ 

There is no contradiction here, of course. Rather, we see an explicit example where “formal” application of a theorem to a model which is short of the required regularity leads to wrong conclusion (at least at the fine level of algebraic factors). Remark that one can trace the origin of this unexpected $y^{-3/4+a/c^2}$ factor to the behaviour of the one-dimensional variance process $V$; also known as Feller- or Cox-Ingersoll-Ross diffusion. Curiously then even a large deviation principle for $V^\epsilon$ as given above presently lacks justification, despite the recent advances in [24], [9]. Clearly then, we are not anywhere near in obtaining the Heston tail result of [34, 28] with the present methods.

However, in the special case when $\alpha = c^2/4$ it is an easy exercise to see that the Heston model can be realized as Stein-Stein model (take $V = Z^2$, where $Z$ is the volatility component of the Stein-Stein model), the resulting expressions are then seen to be consistent with those obtained in [28] and, in particular, $y^{-3/4+a/c^2} = y^{-1/2}$.

Another class of non-smooth, non-affine stochastic vol model with “$\theta = 2$”-scaling was introduced by Lions-Musiela [54]. For $\delta \in [1/2, 1]$ and $\gamma = 1 - \delta$ they consider the 2-dimensional diffusion

$$
\begin{align*}
\frac{dY_t}{dt} &= -\frac{1}{2} Z^{2\delta} \frac{dZ_t}{dt} + Z^\delta \frac{dW_1^1}{dt}, \quad Y_0 = 0 \\
\frac{dZ_t}{dt} &= bZ_t \frac{dt}{dt} + cZ^\gamma \frac{dW_2}{dt}, \quad Z_0 = z_0 > 0.
\end{align*}
$$

And indeed with $Y^\epsilon = \epsilon^2 Y$ and $Z^\epsilon = \epsilon^{-\delta} Z$ this becomes a small noise problem;

$$
\begin{align*}
\frac{dY^\epsilon}{dt} &= -\frac{1}{2} (Z^{\epsilon})^{2\delta} \frac{dZ^\epsilon}{dt} + (Z^\epsilon)^\delta \epsilon \frac{dW_1^1}{dt}, \quad Y^\epsilon_0 = 0 \\
\frac{dZ^\epsilon}{dt} &= bZ^\epsilon \frac{dt}{dt} + c(Z^\epsilon)^\gamma \epsilon \frac{dW_2}{dt}, \quad Z^\epsilon_0 = \epsilon^{-\delta} z_0.
\end{align*}
$$

In their paper they establish the critical exponential moments of $Y^\epsilon_T$. It is tempting to use corollary 6.1.5 at least to leading large deviation order, to obtain the exponential tail of $Z$ for models that scale with $\theta = 2$. Of course, as was discussed in the Heston case, such a “formal” application can be wrong. Further work, building on [24], [9], will be necessary to deal with such degenerate models directly.
6.4. Comments on Heston (39) and Lions–Musiela (54)
Bibliography


Appendix A

A.1 Feller’s test for explosions

In this section we follow the presentation of Feller’s test in [42], from where the definitions and propositions below are taken. We consider a one dimensional SDE defined on an interval $I = (l, r); \ -\infty \leq l < r \leq +\infty$

\begin{equation}
    dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{A.1.1}
\end{equation}

\begin{equation}
    X_0 \in I
\end{equation}

and assume that the coefficients $\sigma : I \rightarrow \mathbb{R}, b : I \rightarrow \mathbb{R}$ satisfy

\begin{equation}
    \sigma^2(x) > 0; \ \forall x \in I, \tag{A.1.2}
\end{equation}

\begin{equation}
    \forall x \in I, \exists \epsilon > 0 \text{ such that } \int_{x-\epsilon}^{x+\epsilon} \frac{1 + |b(y)|}{\sigma^2(y)} dy < \infty. \tag{A.1.3}
\end{equation}

Define the scale function, $p(x)$

\begin{equation}
    p(x) := \int_c^x \exp \left( -2 \int_c^\xi \frac{b(\zeta)}{\sigma^2(\zeta)} d\zeta \right) d\xi, \tag{A.1.4}
\end{equation}

where $c$ can be any element of $I$ (since Feller’s test only depends on whether $p(l^-) := \lim_{x \downarrow l} p(x)$, and $p(r^+) := \lim_{x \uparrow r} p(x)$ are finite or not).

**Definition A.1.1.** A weak solution in the interval $I = (l, r)$ of equation (A.1.1) is a triple $(X, W), (\Omega, \mathcal{F}, P), \{\mathcal{F}_t\}$, where

1. $(\Omega, \mathcal{F}, P)$ is a probability space, and $\{\mathcal{F}_t\}$ is a filtration of sub-$\sigma$-fields of $\mathcal{F}$

135
satisfying the usual conditions,

2. \( X = \{ X_t, \mathcal{F}_t; 0 \leq t \leq \infty \} \) is a continuous, adapted, \([l, r]\)-valued process with \( X_0 \in I \) a.s. and \( \{ W_t, \mathcal{F}_t; 0 \leq t < \infty \} \) is a standard, one-dimensional Brownian motion,

3. with \( \{ l_n \}_{n=1}^{\infty} \) and \( \{ r_n \}_{n=1}^{\infty} \) strictly monotone sequences satisfying \( l < l_n < r_n < r \), \( \lim_{n \to \infty} l_n = l \), \( \lim_{n \to \infty} r_n = r \), and

\[ S_n := \inf\{ t \geq 0 : X_t \notin (l_n, r_n) \}; \quad n \geq 1 \]

the equations

\[
\mathbb{P} \left[ \int_0^{t \wedge S_n} \{|b(X_s)| + \sigma^2(X_s)\} ds < \infty \right] = 1; \quad \forall 0 \leq t < \infty
\]

and

\[
\mathbb{P} \left[ X_{t \wedge S_n} = X_0 + \int_0^t b(X_s) \mathbb{1}_{\{s \leq S_n\}} ds + \int_0^t \sigma(X_s) \mathbb{1}_{\{s \leq S_n\}} dW_s; \quad \forall 0 \leq t < \infty \right] = 1
\]

valid for every \( n \geq 1 \).

We refer to

\[ S = \inf\{ t \geq 0 : X_t \notin (l, r) \} = \lim_{n \to \infty} S_n \]  \hspace{1cm} (A.1.5)

as the exit time from \( I \).

**Proposition A.1.2.** Assume that \( (A.1.2) \) and \( (A.1.3) \) hold, and let \( X \) be a weak solution of \( (A.1.1) \) in \( I \), with nonrandom initial condition \( X_0 = x \in I \). Let \( p \) be given by \( (A.1.4) \) and \( S \) by \( (A.1.5) \). We distinguish four cases:

a) \( p(l+) = -\infty, p(r^-) = \infty \). Then

\[ \mathbb{P}[S = \infty] = \mathbb{P}[\sup_{0 \leq t < \infty} X_t = r] = \mathbb{P}[\inf_{0 \leq t < \infty} X_t = l] = 1. \]

In particular, the process is recurrent: for every \( y \in I \), we have

\[ \mathbb{P}[X_t = y; \text{ some } 0 \leq t < \infty] = 1. \]
b) \( p(l+) > -\infty, p(r-) = \infty. \) Then
\[
P[\lim_{t \uparrow S} X_t = l] = \mathbb{P}[\sup_{0 \leq t < S} X_t < r] = 1.
\]

c) \( p(l-) = -\infty, p(r-) < \infty. \) Then
\[
P[\lim_{t \uparrow S} X_t = r] = \mathbb{P}[\inf_{0 \leq t < S} X_t > l] = 1.
\]

d) \( p(l+) > -\infty, p(r-) < \infty. \) Then
\[
P[\lim_{t \uparrow S} X_t = l] = 1 - \mathbb{P}[\lim_{t \uparrow S} X_t = r] = \frac{p(r-) - p(x)}{p(r-) - p(l+)}.\]

Remark A.1.3. As observed by Karatzas and Shreve, b), c), d) make no claim concerning the finiteness of \( S. \) So for instance the lognormal martingale with \( b \equiv 0, \sigma(x) = x, \) is case b). However, the sublinear growth condition \( (4.3.2) \) is a sufficient condition to prevent explosion at \( \infty \) (see e.g. [42], Remark 5.5.19).

A.2 Wentzell-Freidlin Sample Path Large Deviations by Azencott’s method

Here we report the proof as presented in [9]. As noted in Chapter 4, we removed the assumptions that quasi-continuity was uniform in \( |x| \leq c \) and \( |h| \leq a \) (compare 4.2.1 to [9] A2.3) since they are not necessary for the proof (for fixed initial condition \( x_0 \)). Azencott’s original theorem of which this is a refinement is (8, Chapter 3, Theorem 2.13).

Theorem (4.2.2). (c.f.[9], Theorem 2.4) If Assumption 4.2.1 holds, the family \( \{Y^\epsilon\}_\epsilon \) satisfies a Large Deviations Principle on \( \mathcal{C}_x^d \) with inverse speed \( \epsilon^2 \) and (good) rate function
\[
J(g) = \inf\{I(h); \phi_x^h = g\}, \tag{A.2.1}
\]
with the understanding \( J(g) = +\infty \) if \( \{h : \phi_x^h = g\} \) is empty. In other words
\[
\limsup_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(Y^\epsilon \in F) \leq -\inf_{\psi \in F} J(\psi)
\]
\[
\liminf_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(Y^\epsilon \in G) \geq -\inf_{\psi \in G} J(\psi)
\]
for every closed set \( F \subset \mathcal{C}_x^d \) and open set \( G \subset \mathcal{C}_x^d \) and that the level sets of \( J \) are compact.
Proof. Take $J(g) = a < \infty$. The set $\{ \|h\|_H \leq a + 1 \}$ is compact in $\mathcal{C}_x^m$ (in the uniform topology). By Assumption 4.2.1b) $\{ \phi_{x}^{h} = g \}$ is closed. Furthermore $h \rightarrow \frac{1}{2}\|h\|^2_H$ is (even weakly) lower semicontinuous. Assumption 4.2.1b) ensures that the infimum is attained assuming the infimum is finite. Similarly, $J(g)$ is lower semicontinuous.

**Lower Bound.**

$$\limsup_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(X^\epsilon \in G) \geq - \inf_{\psi \in G} J(\psi)$$

For any Borel set $A \subset \mathcal{C}_x^d$, define $\Lambda(A) = \inf_{a \in A} J(a)$. For any $\delta > 0$, Let $g \in G : J(g) < \Lambda(G) + \delta$. Let $B(f, r) := \{ g \in \mathcal{C}_x^d : \|f - g\|_T \leq r \}$. Take $r > 0 : B(g, r) \subset G$. Then for any $\alpha > 0$,

$$\mathbb{P}(X^\epsilon \in G) \geq \mathbb{P}(\|X^\epsilon - g\|_T < r) \geq \mathbb{P}(\|X^\epsilon - g\|_T < r, \|\epsilon B - h\|_T < \alpha)$$

$$= \mathbb{P}(\|\epsilon B - h\|_T < \alpha) - \mathbb{P}(\|X^\epsilon - g\|_T \geq r, \|\epsilon B - h\|_T < \alpha).$$

Using the LDP for Brownian motion, Theorem 2.1.3 ([68, 8] proposition II.3.6),

$$\lim_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(\|\epsilon B - h\|_T < \alpha) \geq -1/2\|h\|_H^2 = -J(g) \geq \Lambda(G) + \delta.$$

Then by Assumption 4.2.1b) we can find $\alpha, \epsilon_0$:

$$\epsilon^2 \log \mathbb{P}(\|X^\epsilon - g\|_T \geq r, \|\epsilon B - h\|_T < \alpha) < \Lambda(G) + 1.$$  

Since $\delta$ can be made arbitrarily small, we have the result.

**Upper Bound.**

$$\liminf_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(X^\epsilon \in F) \leq - \inf_{\psi \in F} J(\psi).$$

If $\Lambda(F) = 0$ there is nothing to prove, otherwise let $0 < a < \Lambda(F)$, and consider the compact sets (in $\mathcal{C}_x^d$ and $\mathcal{C}_x^m$ respectively)

$$K_a = \{ g \in \mathcal{C}_x^d : J(g) \leq a \}, C_a = \{ h \in \mathcal{C}_x^m : \frac{1}{2}\|h\|_H^2 \leq a \},$$

Then $K_a$ is a subset of the complement of $F$, $F^c$, so for every $g \in K_a$, there exists $r = r_g$ such that $B(g, r) \cap F = \emptyset$. For every $h \in C_a$ the path $g = \phi_{x}^{h}$ belongs to $K_a$, and by Assumption 4.2.1c), there exists $\alpha = \alpha_h$ such that

$$\mathbb{P}(\|Y^\epsilon - g\|_T > r, \|\epsilon B - h\|_T < \alpha) \leq e^{-R/\epsilon^2}$$

for $\epsilon \leq \epsilon_0 = \epsilon_{0,h}$. Since the balls $B(h, \alpha_h)$ form an open cover of the compact set $C_a$, there exists a finite subcover $B(h_i, \alpha_i), i = 1, 2, \ldots, r$. Let $A = \bigcup_{i=1}^r B(h_i, \alpha_i) and
\( g_i = \phi_{hi} \). Then
\[
\mathbb{P}(Y^\epsilon \in F) \leq \mathbb{P}(Y^\epsilon \in F, \epsilon B \in A) + \mathbb{P}(\epsilon B \in A^c)
\]
Again by the LDP for Brownian motion, \( \mathbb{P}(\epsilon B \in A^c) \leq e^{-a/\epsilon^2} \) for sufficiently small \( \epsilon \), whilst if \( g_i = \phi_{hi} \),
\[
\mathbb{P}(Y^\epsilon \in F, \epsilon B \in A) \leq \sum_{i=1}^r \mathbb{P}(Y^\epsilon \in F, \|\epsilon B \|_T < \alpha_i)
\]
\[
\leq \sum_{i=1}^r \mathbb{P}(\|\epsilon B - h_i\|_T > r_i, \|\epsilon B \|_T < \alpha_i)
\]
So for sufficiently small \( \epsilon \) and possibly smaller \( \alpha_i \), \( \mathbb{P}(Y^\epsilon \in F) \leq r e^{-R/\epsilon^2} + e^{-a/\epsilon^2} \), which for \( R > a \), gives
\[
\limsup_{\epsilon \to 0} \epsilon^2 \log \mathbb{P}(Y^\epsilon \in F) \leq -a
\]
for all \( a < \Lambda(F) \).

