Identification in Financial Models with Time-Dependent Volatility and Stochastic Drift Components
Contents

1 Introduction 7

2 Stochastic Preliminaries 10

3 The Bivariate Ornstein-Uhlenbeck model 22
   3.1 Solution of the stochastic differential equation . . . . . . . . . . . . . . . . . . 24
   3.2 Pricing of European Call Options . . . . . . . . . . . . . . . . . . . . . . . . 26

4 Volatility estimation by wavelet methods 28
   4.1 Introduction to wavelet analysis . . . . . . . . . . . . . . . . . . . . . . . . 28
   4.2 The situation and the estimator . . . . . . . . . . . . . . . . . . . . . . . . 34
   4.3 Asymptotic study of the estimator . . . . . . . . . . . . . . . . . . . . . . . 36
      4.3.1 Weak convergence of the estimator . . . . . . . . . . . . . . . . . . . 36
      4.3.2 Mean integrated square error . . . . . . . . . . . . . . . . . . . . . . 46
   4.4 Numerical case studies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
      4.4.1 The L-method as criteria for the choice of the resolution level . . . . . . 62
   4.5 Outlook: Wavelet thresholding . . . . . . . . . . . . . . . . . . . . . . . . . 68

5 Operator equations with Nemytskii operators 70
   5.1 Inverse problems and regularization methods . . . . . . . . . . . . . . . . . . . 70
   5.2 Nemytskii operators: Acting conditions and continuity . . . . . . . . . . . . 76
   5.3 Nemytskii operators with monotone generator functions . . . . . . . . . . . . 79
6 Identification of the time-dependent volatility using option prices

6.1 Inverse option pricing: Tikhonov-Regularization

6.2 The outer problem

6.2.1 Analytical studies

6.2.2 Numerical case studies concerning ill-conditioning effects

6.2.3 Regularization by monotonization – Algorithm

6.2.4 Discrete Variant

7 Identification of the drift parameters

7.1 Preliminary considerations

7.2 State space representation and Kalman filter

7.3 Likelihood function

7.4 Numerical Case Studies
Notations

Spaces

$\mathcal{P}_n$ polynomials of degree $\leq n$

$L^p(I), L^p(I; \mathbb{R}^d)$ Lebesgue spaces of $p$-power integrable functions

$L^p(I)$ contains functions $f : I \to \mathbb{R}$, whereas

$L^p(I; \mathbb{R}^d)$ contains functions $f : I \to \mathbb{R}^d$

$C^k(I)$ space of $k$-times continuously differentiable functions $f : I \to \mathbb{R}$

If $I = [a, b]$ we also write $C[a, b]$.

Furthermore, we use the abbreviation $C$ when appropriate.

$l^2(\mathbb{Z}) := \{\{x_k\}_{k \in \mathbb{Z}} : x_k \in \mathbb{R} \text{ and } \sum_{k \in \mathbb{Z}} x_k^2 < \infty\}$

$l^2(\mathbb{N}) := \{\{x_k\}_{k \in \mathbb{N}} : x_k \in \mathbb{R} \text{ and } \sum_{k \in \mathbb{N}} x_k^2 < \infty\}$

$H^s(\mathbb{R})$ Sobolev spaces cf. (4.5)

$\langle \cdot , \cdot \rangle_H$ scalar product in the Hilbert space $H$

Functions

$[r]$ entier function ($[r]$ is the largest integer that is less or equal than $r$)

$\chi_A$ indicator function of the set $A$, i.e. $\chi_A(s) := \begin{cases} 1 & s \in A \\ 0 & s \notin A \end{cases}$

Operators

$\mathcal{F}f(\xi) = \hat{f}(\xi) := \int_{-\infty}^{\infty} e^{-ix\xi} f(x) \, dx$ Fourier transform of $f$

$F^*$ adjoint operator of $F$
Norms

For a matrix $A = (a_{i,j})_{i,j=1}^{m,n}$ we set $|A|^2 := \sum_{i,j} |a_{i,j}|^2$

$$\|f\|_{C(I)} := \max_{x \in I} |f(x)|$$

$$\|f\|_{L^p(I)} := (\int_I |f(x)|^p \, dx)^{\frac{1}{p}}$$

Miscellaneous

$\wedge \quad r_1 \land r_2 := \min\{r_1, r_2\}$

$\oplus \quad$ orthogonal sum

$\# \quad$ cardinality

Vectors and Matrices

$I \quad$ identity matrix, $I = (\delta_{i,j})_{i,j=1}^n$

$0 \quad$ vector that contains only zeros

Stochastics

$E \xi$ 

$D^2 \xi := E(\xi - E\xi)(\xi - E\xi)^T$ 

$\text{Cov} (\xi, \eta) := E(\xi - E\xi)(\eta - E\eta)^T$ 

$\mathcal{A}(\{\xi_c\}_{c \in \mathcal{C}})$ 

$\xi \sim \mathcal{N}(m, R)$ 

$F_\xi(x)$ 

$\phi_\xi(t) = \int_{\mathbb{R}^n} e^{it \cdot x} dF_\xi(x)$ 

$\xi_n = o_P(1)$

Expectation of the random vector $\xi$

The sigma-algebra generated by the random variables $\xi_c \ (c \in \mathcal{C})$

$\xi$ is normally distributed with expectation $m$ and covariance matrix $R$

distribution function of the random vector $\xi$

characteristic function of the random vector $\xi$

$\{\xi_n\}$ converges to zero in probability, i.e. $\forall \varepsilon > 0$ it holds $P(|\xi_n| > \varepsilon) \to 0$ for $n \to \infty$.

Equivalences

If $A(u)$ and $B(u)$ are positive functions of a set of parameters, the notation

$$A(u) \lesssim B(u)$$
means, that there exists a constant $C > 0$ such that $A(u) \leq C B(u)$ independently of $u$. Furthermore, the notation

$$A(u) \sim B(u)$$

means $A(u) \lesssim B(u)$ and $B(u) \lesssim A(u)$. Furthermore, we use the Landau symbols $O$ and $o$ to describe the asymptotic behavior of functions. To be precise, for two functions $f$ and $g$ we write

$$f(x) = O(g(x)) \quad \text{for} \quad x \to \infty$$

if and only if there exists an $x_0$ and a constant $M > 0$ such that

$$|f(x)| \leq M |g(x)| \quad \text{for} \quad x > x_0.$$  

Besides, we write

$$f(x) = o(g(x)) \quad \text{for} \quad x \to \infty$$

if and only if $\frac{f(x)}{g(x)} \to 0$ for $x \to \infty$.

In order to distinguish between results that are cited from the literature and own contributions we use the term "proposition" when we reformulate facts that are found in the literature. As opposed to that "lemmas" and "theorems" state and prove assertions that we could not find in the literature. As usual we use the term "lemma" for auxiliary results that are mainly used to prove a theorem.
Chapter 1

Introduction

During the last decades a great diversity of price models for financial assets has been developed. It is well-known that as long as it is only possible to observe asset prices (or the corresponding returns) in a discrete scheme, it is always possible to find a model based on a geometric Brownian motion with constant volatility coefficient and stochastic drift term which has identical distributions as the observed returns (cf., e.g. [46]). Clearly, due to this fact one must not argue that the empirically observed returns which fail to have independent normal distributions require extensions of the classical model in order to price options accurately.

On the other hand it is obvious that by introducing further random effects into the corresponding models via a drift for a given (fixed) behaviour of the observed data there are changes in option prices, even though the option price formula itself is unaffected by changes in the drift. Consequently, the study of corresponding models is meaningful. In this context the estimate of volatility has to be reinterpreted in the light of the specific model which is assumed.

Speaking generally, these models are based on stochastic processes which are specified by several model parameters and these parameters have to be calibrated to observed market data. Obviously, a correct identification of these parameters is of core importance as otherwise the models do not yield a good approximation of the real price processes. Moreover, the model parameters are also necessary for pricing derivatives. A computation of these prices with wrong parameters can lead to results which are far away from the prices observed on real markets even if the correct formulas have been used.

Here and in what follows the term parameter means either a finite dimensional vector or a function that specifies a model. This manner of speech is common in the literature of inverse problems but it differs from the statistical literature. There, this term is generally used in the meaning of a finite dimensional parameter vector. Consequently, the branch of statistics which is concerned with the identification of finite dimensional parameter vectors is called parametric statistics, whereas nonparametric statistics aims at the calibration of models containing unknown functions (for example a volatility function). If we want to stress that certain parameters are finite dimensional, we speak of finite dimensional parameter vectors or real-valued parameters.
The focus of this thesis is on parameter identification in market models with partial information in which the stochastic drift of the logarithmic asset price process depends on an unobservable state process. The asset price process is assumed to have a time-dependent but deterministic volatility, which has to be identified. Furthermore, the stochastic drift and the underlying state process are characterised by a finite number of real-valued parameters which are assumed to be constant with respect to time. The aim is an analysis of several calibration techniques which are suitable for the identification of the described parameters. As a toy example we consider a slightly modified version of the Bivariate Trending Ornstein-Uhlenbeck model which has been introduced by Lo and Wang in [37].

In the literature several calibration techniques are discussed. Speaking generally, there are on the one hand statistical approaches which aim at estimating the parameters from observed asset prices. On the other hand there are approaches which use prices of observed derivatives (e.g. observed option price data). In general, the last approach leads to inverse problems.

With respect to the calibration of a time-dependent volatility function there exist methods of nonparametric statistics which use high-frequency asset price data. As an example we consider the method of wavelets, which performs a projection on an orthonormal wavelet basis (cf., e.g. [7], [16], [45, p. 268ff]). With this approach the volatility function can be identified on the time interval on which asset prices are observed, i.e. on a time interval in the past.

However, for pricing options and other derivatives the volatility function (or deviated quantities) has to be known on a time interval \([t_{\text{now}}, t_{\text{future}}]\) starting at the current time point \(t_{\text{now}}\) and ending at some future time instant \(t_{\text{future}}\). To calibrate the volatility over this time interval one can observe prices of options with maturities varying in \([t_{\text{now}}, t_{\text{future}}]\). In this case the identification leads to the inverse problem of option pricing (cf., e.g. [9]), which is known to be ill-posed, i.e. the solution does not depend continuously on the data. In order to overcome these ill-posedness effects many papers have been concerned with the applicability of several regularization methods (cf., e.g. [11], [23], [27]).

Furthermore, with respect to optimisation of the utility from terminal wealth (cf., e.g. [34] in the context of partial information) the necessity of a proper identification of all model parameters is obvious. In other words, for utility optimisation the real-valued parameters that characterise the stochastic drift terms have to be identified too. For a given volatility function this can be done by maximum likelihood estimation (cf., e.g. [20], [36] and [39] for a general introduction). Clearly, if the volatility input is itself a result of the above-mentioned estimation methods, it becomes necessary to discuss the question which effects are caused by small errors in this input.

This thesis combines the above-mentioned approaches. To keep notations consistent with the literature it is inevitable to use the time interval \([0, T]\) in different meanings. To be precise, in the chapters that are concerned with statistical methods the notation \([0, T]\) is used for some time interval in the past. In this situation \(T\) denotes the current time point and 0 the instant where the observation of the asset prices started. As opposed to that, in the context of the inverse problem of option pricing the interval \([0, T]\) denotes some future time interval starting at the current time point \(t_{\text{now}} = 0\).

The thesis is organised as follows. In Chapter 2 we repeat some basic concepts and propositions from probability theory which are used at several places throughout the thesis. In
Chapter 3 we introduce a generalised Ornstein-Uhlenbeck model as a toy example for our numerical case studies. After the presentation of the model and a short motivation we prove unique solvability of the corresponding system of stochastic differential equations for the logarithmic price-process and review results concerning option pricing in this model.

In Chapter 4 we present a nonparametric estimator of the squared volatility function which performs a projection onto an orthonormal wavelet basis. We start with a short introduction into wavelet analysis. After that we generalise convergence results for the considered estimator to the situation of market models with incomplete information in which the stochastic drift depends on an unobservable (possibly multidimensional) state process. Convergence is studied in the weak sense as well as in terms of the mean integrated square error. Moreover, for the mean integrated square error a convergence rate is proven. This rate is also illustrated by means of a numerical case study. Furthermore, the data-driven choice of the resolution level according to the L-method is discussed.

Chapter 5 is devoted to inverse problems that can be formulated as operator equations with Nemytskii operators. After a general introduction into inverse problems and regularization methods we review results about Nemytskii operators. Unfortunately, the literature is only concerned with properties of the Nemytskii operators itself (such as acting conditions, continuity, differentiability) but does not address questions about existence and properties of the corresponding inverse operators. Restricting our considerations to a certain type of Nemytskii operators, namely Nemytskii operators generated by monotone functions we answer some of these open questions in Section 5.3.

Chapter 6 addresses the calibration of the time-dependent volatility function $\sigma$ (or deviated quantities) from observed option price data. In a first part we review results concerning the applicability of Tikhonov regularization to the inverse problem of option pricing, which is concerned with the identification of the squared volatility function $a(t) := \sigma^2(t)$ ($t \in [0,T]$). After that we concentrate on the identification of an antiderivative $S$ of $a$, which leads to an operator equation with a Nemytskii operator generated by a monotone function. Applying results of Chapter 5 we prove well-posedness in a $C$-space setting and discuss ill-conditioning effects which lead to strongly oscillating solutions. Therefore, we discuss the applicability of monotonicity information for stabilising the solution process. As a result we propose a numerically effective algorithm for the computation of a strictly monotonically increasing solution and illustrate its performance by means of a numerical case study.

Finally, Chapter 7 is concerned with the estimation of the real-valued parameters in the considered generalised Ornstein-Uhlenbeck model. Replacing the unobservable state process by a scaled version we can reduce the number of parameters. After deriving the state space representation of the considered model we use the Kalman filter to obtain certain conditional expectation vectors. Using these quantities we can set up the log-likelihood function. Furthermore, we present a short numerical case study and discuss briefly chances and limitations of the method. In this context the effects of small noise in the volatility input are investigated.
Chapter 2

Stochastic Preliminaries

This chapter is intended as survey over a wide range of stochastic topics which will be used throughout the thesis. We start with an elementary inequality and properties of Gaussian random variables. After that we move on to the central limit theorem, which is used in the proof of Theorem 4.3.2. Next, we introduce random processes, especially the Wiener process with respect to a filtration. In order to be brief we abstain from defining the Stochastic Itô integral and refer simply to the literature, for instance [30]. Nevertheless, we define the Stochastic differential and formulate two versions of the Itô formula which are used quite frequently in Chapter 4.

Lemma 2.1.1
Let \( \xi_i \) \((i = 1, \ldots, n)\) be random variables with finite variances. Then it holds

\[
\mathbb{D}^2 \left( \sum_{i=1}^{n} \xi_i \right) \leq n \sum_{i=1}^{n} \mathbb{D}^2 \xi_i. \tag{2.1}
\]

Proof: Cauchy-Schwartz inequality gives

\[
\left( \sum_{i=1}^{n} \frac{1}{n} (\xi_i - \mathbb{E} \xi_i) \right)^2 \leq \left( \sum_{i=1}^{n} \frac{1}{n^2} \right) \left( \sum_{i=1}^{n} (\xi_i - \mathbb{E} \xi_i)^2 \right).
\]

Hence we have

\[
\frac{1}{n^2} \mathbb{D}^2 \left( \sum_{i=1}^{n} \xi_i \right) = \mathbb{E} \left( \sum_{i=1}^{n} \frac{1}{n} (\xi_i - \mathbb{E} \xi_i) \right)^2 \leq \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \mathbb{E} \xi_i)^2 \right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{D}^2 \xi_i.
\]

Multiplying both sides by \( n^2 \) gives the assertion.

Properties of Gaussian random variables

Gaussian random variables play an extremely important role in probability theory and consequently in mathematical finance. One reason for this is that they have properties that
2. STOCHASTIC PRELIMINARIES

make them nicely treatable in a mathematical way, for example the sum of the components of a normally distributed random vector is normally distributed. Besides, the components of a normally distributed random vector are independent if and only if they are uncorrelated.

We start this paragraph by defining Gaussian random vectors in terms of the characteristic function. Then we formulate a proposition concerning expectation, variance and density of Gaussian random vectors. Finally we concentrate ourselves onto moments of one-dimensional Gaussian random variables and deviated random variables, as they will occur in the proof of Theorem 4.3.2.

**Definition 2.1.2 (Gaussian distribution)**
A random vector \( \xi = (\xi_1, \ldots, \xi_n)^T \) is called Gaussian or normally distributed if its characteristic function has the form
\[
\phi_\xi(t) = e^{i<t, m>-\frac{1}{2}t^T R t},
\]
where \( m \in \mathbb{R}^n \) is a vector, \( R \in \mathbb{R}^{n \times n} \) is a symmetric nonnegative definite matrix and \( \langle \cdot, \cdot \rangle \) denotes the scalar product in \( \mathbb{R}^n \). We use the abbreviation \( \xi \sim \mathcal{N}(m, R) \).

**Corollary 2.1.3**
Let \( \xi \sim \mathcal{N}(m, R) \) with \( m = (m_1, \ldots, m_n)^T \) and \( R = (r_{kl})_{k,l=1}^n \). Then the expectation and covariances of the components of \( \xi \) are given by
\[
\mathbb{E}_\xi \xi_k = m_k \quad \text{Cov} \left( \xi_k, \xi_l \right) = r_{kl} \quad (k, l = 1, \ldots, n).
\]
Furthermore, if \( R \) is nonsingular we can define the inverse \( A = R^{-1} \). In this situation the random vector has the density
\[
f(x) = \sqrt{\frac{\det A}{(2\pi)^{n/2}}} \exp \left\{ -\frac{1}{2} \langle A(x - m), (x - m) \rangle \right\}
\]
The distribution function of a \( \mathcal{N}(0,1) \)-distributed random variable is denoted by \( \Phi \). Hence,
\[
\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.
\]

In [35, p. 19] and [42, Section 2.2] we find the following results concerning the moments of a centred, normally distributed random variable.

**Proposition 2.1.4**
Let \( X \sim \mathcal{N}(0, \sigma^2) \). Then it holds
\[
\mathbb{E} X^{2k} = (2k-1)!! \sigma^{2k} \quad \text{and} \quad \mathbb{E} X^{2k+1} = 0 \quad \forall k \in \mathbb{N}.
\]
Thereby we use the notation \((2k-1)!! = \prod_{n=1}^{k} (2n-1)\).
**Corollary 2.1.5**

Let $X \sim \mathcal{N}(0, \sigma^2)$ and $Y = X^2 - \mathbb{E}X^2 = X^2 - \sigma^2$. Then the moments of $Y$ satisfy

\begin{align*}
\mathbb{E}Y &= 0, \quad (2.3a) \\
\mathbb{D}^2Y &= 2\sigma^4 \quad \text{and} \\
\mathbb{E}|Y|^3 &\leq 28\sigma^6. \quad (2.3b)
\end{align*}

**Proof:** We start by the remark that Proposition 2.1.4 gives especially

\begin{align*}
\mathbb{E}X^2 &= \sigma^2, \\
\mathbb{E}X^4 &= 3\sigma^4 \quad \text{and} \\
\mathbb{E}X^6 &= 15\sigma^6.
\end{align*}

Equation (2.3a) is obvious. Furthermore, we have

\begin{align*}
\mathbb{D}^2Y &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = \mathbb{E}(X^4 - 2\sigma^2X^2 + \sigma^4) - 0 \\
&= \mathbb{E}X^4 - 2\sigma^2\mathbb{E}X^2 + \sigma^4 = 3\sigma^4 - 2\sigma^4 + \sigma^4 = 2\sigma^4.
\end{align*}

Finally, we get the following upper bound

\begin{align*}
\mathbb{E}|Y|^3 &= \mathbb{E}|X^2 - \sigma^2|^3 = \mathbb{E}|X^6 - 3X^4\sigma^2 + 3X^2\sigma^4 - \sigma^6| \\
&\leq \mathbb{E}X^6 + 3\sigma^2\mathbb{E}X^4 + 3\sigma^4\mathbb{E}X^2 + \sigma^6 \\
&= 15\sigma^6 + 9\sigma^6 + 3\sigma^6 + \sigma^6 = 28\sigma^6.
\end{align*}

\[\blacksquare\]

**Central limit theorem**

In Section 4.3 we will apply the central limit theorem for sums of independent random variables, which states roughly speaking that the distribution of these sums is closely approximated by a normal distribution. For the convenience of the reader we give here the Lindeberg theorem with the sufficient Lyapunov condition. For the proofs see [6, p. 114] and [48, p. 329 ff.].

We consider the following situation. For each $n \in \mathbb{N}$ let $\xi_{i,n}, \ldots, \xi_{n,n}$ be independent random variables with

\[\mathbb{E}\xi_{i,n} = 0 \quad (i = 1, \ldots, n)\]

and

\[\sigma^2_{i,n} = \mathbb{D}^2\xi_{i,n} \in (0, \infty).\]

We set

\[D_n := \sqrt{\sum_{i=1}^{n} \sigma^2_{i,n}}.\]

First we formulate the central limit theorem with Lindeberg condition.
Proposition 2.1.6
If for all $\varepsilon > 0$

$$\lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{E} \left( \left( \frac{\xi_{i,n}}{D_n} \right)^2 \chi_{|\xi_{i,n}| \geq \varepsilon} \right) = 0$$

(2.4)

holds, then

$$\frac{1}{D_n} \sum_{i=1}^{n} \xi_{i,n}$$

converges in distribution to a $\mathcal{N}(0,1)$ variable, i.e. for all $t \in \mathbb{R}$ it holds

$$\mathbb{P} \left( \frac{1}{D_n} \sum_{i=1}^{n} \xi_{i,n} < t \right) \to \Phi(t) \quad \text{for } n \to \infty.$$  

(2.5)

The next corollary states that the Lyapunov condition is sufficient for the Lindeberg condition.

Corollary 2.1.7
Let for some $\delta > 0$ and all $n \in \mathbb{N}$ the expectations $\mathbb{E}|\xi_{i,n}|^{2+\delta} < \infty$ exist for $i = 1, \ldots, n$. Let furthermore

$$\frac{1}{D_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E} |\xi_{i,n}|^{2+\delta} \to 0 \quad \text{for } n \to \infty.$$ 

Then (2.4) holds with any $\varepsilon > 0$. Thus, $\frac{1}{D_n} \sum_{i=1}^{n} \xi_{i,n}$ converges in distribution to a $\mathcal{N}(0,1)$ variable.

Random Processes

Modern mathematical finance is concerned with the development of mathematical models that describe the evolution of asset prices. For continuous time models, which are considered in this thesis, a main tool is the concept of random processes. Therefore, we will consider this and some closely related concepts in this section.

To model the uncertainties we introduce a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A subset $A \in \mathcal{F}$ is called null set if $\mathbb{P}(A) = 0$. We will call the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ complete if every subset of a null set is measurable.

Definition 2.1.8 (Random process)
Let $\mathcal{I}$ be a subset of the real line. A family of random variables $\xi = (\xi_t)_{t \in \mathcal{I}}$ is called random process with time domain $\mathcal{I}$. Note, that for each fixed $t \in \mathcal{I}$ the mapping

$$\omega \mapsto \xi_t(\omega)$$

is a random variable. On the other hand, fixing $\omega \in \Omega$ defines a function $(\xi_t)_{t \in \mathcal{I}}$, which is called realization or trajectory of the process, corresponding to the outcome $\omega$. Thus, a stochastic process can be considered as a function of two variables $t$ and $\omega$. To emphasise this fact also the notation $\xi(t,\omega)$ can be found in the literature. However, in order to keep notation simple we will write $\xi_t(\omega)$ or just $\xi_t$.  


Definition 2.1.9 (Filtration)
A filtration is a family \( \{F_t\}_{t \geq 0} \) of non-decreasing sub-\( \sigma \)-algebras of \( \mathcal{F} \) (i.e. \( F_t \subset F_s \subset \mathcal{F} \) for all \( 0 \leq t < s < \infty \)).

Definition 2.1.10 (Brownian Motion)
Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. A (standard) Brownian Motion is a real-valued process \((B_t)_{t \geq 0}\) with the following properties:

1. \( B_0 = 0 \) a.s.
2. The increments of \( B_t \) are independent, i.e. for any finite set of times \( 0 \leq t_1 < t_2 < \cdots < t_n < T \) the random variables
   \[ B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_n} - B_{t_{n-1}} \]
   are independent.
3. For \( 0 \leq s < t < \infty \), the increment \( B_t - B_s \) is normally distributed with zero mean and variance \( t - s \).
4. For all \( \omega \) in a set of probability one, \( B_t \) is a continuous function of \( t \).

A \( d \)-dimensional process \( (B_t)_{t \geq 0} = (B^1_t, \ldots, B^d_t) \) is called a \( d \)-dimensional Brownian Motion if every \( (B^i_t) \) is a one-dimensional Brownian Motion and \( (B^1_t), \ldots, (B^d_t) \) are independent.

Definition 2.1.11 (Wiener Process with respect to a filtration)
Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a complete probability space with a filtration \( \{F_t\}_{t \geq 0} \). A Brownian Motion \( (B_t)_{t \geq 0} \) is called Wiener process with respect to the filtration \( \{F_t\}_{t \geq 0} \) if the following two conditions are satisfied.

- \( B_t \) is \( F_t \)-measurable, i.e. \( B_t \) is adapted to \( F_t \).
- For \( 0 \leq s < t < \infty \) the increment \( B_t - B_s \) is independent of \( F_s \).

In the following we will use the notation \( (W_t) \) for Wiener processes. We give a definition of the standard Brownian filtration.

Definition 2.1.12 (Brownian Filtration)
We consider a time-interval \( I = [0, T] \) or \( I = [0, \infty) \). Let \( (B_t)_{t \in I} \) denote a Brownian motion on a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Let \( \mathcal{F}_t \) be the smallest \( \sigma \)-algebra such that

1. \( \mathcal{F}_t \) contains all null sets and
2. \( B_t \) is \( \mathcal{F}_t \)-measurable.

Then \( \{\mathcal{F}_t\} \) is called Brownian Filtration.
We remark that a Brownian motion \((B_t)_{t \in I}\) is always a Wiener process with respect to the Brownian Filtration.

In [52, p. 16] we find the following result concerning Hölder continuity of the sample paths of a one-dimensional Wiener process. Clearly, the extension to a multi-dimensional Wiener process is obvious, as a \(d\)-dimensional process is Hölder continuous with Hölder exponent \(\delta\) provided each component has this property.

**Proposition 2.1.13**

Let \(W_t\) denote a Wiener process. For almost all \(\omega\) and any \(T > 0\), the sample path \(W_t\) \((t \in [0, T])\) is uniformly Hölder continuous for each exponent \(0 < \delta < \frac{1}{2}\).

### Stochastic Integral and Itô Formula

Many random processes which occur in asset price models are described by stochastic differential equations. In order to define these equations the concept stochastic or Itô integral is of core importance. In analogy to the definition of the Lebesgue integral the definition of the Itô integral consists of several steps. Roughly speaking the Itô integral is first defined only for certain piecewise constant processes (the so-called simple processes) and then extended to a general class of random processes. For the precise formulation of this procedure we refer to [30].

In the remaining part of this paragraph we will define the Stochastic differential and formulate two versions of the Itô formula which will be frequently used in Chapter 4. Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a (complete) probability space with a filtration \(\{\mathcal{F}_t\}_{t \geq 0}\) and let \((W_t)_{t \geq 0}\) be an \(m\)-dimensional Wiener process with respect to this filtration. We consider a general stochastic process of the form

\[
X_t = X_{t_0} + \int_{t_0}^t f(s, \omega) \, ds + \int_{t_0}^t G(s, \omega) \, dW_s. \tag{2.6}
\]

In this context \(G\) is a \(d \times m\) matrix-valued function, adapted to \(\mathcal{F}_t\) and such that with probability 1 it holds

\[
\int_{t_0}^T |G(s, \omega)|^2 \, ds < \infty.
\]

With respect to the initial value \(X_{t_0}\) and the function \(f : [t_0, T] \times \Omega \to \mathbb{R}^d\) we make the following assumptions.

- \(X_{t_0}\) is an \(\mathcal{F}_{t_0}\)-measurable random variable.
- The function \(f\) is measurable in \((s, \omega)\) and \(\mathcal{F}_t\)-adapted. Furthermore, we have with probability 1
  \[
  \int_{t_0}^T |f(s, \omega)| \, ds < \infty.
  \]

Under these assumptions the terms in (2.6) are completely defined for \(t_0 \leq T\).
Definition 2.1.14 (Stochastic differential)
We shall say that the stochastic process $X_t$ defined by Equation (2.6) possesses the stochastic differential $f(t) \, dt + G(t) \, dW_t$ and we shall write

$$dX_t = f(t) \, dt + G(t) \, dW_t.$$ 

The next proposition is a rather general version of the Itô formula.

**Proposition 2.1.15**
Let $u = u(t, x)$ denote a continuous function defined on $[t_0, T] \times \mathbb{R}^d$ with values in $\mathbb{R}^k$ and let the partial derivatives

$$\frac{\partial}{\partial t}u(t, x) = u_t,$$
$$\frac{\partial}{\partial x_i}u(t, x) = u_{x_i}, \quad x = (x_1, \ldots, x_d)^T,$$
$$\frac{\partial^2}{\partial x_i \partial x_j}u(t, x) = u_{x_i x_j},$$

which are for every $i, j \leq d$ $k$-dimensional vectors, be continuous. If the $d$-dimensional stochastic process $X_t$ is defined on $[t_0, T]$ by (2.6) then the $k$-dimensional process

$$Y_t = u(t, X_t) \quad (t \in [t_0, T])$$

with initial value $Y_{t_0} = u(t_0, X_{t_0})$ possesses a stochastic differential with respect to the same Wiener process $W_t$, and we have

$$dY_t = \left( u_t(t, X_t) + u_x(t, X_t)f(t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d u_{x_i x_j}(t, X_t)(G(t)G(t)^T)_{i,j} \right) dt$$
$$+ u_x(t, X_t)G(t) \, dW_t.$$ 

Here, $u_x = (u_{x_1}, \ldots, u_{x_d})$ is a $k \times d$ matrix and $u_{x_i x_j}$ are $k$-dimensional column vectors $(i, j = 1, \ldots, d)$.

In most situations we are concerned with the special case $m = d = 1$, where the matrix $G$ reduces to a scalar.

**Proposition 2.1.16**
Let $u = u(t, x)$ denote a scalar continuous function defined on $[t_0, T] \times \mathbb{R}$ with continuous partial derivatives $u_t$, $u_x$, and $u_{xx}$. If $X_t$ is a process defined on $[t_0, T]$ with stochastic differential

$$dX_t = f(t) \, dt + G(t) \, dW_t,$$

where $f, G$ and $W_t$ are scalar functions, then $Y_t = u(t, X_t)$ possesses on $[t_0, T]$ the stochastic differential

$$dY_t = \left( u_t(t, X_t) + u_x(t, X_t)f(t) + \frac{1}{2} u_{xx}(t, X_t)G(t)^2 \right) dt + u_x(t, X_t)G(t) \, dW_t.$$
Furthermore, using Proposition 2.1.15 one can easily derive the rule for integration by parts for two stochastic integrals driven by the same Wiener process \((W_t)_{t \in [0,T]}\)

\[
X_t^1 = c_1 + \int_{t_0}^t f_1(s) \, ds + \int_{t_0}^t G_1(s) \, dW_s \\
X_t^2 = c_2 + \int_{t_0}^t f_2(s) \, ds + \int_{t_0}^t G_2(s) \, dW_s .
\]

Indeed, combining the processes \(X^1\) and \(X^2\) into a two-dimensional random process and applying Proposition 2.1.15 with \(u(t,x) = u(x_1, x_2) = x_1 \cdot x_2\) one gets

\[
X_t X_t = c_1 c_2 + \int_{t_0}^t X_s^1 dX_s^2 + \int_{t_0}^t X_s^2 dX_s^1 + \int_{t_0}^t G_1(s) G_2(s) \, ds . \quad (2.7)
\]

In comparison with the corresponding formulas for ordinary integrals or differentials there is an extra term

\[
\int_{t_0}^t G_1(s) G_2(s) \, ds .
\]

As a special case we obtain the following proposition (cf. also [4, Corollary 4.5.10]). This result is useful for proving Corollary 2.1.18, which tells us how stochastic integration and usual Riemann integration can be interchanged. We will make use of it in the proof of Lemma 4.3.7.

**Proposition 2.1.17**

Let \(G(\cdot)\) be adapted to the filtration \(\{\mathcal{F}_t\}_{t \geq 0}\) and be almost certainly continuously differentiable on \([t_0, t]\) with derivative \(G'\). Then it holds

\[
\int_{t_0}^t G(s) \, dW_s = G(t)W_t - G(t_0)W_{t_0} - \int_{t_0}^t G'(s)W_s \, ds.
\]

**Corollary 2.1.18**

Let \(g \in C^1[a, b]\) and \(f \in C[a, b]\) be deterministic functions. Then it holds

\[
\int_a^b f(s) \left(\int_a^s g(u) \, dW_u\right) \, ds = \int_a^b g(u) \left(\int_u^b f(s) \, ds\right) \, dW_u \quad (2.8)
\]

**Proof:** Applying Proposition 2.1.17 gives

\[
\int_a^s g(u) \, dW_u = g(s)W_s - g(a)W_a - \int_a^s g'(u)W_u \, du
\]
Hence, it holds
\[
\int_a^b f(s) \left( \int_a^s g(u) \, dW_u \right) \, ds = \int_a^b f(s) g(s) W_s \, ds - g(a) W_a \int_a^b f(s) \, ds
- \int_a^b f(s) \left( \int_a^s g'(u) W_u \, du \right) \, ds,
\]
where the iterated integral on the right hand side is to be understood in the usual Riemann-sense. The integrand is almost certainly continuous, thus by Fubini’s theorem we get
\[
\int_a^b f(s) \left( \int_a^s g(u) \, dW_u \right) \, ds = \int_a^b f(s) g(s) W_s \, ds - g(a) W_a \int_a^b f(s) \, ds
- \int_a^b g'(u) W_u \left( \int_a^b f(s) \, ds \right) \, du,
\] (2.9)

On the other hand, defining \( h(u) = g(u) \int_u^b f(s) \, ds \) we can compute
\[
\int_a^b h(u) \, dW_u = h(b) W_b - h(a) W_a - \int_a^b h'(u) W_u \, du
\]

Inserting the definition of \( h \) and \( h'(u) = g'(u) \int_u^b f(s) \, ds - g(u) f(u) \) gives
\[
\int_a^b g(u) \left( \int_u^b f(s) \, ds \right) \, dW_u = -g(a) W_a \int_a^b f(s) \, ds
- \int_a^b \left[ g'(u) \left( \int_u^b f(s) \, ds \right) - g(u) f(u) \right] W_u \, du.
\] (2.10)

As the right hand sides of (2.9) and (2.10) are equal the assertion is proven.

**Quadratic Variation**

In Chapter 4 we will identify the time-dependent volatility function from high frequency asset price data. From the stochastical point of view the main idea of the corresponding estimator is to use the quadratic variation of the logarithmic asset price process.

For the formal introduction of the quadratic variation it is necessary to introduce a few basic definitions. We start by the concept partition of an interval \([0, t]\), which is defined as a finite ordered set of time instants \( \{t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n\} \) with \( t_0 = 0 \) and \( t_n = t \). The mesh \( \mu(\pi) \) of a partition \( \pi \) is then defined to be the length of the biggest gap between any pair of successive times \( t_i \) and \( t_{i+1} \) in \( \pi \).
Now, for any partition \( \pi = \{ t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \} \) of \([0,t]\) \( \subset [0,T] \) and for any process \((X_t)_{t \in [0,T]}\), the \( \pi \)-quadratic variation on \([0,t]\) is defined to be the random variable
\[
Q^{[0,t]}_{\pi}(X) = \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2.
\]
Furthermore, if there is a monotonically increasing process \((V_t)\) such that \(Q^{[0,t]}_{\pi_n}(X)\) converges in probability to \(V_t\) for any sequence of partitions \(\pi\) of \([0,t]\) such that \(\mu(\pi) \to 0\) as \(n \to \infty\), then \((V_t)\) is said to be the quadratic variation of \((X_t)\). In this case the quadratic variation is denoted by \(\langle X \rangle_t\).

**Stochastic differential equations**

The rest of this paragraph is devoted to stochastic differential equations. As already remarked above, these equations will play a crucial role in the Chapters 3 and 4 where we consider asset price models which are described by a system of stochastic differential equations.

After the definition of the concept solution of a stochastic differential equation we formulate sufficient conditions for existence and uniqueness of the solution. Finally we confine our consideration to linear stochastic differential equations in the narrow sense, where the solution can be given explicitly.

Let \(\{\mathcal{F}_t\}_{t \geq t_0}\) be a filtration on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Let \(W_t = (W^1_t, \ldots, W^m_t)^T\) be an \(m\)-dimensional Wiener process with respect to this filtration. Let furthermore \(c\) be an \(\mathbb{R}^d\)-valued random variable, adapted to \(\mathcal{F}_{t_0}\) such that \(\mathbb{E}|c|^2 < \infty\).

We will consider a \(d\)-dimensional stochastic differential equation (SDE) of Itô type
\[
dX_t = f(t, X_t) \, dt + G(t, X_t) \, dW_t \quad \text{on} \quad t_0 \leq t \leq T, \tag{2.11}
\]
with initial value \(X_{t_0} = c\). Here, \(f : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}^d\) and \(G : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}\) are assumed to be Borel measurable functions.

By the definition of stochastic differential, this equation is a short hand notation for the stochastic integral equation
\[
X_t = c + \int_{t_0}^{t} f(s, X_s) \, ds + \int_{t_0}^{t} G(s, X_s) \, dW_s \quad \text{on} \quad t_0 \leq t \leq T. \tag{2.12}
\]

**Definition 2.1.19 (Solution of an SDE)**

An \(\mathbb{R}^d\)-valued stochastic process \((X_t)_{t_0 \leq t \leq T}\) is called (strong) solution of Equation (2.11) if it has the following properties:

1. \((X_t)_{t_0 \leq t \leq T}\) is continuous and \(\mathcal{F}_t\)-adapted.
2. For almost all \(\omega \in \Omega\) it holds
   \[
   (f(t, X_t)) \in L^1([t_0, T]; \mathbb{R}^d) \quad \text{and} \quad (G(t, X_t)) \in L^2([t_0, T]; \mathbb{R}^{d \times m}).
   \]
3. Equation (2.12) holds for every $t \in [t_0, T]$ with probability 1.

A solution $(X_t)$ is said to be unique if any other solution $(\tilde{X}_t)$ is indistinguishable from $(X_t)$, i.e.

$$\mathbb{P} \left( \sup_{t_0 \leq t \leq T} |X_t - \tilde{X}_t| > 0 \right) = 0.$$ 

Now we are able to formulate the most common conditions that guarantee existence and uniqueness of a solution of a stochastic differential equation. (cf. [4, p. 107]).

**Proposition 2.1.20**

Assume that there exist two positive constants $C$ and $D$ such that $f(\cdot, \cdot)$ and $G(\cdot, \cdot)$ satisfy

1. (Lipschitz condition)

$$|f(t, x) - f(t, y)| + |G(t, x) - G(t, y)| \leq D |x - y|; \quad \forall x, y \in \mathbb{R}^d, t \in [t_0, T] \quad (2.13)$$

2. (Linear growth condition)

$$|f(t, x)|^2 + |G(t, x)|^2 \leq C \left( 1 + |x|^2 \right); \quad \forall x \in \mathbb{R}^d, t \in [t_0, T] \quad (2.14)$$

Let furthermore $c$ be measurable with respect to $\mathcal{F}_{t_0}$. Then the stochastic differential equation (2.11) with initial value $c$ has an unique solution $X_t$.

Now we will present some properties of a uniquely defined solution $X_t$ of a general stochastic differential equation (2.11), which will be used later on (especially in Chapter 4).

Combining Theorem 4.1 and Corollary 4.2 from [52, Chapter 2] we can formulate the following proposition, which gives us a bound for the $p$-th moment of the solution $X_t$.

**Proposition 2.1.21**

Let $X_t$, $t_0 \leq t \leq T$ be the unique solution of Equation (2.11) with initial value $X_{t_0} = c$, where $c$ is an $\mathbb{R}^d$-valued random variable that satisfies $\mathbb{E} |c|^p < \infty$ for some $p \geq 2$. Assume that the linear growth condition (2.14) is satisfied. Then it holds

$$\mathbb{E} |X_t|^p \leq 2^{\frac{p-2}{2}} \left( 1 + \mathbb{E} |c|^p \right) e^{p\alpha (t-t_0)} \quad \forall t \in [t_0, T]$$

with $\alpha = \sqrt{C} + C^{\frac{p-2}{2}}$.

**Linear Stochastic Differential Equation in the narrow sense**

In this section we review some results from [4, p. 128 ff.] concerning linear stochastic differential equations in the narrow sense. For these equations the solution can be given explicitly, provided the fundamental matrix of the corresponding deterministic equation is known. We start with a definition.
Definition 2.1.22 (Linear stochastic differential equation in the narrow sense)

A stochastic differential equation (2.11) is said to be linear if the functions \( f(t, x) \) and \( G(t, x) \) are linear functions of \( x \in \mathbb{R}^d \) on \([t_0, T] \times \mathbb{R}^d \). In other words, the stochastic differential equation (2.11) is called linear if it holds

\[
f(t, x) = A(t)x + a(t),
\]

and

\[
G(t, x) = (B_1(t)x + b_1(t), \ldots, B_m(t)x + b_m(t))
\]

with matrix-valued functions \( A : [t_0, T] \to \mathbb{R}^{d,d} \), \( B_k : [t_0, T] \to \mathbb{R}^{d,d} \) and vector-valued functions \( a : [t_0, T] \to \mathbb{R}^d \), \( b : [t_0, T] \to \mathbb{R}^d \). It is said to be linear in the narrow sense if \( B_1(t) = \ldots = B_m(t) = 0 \).

In the following we will confine our consideration to linear stochastic differential equations in the narrow sense. To simplify notation we will combine the \( m \) vectors \( b_i \) into a single \( d \times m \) matrix \( \Sigma = (b_1, \ldots, b_m) \). In the rest of this section we will always assume the functions \( A \), \( a \) and \( \Sigma \) to be measurable and bounded on \([t_0, T] \). In this case Proposition 2.1.20 implies that there exists for every \( X_{t_0} = c \) a unique solution. The form of this solution is given by the following proposition.

**Proposition 2.1.23**

The linear (in the narrow sense) stochastic differential equation

\[
dX_t = (A(t)X_t + a(t)) \, dt + \Sigma(t) \, dW_t, \quad X_{t_0} = c
\]

has on \([t_0, T]\) the solution

\[
X_t = \Phi(t) \left( c + \int_{t_0}^t \Phi(s)^{-1}a(s) \, ds + \int_{t_0}^t \Phi(s)^{-1}\Sigma(s) \, dW_s \right).
\]

Here, \( \Phi(t) \) is the fundamental matrix of the deterministic equation \( \dot{X}_t = A(t)X_t \), i.e. the solution of the matrix equation

\[
\dot{\Phi}(t) = A(t)\Phi(t), \quad \Phi(t_0) = I.
\]

For the special case where the function \( A \) is constant the fundamental matrix is simply \( \Phi(t) = e^{A(t-t_0)} \). In this situation the solution of the stochastic differential equation is given by the following corollary.

**Corollary 2.1.24**

If the matrix \( A(t) \equiv A \) in equation (2.15) is independent of \( t \), then we have

\[
X_t = e^{A(t-t_0)}c + \int_{t_0}^t e^{A(t-s)}a(s) \, ds + \int_{t_0}^t e^{A(t-s)}\Sigma(s) \, dW_s
\]

We conclude this section by the following proposition characterising the distribution of the process \( X_t \) defined in (2.15).

**Proposition 2.1.25**

The solution (2.16) is a Gaussian stochastic process if and only if \( c \) is normally distributed.
Chapter 3

The Bivariate Ornstein-Uhlenbeck model

In this chapter we present the Bivariate Trending Ornstein-Uhlenbeck model which was the starting point of our considerations. Although all calibration methods presented in this thesis are applicable to a larger class of models with time-depending volatility and stochastic drift components we will use this model as toy example for the numerical case studies.

In the first part of the chapter we introduce the generalised Ornstein-Uhlenbeck model in form of a stochastic differential equation. After that we compute the explicit solution of this equation. Finally, we discuss pricing of European Call Options in this model.

We consider the price process $P = (P_t)_{t \geq 0}$ of a financial asset during the time interval $[0, T_3]$. By $p = (p_t)_{t \geq 0}$ the logarithmic asset price process is denoted, i.e. $p_t = \ln P_t$. The basis for the model which is analysed in this chapter forms the Bivariate Trending Ornstein-Uhlenbeck model of Lo and Wang, introduced in [37]. The logarithmic asset price process $p$ is assumed to have a linear deterministic trend $\mu t$. Then it is convenient to introduce the process $q = (q_t)_{t \geq 0}$ by

$$q_t := p_t - \mu t. \quad (3.1)$$

In the following, this process will be called detrended log-price process to emphasise that $q_t$ contains no deterministic trend component. Uncertainty is modelled by means of a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The Bivariate Trending Ornstein-Uhlenbeck model of Lo and Wang assumes that $q_t$ satisfies the following pair of stochastic differential equations,

$$\begin{align*}
dq_t &= - (\gamma q_t - \lambda X_t) \, dt + \sigma \, dW^q_t \\
dX_t &= - \beta X_t dt + \sigma X \, dW^X_t,
\end{align*} \quad (3.2)$$

where $\gamma \geq 0$, $\lambda \geq 0$, $\beta \geq 0$, $\sigma > 0$, $\mu \in \mathbb{R}$ and $\sigma X > 0$ are real-valued parameters, the initial conditions $q_0 = c_q$, $X_0 = c_X$ hold and $W^q$ and $W^X$ are correlated Wiener processes with correlation coefficient $\kappa$, i.e. $\mathbb{E} \left( W^q_t W^X_t \right) = \kappa t$.

As a motivation for considering this model Lo and Wang argue that empirical observations have indicated that the returns $r_{\tau}(t) = \ln \left( \frac{P_{t+\tau}}{P_t} \right)$ show certain correlation patterns, which means that the classical Black-Scholes model is inappropriate for describing the price process of these assets. In empirical observations the correlation of the returns

$$\text{Corr} \left( r_{\tau}(t), r_{\tau}(t + \tau) \right)$$
3. THE BIVARATE ORNSTEIN-UHLENBECK MODEL

is typically positive for short horizons $\tau$ and negative for large $\tau$.

The Bivariate Trending Ornstein Uhlenbeck model allows to model a wide variety of such correlation patterns. In order to illustrate this we plotted in Figure 3.1 for different values $\gamma$ and $\lambda$ the first order autocorrelation in dependence of the length of the holding period $\tau$. Thereby we have chosen the initial values $q_0$ and $X_0$ in such a way that the process $q$ (and therefore also $p$) is stationary. In this situation the autocorrelation $\rho_\tau(k) := \text{Corr}(r_\tau(t), r_\tau(t + k\tau))$ is independent of $t$. For the figure we fixed the parameters $\kappa = 0$, $\beta = 0.6$, $\sigma^2 = 0.5$ and $\sigma^2_X = 2$.

In order to get an appropriate toy example for our numerical case studies we are now going to make some modifications in the Model 3.2. First of all we restrict our considerations to the case of independent Wiener processes $W^q$ and $W^X$, i.e. $\kappa = 0$. We remark that the resulting models still exhibit all correlation patterns shown in Figure 3.1.

On the other hand with respect to several effects in option pricing we are interested in a more general behaviour of the asset prices in the risk neutral measure as the constant volatility coefficient $\sigma$ would admit. Therefore we generalise this model inasmuch as we allow the volatility $\sigma$ to be time-dependent (but still non-random). Furthermore, in order to allow a scaling of the prices (which plays the role of adjusting the monetary unit) we introduce an additive constant $d$ to $q_t$. Obviously, this leads to a multiplication of the asset prices by $\exp(d)$.

Summarising, the model which we want to call generalised Ornstein-Uhlenbeck Model is described as follows.

**Model 3.0.26** The price process $P = (P_t)_{t \geq 0}$ of a tradable financial asset is described in terms of the detrended log-price process $q = (q_t)_{t \geq 0}$, which is defined by

$$q_t := \ln(P_t) - \mu t - d.$$  \hspace{1cm} (3.3)
We assume that \( q \) satisfies
\[
\begin{align*}
\frac{d}{dt}q_t &= - (\gamma q_t - \lambda X_t) + \sigma(t) dW^q_t \\
\frac{d}{dt}X_t &= -\beta X_t + \sigma_X dW^X_t.
\end{align*}
\] (3.4)

with \( \mu, d \in \mathbb{R}, \gamma \geq 0, \lambda \geq 0, \beta \geq 0, \sigma_X > 0 \) and a time-dependent, continuous volatility function \( \sigma \) with \( \sigma(t) > 0, 0 \leq t \leq T_3 \). The initial values \( q_0 = c_q \) and \( X_0 = c_X \) are assumed to be stochastic variables with finite second order moments, i.e.
\[
\mathbb{E}c_q^2 < \infty \quad \text{and} \quad \mathbb{E}c_X^2 < \infty.
\]
Furthermore, we assume the vector \((c_q, c_X)^T\), which contains the initial values of the processes \( q_t \) and \( X_t \), and the independent Wiener processes \( W^q, W^X \) to be mutually independent.

**Remark 3.0.27** For our numerical case studies we will use (unless stated otherwise) the time unit days and the parameters \( \mu = 1.6483 \cdot 10^{-4}, \gamma = 0.3748, \lambda = 0.004797, \beta = 0.006, \sigma_X = 1 \) together with a volatility function \( \sigma(t) \) which varies around \( 0.01 \). Note that these parameters (together with a constant volatility \( \sigma = 0.01 \) and initial values taken from the stationary distribution) yield the following (daily) logarithmic returns
\[
\begin{align*}
\mathbb{D}^2r(t) &= 0.1098 \cdot 10^{-3}, \quad \rho_r(1) = 0.0681, \quad \rho_r(5) = 0.0094, \quad \rho_r(10) = -0.00045.
\end{align*}
\]
For simplicity we set the initial values and the constant \( d \) to zero, i.e. \( q_0 = X_0 = 0 \).

Finally we remark that it is easy to prove (using Itô’s formula) that under the assumptions of Model 3.0.26 the price process \( P_t = e^{\mu t} \) satisfies the stochastic differential equation
\[
dP_s = P_s \left( -\gamma \ln P_s + \gamma \mu s + \gamma d + \lambda X_s + \mu + \frac{\sigma^2(s)}{2} \right) ds + \sigma(s) P_s dW^q_s \] (3.5)
with initial condition \( P_0 = e^{d} e^{c_q} \).

### 3.1 Solution of the stochastic differential equation

In this section we will apply results of Subsection 2 concerning linear (in the narrow sense) stochastic differential equations to prove the existence and uniqueness of the solutions of (3.4). Furthermore we give an explicit formula for \( q_t \) and \( X_t \) on the interval \([0, T_3] \).

We start by combining the processes \( q_t \) and \( X_t \) into an \( \mathbb{R}^2 \)-valued random process \( \alpha \) and the independent Wiener processes \( W^q \) and \( W^X \) into a 2-dimensional Wiener process \( W \), i.e. we set
\[
\alpha_t := \begin{pmatrix} q_t \\ X_t \end{pmatrix} \quad \text{and} \quad W_t = \begin{pmatrix} W^q_t \\ W^X_t \end{pmatrix}.
\]
Thus, the system (3.4) attains the form
\[
\frac{d}{dt} \alpha_t = \begin{pmatrix} -\gamma & \lambda \\ 0 & -\beta \end{pmatrix} \alpha_t dt + \begin{pmatrix} \sigma(t) & 0 \\ 0 & \sigma_X \end{pmatrix} dW_t
\] (3.6)
with initial value $\alpha_0 = c_\alpha := (c_q \ c_X)^T$, which is a linear stochastic equation in the narrow sense (cf. Definition 2.1.22). Note, that the matrix-valued function $\Sigma : [0, T_3] \to \mathbb{R}^{2 \times 2}$, defined by

$$\Sigma(t) := \begin{pmatrix} \sigma(t) & 0 \\ 0 & \sigma_X \end{pmatrix},$$

is measurable and bounded on $[0, T_3]$. Thus, Proposition 2.1.20 implies that for every initial value $c$ there exists a unique solution $\alpha_t$ of (3.6).

Furthermore, Corollary 2.1.24 states that this solution is given by

$$\alpha_t = e^{t A} c_\alpha + \int_0^t e^{(t-s) A} \Sigma(s) \, dW_s, \tag{3.7}$$

where we have introduced the notation

$$A := \begin{pmatrix} -\gamma & \lambda \\ 0 & -\beta \end{pmatrix}.$$

In order to compute the exponential of the matrix $A$ it is necessary to consider the two cases $\gamma \neq \beta$ and $\gamma = \beta$ separately. The necessary computations can be found in [32].

1. For $\gamma \neq \beta$ we obtain

$$e^{t A} = \begin{pmatrix} e^{-\gamma t} & \frac{\lambda}{\gamma - \beta} (e^{-\beta t} - e^{-\gamma t}) \\ 0 & e^{-\beta t} \end{pmatrix}. $$

Therefore, the solution $\alpha_t$ of (3.6) is given by

$$\alpha_t = \begin{pmatrix} e^{-\gamma t} & \frac{\lambda}{\gamma - \beta} (e^{-\beta t} - e^{-\gamma t}) \\ 0 & e^{-\beta t} \end{pmatrix} c + \int_0^t \begin{pmatrix} e^{-\gamma (t-s)} & \frac{\lambda}{\gamma - \beta} (e^{-\beta (t-s)} - e^{-\gamma (t-s)}) \\ 0 & e^{-\beta (t-s)} \end{pmatrix} \begin{pmatrix} \sigma(s) & 0 \\ 0 & \sigma_X \end{pmatrix} \, dW_s.$$

2. For the situation $\gamma = \beta$ it holds

$$e^{t A} = \begin{pmatrix} e^{-\gamma t} & \lambda t e^{-\gamma t} \\ 0 & e^{-\gamma t} \end{pmatrix}. $$

Therefore, in this situation we have

$$\alpha_t = \begin{pmatrix} e^{-\gamma t} & \lambda t e^{-\gamma t} \\ 0 & e^{-\gamma t} \end{pmatrix} c + \int_0^t \begin{pmatrix} e^{-\gamma (t-s)} & \lambda (t-s) e^{-\gamma (t-s)} \\ 0 & e^{-\gamma (t-s)} \end{pmatrix} \begin{pmatrix} \sigma(s) & 0 \\ 0 & \sigma_X \end{pmatrix} \, dW_s.$$

**Corollary 3.1.1**

The process $q(t)$ is given by
\[ q_t = e^{-\gamma t} c_q + \frac{\lambda}{\gamma - \beta} (e^{-\beta t} - e^{-\gamma t}) c_X + \frac{\lambda \sigma_X}{\gamma - \beta} \int_0^t [e^{-\beta (t-s)} - e^{-\gamma (t-s)}] dW_s^X \]
\[ + \int_0^t \sigma(s)e^{-\gamma (t-s)} dW_s^q \]
for \( \gamma \neq \beta \) and by
\[ q_t = e^{-\gamma t} c_q + \lambda t e^{-\gamma t} c_X + \lambda \sigma_X \int_0^t (t-s)e^{-\gamma (t-s)} dW_s^X + \int_0^t \sigma(s)e^{-\gamma (t-s)} dW_s^q \]
for \( \gamma = \beta \).