\section{A.3 Implicit Function Theorem on a Hilbert Space}

We use the statement of the Theorems in the appendix of [50] where proofs are also given. Given a separable Hilbert space \( X, N \in \mathbb{Z}^+ \), and an \( F \in C^1(X; \mathbb{R}^N) \) whose first derivative \( F^{(1)} \) at the origin has maximal rank, we will use the notation
\[
U(F) \equiv F^{(1)}(0)^* (F^{(1)}(0) F^{(1)}(0)^*)^{-1/2},
\]
\[
\Pi(F) \equiv U(F) U(F)^* , \quad \text{and} \quad \Pi^\perp(F) \equiv I_X - \Pi(F).
\]
\( U(F) \) is a unitary mapping from \( \mathbb{R}^N \) onto \( \ker(F^{(1)}(0))^\perp \), and \( \Pi^\perp(F) \) is orthogonal projection onto \( \ker(F^{(1)}(0)) \).

\textbf{Theorem A.3.1.} Given \( M \in [1, \infty) \), there exist positive numbers \( r_1(M), r_2(M) \) and \( \varepsilon(M) < r_2(M) \) such that to every element \( F \) of
\[
\mathcal{F}(M) \equiv \left\{ F \in C^2(X; \mathbb{R}^N) : \|F\|_{C^2(B_X(0,1); \mathbb{R}^N)} \leq M, \right. \\
\left. |F(0)| \leq \varepsilon(M); \quad \text{and} \quad F^{(1)}(0) F^{(1)}(0)^* \geq \frac{I_{\mathbb{R}^N}}{M} \right\}
\]
there corresponds a mapping \( \mathcal{M}[F] \in C^2(X \times \mathbb{R}^N; X) \) with the property that for each \( (x, \xi) \in X \times \mathbb{R}^N \) satisfying \( \|\Pi^\perp(F) x + U(F) \xi\|_X < r_2(M) \), \( \mathcal{M}[F](x, \xi) \) is the one and only element \( y \) of \( \ker(F^{(1)}(0))^\perp \) satisfying both \( \|\Pi^\perp(F) x + y\|_X < r_1(M) \) and \( F(\Pi^\perp(F) + y) = \xi \). Moreover, \( \mathcal{M} \) can be chosen so that, for each \( m \in \mathbb{N} \), and \( n \in \mathbb{Z}^+ \),
the restriction of $\mathcal{F}(M) \cap C_{\text{m+n+1}}^m(X;\mathbb{R}^N)$ has $m$ continuous derivatives as a mapping with values in $C^n_c(X \times \mathbb{R}^N;X)$

**Corollary A.3.2.** Let $X$ be a separable Hilbert space and $F: X \to \mathbb{R}^N$ a twice continuously differentiable map with the properties that $F(0) = 0$ and that $F^{(1)}(0)$ has maximal rank. Then for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any $G \in C^2(X;\mathbb{R}^N)$ satisfying $\|G - F\|_{C^b_b(B_X(0,\varepsilon);\mathbb{R}^N)} < \delta$, $G(x) = 0$ for some $x \in B_X(0,\varepsilon)$. In addition, if $f \in C^1(X;\mathbb{R})$, $r \in (0,\infty)$, and $f(0) = \max\{f(x) : x \in B_X(0,r) \text{ and } F(x) = 0\}$, then $f^{(1)}(0) \perp \ker(F^{(1)}(0))$.

### A.4 Non-focality and infinite-dimensional non-degeneracy

Recall $\mathcal{T}_{h_0}K_a = \ker D\Pi_{\lambda}^{h_0}(x_0) : H_0$. Since $h_0 \in K^\text{min}_a$, it is critical in the sense that $I'(h_0) = D I(h_0) = 0$ on $\mathcal{T}_{h_0}K_a = H_0$.

Also recall $x_T = \phi_{T(0)}^{h_0}(x_0) = \phi_T^{h_0}$, notation used when $x_0$ is fixed. Given $q \in \mathbb{R}^l$ with $1 \leq l \leq d$ we shall write\(^1\)

$$(q,0) \in \mathcal{T}^\ast_{x_T}\mathbb{R}^d$$

for $q$ “viewed” as element in $\mathcal{T}^\ast_{x_T}\mathbb{R}^d$. We can describe $H_0$ as the set of those $h = (h^1,\ldots,h^m) \in \mathcal{H}^m$ such that, for any $q \in \mathbb{R}^l$,

$$\int_0^T \langle (q,0), \Phi_{T+t}^{h_0}V_i(x_t) \rangle \dot{h}_t^i dt = 0;$$

where we tacitly use Einstein’s summation convention. We recall our assumption that the deterministic Malliavin covariance matrix $C^{h_0}$ is invertible.

**Lemma A.4.1.** The linear map $\tilde{\rho}^{h_0}: \mathbb{R}^l \to \mathcal{H}^m$ given by

$$\tilde{\rho}^{h_0}(q) := \left( \int_0^T \langle (q,0), \Phi_{T+t}^{h_0}V_1(x_t) \rangle dt \, \cdots \, \int_0^T \langle (q,0), \Phi_{T+t}^{h_0}V_m(x_t) \rangle dt \right)$$

\(^1\)In fancy notation, $(q,0) = (\Pi_t)_{\ast}^{\langle q \rangle}$ where $(\Pi_t)_{\ast}$ is the adjoint of $(\Pi_t): T_{x_T}\mathbb{R}^d \to T_{\Pi x_T}\mathbb{R}^d$, the differential of the projection map $\Pi_t: (x^1,\ldots,x^d) \to (x^1,\ldots,x^l)$.
for \( i = 1, \ldots, m \) and \( t \in [0, T] \) is one-one with range \( H_0^\perp \).

Proof. Since \( H_0 \) is the set of those \( h \in \mathcal{H}^m \) such that, for any \( q \in \mathbb{R}^l \),
\[
\int_0^T \langle (q, 0), \Phi_{T-t}^h V_i(x_t) \rangle \, dt = 0
\]
we see that \( H_0 \) is the orthogonal complement in \( \mathcal{H}^m \) of
\[
\{ \hat{\rho}_h (q) : q \in \mathbb{R}^l \};
\]
i.e. \( H_0^\perp \) is the range of \( \hat{\rho}_h \). Invertibility of the deterministic Malliavin matrix (along \( h_0 \)) then implies \( \ker \hat{\rho}_h = \{0\} \) which shows that \( \hat{\rho}_h \) is one-one (and also that \( H_0^\perp \) has dimension \( l \)).

Lemma A.4.2. For each minimizer \( h_0 \in K_{a, \min}^m \), there exists a unique \( q = q(h_0) \in \mathbb{R}^l \) s.t.
\[
h_0 = D\phi_{T}^{h_0}\ast [(q, 0)].
\]
(Recall \( D\phi_{T}^{h_0} : \mathcal{H}^m \to T_{x_T} \mathbb{R}^d \); its adjoint then maps \( T_{x_T}^* \mathbb{R}^d \to \mathcal{H}^m \) where we identify \( \mathcal{H}^m \) with \( \mathcal{H}^m \).

Proof. By [6.2.5] with \( g(x) = \Pi_{t}(x) \) we see that \( q = \lambda \), and \((q, 0)\) is the uniquely specified \( p(T) \) and
\[
h_{0i} = \langle (q, 0), \Phi_{T-t}^h V_i(x(t)) \rangle.
\]
It remains to see that, for any \( k \in \mathcal{H}^m \),
\[
\langle k, h_0 \rangle_H = \langle k, D\phi_{T}^{h_0}\ast [(q, 0)] \rangle_H = \langle (q, 0), D\phi_{T}^{h_0} [k] \rangle_H,
\]
but this follows immediately from the computation
\[
\langle k, h_0 \rangle_H = \langle k, \hat{\rho}_h (q) \rangle_H = \int_0^T \hat{k}_i^{\perp} \langle (q, 0), \Phi_{T-t}^h V_i(x_t) \rangle \, dt = \langle (q, 0), \int_0^T \Phi_{T-t}^h V_i(x_t) \, dt \rangle.
\]
Lemma A.4.3. $I''(h_0)$ is a bilinear form on $H_0$ given by

$$I''(h_0) [k, l] = \langle k, l \rangle_H - \left( (q(h_0), 0), D^2 \phi_{h_0}^T [k, l] \right)$$

where $(q(h_0), 0) \in T^*_\mathcal{X}_T \mathbb{R}^d$ was constructed in lemma A.4.2. In particular, an element $k \in H_0$ is in the null-space $\mathcal{N}(h_0)$ of $I''(h_0)$,

$$k \in \mathcal{N}(h_0) := \{ k \in H_0 : I''(h_0) [k, k] = 0 \}$$

if and only if (identifying $\mathcal{H}^m \ast$ with $\mathcal{H}^m$)

$$\langle k, \cdot \rangle_H - \left( q, D^2 \left( \Pi_1 \phi_{h_0}^T \right) [k, \cdot] \right) \in H_0^\perp.$$

Proof. The calculation is performed in [76] section 4. Let us just remark that $\mathcal{N}(h_0)$ is indeed equal to the space $\{ k \in H_0 : I''(h_0) [k, \cdot] \equiv 0 \text{ on } H_0 \}$ as is easily seen from the fact that $I''(h_0)$ is positive semi-definite, since $h_0$ is (by assumption) a minimizer. □

If $U$ is a vector field on $\mathbb{R}^d$ we define the push-forward, under the diffeomorphism $(\phi_{s+T})^{-1}$, by

$$(\phi_{s+T})^{-1} U(z) := (\phi_{s+T})^{-1} U(\phi_{s+T}(z)) \in T_z \mathbb{R}^d$$

We shall then need the following known formula, cf. [17, 1.21] combined with trivial time reparameterization $t \sim T - t$;

$$D \left( \phi_{t-T}^h \right)_*^{-1} U(z) [h] = \int_T^0 \left( \phi_{s+T}^h \right)_*^{-1} V_j \left( \phi_{t-T}^h \right)_*^{-1} U(z) \dot{h}_j^s ds. \tag{A.4.1}$$

Lemma A.4.4. For $h, l \in \mathcal{H}^m$ we have, with $x_T = \phi_{h_0}^T$,

$$D^2 \phi_{h_0}^T [h, l] = \int_0^T \int_T^0 \left( \phi_{s+T}^h \right)_*^{-1} V_j \left( \phi_{t-T}^h \right)_*^{-1} V_i (x_T) \dot{h}_j^s \dot{h}_i^t ds dt + \int_0^T \phi_{T-t}^h \partial_x V_i (x_t) \phi_{t-T}^h D \phi_{h_0}^T [h] \dot{h}_i^t dt.$$
A.4. Non-focality and infinite-dimensional non-degeneracy

Proof. The computation is again performed in [76] section 4. 

\[ D^2 \phi_T^{h_0} [h, l] = \int_0^T D \left\{ \phi_T^{h_0} v_i (x_t) \right\} [h] \dot{l}_i dt + \int_0^T \phi_T^{h_0} \partial_x v_i (x_t) \Phi_T^{h_0} [h, l] \dot{l}_i dt. \]

The proof is then finished using (A.4.1). \hfill \square

Given \( h \in H_0 \), set

\[ \left( \begin{array}{c} 0 \\ \eta \end{array} \right) := D \phi_T^{h_0} [h] \tag{A.4.2} \]

where the notation is meant to suggest that

\[ \eta \in T_{x_T} N_a \text{ where } N_a = (a, \cdot) \subset \mathbb{R}^l \times \mathbb{R}^{d-l} \cong \mathbb{R}^d. \]

**Proposition A.4.5.** Elements \( h \in \mathcal{N} (h_0) \subset H_0 \) are characterized by (inhomogeneous, linear “backward”) Volterra equation

\[ \dot{h}_i = \left\langle (q (h_0), 0), \int_t^T \left( \begin{array}{c} \phi_T^{h_0} v_j (s) \\ \phi_T^{h_0} v_i (s) \end{array} \right) -1 \right\rangle \left( \begin{array}{c} 0 \\ \eta \end{array} \right) \dot{h}_s ds + \left\langle (q (h_0), 0), \Phi_T^{h_0} \partial_x v_i (x_t) \Phi_T^{h_0} \left( \begin{array}{c} 0 \\ \eta \end{array} \right) \right\rangle + \left\langle (\theta, 0), \Phi_T^{h_0} \partial_x v_i (x_t) \right\rangle. \]

where

\[ \eta = \eta (h) \in \text{span} \{ \partial_{l+1} |_{x_T}, \ldots, \partial_d |_{x_T} \} = T_{x_T} N_a \]

is given by (A.4.2) and

\[ \theta = \theta (h) \in \mathbb{R}^l = T_{x_T}^* N_a. \]

**Remark A.4.6.** When \( h \in \mathcal{N} (h_0) \) is also in \( H_1 = \ker D \phi_T^{h_0} \) (which is always true in the point-point setting!) we have \( \eta = 0 \); the equation for \( h \) simplifies accordingly and matches precisely the Bismut’s equation [17, 1.65].

**Remark A.4.7.** It is an important step in our argument to single out \( \eta \). In fact, we must not use

\[ \left( \begin{array}{c} 0 \\ \eta \end{array} \right) = \int_0^T \Phi_T^{h_0} v_j (\phi_T^{h_0} |_{x_T}) \dot{h}_s ds \]

\footnote{It should be noted that the term \( D \phi_T^{h_0} [h] \) is zero for \( h \in H_1 = \ker D \phi_T (h_0) \); in particular the second summand will vanish when \( D^2 \phi_T^{h_0} [\cdot, \cdot] \) is restricted to \( H_1 \) i.e. when considering the point-point case \( l = d \).}

\[ \ldots \text{ which takes the usual form upon reparameterizing time } \tau \leftarrow T - t \ldots \]
as integral term for \( \dot{h} \) in the above integral equation for \( \dot{h} \). Indeed, doing so would lead to a Fredholm integral equation (of the second kind) for \( \dot{h} \) whereas it will be crucial for the subsequent argument to have a Volterra structure. (Solutions to such Volterra equations are unique; the same is not true for Fredholm integral equations.)