Finally we want to remark that the vector process \( \alpha = (q \ X)^T \) is a Gaussian process, provided the vector \( (c_q \ c_X)^T \) is normally distributed. In this context it should be taken in mind that normally distributed random vectors comprehend the set of constant vectors (cf. Definition 2.1.2).

### 3.2 Pricing of European Call Options

We consider now an European vanilla call option with strike \( K \) and expiry \( t_{mat} \). We are interested in the fair price of this option at the timepoint \( t_{buy} \). We start by defining the term European Call Option. Next we review results of [32] stating that the option price formula derived from the Black Scholes model holds also true if the price of the underlying asset follows the generalised Ornstein-Uhlenbeck model.

**Definition 3.2.1 (European call option)**

A European call option is a contract that gives the owner the right to buy a share of a prescribed asset, known as the underlying asset, for a fixed price \( K > 0 \) (strike or exercise price) on a given date \( t_{mat} \).

The market which we consider consists of an asset, whose price process \( P \) follows Model 3.0.26 and the bond whose price process \( B = (B_t)_{t \geq 0} \) is given by \( B_t = e^{rt} \). We introduce the Black-Scholes function \( U_{BS} \) as follows.

**Definition 3.2.2 (Black-Scholes function)**

For parameters \( \tilde{P} > 0, K > 0, r \geq 0, t \geq 0 \) and \( s \geq 0 \) the Black-Scholes function is defined as

\[ U_{BS}(\tilde{P}, K, r, t, s) := \begin{cases} \tilde{P} \Phi(d_1) - K e^{-rt} \Phi(d_2) & \text{if } s > 0 \\ \max(\tilde{P} - K e^{-rt}, 0) & \text{if } s = 0 \end{cases} \]  

(3.10)

with

\[ d_1 := \frac{\ln(\frac{\tilde{P}}{K}) + rt + \frac{s}{2}}{\sqrt{s}}, \quad d_2 := d_1 - \sqrt{s}. \]  

(3.11)

In (3.10) \( \Phi \) denotes the distribution function of the standard normal distribution.
Furthermore we introduce the function $S : [t_{\text{buy}}, T] \rightarrow \mathbb{R}$ by

$$S(t) := \int_{t_{\text{buy}}}^{t} \sigma^2(v) \, dv \quad t \in [t_{\text{buy}}, T]. \quad (3.12)$$

The next proposition reviews a result of [32].

**Proposition 3.2.3**

Let the price of the underlying asset follow Assumption 3.0.26 and the asset price at time $t_{\text{buy}} = 0$ be denoted by $P^\dagger$, i.e.

$$P^\dagger = P_{t_{\text{buy}}} = e^{r_{t_{\text{buy}}}}.$$

At time $t_{\text{buy}}$ the fair price of a European Vanilla Call Option with strike $K > 0$ and maturity $t_{\text{mat}}$ is equal to

$$u(P^\dagger, t_{\text{buy}}, K, t_{\text{mat}}) = U_{BS}(P^\dagger, K, r, t_{\text{mat}} - t_{\text{buy}}, S(t_{\text{mat}})) \quad t_{\text{mat}} \in [t_{\text{buy}}, T]. \quad (3.13)$$

From now on we will always set $t_{\text{buy}} := 0$ and denote the fair price of a European Call option with maturity $t \in [0, T]$ by $u(t)$. The Formula (3.13) will be the basis of our considerations in Chapter 6. There we will discuss the inverse problems of calibrating the functions

$$a(t) = \sigma^2(t) \quad \text{or} \quad S(t) = \int_{0}^{t} \sigma^2(v) \, dv \quad t \in [0, T]$$

from option price data $u(t) \quad t \in [0, T]$. 

Chapter 4

Volatility estimation by wavelet methods

In this chapter we present a non-parametric estimator of the diffusion coefficient in a stochastic differential equation. This estimator is based on a projection method on a wavelet orthonormal basis of $L^2(\mathbb{R})$. The idea has first been formulated in the papers [16] and [17] of Genon-Catalot, Laredo and Picard, in which this method has been proposed for a univariate model and convergence as well as error estimates have been proven. Based on these results several other publications have been concerned with this method, especially with numerical tests and comparisons with kernel estimates. In [49] a univariate model with a very smooth time-depending diffusion coefficient (at least $C^3$) is investigated, whereas the articles [7] and [8] are concerned with a diffusion coefficient which is a realization of a stochastic process which is less regular than $C^1$. In 2003 Pinheiro et. al have published a report which compares nonparametric estimation by wavelet projection and thresholding with three kernel estimates (cf. [43]). However, as far as we know this method has not been studied for multivariate models.

In the context of asset price models the stochastic differential equation describes the logarithmic price process and the diffusion coefficient is the volatility. In view of market models with incomplete information and time-depending but deterministic volatility such as Model 3.0.26 a generalisation of the above mentioned theories to multivariate models is necessary. Under the assumptions of Proposition 2.1.20 guaranteeing existence and uniqueness of a solution of the stochastic differential equation we will accomplish this generalisation in this chapter. Furthermore, we will perform some numerical case studies to illustrate the performance of the method. Finally, we discuss chances and limitations of the L-method, which has been suggested in [43] for the choice of the resolution level.

We start this section with a short introduction to wavelet analysis. Thereby we reduce the presentation to those facts which will be needed in the following. For further details we refer to [10], [38] and [41], where also the following definitions and propositions can be found.

4.1 Introduction to wavelet analysis

Definition 4.1.1 (Multiresolution analysis)
A multiresolution analysis of $L^2(\mathbb{R})$ is a sequence of closed sub-spaces of $L^2(\mathbb{R})$, such that the following properties are satisfied:
1. The sequence is nested, i.e. for all $j \in \mathbb{Z}$ it holds
   \[ V_j \subset V_{j+1}. \]

2. The spaces are related to each other by dyadic scaling, i.e.
   \[ f(\cdot) \in V_j \iff f(2^{j}\cdot) \in V_{j+1} \iff f(2^{-j}\cdot) \in V_0. \]
   \[ (4.1) \]

3. The intersection of the spaces is reduced to the null function and the union of the
   spaces is dense in $L^2(\mathbb{R})$, i.e.
   \[ \bigcap_{j=-\infty}^{\infty} V_j = \{0\}, \quad \bigcup_{j=-\infty}^{\infty} V_j = L^2(\mathbb{R}). \]

4. There exists a function $g \in V_0$ such that the family
   \[ g(\cdot - k), \quad k \in \mathbb{Z}, \]
   \[ (4.2) \]
   is a Riesz basis of $V_0$, i.e.
   \[ V_0 = \text{span}\{g(\cdot - k) : k \in \mathbb{Z}\} \]

and there are positive constants $C' \geq C > 0$ such that
\[
C \sum_{k \in \mathbb{Z}} |\alpha_k|^2 \leq \left\| \sum_{k \in \mathbb{Z}} \alpha_k e_k \right\|_{L^2(\mathbb{R})}^2 \leq C' \sum_{k \in \mathbb{Z}} |\alpha_k|^2.
\]

holds for all $\{\alpha_k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$.

In this case the function $g$ is called scaling function. Furthermore, if the family $g(\cdot - k)$ is
an orthonormal basis of $V_0$ we call $g$ orthonormal scaling function.

We remark that Property 4 implies
\[
f(\cdot) \in V_0 \iff f(x) = \sum_{l \in \mathbb{Z}} \alpha_l g(x - l)
\]
\[
\iff f(x - k) = \sum_{l \in \mathbb{Z}} \alpha_l g(x - l - k) \iff f(\cdot - k) \in V_0.
\]

**Definition 4.1.2 (r-regular multiresolution analysis)**
A multiresolution analysis $(V_j)_{j \in \mathbb{Z}}$ is called $r$-regular ($r \in \mathbb{N}$) if the function $g$ in Condition 4 of Definition 4.1.1 can be chosen in such a way that the $l$-th derivatives $D^l g$ of $g$ satisfy
\[
|D^l g(x)| \leq C_m (1 + |x|)^{-m} \quad \forall m \in \mathbb{N}, \quad l = 0, 1, \ldots, r.
\]

The next proposition is concerned with the transformation of the Riesz sequence $(g(\cdot - k))_{k \in \mathbb{Z}}$
into an orthonormal basis of $V_0$ with the same structure. For a proof see [41, p. 26-29].
Figure 4.1: The scaling functions $\phi$ for Daubechies wavelets of different orders $l$.

**Proposition 4.1.3**

Let $(V_j)_{j \in \mathbb{Z}}$ be an $r$-regular multiresolution analysis of $L^2(\mathbb{R})$ and $g$ according to Definition 4.1.2. Furthermore, let $\phi \in L^2(\mathbb{R})$ be defined by

$$\hat{\phi}(\xi) = \hat{g}(\xi) \left( \sum_{k \in \mathbb{Z}} |\hat{g}(\xi + 2k\pi)|^2 \right)^{-1/2}$$

where $\hat{\phi}$ and $\hat{g}$ denote the Fourier transforms of $\phi$ and $g$ respectively. Then $(\phi(\cdot - k))_{k \in \mathbb{Z}}$ is an orthonormal basis of $V_0$ which satisfies

$$|D^l \phi(x)| \leq C_m (1 + |x|)^{-m}$$

for all $m \in \mathbb{N}$ and $l = 0, 1, \ldots, r$.

We remark that even if starting with a compactly supported function $g$ the orthogonalisation (4.3) can lead to a globally supported function $\phi$, which is difficult to manage numerically (cf. [10, p. 56]). In Section 4.2 we will always use compactly supported, orthonormal wavelets with some regularity $r > 1$. A class of such wavelets providing multiresolution analyses with arbitrary high regularity are the Daubechies wavelets. For a precise definition of these wavelets we refer to [38, p. 169ff].

As we will use the Daubechies wavelets in Section 4.4 for our numerical case studies it seems appropriate to give a short illustration of their shape. To do this we plotted for three different orders $l = 1, 3, 5$ the corresponding scaling function $\phi$. The result is shown in Figure 4.1. It can clearly be seen that the smoothness of these scaling functions increases for increasing order $l$.

Let now $(\phi(\cdot - k))_{k \in \mathbb{Z}}$ denote an orthonormal basis in $V_0$. We will conclude that the functions

$$\phi_{j,k} = 2^{j/2} \phi(2^j \cdot - k), \ k \in \mathbb{Z}$$

form an orthonormal basis of the space $V_j$. This can be seen as follows. Let $g \in V_j$ be arbitrary. Because of (4.1) there exists $\tilde{g} \in V_0$ such that $g = 2^{j/2} \tilde{g}(2^j \cdot)$. Furthermore, $\tilde{g}$ can be written as $\tilde{g} = \sum_{k \in \mathbb{Z}} c_k \phi(\cdot - k) \in V_0$. Combining these considerations we see $g = \sum_{k \in \mathbb{Z}} c_k \phi_{j,k}$. Moreover, we have

$$\|g\|^2_{L^2(\mathbb{R})} = \|	ilde{g}\|^2_{L^2(\mathbb{R})} = \sum_{k \in \mathbb{Z}} c_k^2.$$
Our aim is to approximate functions $f \in L^2(\mathbb{R})$ by appropriate functions $f_j \in V_j$. To find these appropriate functions we define the projectors $P_j : L^2(\mathbb{R}) \rightarrow V_j$ onto the spaces $V_j$. Using the orthonormal basis $(\phi_{j,k})_{k \in \mathbb{Z}}$ this projector can be written as

$$P_j f = \sum_{k \in \mathbb{Z}} \langle f, \phi_{j,k} \rangle_{L^2(\mathbb{R})} \phi_{j,k}. $$

For $s \geq 0$ we denote by $H^s(\mathbb{R})$ the Sobolev space defined by

$$H^s(\mathbb{R}) = \left\{ f \in L^2(\mathbb{R}) : \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \left(1 + |\xi|^2\right)^s \, d\xi < \infty \right\} \quad (4.5a)$$

with the norm $\| \cdot \|_s$ defined by

$$\|f\|_s^2 = \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \left(1 + |\xi|^2\right)^s \, d\xi. \quad (4.5b)$$

For $s < 0$ we define the Sobolev space $H^s(\mathbb{R})$ as the topological dual space of $H^{-s}(\mathbb{R})$.

Clearly, we are interested, in which spaces and how fast the projections $P_j f$ of a function $f \in H^s(\mathbb{R})$ converge to the function $f$. The answer to this question is given by the following proposition (cf. [41, p. 41]).

**Proposition 4.1.4**

Let $r \in \mathbb{N}$ and let $\{V_j\}_{j \in \mathbb{Z}}$ be an $r$-regular multiresolution analysis of $L^2(\mathbb{R})$. If $f$ belongs to $H^s(\mathbb{R})$ with $0 \leq s \leq r$, then $P_j f$ converges to $f$ in $H^s(\mathbb{R})$, i.e.

$$\|f - P_j f\|_s \rightarrow 0 \quad \text{as} \quad j \rightarrow \infty.$$

Before we state further results we define complement spaces $W_j$ by

$$V_{j+1} = V_j \oplus W_j,$$

i.e. $W_j$ is the orthogonal complement of $V_j$ in $V_{j+1}$. The orthogonal projector $Q_j : L^2(\mathbb{R}) \rightarrow W_j$ onto $W_j$ is then given by

$$Q_j f = P_{j+1} f - P_j f.$$ 

The spaces $W_j$ are also related to each other by dyadic scaling, i.e. it holds $f \in W_0$ if and only if $f(2^j \cdot) \in W_j$.

Furthermore, we can construct an orthonormal basis of $W_0$ as follows. Since $V_0 \subset V_1$ the scaling function $\phi \in V_0$ can be expanded in terms of the basis of $V_1$, i.e. there exists a sequence $(h_n)_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$ such that

$$\phi(x) = \sum_{n \in \mathbb{Z}} h_n \phi(2x - n). \quad (4.6)$$

This equation is called refinement equation.
Using these coefficients one can define new coefficients
\[ g_n = (-1)^n h_{1-n} \]  
and a function \( \psi \in V_1 \) by
\[ \psi(x) = \sum_{n \in \mathbb{Z}} g_n \phi(2x - n). \]
This function \( \psi \) is called wavelet. The next theorem states that its scaled and dilated versions \( \psi_{j,k} \) defined by
\[ \psi_{j,k} = 2^{j/2} \psi(2^j \cdot -k), \quad k, j \in \mathbb{Z} \]
provide an orthonormal basis of the complement spaces \( W_j \) (cf. [10, Theorem 2.6.1] and [38, Satz 2.2.10]).

**Proposition 4.1.5**
The function \( \psi \) satisfies
\[ \langle \psi, \phi_{0,k} \rangle_{L^2(\mathbb{R})} = 0 \quad \forall k \in \mathbb{Z}. \]
Moreover, the projector \( Q_j \) can be expanded into
\[ Q_j f = \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \psi_{j,k} \]
and the functions \( \psi_{j,k}, \ k \in \mathbb{Z} \) constitute an orthonormal basis of the complement space \( W_j \).

Combining the above definitions we see that every function \( f \in L^2(\mathbb{R}) \) allows the decomposition
\[ f = f_{j_0} + \sum_{j=j_0}^{\infty} d_j, \]  
where \( f_{j_0} \) denotes the approximation of \( f \) in the space \( V_{j_0} \) and \( d_j \) is the detail of level \( j \), which can be expanded as
\[ d_j = Q_j f = \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \psi_{j,k}. \]
The following proposition characterises the decay of the norm of these details (cf. [41, p. 48]).

**Proposition 4.1.6**
Let \( \{V_j\}_{j \in \mathbb{Z}} \) be an \( r \)-regular multiresolution analysis of \( L^2(\mathbb{R}) \) and \( f \in H^s(\mathbb{R}) \) for some \( s \in [0, r) \). Then we have
\[ P_0 f \in L^2(\mathbb{R}) \quad \text{and} \quad \|Q_j f\|_{L^2(\mathbb{R})} = \varepsilon_j 2^{-js}, \quad j \in \mathbb{N}, \]
where \( \{\varepsilon_j\}_{j=1}^{\infty} \) is a sequence of \( l^2(\mathbb{N}) \). Moreover, it holds
\[ \|f\|_s \approx \|P_0 f\|_{L^2(\mathbb{R})} + \left( \sum_{j \in \mathbb{N}} \varepsilon_j^2 \right)^{1/2}. \]
In Section 4.3 we will often have to estimate \( \|f - P_j f\|_{L^2(\mathbb{R})} \) for some \( f \in H^s(\mathbb{R}) \). Because of Proposition 4.1.6 we have

\[
\|f - P_j f\|_{L^2(\mathbb{R})}^2 = \left\| \sum_{k \geq j} Q_k f \right\|_{L^2(\mathbb{R})}^2 = \sum_{k \geq j} \|Q_k f\|_{L^2(\mathbb{R})}^2 = \sum_{k \geq j} \varepsilon_k^2 2^{-2ks}
\]

with a sequence \( \{\varepsilon_k\}_{k=1}^{\infty} \subseteq \ell^2(\mathbb{N}) \), which implies \( \varepsilon_k^2 \to 0 \) as \( k \to \infty \). Thus, \( \max_{k \geq j} \varepsilon_k^2 \) exists for all \( j \) and converges to zero as \( j \) tends to infinity. This gives

\[
\|f - P_j f\|_{L^2(\mathbb{R})}^2 \leq \max_{l \geq j} \varepsilon_l^2 \sum_{k \geq j} (2^{-2s})^k = \max_{l \geq j} \varepsilon_l^2 \frac{1}{1 - 2^{-2s}} (2^{-2s})^j,
\]

where \( \tilde{\varepsilon}_j := \max_{l \geq j} \varepsilon_l^2 \frac{1}{1 - 2^{-2s}} \to 0 \) as \( j \to \infty \). Thus, we can formulate the following corollary.

**Corollary 4.1.7**

Let \( \{V_j\}_{j \in \mathbb{Z}} \) be an \( r \)-regular multiresolution analysis of \( L^2(\mathbb{R}) \). Let furthermore \( f \in H^s(\mathbb{R}) \) with \( 0 \leq s < r \) then

\[
\|f - P_j f\|_{L^2(\mathbb{R})} \leq 2^{-js} \tilde{\varepsilon}_j \quad (4.9)
\]

with \( \tilde{\varepsilon}_j \to 0 \) as \( j \to \infty \).

We remark that the sequence \( \varepsilon_j \) in (4.9) depends on \( f \in H^s(\mathbb{R}) \), not only on the norm \( \|f\|_s \). However, combining [10, p. 90 f.] and [10, Theorem 3.3.3] we get the following direct estimate, which gives a uniform convergence of the error \( \|f - P_j f\|_{L^2(\mathbb{R})} \) for all functions \( f \) satisfying \( \|f\|_s \leq K \).

**Proposition 4.1.8**

Let \( \{V_j\} \) be a multiresolution analysis such that the corresponding wavelet \( \psi \) belongs to \( H^r \). Let furthermore \( f \in H^s(\mathbb{R}) \) with \( 0 \leq s \leq r \). Then it holds

\[
\|f - P_j f\|_{L^2(\mathbb{R})} \leq C 2^{-js}\|f\|_s,
\]

where the constant \( C \) is independent of \( f, j \) and \( s \).

In the following we assume that the functions \( \phi \) and \( \psi \) are compactly supported, i.e. it exists an \( A \in \mathbb{R} \) such that

\[
\text{supp } \phi \subset [-A, A] \quad \text{and} \quad \text{supp } \psi \subset [-A, A].
\]

Then the supports of the scaled and dilated versions \( \phi_{j,k} \) and \( \psi_{j,k} \) \((j, k \in \mathbb{Z})\) satisfy

\[
\text{supp } \phi_{j,k} \subset \left[ 2^{-j}(-A + k), 2^{-j}(A + k) \right] \quad (4.10a)
\]

and

\[
\text{supp } \psi_{j,k} \subset \left[ 2^{-j}(-A + k), 2^{-j}(A + k) \right], \quad (4.10b)
\]

which will play an important role in Section 4.3.
4. VOLATILITY ESTIMATION BY WAVELET METHODS

4.2 The situation and the estimator

We consider now a more general model than (3.4). Again, let \((\Omega, \mathcal{F}, \mathbb{P})\) denote a complete filtered probability space. Let \(Y = (Y_t)_{t \geq 0}\) be described by the stochastic differential equation

\[dY_t = f(t, Y_t) \, dt + G(t) \, dW_t \quad \text{on } [0, T].\]  

(4.11)

with initial condition \(Y_0 = c\). Here, \(W_t\) denotes an \(m\)-dimensional Wiener process. The functions \(f : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d\) and \(G : [0, T] \to \mathbb{R}^{d \times m}\) are assumed to be deterministic and we assume that \(G\) can be written as

\[G(t) = \begin{pmatrix} \theta(t) & 0 \\ 0 & G_{2,2} \end{pmatrix}\]

with a (constant), positive definite matrix \(G_{2,2} \in \mathbb{R}^{(d-1) \times (m-1)}\).

With respect to \(f, \theta\) and the initial value \(c\) we make the following assumptions.

**Assumption 4.2.1**

1. \(f \in C^1([0, T] \times \mathbb{R}^d; \mathbb{R}^d)\)

2. For \(T > 0\) exists a positive constant \(K_T\) such that

\[|f(t, y)|^2 \leq K_T (1 + |y|^2) \quad \text{holds for } 0 \leq t \leq T \quad \text{and for all } y \in \mathbb{R}^d.\]

3. \(\theta \in C^m([0, T])\) with \(m \geq 1\) and \(\theta(t) \geq 0 \forall t \in [0, T]\)

4. We assume \(E|c|^p < \infty\) for some \(p \geq 2\).

The first two assumptions guarantee the unique solvability of (4.11) (cf. Proposition 2.1.20). Besides, this solutions \(Y_t\) inherits the Hölder continuity of the Wiener process \(W_t\) (cf. Proposition 2.1.13). Furthermore, the third assumption implies that the function \(\theta\) can be extended to a function \(\theta \in C^m(\mathbb{R})\) satisfying

\[\theta(t) \geq 0 \quad \forall t \in \mathbb{R}, \quad \theta(t) = \theta(t) \quad \forall t \in [0, T] \quad \text{and} \quad \text{supp} \theta \subset [-\varepsilon, T + \varepsilon]\]  

(4.12)

for some positive \(\varepsilon\). This ensures that \(\theta\) has finite support and belongs to the Sobolev space \(H^m(\mathbb{R})\).

Let \(y_t\) denote the first component of the vector \(Y_t\). We assume that we can observe the first component \(y_t\) at discrete time points

\[t_i = i \Delta_n \quad i = 0, 1, \ldots, N := \lceil 2^n T \rceil\]

in the interval \([0, T]\), where we introduced the step width \(\Delta_n := 2^{-n}\). We give a non-parametric estimator \(\hat{\theta}^2(t), t \in [0, T]\) of the function \(\theta^2(t), t \in [0, T]\) from the observations \(y_{t_i}, i = 0, 1, \ldots, N\).

We remark that the model (3.4) is contained in this setting, provided \(\sigma \in C^1\). This smoothness assumption will play a crucial role in the proofs for the convergence rates of Section 4.3.
Remark 4.2.2 It suffices to assume $\tilde{\theta}^2 \in H^m(\mathbb{R}) \cap C^1(\mathbb{R})$ instead of $\theta \in C^m([0,T])$. In fact, the smoothness of $\theta$ will be mainly used when we apply the approximation results of Proposition 4.1.4 to Corollary 4.1.7, for which $\tilde{\theta}^2 \in H^m$ is sufficient. Apart from that we will use at one time Taylor expansion of $\theta^2$ up to order one and a lemma formulated in [17], which makes it necessary to assume $\tilde{\theta} \in C^1(\mathbb{R})$.

Let now $(V_j)_{j \in \mathbb{Z}}$ be an $r$-regular multiresolution analysis of $L^2(\mathbb{R})$ such that the associated orthonormal scaling function $\phi$ and wavelet function $\psi$ are compactly supported and belong to $C^1(\mathbb{R}) \cap H^s(\mathbb{R})$. We remark that these assumptions are satisfied if the orthonormal scaling function $\phi$ and the wavelet $\psi$ are compactly supported and belong to $C^r(\mathbb{R})$ with $r \geq 1$. We assume $r \geq m$.

We can write the function $\tilde{\theta}^2$ as

$$
\tilde{\theta}^2(t) = \sum_{k \in \mathbb{Z}} \mu_{j_1,k}\phi_{j_1,k}(t) + \sum_{j \geq j_1} \sum_{k \in \mathbb{Z}} \nu_{j,k}\psi_{j,k}(t),
$$

with the coefficients

$$
\mu_{j,k} = \langle \tilde{\theta}^2, \phi_{j,k} \rangle_{L^2(\mathbb{R})} \quad \text{and} \quad \nu_{j,k} = \langle \tilde{\theta}^2, \psi_{j,k} \rangle_{L^2(\mathbb{R})}.
$$

At this point it suffices to think of $j_1$ as an arbitrary integer.

Next we define $\hat{\mu}_{j,k}$ by

$$
\hat{\mu}_{j,k} := \sum_{i=0}^{N-1} \phi_{j,k}(t_i) (y_{i+1} - y_t)^2.
$$

In the following we will motivate that these $\hat{\mu}_{j,k}$ are empirical estimators of $\mu_{j,k}$. To see this, let us forget for a moment about the smoothness property of the wavelet basis required above and consider the class of Daubechies wavelets of order $l = 1$, which are also known as Haar wavelets (cf. [10] and also Figure 4.1). In this situation we have $\phi_{j,k} = 2^{j/2} \chi_{[k2^{-j},(k+1)2^{-j}]}$ and the coefficient $\hat{\mu}_{j,k}$ attains the form

$$
\hat{\mu}_{j,k} = 2^{j/2} \sum_{i:k2^{-j} \leq t_i \leq (k+1)2^{-j}} (y_{i+1} - y_i)^2,
$$

provided $j$ and $k$ are such that the support $[k2^{-j},(k+1)2^{-j}]$ of $\phi_{j,k}$ is included in the interval $[0,T]$.

In other words, in this situation $2^{-j/2}\hat{\mu}_{j,k}$ is exactly $Q_{\pi}^{(0,(k+1)2^{-j})}(y) - Q_{\pi}^{[0,k2^{-j})}(y)$, where $\pi$ denotes the partition of $[0,T]$ containing all points $t_i$ and $Q_{\pi}^{[0,t]}$ denotes the $\pi$ quadratic variations of the process $(y_t)$ on the interval $[0,t]$. Furthermore, it is well known that in the situation considered here the quadratic variation of the process $(y_t)$ exists and it holds

$$
\langle y \rangle_{(k+1)2^{-j}} - \langle y \rangle_{k2^{-j}} = \int_{k2^{-j}}^{(k+1)2^{-j}} \theta^2(t) \, dt = 2^{-j/2} \langle \tilde{\theta}^2, \phi_{j,k} \rangle_{L^2(\mathbb{R})}.
$$

Remembering that the $\pi$-quadratic variation converges to the quadratic variation in probability we have thus seen that the coefficient $\hat{\mu}_{j,k}$ is an estimator for $\mu_{j,k}$. The extension to a more
general wavelet basis, which satisfies still the assumptions formulated after Remark 4.2.2, is straightforward as these scaling functions can be approximated by step functions.

The subspace $V_{j_1}$ is not finite dimensional. However, noting that the functions $\tilde{\theta}^2$ and $\phi_{j_1,k}$ are compactly supported, the two sets

$$\{k \in \mathbb{Z} : \mu_{j_1,k} \neq 0\} \quad \text{and} \quad \{k \in \mathbb{Z} : \mu_{j_1,k} \neq 0\}$$

are finite and can be included in a finite set $L_{j_1}$ which depends only on the constant $T$, the support of the function $\phi$ and the resolution level $j_1$ and whose cardinality is $O(2^{j_1})$.

A natural estimator of $\theta^2$ is obtained if the second sum in (4.13), which contains details of levels $j \geq j_1$, is omitted and the exact coefficients $\mu_{j_1,k}$ are replaced by the estimates $\hat{\mu}_{j_1,k}$. In this way we obtain the estimator

$$\hat{\theta}^2(t) = \sum_{k \in L_{j_1}} \hat{\mu}_{j_1,k} \phi_{j_1,k}(t) = \sum_{k \in \mathbb{Z}} \hat{\mu}_{j_1,k} \phi_{j_1,k}(t).$$

(4.17)

Here, $j_1$ denotes the resolution level. As we will see below, this level has to be chosen with care. Clearly, if it is too small the function $\theta^2(t)$ can only be poorly reconstructed. On the other side, if for a fixed number of observations the resolution level is chosen too large, the estimator will have strong oscillations. It is therefore necessary to choose the resolution level in dependence of the number of observations, which is equal to $[2^n T]$. This dependence is expressed by the notation $j_1 = j_1(n)$. The asymptotic properties of the estimator $\theta^2(t)$ will give theoretical results how $j_1(n)$ should be chosen in order to ensure convergence. For a concrete trajectory and possibly unknown smoothness of $\theta$ certain data-driven methods have been proposed. One of them is the L-Method which will be formulated, discussed and illustrated in Section 4.4.1.

### 4.3 Asymptotic study of the estimator

In this section we follow the ideas of [17] and study asymptotic properties of the estimator $\hat{\theta}^2$. In the first subsection we show weak convergence of the estimator $\hat{\theta}^2$ to $\theta^2$ under appropriate conditions. These results will be used in Section 4.3.2 where we analyse the mean integrated square error of the estimator.

#### 4.3.1 Weak convergence of the estimator

In order to prove weak convergence of the estimator $\hat{\theta}^2$ to $\theta^2$ in the interior of the interval $[0,T]$ we consider an arbitrary function $h$ which satisfies

#### Assumption 4.3.1

The function $h$ is continuous on $[0,T]$, with compact support included in $(0,T)$ and belongs to the Sobolev space $H^{m'}(\mathbb{R})$ with $m' > \frac{1}{2}$.
4.3. ASYMPTOTIC STUDY OF THE ESTIMATOR

This assumption implies that there exists some \( \delta > 0 \) such that
\[
\text{supp } h \subset [\delta, T - \delta]. \tag{4.18}
\]

We can write the function \( h \) as follows
\[
h(t) = \sum_{k \in \mathbb{Z}} \alpha_{j_1, k} \varphi_{j_1, k}(t) + \sum_{j \geq j_1} \sum_{k \in \mathbb{Z}} \beta_{j, k} \psi_{j, k}(t) \tag{4.19a}
\]
with
\[
\alpha_{j_1, k} = \langle h, \varphi_{j_1, k} \rangle_{L^2(\mathbb{R})} \quad \text{and} \quad \beta_{j, k} = \langle h, \psi_{j, k} \rangle_{L^2(\mathbb{R})}. \tag{4.19b}
\]

For \( \hat{\theta}^2 \) given in (4.17) we investigate the convergence of the quantity
\[
\int_{\mathbb{R}} h(t) \left( \hat{\theta}^2(t) - \theta^2(t) \right) \, dt = \int_0^T h(t) \left( \hat{\theta}^2(t) - \theta^2(t) \right) \, dt
\]
when suitably normalised. We will prove the following result.

**Theorem 4.3.2**

Let the Assumptions 4.2.1 and 4.3.1 be satisfied. Then, the integral
\[
I_n := 2^{n/2} \int_{\mathbb{R}} h(t) \left( \hat{\theta}^2(t) - \theta^2(t) \right) \, dt \tag{4.20}
\]
converges for \( n \to \infty \) in distribution to a normal variable with zero mean and variance
\[
2 \int_0^T h^2(t) \theta^4(t) \, dt, \quad \text{provided} \, \frac{1}{2} < m' < r, \, m + m' > 2, \, j_1(n) = [\alpha n] \quad \text{with}
\]
\[
\frac{1}{2(m + m')} \leq \alpha < \frac{1}{4}.
\]

Recall that \( r \) is the regularity of the multiresolution analysis, \( m \) (respectively, \( m' \)) denotes the exponent of the Sobolev space of \( \bar{\theta}^2 \) (respectively \( h \)). We set \( h_j = P_j h \).

Before we prove this theorem we will make some preliminary considerations and show some lemmas which are used in the proof of the theorem. Using the compact supports of \( \varphi_{j, k} \) and \( h \) (cf. (4.10a) and (4.18)) we get immediately the following result.

**Lemma 4.3.3**

Whenever
\[
k > 2^j (T - \delta) + A \quad \text{or} \quad k < 2^j \delta - A
\]
holds we have
\[
\alpha_{j, k} = \langle h, \varphi_{j, k} \rangle = 0.
\]
Furthermore we have \( \text{supp } h_j \subset (0, T) \) whenever \( 2^j > \frac{2A}{\delta} \).
In the following we will use the notation \( j_1 = j_1(n) \) to express the dependence of the resolution level from the number of observations. In what follows we will often use the fact \( j_1(n) \to \infty \) for \( n \to \infty \), which is guaranteed by the choice \( j_1(n) = [an] \), cf. Theorem 4.3.2. Let \( K_{j_1(n)} = \{ k \in \mathbb{Z} : \alpha_{j_1(n),k} \neq 0 \} \). Lemma 4.3.3 gives
\[
\# K_{j_1(n)} \le 2A + 1 + 2^{j_1(n)}(T - 2\delta) = O\left(2^{j_1(n)}\right) \quad \text{for} \ n \to \infty.
\]

Defining the modulus of continuity \( \omega_f(\delta) \) of a function \( f \) by
\[
\omega_f(\delta) = \sup \{ |f(s) - f(t)| : |s - t| \le \delta \}
\]
we can bound \( \omega_{\phi_{j_1(n),\cdot}}(\Delta_n) \) and \( \omega_{h_{j_1(n)}}(\Delta_n) \) from above.

**Lemma 4.3.4**
For \( n \to \infty \) it holds
\[
\omega_{\phi_{j_1(n),\cdot}}(\Delta_n) \le O\left(2^{\frac{j_1(n)}{2}}\Delta_n\right) \quad (4.22)
\]
\[
\omega_{h_{j_1(n)}}(\Delta_n) \le O\left(2^{2j_1(n)-n}\right).
\]

**Proof:** Remembering \( \phi_{j_1(n),k} = 2^{j_1(n)/2} \phi(2^{j_1(n)} \cdot -k) \) and the assumption that \( \phi \in C^1(\mathbb{R}) \) and \( \phi \) is compactly supported we see immediately (4.22). Using this result and Cauchy-Schwartz inequality we get with \( |t - s| \le \Delta_n \)
\[
\left| h_{j_1(n)}(t) - h_{j_1(n)}(s) \right| \le \sum_{k \in \mathbb{Z}} \left| \alpha_{j_1(n),k} \right| \left| \phi_{j_1(n),k}(t) - \phi_{j_1(n),k}(s) \right| \le 2^{\frac{j_1(n)}{2}} \Delta_n O(1)
\]
\[
\le \sqrt{\sum_{k \in \mathbb{Z}} \alpha_{j_1(n),k}^2} \sqrt{\# K_{j_1(n)} 2^{\frac{j_1(n)}{2}} \Delta_n O(1)} = O\left(2^{2j_1(n)-n}\right).
\]

Furthermore we will use the following result which gives a decomposition of the squared increments of the process \( y_t \). The proof uses integration by parts.

**Lemma 4.3.5**
It holds
\[
(y_{t_{i+1}} - y_{t_i})^2 = 2 \int_{t_i}^{t_{i+1}} (y_s - y_{t_i}) \, dy_s + \int_{t_i}^{t_{i+1}} \theta^2(u) \, du.
\]

**Proof:** We use (4.11) which are \( d \) linked stochastic differential equations for the \( d \) components of the vector-valued process \( Y_t \). Considering only the first equation we get
\[
dy_t = f^1(t, Y_t) \, dt + \theta(t) \, dW^1_t,
\]
where \( f^1(t, Y_t) \) and \( W^1_t \) denote the first components of the vectors \( f(t, Y_t) \) and \( W_t \) respectively.
For fixed \( i \) (and therefore fixed \( t_i \)) we can write
\[
y_t - y_{t_i} = \int_{t_i}^t f^1(s, Y_s) \, ds + \int_{t_i}^t \theta(s) \, dW_s^1.
\]
(4.25)

Using integration by parts (cf. (2.7)) we obtain the assertion.

In the following we will use the wavelet expansion of \( h \) and \( \bar{\theta}^2 \) (cf. (4.19) and (4.13)) and the definition of the estimator \( \hat{\theta}^2 \) (cf. (4.17)). Noting that the functions \( \theta^2 \) and \( \bar{\theta}^2 \) are equal on \([0, T]\) and the support of \( h \) is contained in \((0, T)\) we obtain the following decomposition
\[
\int_{\mathbb{R}} h(t) \left( \bar{\theta}^2(t) - \theta^2(t) \right) \, dt = \int_{\mathbb{R}} h(t) \left( \hat{\theta}^2(t) - \bar{\theta}^2(t) \right) \, dt = Z_n + T_n
\]

with
\[
Z_n := \langle h, \hat{\theta}^2 - P_{j_1(n)} \bar{\theta}^2 \rangle = \sum_{k \in \mathbb{Z}} \alpha_{j_1(n),k} \left( \hat{\mu}_{j_1(n),k} - \mu_{j_1(n),k} \right)
\]
and
\[
T_n := \langle h, P_{j_1(n)} \bar{\theta}^2 - \bar{\theta}^2 \rangle = - \sum_{j \geq j_1(n)} \sum_{k \in \mathbb{Z}} \beta_{j,k} \nu_{j,k}.
\]

Using the Corollary 4.1.7 we can easily bound the summand \( T_n \).

**Lemma 4.3.6**

The condition
\[
2^{n/2 - j_1(n)(m + m')} = O(1) \quad \text{for } n \to \infty
\]
(4.26)
is sufficient to guarantee \( 2^{n/2} T_n \to 0 \) for \( n \to \infty \).

**Proof:** As \( I - P_{j_1(n)} \) is an orthogonal projector we get together with Corollary 4.1.7 for bounding the approximation error of \( h \) and Proposition 4.1.8 for \( \bar{\theta}^2 \)
\[
|T_n| = \left| \langle h, (I - P_{j_1(n)}) \bar{\theta}^2 \rangle_{L^2(\mathbb{R})} \right| = \left| \langle (I - P_{j_1(n)}) h, (I - P_{j_1(n)}) \bar{\theta}^2 \rangle_{L^2(\mathbb{R})} \right| \\
\leq \left\| h - P_{j_1(n)} h \right\|_{L^2(\mathbb{R})} \left\| \bar{\theta}^2 - P_{j_1(n)} \bar{\theta}^2 \right\|_{L^2(\mathbb{R})} \leq 2^{-j_1(n)(m + m')} e^{-j_1(n)},
\]
with \( e^{-j_1(n)} \to 0 \) for \( j_1(n) \to \infty \).

In order to analyse \( Z_n \) we will use the following decomposition.

**Lemma 4.3.7**

For sufficiently large \( n \), the term \( Z_n \) can be decomposed as follows.
\[
Z_n = C_n + \sum_{i=1}^{5} A_{i,n},
\]
(4.27)
As for sufficiently large $n$ we have $2\hat{h}^{(n)} > \frac{2\lambda}{\delta}$ and therefore $\text{supp} \, h_{j_1(n)} \subset [0, T]$ we get

$$
\sum_{k \in \mathbb{Z}} \alpha_{j_1(n), k} \mu_{j_1(n), k} = \int_{\mathbb{R}} h_{j_1(n)}(t) \bar{\theta}^2(t) \, dt = \int_{0}^{T} h_{j_1(n)}(t) \theta^2(t) \, dt.
$$

Using this equality together with the definition of $\hat{\mu}_{j_1(n), k}$ (cf. (4.15)) we obtain

$$
Z_n = \sum_{i=0}^{N-1} (y_{i+1} - y_i)^2 h_{j_1(n)}(t_i) - \int_{0}^{T} h_{j_1(n)}(t) \theta^2(t) \, dt. \tag{4.29}
$$

Using Lemma 4.3.5 gives

$$
Z_n = 2 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (y_s - y_i) \, dy_s h_{j_1(n)}(t_i) + C_n.
$$

It remains therefore to decompose the integrals $\int_{t_i}^{t_{i+1}} (y_s - y_i) \, dy_s$. We start by substituting $dy_s$ and $y_s - y_i$ (cf. (4.24) and (4.25)). We obtain

$$
\int_{t_i}^{t_{i+1}} (y_s - y_i) \, dy_s = \int_{t_i}^{t_{i+1}} f^1(s, Y_s)(y_s - y_i) \, ds + \int_{t_i}^{t_{i+1}} \theta(s) (y_s - y_i) \, dW^1_s
$$

$$
= \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \left( \int_{t_i}^{s} f^1(u, Y_u) \, du + \int_{t_i}^{s} \theta(u) dW^1_u \right) \, ds
$$

$$
+ \int_{t_i}^{t_{i+1}} \theta(s) \left( \int_{t_i}^{s} f^1(u, Y_u) \, du + \int_{t_i}^{s} \theta(u) dW^1_u \right) \, dW^1_s.
$$

**Proof:** As for sufficiently large $n$ we have $2\hat{h}^{(n)} > \frac{2\lambda}{\delta}$ and therefore $\text{supp} \, h_{j_1(n)} \subset [0, T]$ we get

$$
\sum_{k \in \mathbb{Z}} \alpha_{j_1(n), k} \mu_{j_1(n), k} = \int_{\mathbb{R}} h_{j_1(n)}(t) \bar{\theta}^2(t) \, dt = \int_{0}^{T} h_{j_1(n)}(t) \theta^2(t) \, dt.
$$

Using this equality together with the definition of $\hat{\mu}_{j_1(n), k}$ (cf. (4.15)) we obtain

$$
Z_n = \sum_{i=0}^{N-1} (y_{i+1} - y_i)^2 h_{j_1(n)}(t_i) - \int_{0}^{T} h_{j_1(n)}(t) \theta^2(t) \, dt. \tag{4.29}
$$

Using Lemma 4.3.5 gives

$$
Z_n = 2 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (y_s - y_i) \, dy_s h_{j_1(n)}(t_i) + C_n.
$$

It remains therefore to decompose the integrals $\int_{t_i}^{t_{i+1}} (y_s - y_i) \, dy_s$. We start by substituting $dy_s$ and $y_s - y_i$ (cf. (4.24) and (4.25)). We obtain

$$
\int_{t_i}^{t_{i+1}} (y_s - y_i) \, dy_s = \int_{t_i}^{t_{i+1}} f^1(s, Y_s)(y_s - y_i) \, ds + \int_{t_i}^{t_{i+1}} \theta(s) (y_s - y_i) \, dW^1_s
$$

$$
= \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \left( \int_{t_i}^{s} f^1(u, Y_u) \, du + \int_{t_i}^{s} \theta(u) dW^1_u \right) \, ds
$$

$$
+ \int_{t_i}^{t_{i+1}} \theta(s) \left( \int_{t_i}^{s} f^1(u, Y_u) \, du + \int_{t_i}^{s} \theta(u) dW^1_u \right) \, dW^1_s.
$$
which can also be written as
\[
\int_{t_i}^{t_{i+1}} (y_s - y_{t_i}) \, dy_s = \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{t_i}^{s} f^1(u, Y_u) \, du \, ds + \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{t_i}^{s} \theta(u) dW_u^1 \, ds + \int_{t_i}^{t_{i+1}} \theta(s) \int_{t_i}^{s} f^1(u, Y_u) \, du \, dW_u^1 + \int_{t_i}^{t_{i+1}} \theta(s) \int_{t_i}^{s} \theta(u) dW_u^1 \, dW_s^1.
\]

Furthermore, Corollary 2.1.18 gives
\[
\int_{t_i}^{t_{i+1}} \int_{s}^{t_{i+1}} \theta(u) dW_u^1 \, ds = \int_{t_i}^{t_{i+1}} \theta(u) \int_{u}^{t_{i+1}} 1 \, ds \, dW_u^1 = \int_{t_i}^{t_{i+1}} \theta(u)(t_{i+1} - u) \, dW_u^1,
\]
which shows
\[
\int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{s}^{t_{i+1}} \theta(u) dW_u^1 \, ds = \int_{t_i}^{t_{i+1}} \theta(u)(t_{i+1} - u) \, dW_u^1 f^1(t_i, Y_{t_i}) + \int_{t_i}^{t_{i+1}} (f^1(s, Y_s) - f^1(t_i, Y_{t_i})) \int_{s}^{t_{i+1}} \theta(u) dW_u^1 \, ds.
\]

The estimation of the term \( C_n \) is simple. Indeed, we have
\[
C_n = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \theta^2(s) \left( h_{j_1(n)}(t_i) - h_{j_1(n)}(s) \right) \, ds
\]
and therefore \(|C_n| \leq TC_{\theta \omega_{h_{j_1(n)}}(\Delta_n)}(\Delta_n)\), where we have introduced the constant
\[
C_0 := \max_{t \in [0,T]} \theta^2(t).
\]

Using the estimates for the modulus of continuity \( \omega_{h_{j_1(n)}}(\Delta_n) \) formulated in Lemma 4.3.4 we see that a sufficient condition for \( 2^{n/2}C_n = o(1) \) for \( n \to \infty \) is
\[
2^{2j_1(n) - n/2} = o(1) \quad \text{or equivalently} \quad j_1(n) - \frac{n}{4} \to -\infty \quad \text{for} \quad n \to \infty.
\]
Next we analyse the terms \( A_{i,n} \).

**Lemma 4.3.8**

If the resolution level \( j_1(n) \) is chosen in such a way that
\[
j_1(n) - \frac{n}{2} \to -\infty \quad \text{for} \quad n \to \infty
\]
then the term \( 2^{n/2} \sum_{i=1}^{4} A_{i,n} \) converges to zero in probability when \( n \) tends to infinity.
Proof:

1. We consider first $A_{1,n}$. As

$$\sqrt{\Delta_n} \sup_{0 \leq t \leq T} (f^1(t, Y_t))^2 = o_p(1) \quad \text{for} \quad n \to \infty$$

it suffices to show

$$\frac{1}{N-1} \sum_{i=0}^{N-1} \Delta_n |h_{j_1(n)}(t_i)| \to \int_0^T |h(t)| \, dt. \quad (4.32)$$

Because of $m' > 1/2$ the space $H^{m'}(\mathbb{R})$ is contained in $C(\mathbb{R})$ and one has

$$\sup_{t \in \mathbb{R}} |h_j(t) - h(t)| \leq C_s \|h_j - h\|_{H^{m'}(\mathbb{R})}. \quad (4.33a)$$

with a constant $C_s$. Now, by Proposition 4.1.4 we have

$$\|h_j - h\|_{H^{m'}} \to 0 \quad \text{as} \quad j \to \infty \quad (4.33b)$$

and therefore

$$\left| \frac{1}{N-1} \sum_{i=0}^{N-1} \Delta_n |h_{j_1(n)}(t_i)| - \int_0^T |h(t)| \, dt \right| \leq T \left( \omega_{|h_{j_1(n)}|}(\Delta_n) + \sup_{t \in \mathbb{R}} \left\{ |h_j(t)| - |h(t)| \right\} \right). \quad (4.34)$$

By Lemma 4.3.4 and the considerations (4.33) the right-hand side converges to zero as $n$ tends to infinity, provided $2j_1(n) - n \to -\infty$. Hence we have shown $2^{m/2} A_{1,n} = o_p(1)$ for $n \to \infty$.

2. Now we are going to consider $A_{2,n}$, which can be written as

$$A_{2,n} = \int_0^T \xi_u \, dW_u^1,$$

where $\xi_u$ denotes the process

$$\xi_u = 2 \sum_{i=0}^{N-1} \chi_{(t_i, t_{i+1})}(u) h_{j_1(n)}(t_i) f^1(t_i, Y_{t_i}) \theta(u) (t_{i+1} - u).$$

We consider now the auxiliary process

$$X_t = \left( \frac{1}{\sqrt{\Delta_n}} \int_0^t \xi_u \, dW_u^1 \right)^2.$$ 

We remark $X_T = \left( \frac{1}{\sqrt{\Delta_n}} A_{2,n} \right)^2$. We compute

$$\mathbb{E} X_t = \frac{1}{\Delta_n} \int_0^t \mathbb{E} \left\{ \xi_u^2 \right\} \, du$$

$$= \frac{4}{\Delta_n} \int_0^t \mathbb{E} \left\{ \sum_{i=0}^{N-1} \chi_{(t_i, t_{i+1})}(u) h_{j_1(n)}^2(t_i) f^{12}(t_i, Y_{t_i}) \theta^2(u)(t_{i+1} - u)^2 \right\} \, du$$

$$= \frac{4}{\Delta_n} \int_0^t \sum_{i=0}^{N-1} \chi_{(t_i, t_{i+1})}(u) h_{j_1(n)}^2(t_i) \theta^2(u)(t_{i+1} - u)^2 \mathbb{E} \left\{ f^{12}(t_i, Y_{t_i}) \right\} \, du.$$
Now we use Proposition 2.1.21 with $p = 2$, which gives
\[ E|Y_t|^2 \leq (1 + E|c|^2) e^{2\alpha t} \]
with $\alpha := \sqrt{K_T} + K_T \frac{1}{2}$, where $K_T$ is the constant appearing in the linear growth condition of Assumption 4.2.1. Using this linear growth condition we get
\[ E|f^1(t, Y_t)|^2 \leq E|f(t, Y_t)|^2 \leq K_T \left(1 + E|Y_t|^2\right) \leq K_T \left(1 + (1 + E|c|^2) e^{2\alpha T}\right). \]  

(4.35)

Using the constant $C_\theta$ defined (4.30) we can bound $EX_t$ as follows
\[ EX_t = \frac{4}{\Delta_n} C_T C_\theta \sum_{i=0}^{N-1} h_{j_1(n)}(t_i) \frac{1}{3} \Delta_n^3 \leq \hat{C}_T \Delta_n \sum_{i=0}^{N-1} h_{j_1(n)}^2(t_i) \Delta_n \]
with $\hat{C}_T := \frac{4}{3} C_T C_\theta$. Similar to (4.34) we get
\[ \sum_{i=0}^{N-1} h_{j_1(n)}^2(t_i) \Delta_n \to \int_0^T h^2(t) \, dt. \]  

(4.36)

Hence, there exists a constant $\hat{C}_T$ such that $EX_t \leq \hat{C}_T \Delta_n$ holds $\forall t \in [0, T]$. Thus, we have seen
\[ \mathbb{D}^2 \left\{ \frac{1}{\sqrt{\Delta_n}} A_{2,n} \right\} = EX_T \leq \hat{C}_T \Delta_n \to 0 \quad \text{for} \ n \to \infty. \]

Applying Tschebycheff inequality we get for arbitrary $\varepsilon > 0$
\[ \mathbb{P} \left( \left| \frac{1}{\sqrt{\Delta_n}} A_{2,n} \right| > \varepsilon \right) \leq \frac{\hat{C}_T \Delta_n}{\varepsilon^2}, \]
thus $2n^{2/3} A_{2,n} = o_p(1)$ for $n \to \infty$.

3. We consider now $A_{3,n}$. Remembering Assumption 4.2.1, especially $f^1(\cdot, \cdot) \in \mathcal{C}^1$ together with the fact that the trajectories of $Y_t$ are Hölder continuous with exponent $0 < \beta < \frac{1}{2}$, we can find for every trajectory of $Y_t$ constants $C_f$ and $H(\omega)$ such that
\[ |f^1(s, Y_s) - f^1(t_i, Y_{t_i})| \leq C_f \left( |s - t_i| + \left| Y_s - Y_{t_i} \right| \right) \leq H(\omega)|s - t_i|^\beta \]
holds with a suitable constant $\hat{C}(\omega)$. Furthermore for every trajectory of the stochastic integral $\int_{t_i}^s \theta(u) \, dW_u^1$ there exists a constant $\hat{C}(\omega)$ such that
\[ \left| \int_{t_i}^s \theta(u) \, dW_u^1 \right| \leq \hat{C}(\omega)(s - t_i)^\beta. \]

Thus, we get
\[ |A_{3,n}| \leq 2 \sum_{i=0}^{N-1} |h_{j_1(n)}(t_i)| \int_{t_i}^{t_{i+1}} \hat{C}(\omega) |s - t_i|^{\beta} \hat{C}(\omega) |s - t_i|^{\beta} \, ds \]
\[ \leq \hat{C}(\omega) \Delta_n^{2\beta} \sum_{i=0}^{N-1} |h_{j_1(n)}(t_i)| \Delta_n, \]
with \( \dot{C}(\omega) := 2\dot{\mathcal{C}}(\omega) \dot{C}(\omega) \). Remembering (4.32) and choosing \( \beta \in \left( \frac{1}{4}, \frac{1}{2} \right) \) we obtain

\[
2^{n/2} A_{3,n} \to 0 \quad \text{a.s.}
\]

As almost sure convergence implies stochastic convergence we get the assertion.

4. It remains to analyse \( A_{4,n} \), which can be bounded in a similar way as \( A_{2,n} \). We have

\[
A_{4,n} = \int_0^T \tilde{\xi}_s \, dW_s^1,
\]

where the process \( \tilde{\xi}_s \) is defined as

\[
\tilde{\xi}_s = 2 \sum_{i=0}^{N-1} \chi_{(t_i, t_{i+1})} \theta(s) \left( \int_{t_i}^s f^1(u, Y_u) \, du \right).
\]

Thus, we get

\[
\mathbb{D}^2 \left( \frac{1}{\Delta_n} A_{4,n} \right) = \frac{1}{\Delta_n} \int_0^T \mathbb{E} \left\{ \tilde{\xi}_s^2 \right\} \, ds
\]

\[
= \frac{4}{\Delta_n} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} h_{j_1(n)}^2(t_i) \mathbb{E} \left( \int_{t_i}^s f^1(u, Y_u) \, du \right)^2 \theta^2(s) \, ds
\]

\[
\leq \frac{4}{\Delta_n} \sum_{i=0}^{N-1} h_{j_1(n)}^2(t_i) C \int_{t_i}^{t_{i+1}} \mathbb{E} \left( \int_{t_i}^s f^1(u, Y_u) \, du \right)^2 \, ds,
\]

where we have used the constant \( C_\theta \) defined in (4.30).

Using Cauchy-Schwartz inequality for \( s \in [t_i, t_{i+1}] \) yields

\[
\int_{t_i}^s \left| f^1(u, Y_u) \right| \, du \leq \sqrt{\int_{t_i}^s \left| f^1(u, Y_u) \right|^2 \, du} \sqrt{\Delta_n},
\]

squaring both sides gives

\[
\left( \int_{t_i}^s f^1(u, Y_u) \, du \right)^2 \leq \Delta_n \int_{t_i}^s \left| f^1(u, Y_u) \right|^2 \, du.
\]

Thus,

\[
\mathbb{E} \left( \int_{t_i}^s f^1(u, Y_u) \, du \right)^2 \leq \Delta_n \int_{t_i}^s \mathbb{E} \left| f^1(u, Y_u) \right|^2 \, du \leq \Delta_n^2 C_T,
\]

where we have used the same constant \( C_T \) as in (4.35).

Thus, we can proceed in bounding \( \mathbb{D}^2 \left( \frac{1}{\Delta_n} A_{4,n} \right) \) as follows

\[
\mathbb{D}^2 \left( \frac{1}{\Delta_n} A_{4,n} \right) \leq 4C_\theta C_T \Delta_n \sum_{i=0}^{N-1} h_{j_1(n)}^2(t_i) \Delta_n.
\]

As above we can now apply Tschebyscheff inequality and get \( 2^{n/2} A_{4,n} = o_P(1) \) for \( n \to \infty \).
4.3. ASYMPTOTIC STUDY OF THE ESTIMATOR

Considering the auxiliary process $x^1_t := \int_{t_i}^t \theta(s) dW^1_s$ and integrating the term

$$2 \int_{t_i}^{t_{i+1}} x^1_s dx^1_s = 2 \int_{t_i}^{t_{i+1}} \theta(s) \theta(u) dW_u dW^1_s$$

in $A_{5,n}$ by parts (cf. (2.7)) yields

$$A_{5,n} = \sum_{i=0}^{N-1} h_{j_1(n)}(t_i) \left[ \left( \int_{t_i}^{t_{i+1}} \theta(u) dW^1_u \right)^2 - \int_{t_i}^{t_{i+1}} \theta^2(u) du \right].$$

(4.37)

Combining the above considerations we obtain the following Corollary, which is the main step for the proof of Theorem 4.3.2 given below.