**Proof.** For fixed \( h \in H_0 \), we write

\[
\begin{pmatrix}
0 \\
\eta
\end{pmatrix} := D\phi^{b_0}_T[h].
\]

With slight abuse of notation (Riesz!) the previous result then implies that

\[
\left\{ D^2\phi^{b_0}_T[h, \cdot] \right\}_t^T = \int_t^T \left[ \left( \phi^{b_0}_{s-T} \right)^{-1}_s V_j, \left( \phi^{b_0}_{t-T} \right)^{-1}_s V_i \right] (x_T) \dot{h}_s^j ds + \phi^{b_0}_{t-T} \partial_x V_i (x_t) \phi^{b_0}_t (0, \eta)
\]

(A.4.3)

On the other hand, for \( h \in \mathcal{N}(h_0) \), we know that

\[
\langle h, \cdot \rangle_H - \left\langle (q(h_0), 0), D^2\phi^{b_0}_T[h, \cdot] \right\rangle \in H_0^\perp = \text{range } (\tilde{\rho}_{h_0}).
\]

Hence, recalling

\[
\tilde{\rho}_{h_0}(\theta) = \left( \theta, 0 \right), \Phi^{b_0}_{T-t} V_i (x_t)
\]

it follows from (A.4.3) that

\[
\dot{h}_t^i = \left\langle (q(h_0), 0), \int_t^T \left[ \left( \phi^{b_0}_{s-T} \right)^{-1}_s V_j, \left( \phi^{b_0}_{t-T} \right)^{-1}_s V_i \right] (x_T) \dot{h}_s^j ds + \left( q(h_0), 0 \right), \Phi^{b_0}_{T-t} \partial_x V_i (x_t) \phi^{b_0}_t (0, \eta) \right\rangle + \left( \theta, 0 \right), \Phi^{b_0}_{T-t} V_i (x_t)
\]

Remark A.4.8. If we introduce the orthogonal complement \( H_2 \) so that

\[
H_0 = H_1 \oplus H_2 \text{ (orthogonal)}
\]

the map

\[
h \mapsto D\phi^{b_0}_T[h] = \begin{pmatrix}
0 \\
\eta
\end{pmatrix} \mapsto \eta
\]

is a bijection from \( H_2 \to T_{x_T} N_a \).
A.4.1 Jacobi variation

Again, the starting point is the formula

\[ \dot{h}_{0t}^i = \langle p_T, \Phi_{T-t}^{h_0} V_i(x_t) \rangle \]

\[ = \langle (q(h_0), 0), \Phi_{T-t}^{h_0} V_i(x_t) \rangle \]

where we recall

\[ p_T = (q(h_0), 0), \quad x_T \in (a, \cdot) \equiv N_a. \]

We keep \( p_T \) and \( x_T \) fixed and note that the Hamiltonian (backward) dynamics are such that

\[ \pi H_{t \leftarrow T} (x_T, p_T) = x_t \]

Replace \( p_T \) by \( p_T + \epsilon(\theta, 0) \) above, \( x_T \) by \( x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix} \) and write \( h_0(\epsilon) \) for the corresponding control \(^4\) which satisfies the relation

\[ \dot{h}_0(\epsilon)_t^i = \left( p_T + \epsilon(\theta, 0), \Phi_{T-t}^{h_0(\epsilon)} V_i \left( \phi_{t \leftarrow T}^{h_0(\epsilon)} \left( x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix} \right) \right) \right) \]

Define the Jacobi type variation

\[ g := \partial_{(\theta, \eta)} h_0 := \partial_{\epsilon} h_0(\epsilon) \bigg|_{\epsilon=0} \]

so that

\[ \dot{g}_t^i = \left( p_T, D \left\{ \Phi_{T-t}^{h_0} V_i (x_t) \right\} [g] \right) \]

\[ + \left( p_T, \Phi_{T-t}^{h_0} \partial_x V_i (x_t) \Phi_{t \leftarrow T}^{h_0} \left( \begin{pmatrix} 0 \\ \eta \end{pmatrix} \right) \right) \]

\[ + \left( (\theta, 0), \Phi_{T-t}^{h_0} V_i \left( \phi_{t \leftarrow T}^{h_0(\epsilon)} (x_T) \right) \right). \]

With \( p_T = (q(h_0), 0) \) and formula (A.4.1) we see that \( \dot{g} \) satisfies the identical (inhomogeneous, linear backward Volterra equation) as the one given for \( \dot{h} \) in proposition A.4.5. By basic uniqueness theory for such Volterra equations we see that \( \dot{g} = \dot{h} \) as

\[ \text{... which can be constructed explicitly from the Hamiltonian (backward) flow} \]

\[ (x_t(\epsilon), p_t(\epsilon)) := H_{t \leftarrow T} \left( x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix}, p_T + \epsilon(\theta, 0) \right) \]

and the usual formula \( h_0(\epsilon)_t^i = (p_t(\epsilon), V_i(x_t(\epsilon))). \)

\[ ^5\text{Trivial reparameterization } t \sim T - t \text{ will bring it in standard “forward” form.} \]
elements in \( L^2([0,T], \mathbb{R}^d) \), and hence \( g = h \) as elements in \( \mathcal{H}^m \).

**Proposition A.4.9.** Let \( h \in \mathcal{N}(h_0) \subset H_0 \) with associated parameters

\[
\begin{align*}
\theta &\in \mathbb{R}^l = T_{x_T}^* N_{a}
\eta &\in \text{span} \{\partial_{l+1}|_{x_T}, \ldots, \partial_d|_{x_T}\} = T_{x_T} N_a
\end{align*}
\]

provided by proposition A.4.5 (In particular, \( \eta \) is given by \( D\phi_{T}^{h_0} [h] \), cf. (A.4.2).) Then \( h \) can be written in terms of a Jacobi type variation

\[
h = \partial_{(\theta, \eta)} h_0.
\]

Conversely, any Jacobi type variation, with \( \theta \in T_{x_T}^* N_{a}, \eta \in T_{x_T} N_a \) yields an element in \( \mathcal{N}(h_0) \).

**Proof.** The first part follows from the above discussion and it only remains to prove the converse part. Since we have seen that every Jacobi type variation \( g := \partial_{(\theta, \eta)} h_0 \) satisfies the appropriate Volterra equation, cf. proposition A.4.5, we only need to check

\[
\begin{pmatrix}
0 \\
\eta
\end{pmatrix} = D\phi_T^{h_0} [g] x.
\]

\( \square \)

Recall that we say that \( x_0 \) is non-focal for \( (a, \cdot) \equiv N_a \) along \( h \) if for all \( \theta \in T_{x_T}^* N_{a}, \eta \in T_{x_T} N_a \)

\[
\partial_{\epsilon} \pi H_{0_{-T}} \left( x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix}, p_T + \epsilon (\theta, 0) \right) \Big|_{\epsilon=0} = 0 \implies (\theta, \eta) = 0.
\]

In the point-point setting (i.e. \( l = d \) so that \( \theta \in T_{x_T}^* \mathbb{R}_d, \eta = 0 \)) the criterion reduces to

\[
\partial_{\epsilon} \pi H_{0_{-T}} (x_T, p_T + \epsilon \theta) \Big|_{\epsilon=0} = 0 \implies \theta = 0;
\]

disregarding time reparameterization \( t \leftarrow T - t \) and the fact that our setup allows for a non-zero drift vector field, this is precisely Bismut’s non-conjugacy condition [17, p.50].

**Corollary A.4.10.** The point \( x_0 \) is non-focal for \( (a, \cdot) \equiv N_a \) along \( h_0 \) if and only if \( I''(h_0) \), i.e. the second derivative of \( \|\cdot\|_{H}^2 \big|_{K_a} \) at the minimizer \( h_0 \), viewed as quadratic
form on \( H_0 = \ker D (\Pi \varphi_T^{h_0}) \), is non-degenerate, i.e.

\[
N (h_0) \equiv \{0\}.
\]

**Proof.** “⇒”: Take \( h \in N (h_0) \); from proposition A.4.9

\[
h = \partial_{(\theta, \eta)} h_0 \equiv \partial_0 h_0 (\epsilon) |_{\epsilon=0}
\]

for suitable \( \theta \in T^*_x N_{a^\perp}, \eta \in T_x N_a \); in fact,

\[
\begin{pmatrix}
0 \\
\eta
\end{pmatrix} = D \Phi_T^{h_0} (x_0) [h].
\]

The criterion says that if

\[
\partial_\epsilon \pi H_{0+T} \left( x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix}, p_T + \epsilon (\theta, 0) \right) |_{\epsilon=0} = \partial_\epsilon \left( \phi_T^{h_0} (x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix}) \right) |_{\epsilon=0}
\]

equals zero then \((\theta, \eta)\) must be zero. But this is indeed the case here since

\[
\begin{align*}
\partial_\epsilon \left( \phi_T^{h_0} (x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix}) \right) |_{\epsilon=0} &= D \left\{ \phi_T^{h_0} (x_T) \right\} [\partial_\epsilon h_0 (\epsilon) |_{\epsilon=0}] + \phi_T^{h_0} \left( \begin{pmatrix} 0 \\ \eta \end{pmatrix} \right) \\
&= D \left\{ \phi_T^{h_0} (x_T) \right\} [h] + \phi_T^{h_0} [D \phi_T^{h_0} (x_0) [h] \\
&= D \left\{ \phi_T^{h_0} \circ \phi_T^{h_0} (x_0) \right\} [h] \\
&= 0.
\end{align*}
\]

We thus conclude that the directional derivative \(\partial_{(\theta, \eta)} h_0\), which of course depends linearly on \((\theta, \eta)\), vanishes. It then follows that \( h = \partial_{(\theta, \eta)} h_0 = 0 \) which is what we wanted to show.

“⇐”: Assume there exists \((\theta, \eta) \neq 0\) so that

\[
\partial_\epsilon \pi H_{0+T} \left( x_T + \epsilon \begin{pmatrix} 0 \\ \eta \end{pmatrix}, p_T + \epsilon (\theta, 0) \right) |_{\epsilon=0} = 0.
\]

Then \( h := \partial_{(\theta, \eta)} h_0 \) yields an element in the null-space \( N (h_0) \). We need to see that \( h \) is non-zero. Assume otherwise, i.e. \( h = 0 \). Then \( D \Phi_T^{h_0} (x_0) [h] = 0 \) and hence also \( \eta = 0 \).
From the Volterra equation for $h$ we see that

$$0 = \left( (\theta, 0), \Phi^{b_0}_{T-t} V_i \left( \phi^{b_0}_{t-T} (x_T) \right) \right) = \tilde{\rho}_{b_0}((\theta, 0)).$$

But $\ker \tilde{\rho}_{b_0}$ was seen to be trivial and so $\theta = 0$; in contradiction to assumption $(\theta, \eta) \neq 0$.

## A.5 Asymptotic expansion of the marginal density

**Proof.** Assume $\#K^\text{min}_a = 1$. The key observation is that essentially all geometric concepts channel through the (non-geometric, but infinite-dimensional) condition (iii) of proposition 6.2.7 into the application of Laplace’s method. Now, the whole point of proposition 6.2.7 was to provide check-able conditions for $x_0, y$ to satisfy (iii). Having made these part of our assumption we are in fact ready to proceed along the lines of Ben Arous [11]. Fix $y$ and note that for any $C^\infty$-bounded function $z \mapsto F(z)$ on $\mathbb{R}^l$, by Fourier inversion,

$$f_\epsilon(y, T) e^{-F(y)/\epsilon^2} = \frac{1}{(2\pi\epsilon)^l} \int_{\mathbb{R}^l} \mathbb{E} \left[ \exp \left( i \zeta \cdot (Y^\epsilon_T - y) - \frac{F(Y^\epsilon_T)}{\epsilon^2} \right) \right] d\zeta \quad \text{(A.5.1)}$$

$$= \frac{1}{(2\pi\epsilon)^l} \int_{\mathbb{R}^l} \mathbb{E} \left[ \exp \left( i \zeta \cdot \left( Y^\epsilon_T - y \right) - \frac{F(Y^\epsilon_T)}{\epsilon^2} \right) \right] d\zeta. \quad \text{(A.5.2)}$$

In particular, the last integrand can be computed, as asymptotic expansion in $\epsilon$ for fixed $\zeta$, by the Laplace method in Wiener space, cf. [11], [6], based on the full (Markovian) process $X^\epsilon_T$. We pick $F$ (for fixed $y$) such that $F(\cdot) + \Lambda(\cdot)$ has minimum at $y$, i.e.

$$\Lambda(y) = \inf \left\{ F(z) + \Lambda(z) : z \in \mathbb{R}^l \right\}$$

and such that this minimum is non-degenerate; a natural candidate for $F(z)$ would then be given (at least for $z$ near $y$) by

$$z \mapsto \lambda |z-y|^2 - \Lambda(z), \text{ some } \lambda > 0;$$

or

$$z \mapsto \lambda |z-y|^2 - [\Lambda(z) - \Lambda(y)],$$

since adding constants is irrelevant here (recall that $y$ is kept fix). The trouble with the above candidate is their potential lack of (global) smoothness of $\Lambda$; even in the classical Riemannian setting $\Lambda$ will not be smooth at the cut-locus. On the other hand, $\Lambda(\cdot)$ is smooth near $y$ in case $\#K^\text{min}_a = 1$; this is seen exactly as in [17] Thm
1.26]. (In the case \(1 < \#K_{a_{\min}}\), smoothness of \(\Lambda (\cdot)\) near \(y\) was in fact part of our assumptions.) It is thus natural to localize the above candidates around \(y\) which leads us to define \(F\), at least in a neighbourhood of \(y\), by

\[
F(z) = \lambda |z - y|^2 - \left[ \sum_{i}^{d} \partial_{y_i} \Lambda (y) (y^i - z^i) + \frac{1}{2} \sum_{i,j=1}^{d} \partial_{y_i} \partial_{y_j} \Lambda (y) (y^i - z^i)(y^j - z^j) \right];
\]

a routine modification of \(F\), away from \(y\), then guarantees \(C^\infty\)-boundedness of \(F\). (Since \(F(y) = 0\) with this last choice of \(F\), the l.h.s. of (A.5.1) is actually precisely \(f^\epsilon(y,T)\)).

Non-degeneracy of the minimum \(y\) of \(F\) entails that the functional \(H \ni h \mapsto F (\phi_T^h) + \frac{1}{2} \| h \|^2_H\) has a non-degenerate minimum at \(h_0 \in H\). (The argument is identical to [11, Thm 2.6] and makes crucial use of proposition 6.2.7.) The Laplace method is then applicable: we replace \(\epsilon dW\) by \(\epsilon dW + dh_0\) in (6.1.1) and call the resulting diffusion process \(Z^\epsilon\). The integrand of (A.5.2) can then be expressed in terms with \(X^\epsilon\) replaced by \(Z^\epsilon\); of course at the price of including the Girsanov factor

\[
\mathcal{G} := \exp \left( -\frac{1}{\epsilon} \int_0^T \dot{h}_0(t) dW_t - \frac{1}{2\epsilon^2} \int_0^T |\dot{h}_0(t)|^2 dt \right) = \exp \left( -\frac{1}{\epsilon} \int_0^T \dot{h}_0(t) dW_t - \frac{1}{\epsilon^2} \Lambda (y) \right).
\]

A stochastic Taylor expansion of \(Z^\epsilon\), noting right away that

\[
F (\Pi_t Z^\epsilon_T)|_{\epsilon=0} = F \left( \Pi_t \phi_T^h \right) = F (y) = 0,
\]

then leads to (cf. [6, Lemme 1.43])

\[
\begin{align*}
\exp \left( -\frac{1}{\epsilon^2} F (\Pi_t Z^\epsilon_T) \right) &= \exp \left( -\frac{1}{\epsilon^2} \left[ F (y) - \epsilon \int_0^T \dot{h}_0(t) dW_t - \epsilon \Pi_t \bar{X}_T \cdot \partial_y \Lambda (y) + O (\epsilon^2) \right] \right) \\
&= \exp \left( \frac{1}{\epsilon} \int_0^T \dot{h}_0(t) dW_t + \frac{1}{\epsilon} (\bar{Y}_T) \cdot \partial_y \Lambda (y) + O (1) \right).
\end{align*}
\]

(A.5.3)
Putting things together, we have, using $F(y) = 0$, and noting cancellation of $\int_0^T h_0(t) \, dW_t$ in (A.5.3) with the identical term in the Girsanov factor $G$,

$$f_\epsilon(y, T) = \frac{1}{(2\pi \epsilon)^d} \int_\mathbb{R}^d \mathbb{E} \left[ G \times \exp \left( i \left( \zeta, 0 \right) \cdot \left( \frac{Z_T^\epsilon - (y, 0)}{\epsilon} \right) \right) e^{-F(h_0 z_T^\epsilon)} \right] \, d\zeta =: c_0$$

where $O(1)$ denotes the term, bounded as $\epsilon \downarrow 0$, from (A.5.3). What is left to show, of course, is that $c_0$, i.e. the final factor in the above expression, is indeed a strictly positive and finite real number. But since our analysis is based on the full Markovian process $X_T$ (resp. $Z_T^\epsilon$ after change of measure), the arguments of [11, Lemme (3.25)] apply with essentially no changes. In particular, one uses large deviations as in [11, Lemme (3.25)] and, crucially, non-degeneracy of the minimizer $h_0 \in H$, guaranteed by proposition 6.2.7. Finally, integrating the asymptotic expansion with respect to $\zeta \in \mathbb{R}^l$ is justified using the estimates of [11, Lemme 3.48], obtained using Malliavin calculus techniques. At last one sees $c_0 > 0$, as in [11, p. 330].

\[\square\]

Remark A.5.1. [Finitely many multiple minimizers] The case $1 < \# K_{a_{\min}} < \infty \in \{2, 3, \ldots\}$ is handled as in [6]. If

$$K_{a_{\min}} = \left\{ h_0^{(1)}, \ldots, h_0^{(n)} \right\},$$

and invertibility of the Malliavin matrix as well as non-focality holds along each of these, the expansion for $f_\epsilon(y, T)$ as given in theorem 6.1.3 remains valid. Indeed, after localization around each of these $n$ minimizers,

$$f_\epsilon(y, T) = \sum_{h_0 \in K_{a_{\min}}} \epsilon^{-\frac{\Lambda(y)}{2}} e^{\frac{X(y)}{\epsilon} \frac{\hat{Y}_T(h_0)}{\epsilon} - \epsilon^{-\ell} c_0 (h_0)} (1 + O(\epsilon))$$

$$\sim (\text{const}) \epsilon^{-\frac{\Lambda(y)}{2}} e^{\max_{h_0 \in K_{a_{\min}}} \left\{ \frac{X(y)}{\epsilon} \frac{\hat{Y}_T(h_0)}{\epsilon} : h_0 \in K_{a_{\min}} \right\} \epsilon^{-\ell}}$$

where $\hat{Y}_T(h_0)$ denotes the solution of (6.1.9).

Remark A.5.2. [Localization] The assumptions on the coefficients $b, V$ in theorem 6.1.3 (smooth, bounded with bounded derivatives of all orders) are typical in this context (cf. Ben Arous [11, 6] for instance) but rarely met in practical examples from finance. This difficulty can be resolved by a suitable localization which we now out-
line. Set \( \tau_R := \inf \left\{ t \in [0, T] : \sup_{s \in [0, t]} |X_s^\epsilon| \geq R \right\} \) and assume

\[
P [\tau_R \leq T] \lesssim e^{-J_R/\epsilon^2} \quad \text{as} \quad \epsilon \downarrow 0
\]

with \( J_R \to \infty \) as \( R \to \infty \) by this we mean, more precisely,

\[
\lim_{R \to \infty} \limsup_{\epsilon \to 0} \epsilon^2 \log P [\tau_R \leq T] = -\infty. \tag{A.5.5}
\]

In that case, we can pick \( R \) large enough so that \( \Lambda (y) < J_R \), uniformly for \( \epsilon \) near \( 0^+ \), and can expect that the behaviour beyond some big ball of radius \( R \) will not influence the expansion. In particular, if the coefficients \( b, V \) are smooth, but fail to be bounded resp. have bounded derivatives, we can modify them outside a ball of radius \( R \) such as to have this property; call \( \tilde{b}, \tilde{V} \) these new coefficients and \( \tilde{X}^\epsilon \) the associated diffusion.

To illustrate the localization, consider \( l = 1 \), i.e. \( Y^\epsilon_T \equiv X^\epsilon_{T,1} \), and the distribution function for \( Y^\epsilon_T \). Clearly, one has the two-sided estimates

\[
P [Y^\epsilon_T \geq y; \tau_R > T] \leq P [Y^\epsilon_T \geq y] \leq P [Y^\epsilon_T \geq y; \tau_R > T] + P [\tau_R \leq T],
\]

and similar for \( \tilde{Y}^\epsilon_T \equiv \tilde{X}^\epsilon_{T,1} \). Since \( P [Y^\epsilon_T \geq y; \tau_R > T] = P \left[ \tilde{Y}^\epsilon_T \geq y; \tau_R > T \right] \) it then follows

\[
\left| P [Y^\epsilon_T \geq y] - P \left[ \tilde{Y}^\epsilon_T \geq y \right] \right| \leq P [\tau_R \leq T] \lesssim e^{-J_R/\epsilon^2}.
\]

In particular, any expansion for \( \tilde{Y}^\epsilon_T \) of the form

\[
P \left[ \tilde{Y}^\epsilon_T \geq y \right] = e^{-c_1/\epsilon^2} e^{c_2/\epsilon^2} e^{-1} c_0 (1 + O (\epsilon))
\]

leads, upon taking \( R \) large enough so that \( J_R > c_1 \), to the same expansion for \( P [Y^\epsilon_T \geq y] \). With more work of routine type, this localization also be employed for the density expansion in theorem 6.1.3.
A.6  C++ Code for Chapter 4
```cpp
#include <math.h>
#include <omp.h>
#include <vector>
#define CMINPACK_NO_DLL
#include <cminpack.h>
#include "mkl_rci.h"
#include "mkl_lapack.h"
#include "mkl_types.h"
#include "mkl_service.h"

typedef std::vector<double> ArrayDouble;

class DCEV{
public:
    enum Vars {eVp,eV,nDims};
    DCEV(
        double v0,
        double vp0,
        double z3,
        double kappa,
        double c,
        double xi1,
        double xi2,
        double alpha,
        double beta,
        double rho,
        double rho1,
        double rho2,
        double strike,
        double T,
        double deltaT,int nPts,bool pcConst=false);
    //* calculates the forward for epsilon=0 at T, required to get epsilon zero implied vol*/
    void setT(double T){
        T_=T;
        dt_=T/(nPts_-(pcConst?0:1)); // nPts includes zero
        sqrtdt_=_sqrt(dt_);
    }
    void setStrike(double strike){strike_=strike;}
    double getForwarde0() const;

    // we write for v' then v, because last v is fixed
    /*!
    xdot(nPts*nDims), but last is output rather than input, */
    void getGuess(double * xdotsguess) const;
    int calcf$s(double * fs, double * xdots);
    /*!
    dfdsxdots has (nPts*nDims) rows and (nPts*nDims-1) columns we use column major order to be consistent with fortran */
    int calcdfsxdots(double * fs, double * xdots);
private:
    struct Vec2D{double coeffs_[nDims];};
    struct Mat2D{double coeffs_[nDims][nDims];};
    typedef std::vector <Vec2D> ArrayVec2D;
    typedef std::vector <Mat2D> ArrayMat2D;

    double b(const Vec2D & x,int dim) const {
        return (dim ==eVp)? mrs_[dim]*(z3_-.coeffs_[eVp]):mrs_[dim]*(x.coeffs_[eVp]-x.coeffs_[eV]);
    }
    double sigma(const Vec2D& x, int dim) const {
        return xis_[dim]*pow(x.coeffs_[dim],betas_[dim]);
    }
    double xinit_[nDims];
    double z3;  
    double mrs_[nDims];
};
```

double xis_[nDims];
double betas_[nDims];
double rho_;
double rhobar_;
double rho1_;
double rho2_;
double deltaT_;
double a1_;
double a2_;
double a3_;

double T_;
double dt_;
double sqrtddt_;
double strike_;
double T;

DCEV::DCEV(
  double v0,
  double vp0,
  double z3,
  double kappa,
  double c,
  double x11,
  double x12,
  double alpha,
  double beta,
  double rho,
  double rho1,
  double rho2,
  double strike,
  double T,
  double deltaT,
  int nPts,
  bool pcConst):xs(nPts),dfsdxs(nPts){
  xinit_[eVp]=vp0;
  xinit_[eV]=v0;
  z3_=z3;
  mrs_[eVp]= c;
  mrs_[eV]=kappa;
  xis_[eVp]= x12;
  xis_[eV]= x11;
  betas_[eVp]=beta;
  betas_[eV]= alpha;
  rho_= rho;
  rhobar_=sqrt(1-rho_*rho_);
  rho1_= rho1;
  rho2_= rho2;
  /*
   phi_v_= phi_v;
   phi_x_= phi_x;
  */
  deltaT_=deltaT;
  a1_=(1-exp(-kappa * deltaT_))/(kappa * deltaT_);
  a2_=(kappa/(kappa - c)*(1-exp(-c * deltaT_))/(c * deltaT_)-a1_);
  a3_=a1_-a2_;
  strike_=strike;
nPts = nPts;
pcConst = pcConst;
setT();
}

// given gradients at timepoints we recover x as x(t-1) + average grad
for (iDims = 0; iDims < nDims; iDims++) {
  // we write for v rather than x because last v is fixed
  if (iPts == 0) {  // iPts<nPts
    xLast = xinit + EXP(-mrs * dt) * z3 * (1 - EXP(-mrs * T_))
    ydot = strike * xLast - 1 * xLast * xinit / (a_1 * T_);
    for (iPts = 0; iPts < nPts; iPts++) {
      xdotNew = mrs * dt * (iPts + 0.5);
      xdotguess[iPts*nDims+ev] = xdotNew;
    }
  }
  else {  // Vec2D xt; xt.coeffs[ev] = xinit + EXP(-mrs * dt) * z3 * (1 - EXP(-mrs * T_))
    ydot = strike * xLast - 1 * xLast * xinit / (a_1 * (nPts - 1) * dt);
    for (iPts = 0; iPts < nPts; iPts++) {
      xNew = xinit + EXP(-mrs * iPts * dt) * z3 * (1 - EXP(-mrs * T_))
      xdotNew = xNew - 01d * (xNew - x01d) * dt - xdotOld;
      xdotguess[iPts*nDimensions + ev] = xdotNew;
    }
    xdotguess[0*nDimensions + ev] = ydot;
  }
}

for (iPts = 0; iPts < nPts; iPts++) {
  xNew = xinit + EXP(-mrs * iPts * dt) * z3 * (1 - EXP(-mrs * T_))
  xdotNew = xNew - 01d * (xNew - x01d) * dt - xdotOld;
  xdotguess[iPts*nDimensions + ev] = xdotNew;
}

// we order h12, j12, h22, h31, h32...
// but xs 11 12 13 14
// this is because we want vps, then vs because of fixed last v
// we write for v' then v, because last v is fixed

int iPts, iDimensions;
if (pcConst) {
  for (iDimensions = 0; iDimensions < nDimensions; iDimensions++) {
    xs[iDimensions] = xinit + 0.5 * xdotGuess[0*nDimensions + iDimensions] * dt;
    // given gradients at timepoints we recover x as x(t-1) + average grad
    for (iDimensions = 0; iDimensions < nDimensions; iDimensions++) {
for (iPts=1;iPts<nPts-IDims;iPts++){
    xs[iPts].coeffs_[iDims]=xs[iPts-1].coeffs_[iDims]+.5*(xdots[(iPts-1)*nDims+iDims]+xdots[(iPts)*nDims+iDims])*dt_;
}
doctor {vfinal=(strike_*strike_*a2_*xs[nPts-1].coeffs_[eVp]-a3_*z3_)/a1_; // last element fixed
    xdots[(nPts-1)*nDims+eV]=(vfinal-xs[nPts-2].coeffs_[eV])/dt_-.5*xdots[(nPts-2)*nDims+eV];
    xs[nPts-1].coeffs_[eV]=vfinal-.5*xdots[(nPts-1)*nDims+eV]*dt_;}
else {
    for (iDims=0;iDims<nDims;iDims++){
        xs[0].coeffs_[iDims]=xinit_[iDims];
    }
    // given gradients at timepoints we recover x as x(t-1) +average grad
    for (iDims=0;iDims<nDims;iDims++){
        for (iPts=1;iPts<nPts -iDims;iPts++){
            xs[iPts].coeffs_[iDims]=xs[iPts-1].coeffs_[iDims]+.5*(xdots[(iPts-1)*nDims+iDims]+xdots[(iPts)*nDims+iDims])*dt_;
        }
    }
    x[nPts-1].coeffs_[eV]=(strike_*strike_*a2_*xs[nPts-1].coeffs_[eVp]-a3_*z3_)/a1_; // last element fixed
}