**Corollary 4.3.9**

*Under the assumptions of Theorem 4.3.2, the integral $I_n$ defined in (4.20) can be decomposed as follows*

$$I_n = 2^{n/2} \sum_{i=0}^{N-1} h_{j_1(n)}(t_i) \left[ \left( \int_{t_i}^{t_{i+1}} \theta(u) dW^1_u \right)^2 - \int_{t_i}^{t_{i+1}} \theta^2(u) du \right] + o_P(1) \text{ for } n \to \infty.$$  

**Proof of Theorem 4.3.2.**

Defining

$$\xi_{i,n} = \Delta_n^{-1/2} h_{j_1(n)}(t_i) \left[ \left( \int_{t_i}^{t_{i+1}} \theta(u) dW^1_u \right)^2 - \int_{t_i}^{t_{i+1}} \theta^2(u) du \right]$$

it remains to show

$$S_N := \sum_{i=0}^{N-1} \xi_{i,n} \xrightarrow{d} N \left( 0, 2 \int_0^T \theta^4(t) h^2(t) dt \right) \text{ for } n \to \infty.$$  

where $\xrightarrow{d}$ denotes convergence in distribution. Note, that from $\eta_n \xrightarrow{d} \eta$ and $\zeta_n \xrightarrow{P} 0$ it follows $\eta_n + \zeta_n \xrightarrow{d} \eta$.

For fixed $n \in \mathbb{N}$ the random variables $\xi_{i,n} (i = 0, \ldots, N - 1)$ are independent, centred and have the variance

$$\mathbb{D}^2 \xi_{i,n} = \frac{2}{\Delta_n} h^2_{j_1(n)}(t_i) \left( \int_{t_i}^{t_{i+1}} \theta^2(u) du \right)^2$$

(cf. Corollary 2.1.5). We shall apply to $S_N = \sum_{i=0}^{N-1} \xi_{i,n}$ the Lindeberg theorem with Lya-
punov condition (cf. Proposition 2.1.7) with $\delta = 1$. For this, it suffices to prove

$$\sum_{i=0}^{N-1} \mathbb{D}^2 \xi_{i,n} \to 2 \int_0^T \theta^4(t) h^2(t) dt$$

(4.38)

$$\sum_{i=0}^{N-1} \mathbb{E} |\xi_{i,n}|^3 \to 0.$$  

(4.39)
(4.38) may be obtained in a similar same way as the convergence in (4.32). For (4.39) we use
\[
\sum_{i=0}^{N-1} E[\xi_{i,n}^3] \leq 28\Delta_n^{-3/2} \sum_{i=0}^{N-1} |h_{j_1(n)}(t_i)|^3 \left( \int_{t_i}^{t_{i+1}} \theta^2(u) \, du \right)^3 \leq C'\Delta_n^{1/2},
\]
cf. Lemma 2.1.5. Therefore, Theorem 4.3.2 has been proved under the conditions (4.26), (4.31) and \( m' > \frac{1}{2} \). If we seek \( j_1(n) \) under the form \( j_1(n) = [\alpha n] \), combining (4.26) and (4.31) yields that \( \alpha \) must satisfy
\[
\frac{1}{2(m + m')} \leq \alpha < \frac{1}{4}.
\]
In particular, it is necessary that \( m + m' > 2 \) holds.

### 4.3.2 Mean integrated square error

We study in this section the mean integrated square error of the estimator \( \hat{\theta}^2(t) \), which is defined by
\[
\mathcal{R}_n := \mathbb{E} \int_\mathbb{R} \left( \hat{\theta}^2(t) - \theta^2(t) \right)^2 h(t) \, dt.
\] (4.40)

In this context \( h(t) \) denotes a nonnegative continuous function with support in \([\alpha_1, \alpha_2] \subset (0, T)\). Following the classical method, we write \( \mathcal{R}_n \) as the sum of a squared bias term and a variance term
\[
\mathcal{R}_n = \mathcal{B}_n^2 + \mathcal{V}_n
\] (4.41)
with
\[
\mathcal{B}_n^2 := \int_\mathbb{R} \left( \mathbb{E}\hat{\theta}^2(t) - \theta^2(t) \right)^2 h(t) \, dt \quad \text{and} \quad \mathcal{V}_n := \mathbb{E} \int_\mathbb{R} \left( \hat{\theta}^2(t) - \mathbb{E}\hat{\theta}^2(t) \right)^2 h(t) \, dt.
\]

Of course we have \( \mathcal{V}_n \leq C'D_n \), where \( C' := \sup_{t \in \mathbb{R}} h(t) \) and
\[
D_n = \mathbb{E} \int_0^T \left( \hat{\theta}^2(t) - \mathbb{E}\hat{\theta}^2(t) \right)^2 \, dt.
\] (4.42)

In the following we will find bounds for the squared bias and the variance terms \( \mathcal{B}_n^2 \) and \( D_n \). Let us start by studying the bias term. To do this, we have to compute \( \mathbb{E}\hat{\theta}^2(t) - \theta^2(t) \).

**Lemma 4.3.10**

*Introducing the notation*

\[
B_1(t) := \sum_{k=-A}^{2j_1(n)T+A} \phi_{j_1(n),k}(t) \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds - \mu_{j_1(n),k} \right)
\]
\[
B_2(t) := 2 \sum_{k=-A}^{2j_1(n)T+A} \phi_{j_1(n),k}(t) \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \mathbb{E} \left( \int_{t_i}^{t_{i+1}} (y_s - y_{t_i}) f^1(s, Y_s) \, ds \right) \right)
\]
\[
B_3(t) := P_{j_1(n)} \hat{\theta}^2 - \theta^2
\]
we get the decomposition
\[ \mathbb{E}\hat{\theta}^2(t) - \theta^2(t) = B_1(t) + B_2(t) + B_3(t). \] (4.43)

**Proof:** For \( t \in [0, T] \) we can write \( \mathbb{E}\hat{\theta}^2(t) - \theta^2(t) \) as sum of \( \mathbb{E}\hat{\theta}^2(t) - P_{j_1(n)}\bar{\theta}^2(t) \) and \( B_3(t) \). The main point is therefore to prove that the first summand can be decomposed into \( B_1(t) \) and \( B_2(t) \). To do this we remember the definition of the estimator and the wavelet-decomposition of \( \bar{\theta}^2 \)

\[ \hat{\theta}^2(t) = \sum_{k \in \mathbb{Z}} \mu_{j_1(n),k} \phi_{j_1(n),k}(t) \quad \text{and} \quad P_{j_1(n)}\bar{\theta}^2(t) = \sum_{k \in \mathbb{Z}} \mu_{j_1(n),k} \phi_{j_1(n),k}(t). \] (4.44a)

In view of the compact support of the functions \( \phi_{j_1(n),k} \) (cf. (4.10a)) we see that for \( t \in [0, T] \) the summands in \( \hat{\theta}^2(t) \) and in \( P_{j_1(n)}\bar{\theta}^2(t) \) are relevant only if \(-A \leq k \leq 2^j(n)T + A \). Thus, we have

\[ \mathbb{E}\hat{\theta}^2(t) - P_{j_1(n)}\bar{\theta}^2(t) = \sum_{k = -A}^{2^j(n)T + A} \phi_{j_1(n),k}(t) \left( \mathbb{E}\hat{\mu}_{j_1(n),k} - \mu_{j_1(n),k} \right) \]
\[ = \sum_{k = -A}^{2^j(n)T + A} \phi_{j_1(n),k}(t) \left( \mathbb{E} \left\{ \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \left( y_{i+1} - y_i \right)^2 \right\} - \mu_{j_1(n),k} \right). \]

Applying Lemma 4.3.5 we see

\[ \mathbb{E} \left( y_{i+1} - y_i \right)^2 = 2\mathbb{E} \int_{t_i}^{t_{i+1}} (y_s - y_i) dy_s + \int_{t_i}^{t_{i+1}} \theta^2(s) ds. \]

Using \( dy_s = f^1(s, Y_s) ds + \theta(s) dW^1_s \) and \( \mathbb{E} \int_{t_i}^{t_{i+1}} (y_s - y_i) \theta(s) dW^1_s = 0 \) gives

\[ \mathbb{E} \left( y_{i+1} - y_i \right)^2 = 2\mathbb{E} \left\{ \int_{t_i}^{t_{i+1}} (y_s - y_i) f^1(s, Y_s) ds \right\} + \int_{t_i}^{t_{i+1}} \theta^2(s) ds. \]

Hence, we have

\[ \mathbb{E}\hat{\theta}^2(t) - P_{j_1(n)}\bar{\theta}^2(t) = \sum_{k = -A}^{2^j(n)T + A} \phi_{j_1(n),k}(t) \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \int_{t_i}^{t_{i+1}} \theta^2(s) ds - \mu_{j_1(n),k} \right) \]
\[ + 2 \sum_{k = -A}^{2^j(n)T + A} \phi_{j_1(n),k}(t) \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \mathbb{E} \left( \int_{t_i}^{t_{i+1}} (y_s - y_i) f^1(s, Y_s) ds \right) \right). \]

After the proof of Theorem 4.3.15 we will interpret the summands \( B_i \) defined in Lemma 4.3.10 and discuss their behaviour for increasing resolution level \( j_1 \) if the number of observations is fixed. At this point we only want to remark that the first two terms \( B_1(t) \) and \( B_2(t) \) arise from the fact that in general \( \mathbb{E} \hat{\mu}_{j,k} \neq \mu_{j,k} \). The third term \( B_3(t) \) is the approximation error.
Clearly, Proposition 4.1.8 tells us how to bound the approximation error. We obtain
\[ \int_{\mathbb{R}} B_3^2(t) \, dt = \left\| \theta^2 - P_j \theta^2 \right\|_{L^2(\mathbb{R})}^2 \leq C 2^{-2j_1 m} \left\| \theta^2 \right\|_{H^m(\mathbb{R})}, \] (4.45)
where the constant \( C \) is independent of \( j_1 \) and of \( f \).

In order to bound the integrals \( \int_{\mathbb{R}} B_3^2(t) \hat{h}(t) \, dt \) and \( \int_{\mathbb{R}} B_2^2(t) \, dt \) some more computations are necessary. The result is formulated in the following lemma.

**Lemma 4.3.11**

It holds
\[ \int_{\mathbb{R}} B_1^2(t) \hat{h}(t) \, dt \leq C_{B_1} 2^{4j_1(n) - 2n}, \] (4.46)
\[ \int_{\mathbb{R}} B_2^2(t) \, dt \leq C_{B_2} \Delta_n, \] (4.47)
where the constants \( C_{B_1} \) and \( C_{B_2} \) depend only on \( \phi, \theta^2 \) and \( \hat{h} \).

**Proof:** We start by studying \( B_1(t) \). Using Definition (4.14) we see that
\[ B_1(t) = R(t) + \sum_{k=-A}^{2j_1(n)T+A} \phi_{j_1(n),k}(t) \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds - \int_0^T \theta^2(s) \phi_{j_1(n),k}(s) \, ds \right), \]
where we have set
\[ R(t) = - \sum_{k=-A}^{2j_1(n)T+A} \phi_{j_1(n),k}(t) \int_{\mathbb{R}[0,T]} \theta^2(s) \phi_{j_1(n),k}(s) \, ds. \]

Considering precisely the supports of the functions \( \theta^2 \) and \( \phi_{j_1(n),k} \), we obtain that if
\[ 2^{j_1(n)} \geq 2 \frac{A}{\alpha_1} + 2 \frac{A}{T - \alpha_2}, \] (4.48)
then we have \( R(t) = 0 \) for all \( t \in [\alpha_1, \alpha_2] \), where \( [\alpha_1, \alpha_2] \) is the support of the weight function \( \hat{h}(t) \).

Let now
\[ \lambda_k := \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds - \int_0^T \theta^2(s) \phi_{j_1(n),k}(s) \, ds. \] (4.49)

Then it holds
\[ |\lambda_k| \leq C_1 \omega_{\phi_{j_1(n),k}}(\Delta_n) \leq C_2 2^{3j_1(n)/2} \Delta_n = C_2 2^{3j_1(n)/2 - n}. \]

Furthermore, because of
\[ B_1(t) = R(t) + \sum_{k=-A}^{2j_1(n)T+A} \lambda_k \phi_{j_1(n),k}(t) \]
we get
\[
\int_{\mathbb{R}} B_i^2(t) \hat{h}(t) \, dt \leq C_3 \sum_{k=-A}^{2^{j_i} T + A} \lambda_k^2 \leq C_4 2^{4j_i - 2n},
\]
provided (4.48) is satisfied. Here, \( C_i, i = 1, \ldots, 4 \) are constants depending only on \( \phi, \theta^2 \) and \( \hat{h} \).

In order to study \( B_2(t) \) we apply the formula (4.25) and obtain
\[
B_2(t) = 2 \sum_{k=-A}^{2^{j_i} T + A} \phi_{ji(n),k}(t) (S_k' + S_k''),
\]
where we have set
\[
S_k' = \sum_{i=0}^{N-1} \phi_{ji(n),k}(t_i) \mathbb{E} \left\{ \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{t_i}^{s} \theta(u) \, dW_u^1 \, ds \right\},
\]
\[
S_k'' = \sum_{i=0}^{N-1} \phi_{ji(n),k}(t_i) \mathbb{E} \left\{ \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{t_i}^{s} f^1(u, Y_u) \, du \, ds \right\}.
\]

We shall now bound \( S_k' \) and \( S_k'' \). Assumption 4.2.1 and Proposition 2.1.21 give
\[
\mathbb{E}[f^1(s, Y_s)]^2 \leq \mathbb{E} \left\{ K_T (1 + |Y_s|^2) \right\} \leq C_T
\]
(cf. also (4.35)). Using Cauchy-Schwartz inequality for the expectation we get
\[
\left| \mathbb{E} \left\{ \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{t_i}^{s} \theta(u) \, dW_u^1 \, ds \right\} \right| \leq \int_{t_i}^{t_{i+1}} \left| \mathbb{E} \left\{ f^1(s, Y_s) \int_{t_i}^{s} \theta(u) \, dW_u^1 \right\} \right| \, ds
\]
\[
\leq \int_{t_i}^{t_{i+1}} \sqrt{\mathbb{E}[f^1(s, Y_s)]^2} \sqrt{\int_{t_i}^{s} \theta^2(u) \, du} \, ds
\]
\[
\leq \hat{C}_1 \Delta_n \sqrt{\Delta_n} = \hat{C}_1 \Delta_n^{3/2}. \tag{4.51}
\]

Furthermore, using Jensen inequality for integrals we obtain
\[
\left| \mathbb{E} \left\{ \int_{t_i}^{t_{i+1}} f^1(s, Y_s) \int_{t_i}^{s} f^1(u, Y_u) \, du \, ds \right\} \right|
\leq \int_{t_i}^{t_{i+1}} \left| \mathbb{E} \left\{ f^1(s, Y_s) \int_{t_i}^{s} f^1(u, Y_u) \, du \right\} \right| \, ds
\leq \int_{t_i}^{t_{i+1}} \sqrt{\mathbb{E}[f^1(s, Y_s)]^2} \mathbb{E} \left\{ \int_{t_i}^{s} f^1(u, Y_u) \, du \right\}^2 \, ds
\leq \int_{t_i}^{t_{i+1}} \sqrt{C_T} \sqrt{\mathbb{E} \left\{ (s - t_i) \int_{t_i}^{s} f^2(u, Y_u) \, du \right\}} \, ds \leq \hat{C}_2 \Delta_n^2. \tag{4.52}
\]

Thus, we have
\[
|S_k' + S_k''| \leq \hat{C}_3 \sqrt{\Delta_n} \sum_{i=0}^{N-1} \Delta_n |\phi_{ji(n),k}(t_i)|.
\]
Remembering (4.4) and \( \Delta_n = 2^{-n} \) gives
\[
\Delta_n \left| \phi_{j_1(n),k}(t_i) \right| = 2^{-n} 2^{j_1(n)/2} \left| \phi(2^{j_1(n)} t_i - k) \right| = 2^{j_1(n)/2} 2^{-n} \left| \phi(2^{j_1(n)} t_i - k) \right| .
\]
Defining \( u_{i,k} = 2^{j_1(n)} t_i - k \) and computing
\[
u_{i+1,k} - u_{i,k} = 2^{j_1(n)} (t_{i+1} - t_i) = 2^{j_1(n)} \Delta_n = 2^{j_1(n)-n}
\]
we obtain
\[
\left| S'_k + S''_k \right| \leq \hat{C}_3 \sqrt{\Delta_n} \sum_{i=0}^{N-1} 2^{-\frac{j_1(n)}{2}} (u_{i+1,k} - u_{i,k}) \left| \phi(u_{i,k}) \right|
\leq \hat{C}_3 \sqrt{\Delta_n} 2^{-\frac{j_1(n)}{2}} \sum_{i=0}^{N-1} (u_{i+1,k} - u_{i,k}) \left| \phi(u_{i,k}) \right|
\leq \hat{C}_4 \sqrt{\Delta_n} 2^{-\frac{j_1(n)}{2}} .
\]
Finally we have
\[
\int_{\mathbb{R}} B^2_2(t) dt \leq \hat{C}_5 \Delta_n = O(2^{-n}) \quad \text{for } n \to \infty . \tag{4.53}
\]
Combining the above considerations (see (4.45) and Lemma 4.3.11) we obtain the following estimation of the bias term
\[
B^2_n \leq C_B \left( 2^{4j_1(n)-2n} + 2^{-2j_1(n)\hat{m}} + 2^{-n} \right) \tag{4.54}
\]
with a constant \( C_B \) which depends only on \( \phi, \theta^2 \) and \( \hat{h} \).

We are now going to consider the variance term. Clearly, Definition (4.42) implies
\[
\mathcal{D}_n = \mathbb{E} \int_{0}^{T} \left( \sum_{k=-A}^{2^{j_1(n)T+A}} \left( \hat{\mu}_{j_1(n),k} - \mathbb{E} \hat{\mu}_{j_1(n),k} \right) \phi_{j_1(n),k}(t) \right)^2 dt \tag{4.55}
\]
\[
= \mathbb{E} \sum_{k=-A}^{2^{j_1(n)T+A}} \left( \hat{\mu}_{j_1(n),k} - \mathbb{E} \hat{\mu}_{j_1(n),k} \right)^2 = \sum_{k=-A}^{2^{j_1(n)T+A}} \mathbb{D}_n^2 \hat{\mu}_{j_1(n),k} . \tag{4.56}
\]

Defining
\[
\xi_{i} := \left( \int_{t_i}^{t_{i+1}} \theta(u) dW^1_u \right)^2
\]
and
\[
\eta_{i} := \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) du \right)^2 + 2 \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) du \right) \left( \int_{t_i}^{t_{i+1}} \theta(u) dW^1_u \right)
\]
and remembering the Definition (4.15) of \( \hat{\mu}_{j_1(n),k} \) as well as the formula (4.25) we can write \( \hat{\mu}_{j_1(n),k} \) as follows
\[
\hat{\mu}_{j_1(n),k} = \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) (\xi_{i} + \eta_{i}) .
\]
Therefore, the variance $D^2 \hat{\mu}_{j_1(n),k}$ is then given by

$$D^2 \hat{\mu}_{j_1(n),k} = D^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \xi_i \right) + 2 \text{Cov} \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \xi_i, \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \eta_i \right)$$

$$+ D^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \eta_i \right).$$

(4.57)

**Lemma 4.3.12**  
The variance of the random variables $\eta_i$ is bounded by

$$D^2 \eta_i \leq E \eta_i^2 \leq c \Delta^3_n,$$

(4.58)

where $c$ denotes a constant.

**Proof:** It holds

$$E(\eta_i)^2 = E \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^4 + 4E \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^3 \left( \int_{t_i}^{t_{i+1}} \theta(u) \, dW_u^1 \right)$$

$$+ 4E \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^2 \left( \int_{t_i}^{t_{i+1}} \theta(u) \, dW_u^1 \right)^2.$$

Using Jensen inequality for $g(x) = x^4$ we get

$$\left( \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^4 \leq \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} f^{14}(u, Y_u) \, du$$

and therefore

$$\left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^4 \leq (t_{i+1} - t_i)^3 \int_{t_i}^{t_{i+1}} f^{14}(u, Y_u) \, du.$$

Thus, we have

$$E \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^4 \leq \Delta^3_n \int_{t_i}^{t_{i+1}} E f^{14}(u, Y_u) \, du \leq c_1 \Delta^4_n.$$  

Furthermore, using Cauchy-Schwartz inequality we get

$$\sqrt{E \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^2 \left( \int_{t_i}^{t_{i+1}} \theta(u) \, dW_u^1 \right)^2} \leq \sqrt{E \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^4} \sqrt{\int_{t_i}^{t_{i+1}} \theta(u) \, dW_u^1} \leq \sqrt{c_1 \Delta^4_n \Delta^2_n}.$$
Analogously, Cauchy-Schwartz inequality gives
\[
\mathbb{E} \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^3 \left( \int_{t_i}^{t_{i+1}} \theta(u) \, dW_u^1 \right) \\
\leq \sqrt{\mathbb{E} \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^4 \mathbb{E} \left( \int_{t_i}^{t_{i+1}} f^1(u, Y_u) \, du \right)^2 \left( \int_{t_i}^{t_{i+1}} \theta(u) \, dW_u^1 \right)^2}.
\]
and therefore the assertion.

In the proof of the next lemma we will apply the following proposition, which is proven in [17, Lemma 4.1].

**Proposition 4.3.13**
For \( h \in \mathcal{C}^1([0, T]) \) it holds
\[
\sum_{k \in \mathbb{Z}} \sum_{i=0}^{N-1} \Delta_n \phi^2_{j_1(n), k}(t_i) \xi_i \to \int_0^T h(s) \, ds \quad \text{for } n \to \infty.
\]

**Lemma 4.3.14**
For \( n \to \infty \) it holds
\[
E_n := \sum_{k=-A}^{2j_1(n)T-A} \mathbb{D}^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n), k}(t_i) \xi_i \right) = 2^{j_1(n)-1} \int_0^T \theta^4(t) \, dt + o \left( 2^{j_1(n)-n} \right) \quad (4.59a)
\]
\[
F_n := \sum_{k=-A}^{2j_1(n)T+A} \mathbb{D}^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n), k}(t_i) \eta_i \right) = o \left( 2^{j_1(n)-n} \right). \quad (4.59b)
\]

**Proof:** In order to prove (4.59a) we observe that the random variables \( \xi_i \) are independent and have the variance \( \mathbb{D}^2 \xi_i = \left( \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds \right)^2 \) (see Corollary 2.1.5). Hence, we obtain
\[
E_n = 2^{j_1(n)+1} A N \sum_{k=-A}^{2j_1(n)T-N} \sum_{i=0}^{N-1} \phi^2_{j_1(n), k}(t_i) \left( \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds \right)^2.
\]
Next we apply Proposition 4.3.13 to the sum \( E'_n \) defined by
\[
E'_n := 2^{j_1(n)T+A} N \sum_{k=-A}^{2j_1(n)T-A} \sum_{i=0}^{N-1} \phi^2_{j_1(n), k}(t_i) \theta^4(t_i) \Delta_n^2
\]
and obtain
\[
\frac{E'_n}{2^{j_1(n)T-N}} \to 2 \int_0^T \theta^4(s) \, ds.
\]
In order to prove (4.59a) it remains to show
\[
\frac{E_n - E'_n}{2^{j_1(n)T-N}} \to 0. \quad (4.60)
\]
Using Taylor expansion up to order one of $\theta^2(s)$ we obtain
\[ \theta^2(s) = \theta^2(t_i) + 2\theta'(t_i)\theta(t_i)(s - t_i) + |s - t_i|o(1) \quad \text{for } n \to \infty. \]

Integrating both sides of the equation yields
\[ \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds = \int_{t_i}^{t_{i+1}} \left( \theta^2(t_i) + 2\theta'(t_i)\theta(t_i)(s - t_i) + |s - t_i|o(1) \right) \, ds = \Delta_n \theta^2(t_i) + \Delta_n^2 \theta'(t_i)\theta(t_i) + \Delta_n^2 o(1) \quad \text{for } n \to \infty. \]

Squaring both sides we obtain
\[ \left( \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds \right)^2 = \Delta_n^2 \theta^4(t_i) + \Delta_n^2 \theta^3(t_i)\theta'(t_i) + \Delta_n^2 o(1) \quad \text{for } n \to \infty. \]

Thus, for $n \to \infty$ we get
\[
\left| \frac{E_n - E'_n}{2^{j_1(n)}\Delta_n} \right| \leq 2 \sum_{k=-A}^{2^{j_1(n)}T+A} \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i) \left| \Delta_n^2 \theta^3(t_i)\theta'(t_i) + \Delta_n^2 o(1) \right| = 2\Delta_n \sum_{k=-A}^{2^{j_1(n)}T+A} \sum_{i=0}^{N-1} \Delta_n \phi^2 \left( 2^{j_1(n)}t_i - k \right) \left| \theta^3(t_i)\theta'(t_i) + o(1) \right|.
\]

Bounding $2\theta(t_i)\theta'(t_i)$ by $\|\theta^2\|_{C^1(\mathbb{R})}$ and applying again Proposition 4.3.13 we get (4.59a).

We are now going to show (4.59b). Remembering (4.10a) and Lemma 2.1.1 gives
\[
\mathbb{D}^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\eta_i \right) \leq \mathbb{D}^2 \left( \sum_{i=2^{j_1(n)}(A+k)}^{2^{j_1(n)}(A+k)} \phi_{j_1(n),k}(t_i)\eta_i \right) \leq 2^{n-j_1(n)}(2A + 1) \sum_{i=2^{n-j_1(n)}(-A+k)}^{2^{n-j_1(n)}(A+k)} \phi^2_{j_1(n),k}(t_i) \mathbb{D}^2 \eta_i.
\]

Applying Lemma 4.3.12 we obtain for $n \to \infty$
\[
F_n = \sum_{k=-A}^{2^{j_1(n)}T+A} \mathbb{D}^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\eta_i \right) \leq C2^{-n} \sum_{k=-A}^{2^{j_1(n)}T+A} \sum_{i=0}^{N-1} \Delta_n \phi^2 \left( 2^{j_1(n)}t_i - k \right) \to T \quad \text{(cf. Proposition 4.3.13)}
\]
\[
\leq \mathcal{O}(2^{-n}) = o(2^{j_1(n)-n}).
\]

\[ \square \]

**Theorem 4.3.15**  
Let the Assumption 4.2.1 be satisfied and $j_1(n)$ be chosen in such a way that $j_1(n) - \frac{n}{2} \to -\infty$.
for $n \to \infty$. Then the square bias and the variance term of the mean integrated square error satisfy

$$B_n^2 \leq C_B \left(2^{4j_1(n)-2n} + 2^{-2j_1(n)m} + 2^{-n}\right)$$

$$D_n = 2^{j_1(n)-n}2 \int_0^T \theta^4(t) \, dt + o\left(2^{j_1(n)-n}\right),$$

where the constant $C_B$ depends only on the functions $\phi$, $\hat{h}$ and $\theta^2$ and not on the stochastic drift $f(t,Y_t)$.