#pragma omp parallel default none private(iPts,IDims,k) shared (xs,xdots,fs)
#pragma omp for
for (iPts=0;iPts<nPts;iPts++){for (iDims=0;iDims<nDims;iDims++){
    // this is (xdt - b)/s
    fs[iPts*nDims+iDims]=(xdots[iPts*nDims+iDims]-b(xs[iPts],iDims))/sigma(xs[iPts],iDims)*sqrtdt_;
} // now solve for hdot starting from h2 - triangular structure
// h2=f2, so just have to do h1
fs[iPts*nDims+eV]=-rho_*fs[iPts*nDims+eVp];
fs[iPts*nDims+eVp]=rhobar_;

return(0);
}

extern "C" int _stdcall calcfs(double * fs, double * xdots,
    double v0, double vp0, double z3, double kappa, double c,
    double xi1, double xi2,
    double alpha, double beta, double rho, double rho1, double rho2, double strike, double T,
    double deltaT,int nPts){
    dCEV dCEV(v0,vp0,z3,kappa,c,xi1,xi2,alpha,beta,rho,rho1,rho2,strike,T,deltaT,nPts);
    dCEV.calcfs(fs,xdots);
    return 0;
}

// we actually overwrite the last element of xdots (the constraint point, which is not part of optimisation)
int calcDCEV(void *p, int m, int n, const double *xdots, double *fs, int iflag){
DCEV * pDCEV = static_cast<DCEV *>(p);
pDCEV->calcfs(fs, const_cast<double *>(xdots));
return 0;
}

/* nonlinear system equations without constraints */
/* routine for extended powell function calculation */
/* m in: dimension of function value */
/* n in: number of function variables */
/* xdots in: vector for function calculating */
/* fs out: function value f(x) */

DCEV * pDCEVGlobal;

void calcDCEVIntel (MKL_INT *m, MKL_INT *n, double *xdots, double *fs)
{
pDCEVGlobal->calcfs(fs, xdots);
}

extern "C" void chkdersv(int m, int n, const double *x,
double *fvec, double *fjac, int ldjfjac, double *xp,
double *fvecp, int mode, double *err);

extern "C" int _stdcall calcImpVol(double * fs, double * xdotsguess, double & impVol,
double v0, double vp0, double z3, double kappa,
double c, double xi1, double xi2, double alpha,
double beta, double rho,
double rho1, double rho2, double strike,
double T, double deltaT, int nPts)
{
int info;
int m = nPts*DCEV::nDims;
int n = nPts*DCEV::nDims-1;
int ldjfjac = nPts*DCEV::nDims;
int *ipvt=new int[m];
double tol, fnorm;

double * dfsxdots=new double [nPts*DCEV::nDims*(nPts*DCEV::nDims-1)];
double * fs2=new double [nPts*DCEV::nDims];
double * xdotsguess2=new double [nPts*DCEV::nDims];
double * err=new double [nPts*DCEV::nDims];
int lwa = 5*n+m+m*n; // m*n for numerical jacob
/* lwa is an integer work array of length n. */
/* wa is a work array of length lwa. */
/* lwa is a positive integer input variable not less than */
C:\Users\sv507\Documents\Sean\work\PhD\VIWatanabe\MinPath1rev7.cpp

/*
 * int * iwa=new int[n];
 * double * wa=new double[lwa];
 *
 * DCEV dCEV(v0,vp0,z3,kappa,c,x1,x2,alpha,beta,rho,rho1,rho2,strike,T,deltaT,nPts,false);
 * dCEV.getGuess(xdotsguess);
 */

int tol = 1e-15;//sqrt(dpmpar(1));

//int mode=1;
//for(int i=0;i<nPts;++i) xdotsguess[i*DCEV::nDims+DCEV::eVp]+=xdotsguess[0]; // chk der doesn't like f= x
//
//chkderv( m, n, xdotsguess, fs, dfsdxdots, ldfjac, xdotsguess2,fs2,mode,err);
//calcDCEVjac(static_cast<DCEV*>(&dCEV),m,n,xdotsguess,fs,dfsdxdots,ldfjac,1);
//calcDCEVJ3ac(static_cast<DCEV*>(&dCEV),m,n,xdotsguess,fs,dfsdxdots,ldfjac,2);
//calcDCEVJ3ac(static_cast<DCEV*>(&dCEV),m,n,xdotsguess2,fs2,dfsdxdots,ldfjac,1);
//mode=2;

extern "C" int  _stdcall calcImpVolArray(double * fs, double * xdotsguess, double * impVols,double & timing
,
  double v0,
  double vp0,
  double z3,
  double kappa,
  double c,
  double x1,
  double x2,
  double alpha,
  double beta,
  double rho,
  double rho1,
  double rho2,
  double *strikes, int nstrikes,
  double *Ts,int nTs,
  double deltaT,int nPts){
  double timestamp=dsecnd();
  int info;

  int m = nPts*DCEV::nDims;
  int n = nPts*DCEV::nDims-1;
  int ldfjac = nPts*DCEV::nDims;
  int *ipvt=new int[m];
  double tol, fnorm;

  double * dfsdxdots=new double [nPts*DCEV::nDims*(nPts*DCEV::nDims-1)];
  double * fs2=new double [nPts*DCEV::nDims];
  double * xdotsguess2=new double [nPts*DCEV::nDims];
  double * err=new double [nPts*DCEV::nDims];

  // info = lmdif1(calcDCEV, &dCEV, m, n, xdotsguess, fs, tol, iwa, wa, lwa);
  fnorm = enorm(m, fs);
  // impVol=abs(log(dCEV.getForwade0()/strike))/(fnorm*sqrt(T));
  free(ipvt);
  free(fs2);
  free(xdotsguess2);
  free(dfsdxdots);
  free(iwa);
  free(wa);
  return info;
}
int lwa = 5*n+m + m*n; // m*n for numerical jacob
/*
  lwa is an integer work array of length n. */
/
/*
  lwa is a positive integer input variable not less than */
/*
  m*n+m. */
int *iwa=new int[n];
double * wa=new double[lwa];
tol = sqrt(dpmpar(1));
DCEV dCEV(v0, vp0, z3, kappa, c, xi1, xi2, alpha, rho, roh1, roh2, strikes[0], Ts[0], deltaT, nPts, false);

for (int iTs=0; iTscTs;iTs++){
  dCEV.setT(Ts[iTs]);
  dCEV.setStrike(strikes[0]); // NB getGuess needs strike
  dCEV.getGuess(xdotsguess);
  for (int istrikes=0;istrikes<nstrikes;istrikes++){
    dCEV.setStrike(strikes[istrikes]);
    /*
      set tol to the square root of the machine precision. */
    /*
      unless high solutions are required, */
    /*
      this is the recommended setting. */
    
    //int mode=1;
    //for(int i=0;i<nPts;i++) xdotsguess[i*DCEV::nDims+DCEV::eVp]=xdotsguess[i]; // chk der doesn't like fs=0;
    //chkdersv( m, n, xdotsguess, fs, dfsdxdots, ldfjac, xdotsguess2, fs2, mode, err);
    //calcDCEVJac(static_cast<DCEV *>(&DCEV),m,n,xdotsguess,fs,dfsdxdots,ldfjac,1);
    //calcDCEVJac(static_cast<DCEV *>(&DCEV),m,n,xdotsguess,fs,dfsdxdots,ldfjac,2);
    //calcDCEVJac(static_cast<DCEV *>(&DCEV),m,n,xdotsguess,fs,dfsdxdots,ldfjac,1);
    //mode=2;
    //chkdersv( m, n, xdotsguess, fs, dfsdxdots, ldfjac, xdotsguess2, fs2, mode, err);
    // info = lmdef1(calcDCEVJac, static_cast<DCEV *>(&DCEV), m, n, xdotsguess, fs, tol, iwa, wa, lwa);
    info = lmder1(calcDCEVJac, static_cast<DCEV *>(&DCEV), m, n, xdotsguess, fs, dfsdxdots, ldfjac, tol, ipvpt, wa, lwa);
    fnorm = enorn(m, fs);
    impVols[istrikes+nstrikes*iTs]=abs(log(dCEV.getForwarde0()/strikes[istrikes]))/(fnorm*sqrt(Ts [iTs]));
  }
}
free(ipvt);
free(err);
free(fs2);
free(xdotsguess2);
free(dfsdxdots);
free(lwa);
free(wa);
timing=dsecnd()-timestart;
return info;
}

extern "C" int _stdcall calcImpVolIntel(double * fs, double * xdotsguess, double & impVol,
  double v0,
  double vp0,
  double z3,
  double kappa,
  double c,
  double xi1,
  double xi2,
  double alpha,
  double beta,
double rho,
double rho1,
double rho2,
double strike,
double T,
double deltaT,int nPts){
    /* n - number of function variables
       m - dimension of function value */
    MKL_INT m = nPts*DCEV::nDims;
    MKL_INT n = nPts*DCEV::nDims-1;

double tol;

double * dfsdxdots=new double [nPts*DCEV::nDims*(nPts*DCEV::nDims-1)];

DCEV dCEV(v0,wp0,z3,kappa,c,xi1,xi2,alpha,beta,rho,rho1,rho2,strike,T,deltaT,nPts,true);
dCEV.getGuess(xdottsignature);
    // set tol to the square root of the machine precision. */
    // unless high solutions are required, */
    // this is the recommended setting. */

tol = sqrt(dpmpar(1));
pDCEVGlobal=&dCEV;

    /* precisions for stop-criteria (see manual for more details) */
    double eps[6];

    /* iter1 - maximum number of iterations
       iter2 - maximum number of iterations of calculation of trial-step */
    MKL_INT iter1 = 1000, iter2 = 100;
    /* Initial step bound */
    double rs = 0.0;
    /* reverse communication interface parameter */
    MKL_INT RCI_Request; // reverse communication interface variable
    /* controls of rci cycle */
    MKL_INT successful;

    /* number of iterations */
    MKL_INT iter;
    /* number of stop-criterion */
    MKL_INT st_cr;
    /* Initial and final residuals */
    double r1, r2;
    /* TR solver handle */
    _TRNSP_HANDLE_t handle; // TR solver handle
    /* cycle's counter */
    MKL_INT i;

    /* set precisions for stop-criteria */
    for (i = 0; i < 6; i++)
    {
        eps[i] = 0.0000000001;
    }

    /* initialize solver (allocate mamory, set initial values)
handle in/out: TR solver handle
n in: number of function variables
m in: dimension of function value
x in: solution vector. contains values x for f(x)
eps in: precisions for stop-criteria
iter1 in: maximum number of iterations
iter2 in: maximum number of iterations of calculation of trial-step

rs in: initial step bound */
if (dtrnlsp_init (&handle, &n, &m, xdotsguess, eps, &iter1, &iter2, &rs) != TR_SUCCESS)
{
    /* if function does not complete successful then print error message */
    _RPT0(_CRT_WARN,"| error in dtrnlsp_init\n");
    /* Release internal MKL memory that might be used for computations */
    /* NOTE: It is important to call the routine below to avoid memory leaks */
    /* unless you disable MKL Memory Manager */
    MKL_FreeBuffers();
    free (dfsdxdots);
    /* and exit */
    return 1;
}
/* set initial rci cycle variables */
RCI_Request = 0;
successful = 0;
/* rci cycle */
while (successful == 0)
{
    /* call tr solver */
    handle in/out: tr solver handle
    fvec in: vector
    fjac in: jacobi matrix
    RCI_request in/out: return number which denote next step for performing */
    if (dtrnlsp_solve (&handle, fs, dfsdxdots, &RCI_Request) != TR_SUCCESS)
    {
        /* if function does not complete successful then print error message */
        _RPT0(_CRT_WARN,"| error in dtrnlsp_solve\n");
        /* Release internal MKL memory that might be used for computations */
        /* NOTE: It is important to call the routine below to avoid memory leaks */
        /* unless you disable MKL Memory Manager */
        MKL_FreeBuffers();
        free (dfsdxdots);
        /* and exit */
        return 1;
    }
    /* according with rci_request value we do next step */
    if (RCI_Request == -1 ||
        RCI_Request == -2 ||
        RCI_Request == -3 ||
        RCI_Request == -4 ||
        RCI_Request == -5 ||
        RCI_Request == -6)
    { /* exit rci cycle */
        successful = 1;
    }
    if (RCI_Request == 1)
    {
        /* recalculate function value */
        m in: dimension of function value
        n in: number of function variables
        x in: solution vector
        fvec out: function value f(x) */
        calcDCEVIntel(&m, &n, xdotsguess, fs);
    }
    if (RCI_Request == 2)
    {
        /* compute jacobi matrix */
        extendet_powell in: external objective function
        n in: number of function variables
        m in: dimension of function value
        fjac out: jacobi matrix
        x in: solution vector
        jac_eps in: jacobi calculation precision */
        if (djacobi (calcDCEVIntel, &n, &m, dfsdxdots, xdotsguess, eps) != TR_SUCCESS)
        {
            /* if function does not complete successful then print error message */
            _RPT0(_CRT_WARN,"| error in djacobi\n");
        }
    }
}
Release internal MKL memory that might be used for computations

NOTE: It is important to call the routine below to avoid memory leaks

unless you disable MKL Memory Manager

MKL_FreeBuffers();
free(dfsdxdots);
/* and exit */
return 1;
}
/* get solution statuses */
handle in: TR solver handle
iter out: number of iterations
st_cr out: number of stop criterion
r1 out: initial residuals
r2 out: final residuals

if (dtrnlsp_get(&handle, &iter, &st_cr, &r1, &r2) != TR_SUCCESS)
{
/* if function does not complete successful then print error message */
printf(" | error in dtrnlsp_get\n");
/* Release internal MKL memory that might be used for computations */
/* NOTE: It is important to call the routine below to avoid memory leaks */
/* unless you disable MKL Memory Manager */
MKL_FreeBuffers();
free(dfsdxdots);
/* and exit */
return 1;
}
/* free handle memory */
if (dtrnlsp_delete(&handle) != TR_SUCCESS)
{
/* if function does not complete successful then print error message */
_RPT0 (_CRT_WARN, " | error in dtrnlsp_delete\n");
/* Release internal MKL memory that might be used for computations */
/* NOTE: It is important to call the routine below to avoid memory leaks */
/* unless you disable MKL Memory Manager */
MKL_FreeBuffers();
free(dfsdxdots);
/* and exit */
return 1;
}
/* Release internal MKL memory that might be used for computations */
/* NOTE: It is important to call the routine below to avoid memory leaks */
/* unless you disable MKL Memory Manager */
MKL_FreeBuffers();
free(dfsdxdots);
impVol = abs(log(dCEV.getForward&0() / strike)) / (r2 * sqrt(T));
/* if final residual less then required precision then print pass */
if (r2 < 0.00001)
{
_RPT0 (_CRT_WARN, "| dtrnlsp DCEV.........PASS\n");
return 0;
}
/* else print failed */
else
{
_RPT0 (_CRT_WARN, "| dtrnlsp DCEV.........FAILED\n");
return RCI_Request;
}
extern "C" int _stdcall calcdfsdxdots(double * dfsdxdots, double * xdots, double v0, double vp0, double z3, double kappa, double c, double xi1, double xi2, double alpha, double beta, double rho, double rho1, double rho2, double strike, double T, double deltaT, int nPts)
{
    DCEV dCEV(v0, vp0, z3, kappa, c, xi1, xi2, alpha, beta, rho, rho1, rho2, strike, T, deltaT, nPts);
    dCEV.calcdfsdxdots(dfsdxdots, xdots);
    return 0;
}

int DCEV::calcdfsdxdots(double * dfsdxdots, double * xdots)
{
    // we order h11,j12,h21,h22,h31,h32...
    // but xs 11 12 13 14
    // this is because we want vps, then vs because of fixed last v
    // we write for v' then v, because last v is fixed

    int iPts, iDims, iPtsX, iDimsX;
    for (iDims = 0; iDims < nDims; iDims++)
    {
        xs[0].coeffs_[iDims] = xinit_[iDims];
    }
    // given gradients at timepoints we recover x as x(t-1) + average grad
    for (iDims = 0; iDims < nDims; iDims++)
    {
        for (iPts = 1; iPts < nPts - iDims; iPts++)
        {
            xs[iPts].coeffs_[iDims] = xs[iPts-1].coeffs_[iDims] + .5*(xdots[(iPts-1)*nDims+iDims]+xdots[iPts]*x[0].coeffs_[iDims])*dt_;
        }
    }
    xs[nPts-1].coeffs_[eV] = (strike_*strike_-a2_*xs[nPts-1].coeffs_[eVp]-a3_*z3_)/a1_; // last element fixed
    // zero out first
    for (iPts = 0; iPts < nPts; iPts++)
    {
        for (iDims = 0; iDims < nDims; iDims++)
        {
            dfsdxs[iPts].coeffs_[iDims][iDimsX] = 0;
        }
    }
    return 0;
}
std::fill (&dfsdxdots[0],&dfsdxdots[nPts*nDims *(nPts*nDims-1)],0);

#pragma omp parallel default(none) private(iPts,iDims,k) shared(xs,xdots,fs)
#pragma omp for
for (iPts=0; iPts<nPts; iPts++)
for (iDims=0; iDims<nDims;iDims++){
    // this is (xdot - b)/sigma
    double ftmp=(xdots[iPts*nDims+iDims]-b(xs[iPts],iDims))/sigma(xs[iPts],iDims)*sqrtdt;
    dfsdx[iPts].coeffs[iDims][iDims]=(mrs[iDims]*sqrtdt/sigma(xs[iPts],iDims)- betas[iDims]*
                ftmp/xs[iPts].coeffs[0]);
    dfsdx[iPts].coeffs_[ev][0]=(mrs[ev]/sigma(xs[iPts],ev)*sqrtdt;
    dfsdx[iPts].coeffs_[ev][ev]=0;
}

// now solve for hdot starting from h2 - triangular structure
// h2=2, so just have to do h1
for (iDimsX=0; iDimsX<nDims; iDimsX++)
    dfsdx[iPts].coeffs_[ev][iDimsX]=-rho*dfsdx[iPts].coeffs_[ev][iDimsX];
    dfsdx[iPts].coeffs_[ev][iDimsX]/=rhopbar;
}

// v_last=(K*2-a_3 z_3 - a_2 v'_L )/a_1 ie v_last is generated from constraint rather than from xdots
// so move sensitivity to v'_last
if (iPts==nPts-_1)
    dfsdx[iPts].coeffs_[ev][ev]_1=(a2_/a1_)*dfsdx[iPts].coeffs_[ev][ev];
    dfsdx[iPts].coeffs_[ev][ev]=0;
// add in term from xdot
// \$ \dot{x}_L=(x_L-x_-1/L)*2/dt \dot{x}_L_1 \$
\$ dfsdx[iPts].coeffs_[ev][ev]=(a2_/a1_)/(rhopbar*sigma(xs[iPts],ev)*sqrtdt_2/dt_1);$
}

//now turn sensitivity to x to sensiti to xdot
//sensitivity of f(x) has no sensitivity to xdot(x) via xo
if (iPts>0) for (iDimsX=0; iDimsX<nDims; iDimsX++)
if (iPts==nPts-_1&& (iDimsX==x))
    for (iPtsX=0; iPtsX<nPts-_2;iPtsX++)
    // \$ \dot{x}_L=(x_L-x_-1/L)*2/dt \dot{x}_L_1 \$
    dfsdxdots[(iPtsX*nDims+iDimsX)*(nPts*nPts-_1)]=-2/(rhopbar*sigma(xs[iPts],ev))^sqrtdt_;
    // other term is zero
    }else{
    for (iDims=0; iDims<nDims;iDims++)
    for (iPtsX=0; iPtsX<nPts; iPtsX++)
    dfsdxdots[iPtsX*(nDimX+1)*nDimX+nDimX]=dfsdx[iPts].coeffs_[
    [iDims][iDimX]]^dt_1;
    // diagonal* element,(iPts,iPts) only has 1/2
    // nb assume dfsdxdots contains only terms from dfsdx we add the partial dfsdxdot later...
    dfsdxdots[(iPtsX*nDimX+iDimX)+(iPtsX*nDimX+iDimX)*(nDimX*nDimX)]+=.5; // nb iPts-1
    dfsdxdots[(iPtsX*nDimX+iDimX)+(0*nDimX+iDimX)*(nDimX*nDimX)]+.5.;
    }

}
for (iDims=0;iDims<nDims;iDims++){
    // this is (x dot - b)/sigma
    // we reuse dfsdxs to do sensitivity to dotx
dfsdxs[iPts].coeffs_[iDims][iDims]=sqrtdt_/sigma(xs[iPts],iDims);
dfsdxs[iPts].coeffs_[iDims][(iDims+1)% nDims]=0;
}

// now solve for hdot starting from h2 - triangular structure
// h2=f2, so just have to do h1
for (iDimsX=0;iDimsX<nDims;iDimsX++){
dfsdxs[iPts].coeffs_[eV][iDimsX]=-rho_*dfsdxs[iPts].coeffs_[eVp][iDimsX];
dfsdxs[iPts].coeffs_[eV][iDimsX]/=rhobar_;
if ((iPts==nPts_-1) && (iDimsX==eV)) { /* \f$\dot{x}_L=(x_{L-1})^2/\dot{t} - \dot{x}_{L-1}\f$ */
dfsdxdots[((nPts_-1)*nDims+eV)+((nPts_-2)*nDims+eV)*(nDims*nPts_- )]=dfsdxs[iPts].coeffs_ [eV][eV]; } else { for (iDims=0;iDims<nDims;iDims++){
dfsdxdots[(iPts*nDims+iDims)+(iPts*nDims+iDimsX)*(nDims*nPts_- )]=dfsdxs[iPts].coeffs_ [iDims][iDimsX];
}
}
}
return (0);
A.7 Mathematica Code for Chapter 5
DeltaCV[j_, k_] := CrossVar[j, k];
Clear[One]
One[Except[_Slot]] = 1;
(* mergefunctions combines term consisting of multiple
pure functions into single pure function (of common variable)
All functions must be in slot form ie Sin[#1] not Sin *)

MergeFunctions[functional_] :=
  Replace[Function[functional], Function[body_] := body, -1, Heads -> True];

simplifypure = f_Function := Module[ {t1, fsimple, of},
  fsimple = Simplify[f[t1]] /. t1 -> #1;
  If[FreeQ[fsimple, Slot], fsimple := One[#1]];
  (* if reduce to constant use One to keep in Pure Function form *)
  Evaluate[fsimple & & ];

factorconstantspure =
  Module[ {k},
    Table[MergeFunctions[alphaFactor xFactor],
      ReplacePart[alphadW, k -> 0],
      ReplacePart[alphaIntegrand, k -> MergeFunctions[alphaIntegrand]]],
      k, 1, Length[alphaIntegrand]]];

Protect[Times];
Protect[Power];

A.7. Mathematica Code for Chapter 5 167
Unprotect[Times];
Ap[ito[alphaFactor_, alphadW_List, alphaIntegrand_List],
   factor_, n_Integer, integrand_] := ito[factor, Append[alphadW, n],
   Append[alphaIntegrand, MergeFunctions[alphaFactor integrand]]];
Ap[x_ + y_, factor_, n_, integrand_] :=
   Ap[x, factor, n, integrand] + Ap[y, factor, n, integrand];
Ap[m_ * x_, factor_, n_, integrand_] := m Ap[x, factor, n, integrand] /; FreeQ[m, ito];
ito[alphaFactor_, alphadW_List, alphaIntegrand_List] +
   ito[betaFactor_, betadW_List, betaIntegrand_List] :=
      ito[One[1] &, Drop[betadW, -1], Drop[betaIntegrand, -1]],
      MergeFunctions[alphaFactor betaFactor], Last[betadW], Last[betaIntegrand]] +
   Ap[ito[One[1] &, Drop[alphadW, -1], Drop[alphaIntegrand, -1]] *
      ito[One[1] &, betadW, betaIntegrand],
      MergeFunctions[alphaFactor betaFactor], Last[alphadW], Last[alphaIntegrand]] +
   CrossVar[Last[alphadW], Last[betadW]] *
   Ap[ito[One[1] &, Drop[alphadW, -1], Drop[alphaIntegrand, -1]] *
      ito[One[1] &, Drop[betadW, -1], Drop[betaIntegrand, -1]],
      MergeFunctions[alphaFactor betaFactor], 0,
      MergeFunctions[Last[alphadW], Last[betadW]] Last[betaIntegrand]] /;
   (Length[alphadW] > 1 && Length[alphadW] > 1)
Protect[Times];
$RecursionLimit = Infinity;

(* WARNING BELIEVE ABOVE DEFINITIONS SIGNIFICANTLY SLOW DOWN MATHEMATICA,
so reset Times Power once Ito Integrals have been calculated *)
Unprotect[Times, Power]
Times = .
Power = .
{Times, Power}

IterateItoIntegral[dW_List, integrand_List, n_] := Module[{dWTuple, integrandTuple},
   dWTuple = Tuples[dW, n];
   integrandTuple = Tuples[integrand, n];
   Sum[ito[One[1] &, dWTuple[[1]], integrandTuple[[1]]], {i, 1, Length[dWTuple]}]
   (* dW and integrand correspond to a vector
   single ito integral not iterated integral*)

ItoOrder[x_ito] := Length[x[[2]]]
HermiteOrder[x_HermiteHe1] := x[[1]]

Esp[ito[xFactor_, xDList, xIntegrandList], t_] := If[Union[xDList] == {0},
   ito[xFactor, xD, xIntegrand] /. {simplifypure}, 0]
Esp[a_ + b_, t_] := Esp[a, t] + Esp[b, t]
Esp[n_ * ito[xFactor_, xDList, xIntegrandList], t_] :=
   n Esp[ito[xFactor, xD, xIntegrand], t]
CondExp[ito[xFactor_, xdW_List, xIntegrand_List], x_,
  sigma_, dw_List, integrand_List, t_] := If[Union[xdW] == {0},
  Exp[ito[xFactor, xdW, xIntegrand], t],
  Exp[ito[xFactor, xdW, xIntegrand]] * 
  IterateItoIntegral[dw, integrand, Length[xdW] - Count[xdW, 0], t] / 
  sigma^((Length[xdW] - Count[xdW, 0]) /. simplifypure)
HermiteHe1[Length[xdW] - Count[xdW, 0], x, sigma]
CondExp[a_ + b_, x_, sigma_, dw_List, integrand_List, t_] :=
CondExp[a, x, sigma, dw, integrand, t] + CondExp[b, x, sigma, dw, integrand, t]
CondExp[n_ * ito[xFactor_, xdW_List, xIntegrand_List],
  x_, sigma_, dw_List, integrand_List, t_] :=
n CondExp[ito[xFactor, xdW, xIntegrand], x, sigma, dw, integrand, t]
CondExp[ito[xFactor_, xdW_List, xIntegrand_List],
  x_, sigma_, dw_List, integrand_List, t_] := Module[
{nOrder, ititos, iititos, indexdW, ceExp, elemceExp, c1, f1, z1}, If[Union[xdW] == {0},
  Exp[ito[xFactor, xdW, xIntegrand], t] HermiteHe1[0, x, sigma],
  nOrder = Length[xdW] - Count[xdW, 0];
  ititos = IterateItoIntegral[dw, integrand, nOrder];
  If[Length[dw] == 1, ititos = List[ititos];
  (* need to deal with single element case*)
  (* and ensure standard form Plus[itoa, itob,...]*)
  (* we find the position of all ito integrals (ie dw<>0) in xIntegrand and replace
  by product with corresponding ito term of ititos iterated ito integral *)
  (* need to identify zero's and remove? *)
  indexdW = Position[xdW, Except[0], {1}, Heads -> False][[All, 1]];
  ceExp = Array[elemceExp, Length[indexdW]];
  For[iititos = 1, iititos ≤ Length[indexdW], iititos++,
    cl = CrossVar[xdW[[indexdW]], ititos[[iititos, 2]]];
    (* added if statement to cope with zero correlation*)
    ceExp[[iititos]] = If[Count[cl, 0] > 0, 0,
      f1 = Map[MergeFunctions, cl xIntegrand[[indexdW]] ititos[[iititos, 3]]];
      z1 = xIntegrand;
      z1[[indexdW]] = f1;
      ito[xFactor, ConstantArray[0, Length[xdW]], z1]]; 
    ]
  Apply[Plus, ceExp] HermiteHe1[nOrder, x, sigma] / sigma^nOrder]
CondExp[a_ + b_, x_, sigma_, dw_List, integrand_List, t_] :=
CondExp[a, x, sigma, dw, integrand, t] + CondExp[b, x, sigma, dw, integrand, t]
CondExp[n_ * ito[xFactor_, xdW_List, xIntegrand_List],
  x_, sigma_, dw_List, integrand_List, t_] :=
n CondExp[ito[xFactor, xdW, xIntegrand], x, sigma, dw, integrand, t]
IntegrationVariables[xdW_List] := Unique[Map["W" <> ToString[#1] <> "t" &, xdW]]; (* generate Unique integration variables *)