**Proof:** In equation (4.54) we have already proven the assertion concerning the squared bias term $B_n^2$. It remains to show the estimate for $D_n$. Obviously it holds

$$\left| \text{Cov}\left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\xi_i, \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\eta_i \right) \right| \leq \sqrt{D^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\xi_i \right)} \sqrt{D^2 \left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\eta_i \right)}.$$

Using Cauchy-Schwartz inequality and applying Lemma 4.3.14 this gives

$$\left| \sum_{k=-A}^{2j_1(n)+A} \text{Cov}\left( \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\xi_i, \sum_{i=0}^{N-1} \phi_{j_1(n),k}(t_i)\eta_i \right) \right| \leq \sqrt{E_n} \sqrt{F_n} = o\left(2^{j_1(n)-n}\right) \text{ for } n \to \infty.$$

Together with the decomposition (4.57) this gives the assertion.

**Remark 4.3.16** Theorem 4.3.15 remains valid if the assumptions concerning the multiresolution analysis is relaxed as follows. Let $\{V_j\}_{j \in \mathbb{Z}}$ be a multiresolution analysis with an orthonormal wavelet basis such that the scaling function $\phi$ and the mother wavelet $\psi$ are compactly supported and $\phi, \psi \in C^1(\mathbb{R})$. Furthermore, let $m < l_p + 1$, where $l_p$ denotes the polynomial exactness of $V_j$, i.e. $V_j$ contains all polynomials with order less or equal than $l_p$. Indeed, in the current section we used the smoothness of the scaling function $\phi$ only at two points: First of all, we bounded the approximation error by Proposition 4.1.8, which remains valid if the assumption $\phi \in H^s(\mathbb{R})$ is replaced by $0 < s < l_p + 1$ (cf. [10, Corollary 3.3.1]). Furthermore, we applied [17, Lemma 4.1] which needs the assumption of a continuously differentiable wavelet $\phi$.

After the technical computations which lead to the bounds of the square bias term $B_n^2$ and the variance term $D_n$ it seems interesting to interpret the summands in the decomposition of the bias term. These considerations will give us some insight how these terms behave if $\Delta_n = 2^{-n}$ is fixed and $j_1$ tends to infinity. This insight is in turn useful when we discuss the choice of the resolution level $j_1$.

As we have seen, the bias is due to three different factors:
a) the error $B_1$, which results from the approximation of the integrals

$$
\mu_{j_1,k} = \int_{\mathbb{R}} \phi_{j_1,k}(s)\theta^2(s) \, ds
$$

by the sums $\sum_i \phi_{j_1,k}(t_i) \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds$;

b) the error $B_2$, which is due to the drift $f^1(s,Y_s)$.

c) the approximation error $B_3 = P_{j_1}\bar{\theta}^2 - \bar{\theta}^2$.

Clearly, if the parameter $j_1$ is chosen too small, the function $\bar{\theta}^2$ is poorly approximated by its projection $P_{j_1}\bar{\theta}^2$ and the approximation error $B_3$ is large. In fact, for small $j_1$ the projection $P_{j_1}\bar{\theta}^2$ shows only the main features of $\bar{\theta}^2$. As $j_1$ increases, more and more details are included. In this way, $B_3$ converges to zero when $j_1$ tends to infinity.

Furthermore, the expectation

$$
E\left(\int_{t_i}^{t_{i+1}} (y_s - y_{t_i})f^1(s,Y_s) \, ds\right)
$$

is clearly independent of $j_1$ and we have shown in (4.51) and (4.52) that its absolute value is bounded by $C_3\sqrt{\sum \Delta_n}$ (cf. also (4.25)). As a consequence, we can bound the error $B_2$ by (4.47), which is independent of $j_1$.

Next we interprete $B_1$, which stems from the error made by the approximation

$$
\int_{\mathbb{R}} \phi_{j_1,k}(s)\theta^2(s) \, ds \approx \sum_i \phi_{j_1,k}(t_i) \int_{t_i}^{t_{i+1}} \theta^2(s) \, ds. \tag{4.61}
$$

Whenever $\text{supp} \phi_{j_1,k} \subset [0,T]$ this error is exactly $\lambda_k$ (cf. (4.49)). Remember that for increasing $j_1$ the modulus of continuity of $\phi_{j_1,k}$ increases (cf. Lemma 4.3.4). As a consequence, also the upper bounds

$$
\omega_{\phi_{j_1,k}}(\Delta_n) \int_0^T \theta^2(s) \, ds
$$

of the absolute values $|\lambda_k|$ and of the integral

$$
\int_0^T B_1^2(t) \, dt = \sum_{k=-A}^{2^{(n)T+A}} \lambda_k^2
$$

increase.

Finally, we have to analyse the variance term $V_n$ in (4.41). As we have already seen, for small $j_1$ the estimate $\bar{\theta}^2$ shows only the main features of $\bar{\theta}^2$. As $j_1$ increases, more and more details are included. However, the data $(y_{t_{i+1}} - y_{t_i})^2$, which enter into our estimation procedure, are a noisy picture of the integrals $\int_{t_i}^{t_{i+1}} \theta^2(s) \, ds$. Thus, for increasing $j_1$ also this noise is recovered. As a result, the variance term $D_n^2$ increases without bound and the estimates $\hat{\theta}^2$ become more and more spiky.
Combining these considerations, we conclude that the parameter \( j_1 = j_1(n) \) should be chosen in such a way that the components of the bias terms \( B_1, B_2, B_3 \) and the variance term \( V_n \) are balanced. In mathematical terms this can be formulated as \( B_n^2 \approx D_n \). In view of Theorem 4.3.15 a necessary condition for this is

\[
2^{-2j_1(n)m} \lesssim 2^{j_1(n) - n} \quad \text{which means} \quad 2^{-2j_1(n)m} \leq C 2^{j_1(n) - n}
\]

with an appropriate constant \( C \). This condition reflects the fact that the error made by the approximation should be smaller or of the same size as the variance term \( D_n \). This condition is equivalent to \( j_1(n) \geq \frac{n - \log_4(C)}{2m + 1} \). In view of this restriction, the term \( D_n \) can not have a better asymptotic behaviour than

\[
D_n \lesssim 2^{\frac{n}{2m+1}} = 2^{-\frac{2m}{2m+1}n}.
\]

Furthermore, for the choice

\[
j_1(n) = \frac{n}{2m + 1}
\]

(4.62)

this asymptotic is achieved and also \( B_n^2 \) satisfies \( B_n^2 \lesssim 2^{-\frac{2m}{2m+1}n} \). Thus, we can formulate the following corollary.

**Corollary 4.3.17**

If the parameter \( j_1(n) \) is chosen in accordance with (4.62) the convergence rate

\[
R_n \lesssim 2^{\frac{2m}{2m+1}n},
\]

(4.63)

is achieved.

**Remark 4.3.18** As all the constants in the proof of Theorem 4.3.15 depend only on \( \| \overline{\theta}^2 \|_{C^1(\mathbb{R})}, \| \overline{\theta}^2 \|_{H^m(\mathbb{R})}, T, \psi, \mathbb{E}|c|^2 \) and \( \phi \) we can formulate (4.63) in a little bit stronger form as follows: Considering the time interval \([0, T]\) and the distribution of the initial conditions as fixed we can find for each \( L > 0, m > 0 \) a sufficiently regular wavelet basis and a constant \( C \), such that \( R_n \leq C 2^{\frac{2m}{2m+1}n} \) holds for all \( \overline{\theta} \) that satisfy \( \max \{\| \overline{\theta}^2 \|_{H^m(\mathbb{R})}, \| \overline{\theta}^2 \|_{C^1(\mathbb{R})} \} \leq L \). Furthermore, for \( m > 1 + \frac{1}{2} \) we have \( \| \overline{\theta}^2 \|_{C^1(\mathbb{R})} \leq C_s \| \overline{\theta}^2 \|_{H^m(\mathbb{R})} \) which shows that in this situation the constant \( C \) depends only on \( \| \overline{\theta}^2 \|_{H^m(\mathbb{R})} \).

So far we have shown that (4.63) is the best convergence rate that can be achieved by the estimator defined in (4.15) and (4.17). An interesting question is now, whether there are other estimators which achieve better convergence rates. An answer to this question is given in [24, Section 3.1], where the estimation of a time-depending volatility function \( \sigma(t) \) in a one-dimensional diffusion process over the interval \([0, 1]\) is briefly discussed.

Defining the set \( H^m([0, 1], L) = \{g \in H^m([0, 1]) : \|g\|_{H^m} \leq L\} \) with \( m > 1 + 1/2 \) and the indicator function \( h = \chi_D \) with a compact interval \( D \subset (0, 1) \) we can formulate the following implication of [24, Theorem 4]

\[
l := \lim \inf_{n \to \infty} \inf_{T_n} \sup_{\overline{\theta} \in H^m([0, 1], L)} \mathbb{E}_{\overline{\theta}} \left\{ 2^{\frac{2m}{2m+1}n} \int_0^1 \chi_D \left| \overline{\theta}^2(t) - \overline{T}_n(t) \right|^2 dt \right\}^{1/2} > 0,
\]

(4.64)
where the infimum is taken over all estimators \( \hat{T}_n \) that use the observations \( y_i, \quad i = 0, \ldots, 2^n \)
of the process \( y \) at the time points \( t_i = \frac{i}{2^n} \).

This result shows that at least for the situation \( m > 1 + 1/2 \) there is no admissible estimator \( \hat{T}_n \)
which achieves for all functions \( \theta \) a better convergence rate than \( 2^{-2m+1}n \). In other words, the
estimator \( \hat{\theta}^2 \) defined in (4.15) and (4.17) achieves the optimal rates of convergence, provided
the scaling functions \( \phi \) and corresponding wavelets \( \psi \) of the multiresolution analysis are
chosen sufficiently smooth.

We remark that in [24, Section 3.1] not only the mean integrated square error is investigated,
but the more general concept \( L^p \) minimax risk of an estimator \( \hat{T}_n \) (see also [19, Section 10.4]
for a nice introduction). Furthermore, their theory allows the smoothness assumption on \( \theta \)
to be formulated in terms of Besov spaces \( B_{spq} \), which comprise the spaces \( H^s (B_{2,2} = H^s) \) as
special cases. We believe that it is possible to generalise also these theories for a multivariate
model, however this is beyond the scope of this dissertation.

At the end of this section we want to make some remarks, which are analogous to the remarks
in [17, Section 5].

**Remark 4.3.19** We can generalise the class of models to the following situation: Let \( X_t \)
be a stochastic process that satisfies a stochastic differential equation

\[
dX_t = b(t, X_t) + \begin{pmatrix} \theta(t) h(x_1^1) & 0 \\ 0 & G_{2,2} \end{pmatrix} dW_t,
\]

where \( X_t^1 \) denotes the first component of the vector \( X_t \) and \( h \in C^2(\mathbb{R}) \) is some known function
satisfying \( \inf_{x \in \mathbb{R}} h(x) > 0 \). Assume that we are able to observe the first component of the
process \( X_t \). Then we could introduce the functions

\[
H(v) := \int_0^v \frac{1}{h(\xi)} d\xi
\]

and

\[
u(t, x) = u(x) = (H(x^1) x^2 \cdots x^d)^T, \quad x = (x^1 x^2 \cdots x^d)^T
\]

and consider the process \( Y_t = u(X_t) \). Applying Itô’s Lemma we obtain

\[
dY_t = \tilde{f}(t, X_t) dt + \begin{pmatrix} \theta(t) & 0^T \\ 0 & G_{2,2} \end{pmatrix} dW_t, \tag{4.65}
\]

where we have set

\[
\tilde{f}(t, x) := \begin{pmatrix} \frac{b^1(t,x)}{h(x)} - \frac{1}{2} \theta^2(t) h'(x^1) \\ b^2(t, x) \\ \vdots \\ b^d(t, x) \end{pmatrix}.
\]

We remark that because of

\[
H'(x) = \frac{1}{h(x)} \geq \inf_{x \in \mathbb{R}} \frac{1}{h(x)} > 0 \quad \forall x \in \mathbb{R} \tag{4.66}
\]
the inverse $H^{-1}$ exists. Furthermore, $H^{-1}$ is continuously differentiable and it holds

$$(H^{-1})'(y) = \frac{1}{H'(H^{-1}(y))}.\]

As a consequence, we get by defining

$$f(t, y) := \tilde{f}(t, \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}),$$

the relation $\tilde{f}(t, X_t) = f(t, Y_t)$ and therefore

$$dY_t = f(t, Y_t) \, dt + \begin{pmatrix} \theta(t) \\ 0_T \end{pmatrix}_G \, dW_t. \quad (4.67)$$

If we assume the function $b$ to be in $C^1(\mathbb{R} \times \mathbb{R}^d)$, then we get $\tilde{f} \in C^1(\mathbb{R} \times \mathbb{R}^d)$ and consequently $f \in C^1(\mathbb{R} \times \mathbb{R}^d)$. The growth condition in Assumption 4.2.1 poses further restrictions on $b$ and $h$.

**Remark 4.3.20** It is possible to interpret the presented wavelet estimator as kernel estimator. Indeed, defining the kernels

$$K_j(x, y) := \sum_{k \in \mathbb{Z}} \phi_{j,k}(x) \phi_{j,k}(y) \quad \text{for } j \in \mathbb{Z}, \quad (4.68)$$

the orthogonal projection $P_j : L^2(\mathbb{R}) \to V_j$ onto $V_j$ satisfies

$$[P_j f](x) = \int_{\mathbb{R}} K_j(x, y) f(y) \, dy.$$ 

Furthermore, the estimator $\hat{\theta}^2(t)$ can be written as a kernel estimate of the form

$$\hat{\theta}^2(t) = \sum_{i=0}^{N-1} K_j(t, t_i)(y_{i+1} - y_i)^2. \quad (4.69)$$

However, it should be remarked that the kernel $K_j$ is in general no convolution kernel, i.e. in general an equality of the form $K(x, y) = K(x - y)$ is not valid.

### 4.4 Numerical case studies

In this section we are going to illustrate the performance of the presented wavelet estimator by means of a numerical case study with synthetic data. To produce the data over a finite time interval $[0, T]$ we have simulated the logarithmic asset price process $p(t)$ described by
Model 3.0.26 with parameters introduced in Remark 3.0.27 and the four different squared volatility functions \( \sigma_1^2, \ldots, \sigma_4^2 \), which are illustrated in Figure 4.2.

We have chosen \( T = 16 \) and \( n = 16 \), i.e. we have simulated the asset price data over the time grid \( t_i = i\Delta_n, \quad i = 0, \ldots, N = 2^n T \), with step width \( \Delta_n = 2^{-n} \). For details concerning the simulation we refer to [33]. As we consider in this section the generalised bivariate Ornstein-Uhlenbeck model, in which we have denoted the squared volatility function with \( \sigma \) we use here the notation \( a \) instead of \( \theta \) and \( \hat{a}_{n,j_1} \) instead of \( \hat{\theta}_j \). The indices \( n \) and \( j_1 \) shall denote the dependence of the estimator from the step width \( \Delta_n \) and from the resolution level \( j_1 \). Furthermore, we use the notation \( a^\dagger \) for the true squared volatility function. We use the wavelet basis of Daubechies wavelets with order \( l = 5 \) (see Figure 4.1 for an illustration). We remark that for these wavelets the assumptions concerning the multiresolution analysis introduced above are satisfied and the polynomial exactness of \( V_j \) is \( l_p = 4 \) (cf. [38, p. 184]).

As a start we want to illustrate the mentioned dependence of the estimator from the step width \( \Delta_n \) and from the resolution level \( j_1 \), respectively. For this we consider the squared volatility function \( \sigma_2^2 \) and consider one single trajectory of the asset price process. At first we fix the time grid and vary the resolution level \( j_1 \). The results are shown in Figure 4.3. As we have expected from our theoretical considerations after Theorem 4.3.15, for small resolution level \( j_1 \) the approximation is very poor. For increasing \( j_1 \) we can see two effects, on the one hand side the approximation starts first to get better, as more and more details are included. However, for increasing \( j_1 \) the noise in the estimator increases also, which leads to very spiky estimators for large \( j_1 \). Therefore, the resolution level \( j_1 = 6 \) is clearly too large.

These considerations show that the right choice of the resolution level is of great importance. Clearly, the question which is the best resolution level depends on the norm or metric which measures the distance between the estimator \( \hat{a}_{n,j_1} \) and the exact function \( a^\dagger \). A frequently
used measure for this distance is the integrated square error

\[ ISE_h(\hat{a}_{n,j_1}) := \int_0^T h(t)(\hat{a}_{n,j_1}(t) - a^\dagger(t))^2 \, dt \]  

(4.70)

where \( h \) denotes a nonnegative weight function. In the following we will use this approach with the weight function \( h \) illustrated in Figure 4.4.

Here, a brief remark concerning the weight function \( h \) is in order. The support of this function should be included in \((0, T)\), i.e. this function should vanish in small regions near the boundaries (cf. also the theoretical considerations in Section 4.3). As for the function \( h \) these regions are relatively large it should be remarked that we experimented also with other weight functions where these regions are smaller. In principal the results obtained there were not much worse but for illustrating the proven convergence rates the weight function \( h \) was more suitable.

With respect to the measure (4.70) we can now define the optimal resolution level \( j_1^{\text{opt}}(n) \) as

\[ j_1^{\text{opt}}(n) := \arg\min_{j_1 \in \mathbb{Z}} ISE_h(\hat{a}_{n,j_1}) . \]  

(4.71)

To make notation easier we will write simply \( j_1^{\text{opt}} \) instead of \( j_1^{\text{opt}}(n) \) when appropriate. For the trajectory considered above the optimal resolution level is \( j_1^{\text{opt}}(n) = 3 \). The corresponding estimator \( \hat{a}_{n,j_1^{\text{opt}}} \) is shown in Figure 4.5. Together with the estimators shown in Figure 4.3 we see that for \( j_1 < j_1^{\text{opt}} \) the peaks are not sufficiently well reconstructed as the approximation properties of the spaces \( V_{j_1} \) (\( j_1 < j_1^{\text{opt}} \)) are too poor. On the other hand for \( j_1 > j_1^{\text{opt}} \) the estimators get more and more spiky. In other words, \( j_1^{\text{opt}} \) realizes a good compromise between the approximation properties and the smoothness of the estimator.

Before we discuss the interesting question how the resolution level can be chosen in practice we want to illustrate the dependence of the estimator \( \hat{a}_{n,j_1^{\text{opt}}(n)} \) from the number of observations. In view of the theoretical results we can expect that the estimator becomes better when the number of observations increases. To be precise, we have seen in Section 4.3.2 the
conjecture that the constant before the term $2^{-2^{m-n}}$ of the mean integrated square error, provided $\overline{a} \in H^m(\mathbb{R})$ with $0 < m < l_p + 1$, where $l_p$ denotes the polynomial exactness of $V_j$ (cf. especially Remark 4.3.16 and Corollary 4.3.17). Remember that for Daubechies wavelets of order $l = 5$ the polynomial exactness is $l_p = 4$, thus $l_p + 1 = 5$. Although the case $m = l_p + 1$ is excluded, it is still interesting to note that $2^{-10} \approx 0.5325$. For sufficiently smooth volatility functions ($\overline{a} \in H^m$ with $m = 5$) we can therefore expect the convergence rate to be little slower than $2^{-10}$. Now, some remarks concerning the smoothness of the different volatility functions $a_1, \ldots, a_4$ are in order. Clearly, the function $a_1$ has four jumps and does therefore not satisfy the assumptions made in Section 4.2. As opposed to that the functions $a_2, a_3, a_4$ are all elements of $C'^\infty[0, T]$. Nevertheless, the smoothness of these functions differs inasmuch as $a_2$ has three relatively high peaks, which means that the $L^2$-norms of its derivatives are relatively high.

To illustrate the convergence we have simulated for each of the four squared volatility functions $1000$ trajectories of the asset price processes over the timegrid with $n_1 = 16$ introduced above. By using the data at every grid point, at every second grid point etc. we get from this simulation also data on the coarser time grids with step widths $\Delta_n = 2^{-n_k}$. $n_2 = 15, n_3 = 14, n_4 = 13$. On every time grid we computed for every trajectory the optimal resolution level $j_1^{opt}$ and the corresponding integrated square error $ISE_h(\hat{\alpha}_{n_k,j_1^{opt}})$. The average over all trajectories is denoted by $AISE_h(\hat{\alpha}_{n_k,j_1^{opt}})$ and provides a good approximation of the mean integrated square error.

The results are presented in Table 4.1, which shows the average integrated square errors in dependence of $n_k$. The data illustrate very nicely the convergence of the average integrated square error to zero. As expected, this convergence is relatively slow for the piecewise constant volatility function $a_1$. The quotient $\frac{AISE_h(\hat{\alpha}_{n_k-1,j_1^{opt}})}{AISE_h(\hat{\alpha}_{n_k,j_1^{opt}})}$ is about 0.73 for $a_1 = a_1$, whereas it is about $2^{-10} \approx 0.53$ for the smoothest functions $a_3$ and $a_4$. This fact matches very well our expectations concerning the expected convergence rate.

For the volatility function $a_2$ this asymptotic result can not be seen from the data, as in this situation the mentioned quotient is about 0.6. To explain this, we remark that the $L^2$-norms of the $k$-th derivative, i.e. $\|a_2^{(k)}\|_{L^2(0,T)}$, increase rapidly for increasing $k$. Therefore, one can conjecture that the constant before the term $2^{-2^{m-n}}$ is for large $m$ probably too high to see

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.4870$</td>
<td>$0.3019$</td>
<td>$0.0696$</td>
<td>$0.0871$</td>
</tr>
<tr>
<td>$0.3465$</td>
<td>$0.1935$</td>
<td>$0.0378$</td>
<td>$0.0399$</td>
</tr>
<tr>
<td>$0.2658$</td>
<td>$0.1105$</td>
<td>$0.0184$</td>
<td>$0.0211$</td>
</tr>
<tr>
<td>$0.1932$</td>
<td>$0.0659$</td>
<td>$0.0110$</td>
<td>$0.0104$</td>
</tr>
</tbody>
</table>

Table 4.1: AISE for an optimal choice of the resolution level
the asymptotic rate in the data. Nevertheless, this function will be very illustrative for the outlook on wavelet thresholding presented in Section 4.5.

We are now going to address the question how the resolution level \( j_1 \) should be chosen in practice. Clearly, the choice (4.71) is only possible for academic use, when the exact volatility function \( a^\dagger \) is known. In Section 4.3.2 we have proven that the choice \( j_1(n) = \frac{n}{2^m+1} \) yields the optimal convergence rate. Unfortunately, in practice the asset price data are usually only available for a fixed time grid. Moreover, in general we are not aware of the smoothness of the exact volatility, i.e. \( m \) is in general unknown. In this situation the asymptotic choice is not applicable. It is therefore desirable to find a data-driven method which chooses for given asset price data a resolution level \( j_1^{\text{heur}} \) which is in general close to \( j_1^{\text{opt}} \).

In [43] the so-called L-method is suggested to perform this task. In the following section we will motivate and describe this method. After that we will illustrate and discuss its performance by means of both, theoretical considerations and numerical case studies. Although there exist some pathological volatility functions where this method fails it worked quite well in many situations.

4.4.1 The L-method as criteria for the choice of the resolution level

In order to describe the L-method it is helpful to reformulate the definition of the wavelet estimator (4.17). Using the estimated coefficients \( \hat{\mu}_{j_1,k} \) defined in (4.15) we can define recursively for \( j < j_1 \)

\[
\hat{\mu}_{j-1,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_{n-2k} \hat{\mu}_{j,k} \\
\hat{\nu}_{j-1,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} g_{n-2k} \hat{\mu}_{j,k}
\]

where \( h_n \) denote the filter coefficients occurring in the refinement equation (4.6) and \( g_n \) are defined in (4.7). Now the estimator \( \hat{a} \) can be written as

\[
\hat{a}_{n,j_1}(t) = \sum_{k \in \mathbb{Z}} \hat{\mu}_{j_1,k} \phi_{j_1,k}(t) + \sum_{j=j_0}^{j_1-1} \sum_{k \in \mathbb{Z}} \hat{\nu}_{j,k} \psi_{j,k}(t).
\]

(cf. [10, Section 2.6]). Furthermore, the coefficients \( \hat{\mu}_{j,k} \) and \( \hat{\nu}_{j,k} \) satisfy

\[
\hat{\mu}_{j,k} = \sum_{i=0}^{N-1} \phi_{j,k}(t_i)(p_{t_{i+1}} - p_{t_i})^2 \quad \text{and} \quad \hat{\nu}_{j,k} = \sum_{i=0}^{N-1} \psi_{j,k}(t_i)(p_{t_{i+1}} - p_{t_i})^2.
\]

In this context, Pinheiro, et al. suggest to choose the minimal level of resolution \( j_0 \) as the largest \( j \in \mathbb{Z} \), for which there exists a \( k \in \mathbb{Z} \) such that \( [0,T] \) is covered by the support of \( \phi_{j,k} \), i.e.

\[
\hat{j}_0 := \max \{ j \in \mathbb{Z} : \exists k \in \mathbb{Z} : [0,T] \subset \text{supp} \phi_{j,k} \}.
\]

(4.72)
4.4. NUMERICAL CASE STUDIES

As the support of the scaling function $\phi$ of Daubechies wavelets of order $l$ has the support $\text{supp} \phi = [0, 2^l - 1]$ this rule means that the minimal resolution level $j_0$ is chosen as $j_0 = \lfloor \log_2(2^l - 1) - \log_2(T) \rfloor$.

The $L$-method is based on the concepts level energy and empirical level energy which we introduce in the following.

**Definition 4.4.1 (Level energy)**

The level energy of level $j$ of a function $\overline{a}$ given by the wavelet decomposition

$$\overline{a}(t) = \sum_{k \in \mathbb{Z}} \mu_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \nu_{j,k} \psi_{j,k}(t)$$

is defined as $d_j^2 := \sum_{k \in \mathbb{Z}} \nu_{j,k}^2$. Furthermore, the empirical level energy $\hat{d}_j^2$ of the estimator $\hat{a}$ defined in (4.72) is given by $\hat{d}_j^2 := \sum_{k \in \mathbb{Z}} \hat{\nu}_{j,k}^2$.

Proposition 4.1.8 implies $d_j^2 \leq C 2^{-jm} \| \overline{a} \|_{H^m}$ with an appropriate constant $C$, provided the function $\overline{a} \in H^m(\mathbb{R})$ and the wavelet basis is smooth enough. In other words, the level energies decrease exponentially fast for increasing $j$, which means that the energies are concentrated in few levels of low resolution.

As opposed to that the behaviour of the empirical level energies $\hat{d}_j^2$ for increasing $j$ is fundamentally different in the sense that the empirical level energies can become arbitrarily large and coefficients estimates from high levels tend to be larger than those of low level. This fact is known from wavelet estimation of distribution functions (cf. [44]). In [43] Pinheiro et al. argue that it can be extended to the situation of volatility estimation. Certainly, for insiders this extension is straightforward. However, as the notation in [43] is at some places a little bit confusing (especially the order of Daubechies wavelets and the number of observations have both been denoted by $N$; furthermore the distinction between the empirical level energy, which is a stochastic variable, and its mean is not always clear) it seems appropriate to use the ideas of [44] to proof the mentioned result for the setting of volatility estimation.

In order to be brief and to put weight on the main idea we will confine our considerations to the situation of a model with zero drift, i.e. we set $f \equiv 0$ in (4.11). After the presentation of the theory we will illustrate by means of a numerical case study that the described behaviour occurs also in models that contain stochastic drift components.

As a start we need a small auxiliary result concerning Daubechies wavelets. It states that for every order $l \in \mathbb{N}$ there exists a positive constant $B_{\#l}$ such that for every $x \in \mathbb{R}$ there exists $k \in \mathbb{Z}$ such that $\psi_{0,k}^2(x) \geq B_{\#l}$. This assertion has been formulated in [44, Lemma 3.2].

**Proposition 4.4.2**

Let $\psi$ be a Daubechies wavelet of order $l$. Then the value

$$B_{\#l} := \min_{x \in \mathbb{R}} \max_{k \in \mathbb{Z}} \psi_{0,k}^2(x)$$

is attained and strictly positive.
Now we are able to prove the mentioned result about the empirical level energies.