IntegrationVariablesFixed[xdW_List] := Module[{i},
    Table[Symbol["W" <> ToString[xdW[[i]]]] <> "t" <> ToString[i], {i, 1, Length[xdW]}];

IntegrateFoldList[xIntegrationVariables_List, xIntegrand_List, t_] := Module[{tmp},
    Transpose[{xIntegrand, xIntegrationVariables, Append[Drop[xIntegrationVariables, 1], t]}]; (* generate a list of {Integrands, IntegrationVariables, Integration upper limits}*)

ItoIntegrateTransformRule[t_] = 
    Ito[alphaFactor_, alphadW_, alphaIntegrand_] := alphaFactor[t] Fold[
        Integrate[#[1] #2[[1]][#2[[2]]], {#2[[2]], 0, #2[[3]]}] &, 1,
        IntegrateFoldList[IntegrationVariablesFixed[alphadW], alphaIntegrand, t]];

φ[z_, v_] := \[frac{1}{\sqrt{2 \pi}} e^{\frac{-z^2}{2}}

(* these hermite polynomials are orthogonal with weight function e^{-x^2/2} *)

HermiteHe[n_, z_, Sigma_] := Simplify[HermiteH[n, z/Sqrt[2 Sigma]] (Sigma/2)^n/2]
(* Testing correct behaviour of One*)
One[t]
1
One[100]
One[ Sin[t]]
1
1
One[θ1]
Evaluate[One[θ1]]
One[θ1]

% /. simplifypure
One[θ1] &

(* Require all pure functions to be defined with dummy variables *)
(* ie can't use Sin or One but Sin[θ1] & and One[θ1] & *)
(One[θ1] One[θ1]) + One & /. simplifypure
(One[θ1] One[θ1]) (One[θ1]) & /. simplifypure

One[θ1] &

One[θ1] &

ito[One[θ1] &, {1}, {Exp[θ1] &}] + ito[ Sin[θ1] &, {1, 1}, {Cos[θ1] &}, {Tan[θ1] &} ]
ito[One[θ1] Sin[θ1] &, {1, 1, 1}, {Cos[θ1] &}, {Exp[θ1] &}, {Tan[θ1] &} ] +
ito[One[θ1] Sin[θ1] &, {1, 1, 1}, {Cos[θ1] &}, {Tan[θ1] &}, {Exp[θ1] &} ] +
ito[One[θ1] Sin[θ1] &, {1, 1, 1}, {Exp[θ1] &}, {Cos[θ1] &}, {Tan[θ1] &} ]
IterateItToIntegral[{1, 2}, {one, two}, 2]

ito[One[θ1] &, {1, 1}, {one, one}] + ito[One[θ1] &, {1, 2}, {one, two}] +
ito[One[θ1] &, {2, 1}, {two, one}] + ito[One[θ1] &, {2, 2}, {two, two}]
IterateItToIntegral[{1}, {two}, 2]

ito[One[θ1] &, {1, 1}, {two, two}]

ito[1, (1), {a[θ1] &}] + ito[1, (2), {a[θ1] &}]
CrossVar[2, 1] ito[1 &, {0}, {a[θ1] &}] +
ito[1 &, {1, 2}, {a[θ1] &}, {a[θ1] &}] + ito[1 &, {2, 1}, {a[θ1] &}, {a[θ1] &}]
Table[HermiteHe[i, x, 1], {i, 0, 4}]
{1, x, -1 + x^2, -3 x + x^3, 3 - 6 x^2 + x^4}
CrossVar[0, \_\_] = 0;
CrossVar[\_\_, 0] = 0;
CrossVar[\_\_, \_\_] = 1;
CrossVar[1, 2] = \rho;
CrossVar[2, 1] = \rho;
SetAttributes[CrossVar, Listable]
MatrixForm[Table[CrossVar[i, j], {i, 0, 2}, {j, 0, 2}]]

\( \alpha = 1; \)
\( \beta = 1; \)

\[ V[e\_, t\_] := Module[{s, W1s, W2s}, \]
\[ \text{Exp}\[-\kappa t\] v0 + \text{Exp}\[-\kappa t\] \text{Integrate}[\text{Exp}[s] Vpt[e, s], \{s, 0, t\}] + e \xi_1 \text{Exp}\[-\kappa t\] \text{Integrate}[\text{Exp}[s W1s] Vt[e, W1s]^0, \{W1s, 0, t\}]] \] (* using correlated W1, W2*)

\[ Vp[e\_, t\_] := Module[{s, W1s, W2s}, \]
\[ \text{Exp}\[-ct\] v0 + c \text{Exp}\[-ct\] \text{Integrate}[\text{Exp}[c s] z3, \{s, 0, t\}] + e \xi_2 \text{Exp}\[-ct\] \text{Integrate}[\text{Exp}[c W2s] Vpt[e, W2s]^0, \{W2s, 0, t\}] \] (* define delayed rule, so that dummy variables are unique *)

\[ dVpde[i\_Integer, t\_] := Module[{e}, \]
\[ D[V[e, t], \{e, i\}] /. e \to 0; \]

\[ dVpdeIto[1] = \text{ito}[\xi_2 \text{Exp}\[-c \#1\] \&, \{2\}, \{\text{Exp}[c \#1] \text{Vpt}[0, \#1]^\beta \&\}] /. \text{simplifypure} \]
\[ \text{ito}[\text{e}^{-c \#1} \xi_2 \&, \{2\}, \{\text{e}^{-c \#1} \text{Vpt}[0, \#1] \&\}] \]

\[ dVpdeIto[2] = 2 \beta \xi_2^2 \text{ito}[\text{e}^{-c \#1} \&, \{2, 2\}, \{\text{e}^{-c \#1} \text{Vpt}[0, \#1]^\beta \&, \text{Vpt}[0, \#1]^{-1+\beta} \&\}] /. \text{simplifypure} \]

\[ 2 \text{ito}[\text{e}^{-c \#1} \&, \{2, 2\}, \{\text{e}^{-c \#1} \text{Vpt}[0, \#1] \&, \text{One}[\#1] \&\}] \xi_2^2 \]

\[ dVpde[3, t] \]
\[ 3 \text{e}^{-c t} \left( \int_0^t \text{e}^{c W2s + 9856} \text{Vpt}[2, 0] \text{dW2s + 9856} \right) \xi_2 \]

\[ dVpdeIto[3] = \left( \text{Ap}[\text{dVpdeIto[1]}] \right)^2 3 \beta (\beta-1) \xi_2 \text{e}^{-c \#1} \&, 2, \text{e}^{-c \#1} \text{Vpt}[0, \#1]^{-2+\beta} \& \right) + \left( \text{Ap}[\text{dVpdeIto[2]}] \right)^2 3 \beta \xi_2 \text{e}^{-c \#1} \&, 2, \text{e}^{-c \#1} \text{Vpt}[0, \#1]^{-1+\beta} \& \right) + \left( \text{Ap}[\text{dVpdeIto[1]}] \right)^2 3 \beta \xi_2 \text{e}^{-c \#1} \&, 2, \text{e}^{-c \#1} \& \right) + \left( \text{Ap}[\text{dVpdeIto[2]}] \right)^2 3 \beta \xi_2 \text{e}^{-c \#1} \&, 2, \text{e}^{-c \#1} \& \right) + \left( \text{Ap}[\text{dVpdeIto[1]}] \right) \xi_2^3 3 \text{e}^{-c \#1} \xi_2 \&, 2, \text{e}^{-c \#1} \& \right) \]

\[ \text{dnVdeIto[1]} = \text{ito}[\frac{K - c}{K - c} \xi_2 \text{e}^{-c \#1} \&, \{2\}, \{\text{e}^{c \#1} \text{Vpt}[0, \#1] \&\}] - \]
\[ \text{ito}[\frac{K - c}{K - c} \xi_2 \text{e}^{-c \#1} \&, \{2\}, \{\text{e}^{c \#1} \text{Vpt}[0, \#1] \&\}] + \]
\[ \text{ito}[\xi_1 \text{e}^{-c x} \&, \{1\}, \{\text{e}^{c \#1} \text{Vt}[0, \#1] \&\}] \xi_1 + \text{ito}[\text{e}^{-c \#1} \&, \{2\}, \{\text{e}^{-c \#1} \text{Vpt}[0, \#1] \&\}] \xi_2 - \text{ito}[\text{e}^{-c \#1} \&, \{2\}, \{\text{e}^{-c \#1} \text{Vpt}[0, \#1] \&\}] \xi_2 \]

\[ -c + x \]

Printed by Mathematica for Students
\[
\text{dnVdeIto}[2] = 
\begin{align*}
\text{ito}[2 \alpha & \xi_1^2 e^{-\kappa \xi_1} \alpha, (2, 2, 0), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, Vpt[0, \xi_1]^{1-\alpha} \alpha, e^{(\kappa-c) \xi_1} \alpha\}] + \\
& \text{ito}[2 \alpha \xi_1^2 e^{-\kappa \xi_1} \alpha, (2, 1), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, e^{(\kappa-c) \xi_1 Vt[0, \xi_1]^{1-\alpha}} \alpha\}] - \\
& \text{ito}[2 \alpha \xi_1^2 \alpha e^{-\kappa \xi_1} \alpha, (2, 1), \{e^{e^{\xi_1 Vpt[0, \xi_1]} \alpha, Vt[0, \xi_1]^{1-\alpha} \alpha\}] + \\
& \text{ito}[2 e^{-\kappa \xi_1} \alpha \xi_1^2 \alpha, (1, 1), \{e^{e^{\xi_1 Vt[0, \xi_1]} \alpha, Vt[0, \xi_1]^{1-\alpha} \alpha\}]/. \\
& \text{simpifypure} //. \text{factorconstant}\text{pure}
\end{align*}
\]

\[
2 \text{ito}[e^{\xi_1 \alpha}, (1, 1), \{e^{\xi_1 Vt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha\}] \xi_1^2 + \\
2 \text{ito}[e^{\xi_1 \alpha}, (2, 1), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, e^{e^{\xi_1 Vt[0, \xi_1]} \alpha} \alpha\}] \xi_1 \xi_2 - \\
2 \text{ito}[e^{\xi_1 \alpha}, (2, 2, 0), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha, e^{e^{\xi_1 Vt[0, \xi_1]} \alpha} \alpha\}] \xi_1^2 + \\
2 \text{ito}[e^{\xi_1 \alpha}, (2, 2, 2, 0), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha, e^{e^{\xi_1 Vt[0, \xi_1]} \alpha} \alpha\}] \xi_1 \xi_2
\]

\[
\text{dnVdeIto}[3] = \text{Ap} \left[ \text{dnVdeIto}[1]^2 \alpha \xi_1 (\alpha-1) e^{\xi_1 \alpha} \alpha, 1, e^{\xi_1 \alpha} \alpha, Vt[0, \xi_1]^{1-\alpha} \alpha \right] + \text{Ap} \left[ \text{dnVdeIto}[2], 3 \xi_1 e^{\xi_1 \alpha} \alpha, 1, e^{\xi_1 \alpha} \alpha, Vt[0, \xi_1]^{1-\alpha} \alpha \right]/. \text{simpifypure} //. \text{factorconstant}\text{pure}
\]

\[
6 \text{ito}[e^{\xi_1 \alpha}, (1, 1, 1), \{e^{\xi_1 Vt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha, \text{One}[\xi_1] \alpha\}] \xi_1^3 + \\
6 \text{ito}[e^{\xi_1 \alpha}, (2, 1, 1), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, e^{e^{\xi_1 Vt[0, \xi_1]} \alpha} \alpha, \text{One}[\xi_1] \alpha, \text{One}[\xi_1] \alpha\}] \xi_1^2 \xi_2 - \\
\frac{1}{-c + \alpha} \text{ito}[e^{\xi_1 \alpha}, (2, 1, 1), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha, \text{One}[\xi_1] \alpha\}] \xi_1^3 \xi_2 + \\
6 \text{ito}[e^{\xi_1 \alpha}, (2, 2, 0), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha, \text{One}[\xi_1] \alpha, e^{e^{\xi_1 Vt[0, \xi_1]} \alpha} \alpha\}] \xi_1 \xi_2 + \\
6 \text{ito}[e^{\xi_1 \alpha}, (2, 2, 2, 0), \{e^{\xi_1 Vpt[0, \xi_1]} \alpha, \text{One}[\xi_1] \alpha, \text{One}[\xi_1] \alpha, e^{e^{\xi_1 Vt[0, \xi_1]} \alpha} \alpha\}] \xi_1^2 \xi_2
\]

\[
\text{FlIDW} = (1, 2)
\]

\[
\text{FlIntegrand}[t_] = \left\{ a_1 \xi_1 \text{Exp}[-\alpha (t - \xi_1)] Vt[0, \xi_1]^{\alpha} \alpha, \\
\left( a_2 \xi_2 \alpha \right) \frac{\alpha}{\alpha} \text{Exp}[-\alpha (t - \xi_1)] - \text{Exp}[-\alpha (t - \xi_1)] + a_2 \xi_2 \alpha \text{Exp}[-\alpha (t - \xi_1)] \right\} \\
\text{Vpt[0, \xi_1]^{\alpha} \alpha} /./ \text{simpifypure}
\]

\[
\text{FlIntegrandM}[t_] = \left\{ a_1 \xi_1 \text{Exp}[-\alpha (t - \xi_1)] Vt[0, \xi_1]^{\alpha} \alpha, \\
\left( a_2 \frac{\xi_2}{\alpha} \right) \frac{\alpha}{\alpha} \text{Exp}[-\alpha (t - \xi_1)] - \text{Exp}[-\alpha (t - \xi_1)] + a_2 \frac{\xi_2}{\alpha} \text{Exp}[-\alpha (t - \xi_1)] \right\} \\
\text{Vpt[0, \xi_1]^{\alpha} \alpha} /./ \text{simpifypure}
\]

\[
\{1, 2\}
\]

\[
\left\{ e^{-\kappa \xi_1} a_1 \xi_1 Vt[0, \xi_1] \alpha, \\
\left( e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha + e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha \right) \right\} \\
\left\{ e^{-\kappa \xi_1} a_1 \xi_1 Vt[0, \xi_1] \alpha, \\
\left( e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha + e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha \right) \right\} \\
\left\{ e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha, \\
\left( e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha + e^{-\kappa \xi_1} \alpha a_1 \xi_2 Vpt[0, \xi_1] \alpha \right) \right\}
\]
(* How to deal with repeated integrals and prefactor *)
F1Ito[t_] = IterateItoIntegral[F1dW, F1Integrand[t], 1]

ito[One[π1] &, (1), \{E^{-t κ + π1} a1 ξ1 Var[t]; π1 &, (2), \{E^{-t κ + π1} x a1 ξ2 Var[t]; π1 &\} + ito[One[π1] &, (2), \{E^{-t κ + π1} x a1 ξ2 Var[t]; π1 &\}]

F1ItoTest[t_] := ito[a2 ξ2 e^(-π1) &, (2), \{E^{π1} Var[t]; π1 &, (2, 0), \{E^{π1} Var[t]; π1 &, (π-κ) π1 &, \} + ito[a2 e^{-π1} κ ξ1 &, (1), \{E^{π1} Var[t]; π1 &\}]

(* these are the deterministic limits of v and v' *)
vt0[tb_] = e^{-κ t} v0 + κ (v0 - z3) (e^{-κ t} - e^{-κ t}) + (1 - e^{-κ t}) z3;
vpt0[ta_] = e^{-κ t} (v0 + (-1 + e^{-κ t}) z3);

VptVtRule1 = VptVtRule /. DCEVParams1;
VptVtRule = VptVtRule /. DCEVParams;

subVpVCRules = \{VpC0 \to z3, VpCC \to v0 - z3, VCC \to z3,
VCC \to \frac{K}{K - \xi}, VCC \to \frac{K}{K - \xi}, kdcmc \to \frac{K}{K - \xi}\};

Order2CoefficientList = \{(0, 0, 1, 0, 0, 0), \};
Order3CoefficientList = \{(0, 0, 2, 0, 0, 0), \};
SetDirectory[
  "C:\Users\sv507\Documents\Sean\work\PhD\PhDMathematica\itoform\data2"]

C:\Users\sv507\Documents\Sean\work\PhD\PhDMathematica\itoform\data2

(* decide on order 3 or order 3 list*)

OrderXCoefficientList = Order3CoefficientList

\{(0, 0, 2, 0, 0, 0), (0, 0, 1, 0, 0, 0), 
 (0, 0, 0, 2, 0, 0), (0, 0, 0, 0, 1, 0), (0, 0, 0, 0, 0, 1)\}

Timing[For[i = 1, i <= Length[OrderXCoefficientList], i++,
  nm = "CE1List" <> ToString[#1] & /@ OrderXCoefficientList[[i]]];
  Print[nm];
  Symbol[nm][t_] = CondExp1[
    Times @@ ((v1[1], v2[1], v1[2], v2[2], v1[3], v2[3])^OrderXCoefficientList[[i]]) //.
    {v1 -> dnpdeIto, v2 -> dnpdeIt, simplifyPure} //.
    factorconstantspure, x, 1, F1dW, F1IntegrandN[t], t];
  Save[nm <> ".nb", Evaluate[Symbol[nm]]]
  Clear[Symbol[nm]];
]

\{4.446, Null\}

CE1List000001

CE1List000010

CE1List000200

CE1List001100

CE1List002000

CE1List001000

CE1List000100

\{0.015, Null\}

(* save order 3 integral rules to NM Int *)

(* takes about 5 minutes for 03*)

Timing[For[i = 1, i <= Length[OrderXCoefficientList], i++,
  nm = "CE1List" <> ToString[#1] & /@ OrderXCoefficientList[[i]]];
  Get[nm <> ".nb"];
  Print[nm];
  Symbol[nm]"Int"[t_] =
    Symbol[nm][t] //.
    {VpRule, VtRule, ItoIntegrateTransformRule[t]};
  DumpSave[nm <> "Int" <> ".mx", Evaluate[Symbol[nm] <> "Int"]];
  Clear[Evaluate[Symbol[nm] <> "Int"]];
  Clear[Evaluate[Symbol[nm]]];
]

\{133.927, Null\}
(* -5 minutes if timese/power defn cleared - 40 minutes otherwise!!*)

Timing[For[i = 1, i ≤ Length[OrderXCoefficientList], i++,
   nm = "CE1List" <> ToString[N1] & OrderXCoefficientList[[i]] <> "Int";
   Get[nm <> ".mx"];
   Print[nm];
   A = Symbol[nm][t] /. subExp[t, e];
   (* need simplify to pull out common factor from divisors etc... how to use factor?*)
   (* subexp needed to factor out "constant" exponential terms, ie e^k t*)
   Evaluate[Symbol[Evaluate[nm <> "S"]]] =
   (A /. FactorConstantsIntegrate1 /. e[x_] :> Exp[x t]);
   DumpSave[nm <> "S.mx", Evaluate[nm <> "S"]]
   Remove[Evaluate[nm <> "S"]];
]

CE1List002000Int
CE1List001100Int
CE1List000200Int
CE1List000010Int
CE1List000001Int

Clear[nmvar]

(* now do replace ..array for order 3 coeffs *)
Timing[For[i = 1, i ≤ Length[OrderXCoefficientList], i ++,
        nm = "CE1List" <> ToString[OrderXCoefficientList[[i]]] <> "IntS";
        Get[nm <> ".mx"]; Clear[A1];
        A1 = Symbol[nm];
        heOrder = Union[Cases[A1, HermiteHe1[z_, x, 1] → z, ∞, Heads → True]]; Print[nm, heOrder];
        For[k = 1, k ≤ Length[heOrder], k ++,
            Clear[A];
            nmvar = nm <> "FastHe" <> ToString[heOrder[[k]]];
            A2 = Coefficient[A1, HermiteHe1[heOrder[[k]], x, 1]]; A[1] = ReplaceExponentialIntByArray[A2, W0t1]; Clear[Evaluate[nmvar]]; (* if set then cannot assign!! evaluate returns the value of the symbol *)
            maxindex = 2;
            While[! FreeQ[A2, Symbol["W0t" <> ToString[maxindex]]], maxindex++];
            Evaluate[Symbol[nmvar]] = Reap[
                For[j = 2, j < maxindex, j ++,
                    Sow[A[j - 1][[2, 1]]]; (* Reap returns {ans, {a, c, d} → [2, 1]} *, so {a, c, d} = ..[[2, 1]]*)
                    A[j] = ReplaceExponentialIntByArray[A[j - 1][[1]],
                        Symbol["W0t" <> ToString[j]]];
                    A[j - 1] =.; (* Clear memory *)
                    ClearSystemCache[];
                    Print[MemoryInUse[]];
                ];
                Sow[A[maxindex - 1][[2, 1]]];
                A[maxindex - 1][[1]]
            ];
            Save[nmvar <> ".mx", Evaluate[nmvar]]; Clear[Evaluate[nmvar]];]
        Clear[Evaluate[nm]];]
];
matchvars = (x_); StringMatchQ[SymbolName[Head[x]], "vars" -> NumberString]);

CleanCString[mathc_] :=
StringReplace[ToString[CForm[mathc]],
   {"allpowvar(" -> "allpowvar(" <> x <> "")",
    "vars" -> x : NumberString "(")" -> y,
    "StdF1(t)" -> "stdF1", "κ" -> "k", "o" -> "x",
    "Subscript(a,1)" -> "a1", "Subscript(a,2)" -> "a2", "Subscript(a,3)" -> "a3",
    "Subscript(ξ,1)" -> "ξ1", "Subscript(ξ,2)" -> "ξ2"};

CRule1[Rule[a_, c_]] := ToString[a] <> " = " <> ToString[CForm[c]];  
CRepRule[Rule[a_, c_]] := StringReplace[ToString[a], 
   {"$" -> "_", "[" -> ", "]" -> "]"}] <> 
   " = " <> CleanCString[c] <> " ;\\n";

(* RemoveArray[a_] := StringReplace[ToString[a], 
   {"$" -> "_", "[" -> ", "]" -> "]"}] *)  
(* CRuleNoArray[Rule[a_, c_]] := RemoveArray[a] <> " = " <> CleanCString[c] <> " ;\\n" *)

subVpVCRules = {VpC0 -> z3, VpCC -> vp0 - z3, VCC0 -> z3, 
   VCK -> v0 - z3 - (vp0 - z3) \[Kappa]\[Kappa] - c, 
   VCC -> (vp0 - z3) \[Kappa]\[Kappa] - c, 
   kdkmc -> \[Kappa]\[Kappa] - c};

(* for visual studio vars $[0-9]+$ *)

(* Load Integration definitions *)
Clear[Evaluate[nn]]

nm = "CE1List000100IntSFastHe2"
Print[nm];
Get[nm <> ".mx"];
Head[Symbol[nm][[1]]]

CE1List000100IntSFastHe2

(* Identify variables shared between integrations 
and common computations ( eg exp[kappa T] T^n *)

Clear[varflat, func, vargather, varunique, vardup, varreplace]

Printed by Mathematica for Students
```mathematica
varflat = Flatten[Symbol[nm][[2, 1, 1]]];
func[0] = Symbol[nm][[1]]; Print[Length.Symbol[nm][[2, 1]]];
For[i = 1, i <= Length.Symbol[nm][[2, 1]], i++,
  Print[i];
  vargather[i] = GatherBy[varflat, Last];
  varunique[i] = vargather[i][[All, 1]];
  vardup[i] = Drop[vargather[i], None, 1][[All, All, 1]];
  varreplace[i] = Flatten[Table[Map[Rule[., varunique[i][[j]]], vardup[i][[j]]], {j, 1, Length[vardup[i]]}]];
  vDisp = Dispatch[varreplace[i]];
  If[i < Length.Symbol[nm][[2, 1]],
    varflat = Flatten[Symbol[nm][[2, 1, i + 1]]] /. vDisp;
    func[i] = func[i - 1] /. vDisp;
  ]
  varpow = Union[Flatten[Table[Cases[varunique[i][[All, 2]], Power[a_, b_, _], _]], {i, 1, Length[Symbol[nm][[2, 1]]]}];
  funcpow = Union[Cases[func[Length[Symbol[nm][[2, 1]]]], Power[a_, b_, _], _]];
  Length[funcpow]
  allpow = Union[Join[varpow, funcpow]];
  allpowreplace = Table[Rule[allpow[All], allpowvar[i]], {i, 1, Length[allpow]}];
  DumpSave[nm -> "resultsB.mx",
    {func, vargather, varunique, vardup, varreplace, varpow, funcpow, allpowreplace}];
```

(* using simplified definitions,
create C file to compute particular hermite order coefficient *)

Save[nm -> "resultsB.nb",
    {func, vargather, varunique, vardup, varreplace, varpow, funcpow, allpowreplace}]
sfile = OpenWrite[nm <> " .cpp"];
WriteString[sfile,
    "#include<math.h>\n",
    "\n\n",
    "double " <> nm <>
      "(double stdF1, double k, double c, double xi1, double xi2, double r, double z3, double v0, double vp0, double a1, double a2, double a3, double a4, double t)\n",
    "double VpC0=z3;\n",
    "double VpCC=vp0-z3;\n",
    "double VC0=z3;\n",
    "double VCK=v0-z3-((vp0-z3)*k)/(-c+k);\n",
    "double kdkmc=k/(-c+k);\n",
    "double val=0;\n",
    "double allpowvar=" <> ToString[Length[allpowreplace] + 1] <> "] ;\n"
];
Scan[WriteString[sfile, CRepRule[\n]] & , allpowreplace];

nInt = Length[Symbol[nm][[2, 1]]];
(* why should it be same for each Integral?*)
For[ iInt = nInt, iInt > 0, iInt -- ,
    Scan[WriteString[sfile,
      "double " <> StringReplaceToString[\n], {"$" -> ", "["\n" -> ", "["\n""] <> "] ;\n"
    & , varunique[iInt][All, 1]];]
    WriteString[sfile, "\n"]
];
For[ iInt = 1, iInt ≤ nInt, iInt ++ ,
    Scan[WriteString[sfile, CRule[\n]] & , varunique[iInt] /. allpowreplace];
    WriteString[sfile, "\n"]
];
WriteString[sfile, "\n// now define function\n"]
If[Head[func[nInt]] == Plus,
    Scan[WriteString[sfile, "val=" <> CleanCString[\n] <> "],
        func[nInt] /. allpowreplace],
    WriteString[sfile, "val=" <> CleanCString[func[nInt] /. allpowreplace] <> "] ;\n"
];
WriteString[sfile, "return val;\n"]
];
Close[sfile];