**Lemma 4.4.3**
Let the logarithmic asset price process satisfy the stochastic differential equation

\[ dp_t = \sigma(t) \, dW_t \quad p_0 = c_0. \]

Let \( \{\hat{d}_j\}_{j=j_0}^{\infty} \) be the empirical level energies of the wavelet estimator (4.72) with Daubechies wavelets of order \( l \). Then there exists a constant \( C > 0 \) such that

\[ \mathbb{E} \hat{d}^2_j \geq \frac{B_{\#l} 2^j}{NC} \quad \text{for all } j > j_0. \]

**Proof:** Whenever \( j \) is sufficiently large \( \text{supp} \psi_{j,k} \) contains at most one value \( t_i \). In this situation we have

\[ \psi_{j,k}(t_i)\psi_{j,k}(t_i') = 0 \quad \text{whenever } i \neq i'. \]

Consequently,

\[ \hat{d}^2_{j_1} = \sum_{k \in \mathbb{Z}} \sum_{i=0}^{N-1} \psi^2_{j_1,k}(t_i) \left(p_{t_{i+1}} - p_{t_i}\right)^4 = \sum_{i=0}^{N-1} \left(p_{t_{i+1}} - p_{t_i}\right)^4 \sum_{k \in \mathbb{Z}} \psi^2_{j_1,k}(t_i). \]

Noting

\[ \sum_{k \in \mathbb{Z}} \psi^2_{j_1,k}(t_i) \geq \max_{k \in \mathbb{Z}} \psi^2_{j_1,k}(t_i) \geq \max_{k \in \mathbb{Z}} 2^{j_1} \psi^2_{0,k}(2^{j_1} t_i - k) \geq 2^{j_1} B_{\#l} \]

this gives

\[ \hat{d}^2_{j_1} \geq 2^{j_1} B_{\#l} \sum_{i=0}^{N-1} \left(p_{t_{i+1}} - p_{t_i}\right)^4. \]

Furthermore we have

\[ p_{t_{i+1}} - p_{t_i} \sim \mathcal{N}\left(0, \int_{t_i}^{t_{i+1}} \sigma^2(s) \, ds\right), \]

which gives

\[ \mathbb{E} \left(p_{t_{i+1}} - p_{t_i}\right)^4 = 3 \left(\int_{t_i}^{t_{i+1}} \sigma^2(s) \, ds\right)^2 \geq 3\sigma^4_{\min} \Delta_t^2 \]

and therefore

\[ \mathbb{E} \hat{d}^2_{j_1} \geq \frac{2^{j_1} B_{\#l} 3\sigma^4_{\min} T^2}{N}. \]

In this context we denote by \( \sigma_{\min} \) some positive lower bound of the function \( \sigma \).

We have seen that the level energies \( d_j^2 \) converge exponentially to zero. On the other hand, the empirical coefficients do not match this pattern. Typically, the empirical level energies decrease until they reach their minimum. After this minimum they increase and tend to
4.4. NUMERICAL CASE STUDIES

Figure 4.6: Behaviour of the integrated square error and the empirical level energies for $a_2$, both considered in dependence of the resolution level $j_1$

infinity. Furthermore, the behaviour of $d_j^2$ is generally very similar to the behaviour of the integrated square error, both considered as functions of the level $j$. The L-method suggests now to adjust the maximal level of resolution as follows. After choosing $j_0$ by (4.74) three levels of detail are always maintained and after the third the details are maintained as long as the $j$-th level energy is at most the same as the $(j - 1)$-th level energy. This can be formalised as follows.

$$j := j_0 + 3 \quad \text{while} \quad d_{j+1}^2 \leq d_j^2 \quad \text{do} \quad j++$$

(4.77)

To get an impression of the performance of the L-method, we consider again the numerical case study introduced above. To start with, we illustrate the dependence of the empirical level energies $d_j^2$ as well as the integrated square error of the level $j$. To do this we have chosen a concrete trajectory simulated with the volatility function $a^\dagger = a_2$ and have plotted the integrated square error as well as the empirical level energies $d_j^2$ in dependence if $j$ (cf. Figure 4.6). We can clearly see that both functions decrease in the first part of the plot until they reach their minimum. After that they increase exponentially. (Note that we have used a logarithmic scale on the vertical axis.)

We have to admit that Figure 4.6 shows an ideal situation. In general, the minima of the two curves do not necessarily coincide. However, in most situations which we considered they were relatively close. Clearly, this method is based on the assumption that the level energies $d_j^2$ behave nicely in the sense that they decrease monotonically. However, for every wavelet basis one can construct a function which does not match this pattern. To do this one has simply to choose coefficients $\nu_{j,k}$ in the decomposition (4.75) in such a way that the level energies $d_j^2 = \sum_{k \in \mathbb{Z}} \nu_{j,k}^2$ have more than one local minimum.

For the Daubechies wavelets of order $l = 5$ such a function $a_5$ is displayed in Figure 4.7. The corresponding level energies $d_j^2$ for $j > j_0 = -1$ are shown in Figure 4.8, which illustrates
the local minimum at \( j = 2 \). Therefore, it can be expected that also the empirical level energies show such a minimum, provided the number of observations is sufficiently large. To illustrate this we have simulated several trajectories with the volatility function \( a_5 \), \( n = 16 \) and the remaining parameters as above. For every trajectory we compared the behaviour of the empirical level energies and the integrated square error. Typically, this behaviour was similar to the situation shown in Figure 4.9, i.e. the empirical level energies have a local minimum at \( j = 2 \). Therefore, the L-Method yields \( j^L_1 = 2 \), which is relatively far away from the optimal resolution level \( j_1^{opt} = 5 \).

Although this example is a rather unnatural situation, it indicates that one cannot expect any general convergence results for the estimator \( \hat{a}_j \), which chooses the resolution level by the L-method. In fact, for the volatility function \( a_5 \) the L-method is very unlikely to choose a resolution level larger than \( j^L_1 = 2 \). Therefore, the approximation error does not tend to zero, which means that the mean integrated square error cannot tend to zero.

Apart from this constructed examples, in which the L-Method clearly failed, we observed some problems for very smooth volatility functions. For example for the squared volatility function \( a_4 \) and \( n = 16 \) the optimal resolution level was nearly always \( j_1^{opt} = 1 \). However, according to Pinheiro et al. we have chosen \( j_0 = -1 \) (cf. (4.74)). Then the L-method (4.77) can only yield a level \( j^L_1 \geq 2 \). Furthermore, for some trajectories the L-method yielded also a level \( j^L_1 > 2 \). Furthermore, for smaller \( n \) this effect became even worse, as the optimal resolution level decreased further, therefore the gap between \( j_1^{opt} \) and \( j_1^L \) increased.

To get an impression of the average performance of the L-Method we consider again the numerical case study introduced above. For each of the 1000 trajectories we performed the L-method and computed the average of the corresponding integrated square errors. To compare them with the average integrated square errors obtained with an optimal choice of the resolution level \( j_1^{opt} \) we computed the quotient

\[
\frac{\text{AISE}_h(\hat{a}_{nk,j^L_1})}{\text{AISE}_h(\hat{a}_{nk,j_1^{opt}})}
\]

The results are shown in Table 4.2. Remember that we have set \( n_1 = 16 \), \( n_2 = 15 \), \( n_3 = 14 \) and \( n_4 = 13 \).
It can be seen that the average integrated square error for the L-method was always less than four times of the one obtained with an optimal choice of the resolution level. Furthermore, the quotients are high for the smooth functions $a_3$ and $a_4$ and within these two lines they attain their maximum for $n_4$, which corresponds to a relatively small number of observations. Also for the very smooth functions the quotients tend to get better for increasing number of observations. We remark that we experimented also with other volatility functions. As a conclusion, we would say that the L-method seems to be an interesting parameter choice criterion if no apriori information about the smoothness of the solution is available. However, it should be taken in mind that there are situations where this method can fail and does not yield convergence for $n \to \infty$.

Remark 4.4.4 It is interesting to note that the L-method is very similar to the quasioptimality criterion which is a well known heuristic method for choosing the regularization parameter in the context of inverse problems (cf. for instance [25, p. 142] for a general introduction). To see the similarities let us think of the estimator (4.17) in an abstract way as regularization method (regularization by projection) for the inverse problem

$$Ax = z,$$

where $x := \theta^2$, $A$ denotes the integral operator and the noise free data $z^\dagger$ are given by

$$z^\dagger(t) = \int_0^t \theta^2(s) \, ds.$$

The observable (discrete) noisy data is the $\pi_n$-quadratic variation over intervals $[0, t]$ for varying $t$, where $\pi_n$ denotes the partition $\pi_n := \{t_0 \leq \cdots \leq t_N : t_k = k2^{-n}, k = 0, \ldots, 2^n T\}$.

In this context the regularization parameter is the resolution level $j_1$. Noting

$$\tilde{d}_j = \sum_{k \in \mathbb{Z}} \hat{\nu}_{j,k}^2 = \| \hat{\theta}_{n+1}^2 - \hat{\theta}_{n,j}^2 \|^2_{L^2(\mathbb{R})},$$

the quasioptimality criteria would suggest to choose the regularization parameter $j_1$ as minimum of

$$\tilde{d}_j \to \min_{j \in \mathbb{N}}.$$
which is equal to (4.77) as long as there is only one local minimum.

After explaining the similarities between the quasioptimal criteria and the L-method presented here two remarks are in order. First of all, apparently there is no similarity to Hansen’s L-method, which is another heuristic parameter choice rule in the context of inverse problems, see for instance [18] or also [25, p. 177]. (Although the similar name and the fact that both methods show a plot which resembles an L would suggest a relation.) Secondly, when talking about heuristic choice of the regularization parameter, i.e. a choice which does solely dependent on the noisy data $z^\delta$ but not on the data error $\|z^\dagger - z^\delta\|$, one should always bear in mind the warning of the Bakushinsky veto which states that for linear ill-posed inverse problems a parameter choice according to a heuristic criteria can never lead to a convergent regularization method (cf. [12, p.52]). Although we are not aware of a similar result in the context of nonparametric wavelet estimators we have seen that the specific “heuristic” choice (4.77) does not guarantee convergence of the estimator to the exact volatility.

4.5 Outlook: Wavelet thresholding

We want to finish this section by discussing some advantages of the wavelet estimator (4.17) in comparison to a general kernel estimator (4.69). First of all, wavelet estimators are superior inasmuch as they are faster and need less memory resources (cf. for example [43, p. 22]). Furthermore, it should be mentioned that the decomposition (4.72) offers the possibility of smoothing the estimator by wavelet thresholding. This final section is intended to give a brief motivation for this method. However, it does not intend to answer the arising questions whether soft or hard thresholding is to be preferred or how the threshold should be chosen in practice.

In the discussion after Theorem 4.3.15 and in the numerical case studies presented above we have seen that the linear wavelet estimate (4.17) may have spikes, which is due to the fact that unnecessary many details of the data $(y_{t+1} - y_t)^2$ are recovered. As a consequence of these considerations it seems natural to suppress some of the details, which means that we shrink certain coefficients $\nu_{j,k}$ in the decomposition (4.72).

Loosely spoken, the main idea is that large details are likely to result from a detail of the true volatility function, whereas small details result from stochastic noise. Therefore large coefficients, i.e. coefficients whose absolute value is larger than some threshold $\kappa$ stay or are only slightly damped and smaller coefficients are cut off. As there are various methods how this damping and cutting off can be realized there exist various thresholding procedures, see for example [19, Chapter 11]. Probably the most famous and simplest ones are soft thresholding and hard thresholding. In hard thresholding one simply cuts off the small coefficients, i.e. $\hat{\nu}_{j,k}$ is replaced by

$$\hat{\nu}^H_{j,k} = \begin{cases} 
\hat{\nu}_{j,k} & |\hat{\nu}_{j,k}| > \kappa \\
0 & \text{else}
\end{cases}$$

(4.78)
In soft thresholding all coefficients are damped. In this situation one replaces \( \hat{\nu}_{j,k} \) by

\[
\hat{\nu}^S_{j,k} = (|\hat{\nu}_{j,k}| - \kappa)_+ \text{sign}(\hat{\nu}_{j,k}).
\]  

(4.79)

Furthermore, it is possible to allow the threshold \( \kappa = \kappa_j \) to depend on the level \( j \). In this case one speaks of level-dependent thresholding.

To give an idea of the capabilities of wavelet thresholding we consider the volatility function \( a_2 \) introduced above. We simulated asset price data over a time interval with step width \( \Delta_n = 2^{-12} \) and the remaining parameters as above. Figure 4.10 shows the estimator \( \hat{a}_{j_{opt}} \) with optimal resolution level and Figure 4.11 the corresponding estimator after hard thresholding with an appropriate threshold \( \kappa \). It can clearly be seen that thresholding removed some of the smaller oscillations in the smooth parts of the volatility function without damping the peaks.
Chapter 5

Operator equations with Nemytskii operators

This chapter is intended to provide some theoretical foundations which will be useful in Chapter 6 when we address the inverse problem of option pricing. In the first section we give a short introduction to the theory of inverse and ill-posed problems and discuss briefly some important regularization methods.

As the concept Nemytskii operator is of core importance for the formulation of inverse problems which arise in the context of option pricing we introduce this concept in Section 5.2. Furthermore, we review results about Nemytskii operators.

With respect to an analysis of inverse problems we are not only interested in the properties of the Nemytskii operators itself but especially in the corresponding inverse operators. Unfortunately, in this context there are still many questions open. We give several examples which show that the inverse of a Nemytskii operator needs not to exist and even if it exists it needs not to be continuous.

Restricting our considerations to Nemytskii operators generated by monotonic functions mapping either between $C$-spaces of continuous functions or $L^p$-spaces of Lebesgue measurable functions we address some of these questions in Section 5.3.

5.1 Inverse problems and regularization methods

Many physical-technical and economical processes can be described by a transformation of a certain cause to a certain effect. In mathematical models of such processes the cause $x$ and the effect $y$ are described by functions in appropriate Banach spaces $X$ and $Y$. The transformation of $x$ to $y$ is then described by a linear or nonlinear forward operator

$$F : D(F) \subset X \rightarrow Y.$$  

In this context, the domain of $F$ contains all admissible causes $x$.

Let us illustrate this at the example of the heat equation. In this context the cause can be for example the heat conduction parameter function. The result is the final temperature profile.
The direct problem consists in computing for given cause $x$ and operator $F$ (resulting from the physical laws) the effect $y$.

Corresponding inverse problems are the recovery of the initial temperature either from measurements at a later timepoint or from a desired final temperature profile. Another inverse problem is the calibration of certain model parameters e.g. the heat conduction parameter function. In this sense, inverse problems are concerned with determining a cause for a desired or an observed effect. Using the notation introduced above an inverse problem can be formulated by an operator equation

$$F(x) = y \quad (x \in D(F) \subset X, y \in Y).$$

(5.1)

In the theory of inverse problems the terms well-posedness and ill-posedness in the sense of Hadamard play a crucial role.

**Definition 5.1.1 (well-posedness and ill-posedness)**

We call the operator equation (5.1) well-posed if the following three conditions of Hadamard are satisfied.

(i) For every $y \in Y$ there exists an element $x \in D(F)$, which satisfies (5.1) (existence).

(ii) This solution is unique (uniqueness).

(iii) The solution depends continuously on the data, i.e. every sequence $\{x_n\} \subset D(F)$ satisfying $F(x_n) \xrightarrow{Y} F(x)$ for $n \to \infty$ converges to $x$ in $X$ (stability).

If one or more of these conditions are violated, the equation (5.1) is called ill-posed.

These three conditions can also be formulated by properties of $F$ and $F^{-1}$. The first and second conditions of Hadamard express the surjectivity and the injectivity of $F$, respectively. If they are satisfied, the inverse operator $F^{-1} : Y \to X$ exists. In this situation the third condition is equivalent to the continuity of $F^{-1}$.

The importance of the existence and uniqueness conditions (i) -(ii) is immediately obvious. In order to understand the importance of condition (iii) one has to be aware of the fact that instead of the exact $y$ we can in general only observe noisy data $y^\delta$. Assuming now, as an idealisation that measuring instruments with arbitrary precision are available we get a series of measurements $y^{\delta_n}$ converging to $y$. The stability property guarantees now that the corresponding solutions $x^{\delta_n}$ of $F(x^{\delta_n}) = y^{\delta_n}$ converge to the solution $x$ of $F(x) = y$. In other words, we are theoretically able to identify the function $x$ with arbitrary precision by increasing the precision in the measurement of $y$.

Unfortunately, many inverse problems are ill-posed, i.e. at least one of the conditions (i)-(iii) is violated. In this situation a violation of conditions (i) or (ii) can be overcome by considering least-squares solutions or $x^\ast$-minimum norm-solutions, where $x^\ast$ denotes a reference element. See for example [25, p. 34ff] or [12, p. 32ff and p. 241ff].

If the stability condition is violated, one has to apply a regularization method, which means roughly speaking to add additional information into the solution process. This additional
information can characterise the error of the data (e.g. an error bound or the distribution of a stochastic error) or they can contain information about the searched solution itself. In the latter case these information are also called apriori information.

Apriori information can be divided into two classes. The first class contains objective apriori information, i.e. information about the searched solution that comes from the physical-technical or economical background of the problem and is therefore definitely trustworthy. Examples can be properties such as monotonicity, convexity or nonnegativity of the searched function. The second class contains subjective apriori information, i.e. expectations about the searched solution, e.g. smoothness. As these information are no consequences of (physical) laws but incorporate for example the experience of engineers they should be carefully used.

If there is enough apriori information available to restrict the set of admissible solutions to a compact subset $M$ of $X$ and the forward operator restricted to $M$ is furthermore continuous and injective one can apply the method of quasisolutions. The idea of this regularization method goes back to A. N. Tikhonov and its mathematical background is formulated by the following proposition cf. [5, Lemma 2.2].

**Proposition 5.1.2**

Let $F : M \subset X \to Y$ be a continuous and injective mapping of a compactum $M$ onto $F(M)$. Then the inverse operator $F^{-1} : F(M) \subset Y \to M \subset X$ is continuous.

The method of quasisolutions chooses an approximate solution $x_M^\delta \in M$ of (5.1) which minimizes the discrepancy, i.e. which satisfies

$$\|F(x_M^\delta) - y^\delta\|_Y = \min_{x \in M} \|F(x) - u^\delta\|_Y.$$  \hspace{1cm} (5.2)

It should be noted that the continuity of the operator $F$ and the compactness of the set $F$ guarantee that the minimum in (5.2) is attained. Therefore, the set of quasisolutions is nonempty.

If we assume that the operator equation (5.1) has a unique solution $x^\dagger$ and consider besides the exact data $y^\dagger := F(x^\dagger)$ noisy data $u^\delta$ with noise level $\delta$, i.e. $\|y^\delta - y^\dagger\|_Y \leq \delta$ we can for every quasisolution $x_M^\delta$ bound the defect by $2\delta$ (cf.[5, p.29]). Indeed, we have

$$\|F(x_M^\delta) - y^\dagger\|_Y \leq \|y^\dagger - y^\delta\|_Y + \|F(x_M^\delta) - y^\delta\|_Y \leq 2\delta.$$  

The continuity of the inverse operator $F^{-1}$ on $F(M)$ implies now the convergence $x_M^\delta \to x^\dagger$ for $\delta \to 0$.

In order to describe the speed of convergence we introduce for $\overline{y} \in F(M)$ the local modulus of continuity $\omega_{F^{-1}} : [0, \infty) \to [0, \infty)$ defined by

$$\omega_{F^{-1}}(\delta, \overline{y}) := \sup_{\|y^\delta - \overline{y}\| \leq \delta, y^\delta \in F(M)} \|F^{-1}(y^\delta) - F^{-1}(\overline{y})\| \text{ for } \delta \geq 0.$$  

As the operator $F^{-1} : F(M) \subset Y \to M \subset X$ is continuous at every point $\overline{y} \in F(M)$ and $M$ is a bounded set the local modulus of continuity $\omega_{F^{-1}}(\cdot, \overline{y})$ is a bounded, monotonically
increasing function and satisfies \( \lim_{\delta \to 0} \omega_{F^{-1}}(\delta, \overline{y}) = 0 \). Clearly, given \( \omega_{F^{-1}}(\cdot, \overline{y}) \) we have the error estimate
\[
\| x_{M}^{\delta} - x^{\dagger} \|_{X} \leq \omega_{F^{-1}}(2\delta, y).
\]
(5.3)

It should be noted that the local modulus of continuity \( \omega_{F^{-1}}(\cdot, \overline{y}) \) does not only depend on the operator \( F \) and the point \( \overline{y} \) but also on the set \( M \). Without further information about \( F, \overline{y} \) and \( M \) no assertion about the speed of the convergence (5.3) can be made. In other words, the convergence can be arbitrarily slow.

An interesting question is furthermore, which kind of apriori information is needed for the method of quasisolutions. In order to answer this question, let us consider some examples of compacta in the space \( C[0, 1] \) of continuous functions over the interval \([0, 1]\) and \( L^p(0, 1)\)-spaces of \( p\)-power integrable functions over \((0, 1)\).

**Example 5.1.3** The set
\[
M_C := \{ f \in L^p(0, 1) : 0 \leq f(t) \leq C < \infty \text{ and } f(t_1) \leq f(t_2) \forall t_1 \leq t_2 \}
\]
of all nonnegative, bounded nondecreasing functions is compact in \( L^p(0, 1) \) for all \( p \geq 1 \) (cf. for example [51] and [25, p. 72ff]).

In some applications objective apriori information about the monotonicity of the searched solution can be deduced from physical or economical laws. In order to apply the method of quasisolutions in these situations it is necessary to bound the exact solution pointwise by a constant \( C \). In general this can only be done by using subjective apriori information which result from the experience of engineers but are not really trustworthy.

**Example 5.1.4** An important implication of the Arzelà-Ascoli theorem (cf. [2, p. 49ff]) is that a subset \( M \) of \( C[0, 1] \) is compact if and only if it is closed, uniformly bounded and uniformly equicontinuous, i.e. if there exists a constant \( 0 < c < \infty \) and a function \( \delta : (0, \infty) \to (0, \infty) \) such that
\[
M = \{ f \in C[0, 1] : |f(t)| \leq c \ \forall t \in [0, 1] \text{ and } |f(t_1) - f(t_2)| < \varepsilon \text{ whenever } |t_1 - t_2| < \delta(\varepsilon) \}.
\]

In order to interpret these conditions let us consider the specific example where \( x^{\dagger}(t) \) denotes the place of a certain particle at time \( t \). Then the function \( x^{\dagger} \) is continuous and the function \( \delta \) can be interpreted as maximal velocity of the particle. The constant \( c \) may result from a bounded domain in which the particle moves. In this specific situation the quasisolution method is applicable provided the bounds of the domain and the maximal velocity are known apriori.

However, there exist inverse problems where the constants \( c \) and the function \( \delta \) do not have such a nice interpretation. In these situations there is in general not enough apriori information available to restrict the set of admissible solutions to a compact set.
These examples have indicated that the method of quasisolutions is not always applicable as it requires certain kinds of objective apriori information about the solution which are not always available. In this situation another regularization method has to be used. One very well-known method is Tikhonov regularization, which shall now be briefly presented.

Tikhonov regularization chooses the approximate solution of (5.1) as minimizer of the functional

\[ T_\alpha(x) := \| F(x) - y^\delta \|_Y^2 + \alpha \Omega(x). \]

(5.4)

Here, \( \Omega : X \rightarrow \mathbb{R} \) denotes a functional that measures how well the function \( x \) matches our apriori information. The principal idea of this method is to select an approximate solution which matches on the one hand side well the data (i.e. a function \( x \) for which the term \( \| F(x) - y^\delta \|_Y^2 \) becomes small) and corresponds on the other side well to the apriori information (i.e. for which the functional \( \Omega \) becomes small). In this context, the regularization parameter \( \alpha \) controls the influence of the two summands in (5.4).

The functional \( \Omega \) describes our apriori information concerning the solution. As opposed to the method of quasisolutions here subjective apriori information, which result from the experience of practitioners but need not to be absolutely trustworthy, can be used. One possibility for the choice of the functional \( \Omega \) is \( \Omega(x) := \| x - x^* \|_2^2 \) where \( x^* \) denotes some initial guess, incorporating our apriori information about the solution. In this situation the method (5.4) yields the classical Tikhonov regularization.

If we expect the solution \( x^\dagger \) of (5.1) to be smooth, this apriori information can be expressed by the functional \( \Omega(x) = \| x' \|_2^2 \) or \( \Omega(x) = \| x'' \|_2^2 \). Another choice for \( \Omega \) is the entropy functional

\[ \Omega(x) := \int_a^b x(t) \log \frac{x(t)}{x^*(t)} + x^*(t) - x(t) \, dt \]

which measures a certain distance between \( x \) and \( x^* \). If \( x \) and \( x^* \) are probability density functions this distance is the well-known Kullback-Leibler distance.

In the rest of this section we will review some results from [12, p. 241ff] concerning convergence and convergence rates for classical Tikhonov regularization. Let \( F : \mathcal{D}(F) \subset X \rightarrow Y \) be a (nonlinear) operator between Hilbert spaces \( X \) and \( Y \). Let furthermore the following conditions be satisfied.

1. \( F \) is continuous.
2. \( F \) is weakly (sequentially) closed, i.e. for every sequence \( \{ x_n \}_{n \in \mathbb{N}} \) with elements \( x_n \in \mathcal{D}(F) \) weak convergence \( x_n \rightharpoonup x \) in \( X \) and weak convergence \( F(x_n) \rightharpoonup y \) in \( Y \) imply that \( x \in \mathcal{D}(F) \) and \( F(x) = y \).

Both assumptions are satisfied if \( F \) is continuous and compact and if \( \mathcal{D}(F) \) is weakly closed. For simplicity we confine our considerations here to the case where the considered inverse problem (5.1) has a unique solution which is denoted by \( x^\dagger \). For the situations where the operator equation (5.1) has no solution or the solution is not unique we refer to [12, p. 241ff] and [13].
For given noisy data $y^\delta$ and a fixed regularization parameter $\alpha$ we choose a regularized solution as minimum of the optimisation problem
\[
\|F(x) - y^\delta\|^2 + \alpha\|x - x^*\|^2 \rightarrow \min, \quad x \in \mathcal{D}(F).
\] (5.5)

It should be noted that under the formulated assumptions on $F$ the optimisation problem (5.5) admits a solution. The following proposition states that the regularized solutions depend continuously on the data.

**Proposition 5.1.5**

Let $\alpha > 0$ and let $y_k$ and $x_k$ be sequences where $y_k \rightarrow y^\delta$ and $x_k$ is a minimizer of (5.5) with $y^\delta$ replaced by $y_k$. Then there exists a convergent subsequence of $x_k$ and the limit of every convergent subsequence is a minimizer of (5.5).

The next two propositions are concerned with convergence and convergence rates for the regularized solutions to the true solution.

**Proposition 5.1.6**

Let $y^\delta \in Y$ with $\|y - y^\delta\| \leq \delta$ and let $\alpha(\delta)$ be such that
\[
\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \quad \text{as} \quad \delta \rightarrow 0.
\]

Then
\[
\lim_{\delta \rightarrow 0} x^\delta_{\alpha(\delta)} = x^\dagger.
\]

An interesting implication of Proposition 5.1.6 is that the method (5.5) yields convergence of the approximate solutions $x^\delta_{\alpha(\delta)}$ even if $x^*$ is far away from the exact solution, provided the regularization parameter is chosen appropriately. However, in this situation one can not expect a fast speed of convergence.

In order to guarantee a convergence rate one has to assume certain source conditions. A well known condition that guarantees the convergence rate $O\left(\sqrt{\delta}\right)$ is given by the following proposition.

**Proposition 5.1.7**

Let $\mathcal{D}(F)$ be convex, let $y^\delta \in Y$ with $\|y - y^\delta\| \leq \delta$. Moreover, let the following conditions hold:

1. $F$ is Fréchet-differentiable.
2. There exists $\gamma \geq 0$ such that $\|F'(x^\dagger) - F'(x)\| \leq \gamma\|x^\dagger - x\|$ for all $x \in \mathcal{D}(F)$.
3. There exists an element $\omega \in Y$ satisfying $x^\dagger - x^* = F'(x^\dagger)^* \omega$ and
4. $\gamma\|\omega\| < 1$. 

Then for the choice $\alpha \simeq \delta$ we obtain
\[
\|x^\delta_\alpha - x^\dagger\| = \mathcal{O}\left(\sqrt{\delta}\right) \quad \text{and} \quad \|F(x^\delta_\alpha) - y^\delta\| = \mathcal{O}(\delta).
\]

It should be noted that the source condition $x^\dagger - x^* = F'(x^\dagger)^* \omega$ formulated in Proposition 5.1.7 is very strong. In Section 6.1 we will see at the example of an inverse problem arising in option pricing that it characterises the closeness of the initial guess $x^*$ and the exact solution $x^\dagger$.

### 5.2 Nemytskii operators: Acting conditions and continuity

In this section we consider a special class of operators, Nemytskii operators. First of all we will give a precise definition, where we restrict to Nemytskii operators acting either between spaces of continuous or spaces of $p$-power integrable functions over the interval $[0,1]$. In these spaces we review some results concerning acting conditions and continuity of the forward operator. Finally, we give some examples showing that in order to ensure the existence and continuity of the corresponding inverse operator further assumptions are necessary. Results about the continuity of the inverse operator under appropriate conditions will be derived in Section 5.3.

Let $f : [0,1] \times \mathbb{R} \to \mathbb{R}$ be a given function. Now, for every function $x : [0,1] \to \mathbb{R}$ we get another function $y : [0,1] \to \mathbb{R}$ defined by
\[
y(t) = f(t, x(t)) \quad t \in [0,1]. \tag{5.6}
\]

Loosely speaking, in this way the function $f$ generates an operator $F$
\[
[F(x)](t) := f(t, x(t)), \tag{5.7}
\]

which is usually called Nemytskii operator or superposition operator.

In order to make this definition precise we have to define (Banach) spaces $B_1$ and $B_2$ between which the operator $F$ acts. Here, we will confine our consideration to two situations

(a) $B_1 := C[0,1], \ B_2 := C[0,1]$

(b) $B_1 := L^p(0,1), \ B_2 := L^q(0,1)$ where $1 \leq p, q < \infty$.

After defining the spaces $B_1$ and $B_2$ we have to formulate conditions on the function $f$ guaranteeing that for every element $x$ in $B_1$ the function $F(x)$, defined by (5.7), is indeed an element of $B_2$ and the operator $F : B_1 \to B_2$ is continuous.

In the situation (a) it is clear that for every continuous function $f$ the Nemytskii operator $F$ defined by (5.7) maps the space $C[0,1]$ into itself. Moreover, the following proposition (cf. [2, p. 205]) states that this acting condition on $f$ is not only sufficient but also necessary.
Proposition 5.2.1
The operator (5.7) maps the space $C[0, 1]$ into itself if and only if the function $f$ is continuous. In this situation $F : C[0, 1] \to C[0, 1]$ is continuous.

Now we are going to address the situation (b). We start by a definition.

Definition 5.2.2 (Carathéodory function)
We say that $f : [0, 1] \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function if

- $s \mapsto f(t, s)$ is continuous for almost every $t \in [0, 1]$,
- $t \mapsto f(t, s)$ is measurable for all $s \in \mathbb{R}$.

The following proposition gives a condition guaranteeing that $F : L^p(0, 1) \to L^q(0, 1)$ is well-defined and continuous (cf. [1, p. 16]). We remark that as opposed to Proposition 5.2.1 this condition on $f$ is sufficient but not necessary. For a more general condition and a discussion of acting conditions for other spaces than $C[0, 1]$ and $L^p(0, 1)$ we refer to [3].

Proposition 5.2.3
Let $f : [0, 1] \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function that satisfies

$$|f(t, s)| \leq a + b|s|^q, \quad 1 \leq p, q < \infty$$

for some constants $a, b > 0$. Then the Nemytskii operator $F : L^p(0, 1) \to L^q(0, 1)$ defined by (5.7) is well-defined and continuous.

Speaking generally, the literature about Nemytskii operators contains many results about properties of the forward Nemytskii operators, such as acting conditions, boundedness, special continuity properties (such as Lipschitz, uniform or weak continuity), continuous differentiability and related properties (cf. e.g. [1, Section 1.2], [3] and also [53, Section 25.3]).

However, as we have seen in Chapter 5.1 for solving operator equations we are not only interested in the properties of the forward operator. The question whether the inverse operator exists and is continuous is at least of equal importance. Unfortunately, the literature about Nemytskii operators is not concerned with these topics. Therefore, the rest of this chapter is intended to be a contribution to this interesting subject.

The first questions that arise when thinking about the inverse of a Nemytskii operator are: Do the conditions formulated in Proposition 5.2.1 and 5.2.3 guarantee that the inverse operator $F^{-1}$ exists? Furthermore, provided the inverse operator

$$F^{-1} : \mathcal{R}(F) \subset B_2 \to B_1,$$

defined on the range $\mathcal{R}(F)$ of the operator $F$, exists is $F^{-1}$ necessarily continuous?

The following three examples show that the answer to both of these questions is negative. We remark that we are going to come back to the Example 5.2.5 later, so we present it here quite detailed.
Example 5.2.4 Consider the function \( f(t, s) = t \). Then the operator \( F \) defined by (5.7) is in both settings (a) and (b) well-defined and continuous. However, in both settings it fails to be injective. Thus, there exists no inverse operator.

Example 5.2.5 We consider the function \( f(t, s) = ts \) and restrict our considerations to the setting (a). As \( f \) is continuous, the generated Nemytskii operator \( F : C[0,1] \to C[0,1] \) is well-defined and continuous. Furthermore, we prove that \( F \) is injective. Hence, the inverse operator (5.9) exists, but as we are going to see it is not continuous.

1. We start by proving the injectivity of \( F \). Let \( x_1 \neq x_2 \), i.e. there exists \( \hat{t} \in [0,1] \) such that \( x_1(\hat{t}) \neq x_2(\hat{t}) \). We set \( \varepsilon := |x_1(\hat{t}) - x_2(\hat{t})| > 0 \). As the functions \( x_1 \) and \( x_2 \) are continuous there exists a positive number \( \delta > 0 \) such that

\[
|x_1(t) - x_2(t)| \geq \frac{\varepsilon}{2} \quad \forall t \in U_\delta(\hat{t}) \cap [0,1].
\]

Hence, there exists \( \hat{t} \neq 0 \) such that \( |x_1(\hat{t}) - x_2(\hat{t})| \geq \frac{\varepsilon}{2} > 0 \). As we have for all \( t \neq 0 \)

\[
f(t, s_1) < f(t, s_2) \quad \forall s_1 < s_2 \tag{5.10}
\]

this gives

\[
[F(x_1)](\hat{t}) = f(\hat{t}, x_1(\hat{t})) \neq f(\hat{t}, x_2(\hat{t})) = [F(x_2)](\hat{t}).
\]

This proves \( F(x_1) \neq F(x_2) \) for all \( x_1 \neq x_2 \), i.e. the injectivity of \( F \).

2. As \( F \) is injective, the inverse \( F^{-1} : \mathcal{R}(F) \subset C[0,1] \to C[0,1] \) exists. It remains to show that \( F^{-1} \) is not continuous. To this end, we consider the functions \( S_n \) and \( u_n \) defined by

\[
S_n(t) := \begin{cases} 0 & t \geq \frac{1}{n} \\ 1 - nt & 0 \leq t < \frac{1}{n} \end{cases} \quad n = 1, 2, \ldots
\]

and

\[
u_n := F(S_n) = \begin{cases} 0 & t \geq \frac{1}{n} \\ t(1 - nt) & 0 \leq t < \frac{1}{n} \end{cases} \quad n = 1, 2, \ldots
\]

as well as \( S_0 \equiv 0 \) and \( u_0 = F(S_0) \equiv 0 \). Then we have

\[
\|S_n - S_0\|_{C[0,1]} = 1 \quad \text{and} \quad \|u_n - u_0\|_{C[0,1]} = \frac{1}{4n}.
\]

Thus, it holds

\[
u_n \to u_0 \quad \text{but} \quad F^{-1}(u_n) \not\to F^{-1}(u_0),
\]

which proves that \( F^{-1} \) is not continuous.

Example 5.2.6 Again we consider the function \( f(t, s) = ts \) but now in the setting (b) with \( p = q = 2 \). Obviously, Proposition 5.2.3 implies that the generated Nemytskii operator \( F : L^2(0,1) \to L^2(0,1) \) is in this situation continuous. Besides, we will prove that \( F \) is injective, which implies that the inverse operator (5.9) exists, but as we will see it is not continuous.
1. We start by proving the injectivity of \( F \). Let \( x_1 \neq x_2 \). In other words, the set
\[
I_1 := \{ t \in [0, 1] : x_1(t) \neq x_2(t) \}
\]
has positive Lebesgue measure \( \mu(I_1) > 0 \). Furthermore, (5.10) implies
\[
[F(x_1)](t) = f(t, x_1(t)) \neq f(t, x_2(t)) = [F(x_2)](t) \quad \forall t \in I_1 \setminus \{0\}.
\]
Clearly, it holds \( \mu(I_1 \setminus \{0\}) = \mu(I_1) > 0 \). Thus, \( F(x_1) \neq F(x_2) \).

2. As in Example 5.2.5 we construct elements \( S_n \) such that
\[
S_n \to S_0 \quad \text{but} \quad F(S_n) \to F(S_0) \quad \text{in} \quad L^2(0, 1).
\]
These functions \( S_n \) can be chosen as
\[
S_n(t) := \begin{cases} 
0 & t \geq \frac{1}{n} \\
\sqrt{n - n^2 t} & 0 \leq t < \frac{1}{n}
\end{cases} \quad \text{with } n = 1, 2, \ldots
\]
and \( S_0 \equiv 0 \). Then it holds
\[
\|S_n - S_0\|_{L^2(0,1)} = \frac{1}{\sqrt{2}} \quad \text{and} \quad \|F(S_n) - F(S_0)\|_{L^2(0,1)} = \frac{1}{\sqrt{12n}}.
\]
We remark that the Nemytskii operators \( F \) which we have considered in Examples 5.2.5 and 5.2.6 are multiplication operators, i.e. it holds
\[
[F(x)](t) = m(t)x(t)
\]
with some multiplier function \( m \), here \( m(t) = t \). A composition of these multiplication operators with integral operators arise in many applications cf. \([26, 15]\).

5.3 Nemytskii operators with monotone generator functions

In this section we restrict our consideration to Nemytskii operators generated by a function \( f \) which is strictly monotonically increasing with respect to the second variable, i.e. for which it holds
\[
f(t, s_1) < f(t, s_2) \quad -\infty < s_1 < s_2 < \infty.
\]
For these specific operators we will formulate and prove results about the the structure and the continuity of the corresponding inverse operators.

In order to valuate the relevance of the continuity results it is interesting to remember Proposition 5.1.2, which states that every continuous and injective operator defined on a compact set has a continuous inverse. As we have seen in Section 5.1 this result is the basis of the quasisolution method, which uses apriori information to restrict the set of admissible
solutions to a compact set. It is therefore of core importance for the solution of inverse problems.

The results which we present here, are adapted to the specific class of operators which we consider in this section. They state that in the situations considered here a restriction to a compact set is not necessary. To be specific, in the $C$-space setting the inverse operator is continuous, which means that no regularization method is needed for the stable solution of the inverse problem. Therefore we do not have to restrict the set of admissible solutions.

In the $L^p$-space setting the situation is different. We will see that there exist Nemytskii operators $F : L^2(0,1) \rightarrow L^2(0,1)$ which are generated by a function $f$ that satisfies $(5.12)$ and for which the inverse $F^{-1} : \mathcal{R}(F) \subset L^2(0,1) \rightarrow L^2(0,1)$ is not continuous. In this situation the corresponding inverse problem $(5.1)$ is ill-posed and therefore a regularization method is needed. However, it is not necessary to restrict the set of admissible solutions to a compact subset. A weaker restriction suffices to ensure the continuity of $F^{-1} : F(M) \subset L^2(0,1) \rightarrow L^2(0,1)$. In other words, less apriori information concerning the solution is necessary to ensure the stability.

As a motivation for considering the specific Nemytskii operators we remark that they occur for example in inverse problems of option pricing. Especially, we will apply the obtained results in Chapter 6 when we study the identification of a time-depending volatility (or a deviated quantity) from option price data.

We start by defining the two settings under consideration.

**Assumption 5.3.1** Let $f : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$ be such that $(5.12)$ holds for all $t \in [0, 1]$. We consider one of the following two situations

(A) The function $f$ is assumed to be continuous. We consider the Nemytskii operator $F : D \subset C[0, 1] \rightarrow C[0, 1]$.

(B) The function $f$ is a Carathéodory function and satisfies $(5.8)$. We consider the Nemytskii operator $F : D \subset L^p(0,1) \rightarrow L^q(0,1)$.

In both situations $F$ is defined by $(5.7)$ and $D$ denotes the domain of $F$.

**Remark 5.3.2** The results obtained in the setting (A) can easily be extended to a generating function $f$ which is not defined on the whole set $[0, 1] \times \mathbb{R}$ but only on the set

$$\{(t, s) : b_1(t) \leq s \leq b_1(t)\}$$

where $b_{1/2} : [0, 1] \rightarrow \mathbb{R}$ denote continuous functions. In this situation the domain of the generated Nemytskii operator is

$$D := \{x \in C[0, 1] : b_1(t) \leq x(t) \leq b_2(t) \forall t \in [0, 1]\}.$$ 

Indeed, in this situation one can extend $f$ to $\tilde{f} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\tilde{f}(t, s) := \begin{cases} f(t, s) & b_1(t) \leq s \leq b_2(t) \\ f(t, b_2(t)) + s - b_2(t) & s > b_2(t) \\ f(t, b_1(t)) + s - b_1(t) & s < b_1(t) \end{cases}$$
As it is easy to prove that \( \tilde{f} \) is continuous the generated Nemytskii operator \( \tilde{F} : C[0,1] \to C[0,1] \) is continuous. Furthermore, the Nemytskii operator \( \tilde{F} \) generated by \( \tilde{f} \) restricted to the domain \( D \) is equal to \( F \).

The following lemma states that Assumption 5.3.1 guarantees the injectivity of \( F \) and therefore the existence of the corresponding inverse operator \( N^{-1} \). Besides \( N^{-1} \) is again of Nemytskii type.

**Lemma 5.3.3**

The operators \( F \) defined by situation (A) and (B) in Assumption 5.3.1 respectively are well-defined, continuous and injective. Thus, the inverse operator

\[
F^{-1} : F(D) \subset B_2 \to B_1 \quad (5.13)
\]

exists. Furthermore, there exists some subset

\[
G \subset [0,1] \times \mathbb{R} \quad \text{and a function} \quad g : G \subset [0,1] \times \mathbb{R} \to \mathbb{R}
\]

such that

\[
[F^{-1}(y)](t) = g(t,y(t)) \quad \text{for all} \quad y \in \mathcal{R}(F),
\]

i.e. \( F^{-1} \) is again a Nemytskii operator. Moreover, \( g \) is strictly monotonically increasing with respect to the second variable.

**Proof:** The continuity of \( F \) in the considered situations follows from the Propositions 5.2.1 and 5.2.3. The injectivity of \( F \) can be shown by similar considerations as in the Examples 5.2.5 and 5.2.6.

It remains to show that \( F^{-1} \) is again of Nemytskii type. To this end we consider the functions

\[
\tilde{f}_t : \mathbb{R} \to \mathbb{R} \quad \text{defined by} \quad \tilde{f}_t(s) = f(t,s).
\]

The range of \( \tilde{f}_t \) is denoted by \( \tilde{G}_t \), i.e.

\[
\tilde{G}_t := \{ f_t(s) : s \in \mathbb{R} \}.
\]

Furthermore, we introduce the set

\[
G := \{(t,y) : t \in [0,1], y \in \tilde{G}_t \}.
\]

For fixed \( t \in [0,1] \) the strictly monotonically increasing function \( \tilde{f}_t \) has a strictly monotonically increasing inverse \( \tilde{g}_t : \tilde{G}_t \to \mathbb{R} \). Defining now \( g : G \subset [0,1] \times \mathbb{R} \to \mathbb{R} \) by

\[
g(t,u) = \tilde{g}_t(u)
\]

we have for \( y = F(x) \)

\[
g(t,y(t)) = \tilde{g}_t(y(t)) = \tilde{g}_t(f(t,x(t))) = \tilde{g}_t(\tilde{f}_t(x(t))) = x(t).
\]

In other words, the Nemytskii operator \( F^{-1} \) is given by (5.14). Besides, the strict monotonicity of \( \tilde{g}_t \) implies the strict monotonicity of \( g \) with respect to the second variable.
Remark 5.3.4 It seems noteworthy that in the setting (a) without a restriction like (5.12) the inverse of a Nemytskii operator is not necessarily of Nemytskii type. Indeed, we will see that the inverse of the multiplication operator (5.11) introduced in Example 5.2.5 can not be written in the form (5.14). This is due to the root of the multiplier function \( m(t) = t \) at the point zero.

To be precise, let us consider the functions \( x_1 \equiv 1 \) and \( x_2 \equiv 2 \) as well as \( y_1 := F(x_1) \) and \( y_2 := F(x_2) \). Clearly, \( y_1(t) = t \) and \( y_2(t) = 2t \). Now, if \( F^{-1} \) were of Nemytskii type, i.e. (5.14) would hold, we would get
\[
1 = [F^{-1}(y_1)](0) = g(t, 0) = [F^{-1}(y_2)](0) = 2,
\]
which is clearly a contradiction.

Remark 5.3.5 The proof of Lemma 5.3.3 shows that the set \( G \) is non-degenerate in the following sense. The intersection of the set \( G \) with a parallel \( P_t := \{(t, s) : s \in \mathbb{R}\} \ (t \in [0, 1]) \) is the set \( \tilde{G}_t \) which was defined in (5.15). Now, in the setting (A) this set is always a non-empty, open interval. In the setting (B) this is true for almost every \( t \in [0, 1] \).

Furthermore, it is possible to show that for a continuous function \( f : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R} \) satisfying (5.12) the function \( g : G \subset [0, 1] \times \mathbb{R} \rightarrow \mathbb{R} \), which has been constructed in the proof of Lemma 5.3.3 is again continuous. The ideas of this proof are essentially the same as in the proof of the following Theorem 5.3.6.

We will now discuss the continuity of the inverse operator \( F^{-1} \) in the two situations (A) and (B), respectively. We start by considering the situation (A). In this situation we prove that the inverse operator is again continuous.

Theorem 5.3.6
Let the conditions of setting (A) in Assumption 5.3.1 be satisfied. In this situation the inverse operator \( F^{-1} : F(D) \subset C[0, 1] \rightarrow C[0, 1] \) is continuous.

Proof: As we have assumed that \( f \) is defined on the whole set \([0, 1] \times \mathbb{R}\) we can consider the Nemytskii operator \( \tilde{F} : C[0, 1] \rightarrow C[0, 1] \) generated by \( f \) which is defined on the whole space \( C[0, 1] \). Then \( F \) is the restriction of \( \tilde{F} \) to the set \( D \). Clearly, the continuity of \( F^{-1} \) implies the continuity of \( F^{-1} \).

Let \( y_0 = F(x_0) \) be fixed and \( \varepsilon > 0 \) arbitrary. We will construct \( \delta = \delta(\varepsilon) > 0 \) such that
\[
\|F^{-1}(y_\delta) - F^{-1}(y_0)\|_{C[0,1]} \leq \varepsilon \quad \text{for all } y_\delta \in \mathcal{R}(F) \text{ satisfying } \|y_\delta - y_0\|_{C[0,1]} \leq \delta. \tag{5.16}
\]

To this end, we consider \( \underline{x} \) and \( \overline{x} \) defined by
\[
\underline{x}(t) := x_0(t) - \varepsilon \quad \text{and} \quad \overline{x}(t) := x_0(t) + \varepsilon
\]
as well as \( \underline{y} := \tilde{F}(\underline{x}) \) and \( \overline{y} := \tilde{F}(\overline{x}) \). The relation (5.12) together with \( y_0(t) = f(t, x_0(t)) \) implies
\[
y(t) < y_0(t) < \overline{y}(t) \quad \forall t \in [0, 1].
\]
5.3. NÉMYSKII OPERATORS WITH MONOTONE GENERATOR FUNCTIONS

Furthermore, as \( y, y_0 \) and \( \overline{y} \) are continuous functions over the compact set \([0,1]\) the minima
\[
h_1 := \min_{t \in [0,1]} [y_0(t) - y(t)] \quad \text{and} \quad h_2 := \min_{t \in [0,1]} [\overline{y}(t) - y_0(t)]
\]
are attained and positive. Now, every function \( y^\delta \in \mathcal{R}(F) \) satisfying \( \|y^\delta - y_0\| \leq \min(h_1, h_2) \) is bounded by \( y \) and \( \overline{y} \), i.e.
\[
y(t) \leq y^\delta(t) \leq \overline{y}(t).
\]
As we have seen in Lemma 5.3.3 the inverse operator \( \tilde{F}^{-1} \) is again a Nemytskii operator and its generating function \( g \) is strictly monotonically increasing with respect to the second component. This implies that \( x^\delta := \tilde{F}^{-1}(y^\delta) \) satisfies
\[
x(t) \leq x^\delta(t) \leq \overline{x}(t)
\]
and therefore \( \|x^\delta - x_0\|_{C[0,1]} \leq \varepsilon \). In other words, (5.16) has been shown with \( \delta := \min(h_1, h_2) \).

After proving the continuity of the inverse operator and therefore the implication
\[
y^\delta_k \to y_0 \implies F^{-1}(y^\delta_k) \to F^{-1}(y_0) \quad \text{in} \quad C[0,1]
\]
new questions concerning convergence rates arise. In Theorem 5.3.8 we will formulate a sufficient condition on the generating function \( f \) that guarantees Lipschitz continuity of \( F^{-1} \) and therefore a linear convergence rate. Before we do this we present a lemma which gives us a method how we can compute for fixed \( y_0 \in F(D) \) the local modulus of continuity \( \omega_{F^{-1}}(\cdot, y_0) : [0, c) \to [0, \infty) \) of \( F^{-1} \) defined by
\[
\omega_{F^{-1}}(\delta, y_0) := \sup_{\|y^\delta - y_0\| \leq \delta, y^\delta \in \mathcal{R}(F)} \|F^{-1}(y^\delta) - F^{-1}(y_0)\|, \quad \delta \in [0, c).
\]
numerically. Here, \( c \) denotes a certain positive number \( c > 0 \) (possibly depending on \( y_0 \)) such that the supremum in (5.18) are finite.

In order to illustrate the application of this result we remark that the supremum in definition (5.18) is taken over infinitely many functions and can therefore not be evaluated numerically. The lemma states now that for the specific Nemytskii operators considered in this section \( \omega_{F^{-1}}(\delta, y_0) \) can be represented as maximum of two values \( h_1(\delta) \) and \( h_2(\delta) \), where the values \( h_1(\delta) \) and \( h_2(\delta) \) can be computed with arbitrary precision. In Subsection 6.2.2 we will make use of this fact in order to determine the convergence rate \( h \) in a specific inverse problem of option pricing.

**Lemma 5.3.7**

Let the conditions of setting (A) in Assumption 5.3.1 be satisfied. For fixed \( y_0 = F(x_0) \in F(D) \) we can define positive constants \( c_1 \) and \( c_2 \) by
\[
c_1 := \min_{t \in [0,1]} [y_0(t) - f(t, x_0(t) - 1)] \quad \text{and} \quad c_2 := \min_{t \in [0,1]} [f(t, x_0(t) + 1) - y_0(t)].
\]
Then the minimum \( c := \min(c_1, c_2) \) of these two constants is positive. Furthermore, defining the functions \( h_1 : [0, c) \to \mathbb{R} \) and \( h_2 : [0, c) \to \mathbb{R} \) by

\[
h_1(\delta) := \min_{t \in [0, 1]} [x_0(t) - \tilde{g}_t(y_0(t) - \delta)] \quad \text{and} \quad h_2(\delta) := \min_{t \in [0, 1]} [\tilde{g}_t(y_0(t) + \delta) - x_0(t)] \quad (5.19)
\]

with \( \tilde{g}_t \) introduced in the proof of Lemma 5.3.3 the local modulus of continuity \( \omega_{F^{-1}}(\cdot, y_0) \) defined by (5.18) satisfies

\[
\omega_{F^{-1}}(\delta, y_0) = \max(h_1(\delta), h_2(\delta)). \quad (5.20)
\]

Besides, \( \omega_{F^{-1}}(\cdot, y_0) \) is strictly monotonically increasing and satisfies \( \lim_{\delta \to 0} \omega_{F^{-1}}(\delta, y_0) = 0 \).

**Proof:** Let \( y_0 = F(x_0) \). As in the proof of Theorem 5.3.6 we consider the operator \( \tilde{F} : C[0, 1] \to C[0, 1] \). Furthermore, we will use the sets \( \tilde{G}_t \) and the functions \( \tilde{g}_t \) as introduced in the proof of Lemma 5.3.3. In a first step we have to prove that the values \( c_1 \) and \( c_2 \) are well-defined and positive. To do this, we define \( z_1 \) and \( z_2 \) by

\[
z_1(t) := f(t, x_0(t) - 1) \quad \text{and} \quad z_2(t) := f(t, x_0(t) + 1).
\]

The monotonicity of \( f \) (cf. (5.12)) implies \( z_1(t) < y_0(t) < z_2(t) \) for all \( t \in [0, 1] \). As every continuous function attains its minimum over a compact set the values

\[
c_1 := \min_{t \in [0, 1]} (y_0(t) - z_1(t)) \quad \text{and} \quad c_2 := \min_{t \in [0, 1]} (z_2(t) - y_0(t))
\]

are well-defined and positive.

It remains to show that the function \( h : [0, c) \to [0, \infty) \) defined by (5.18) satisfies (5.20). Let \( \delta \in [0, c) \) be arbitrary we consider the functions \( \underline{y}_\delta \) and \( \overline{y}_\delta \) defined by

\[
\underline{y}_\delta(t) := y_0(t) - \delta \quad \text{and} \quad \overline{y}_\delta(t) := y_0(t) + \delta.
\]

By construction of the constant \( c \) it holds \( \underline{y}_\delta(t) \in \tilde{G}_t \) and \( \overline{y}_\delta(t) \in \tilde{G}_t \) for all \( t \in [0, 1] \). We can therefore define functions \( \underline{x}_\delta : [0, 1] \to \mathbb{R} \) and \( \overline{x}_\delta : [0, 1] \to \mathbb{R} \) by

\[
\underline{x}_\delta(t) = \tilde{g}_t\left(\underline{y}_\delta(t)\right) \quad \text{and} \quad \overline{x}_\delta(t) = \tilde{g}_t\left(\overline{y}_\delta(t)\right).
\]

Besides, the monotonicity of \( \tilde{g}_t \) implies \( \underline{x}_\delta(t) < x_0(t) < \overline{x}_\delta(t) \).

Moreover, as the function \( H : [0, 1] \times \mathbb{R} \to \mathbb{R} \) defined by \( H(t, s) := f(t, s) - \underline{y}_\delta(t) \) is continuous, strictly monotonically increasing with respect to \( s \) and we have \( H(t, \underline{x}_\delta(t)) = 0 \) the implicit function theorem (cf. e.g. [14, p. 421]) states that \( \underline{x}_\delta \) is continuous. Analogously the continuity of \( \overline{x}_\delta \) is proven. As every continuous function attains its minimum over a compact set the definition (5.19) is correct and the functions \( h_1 \) and \( h_2 \) are strictly positive.

Furthermore, by construction all functions \( y_\delta \in F(D) \) with error bound \( \delta \), i.e. \( \|y_\delta - y_0\| \leq \delta \), satisfy the inequality

\[
y_\delta(t) \leq F^{-1}(y_\delta)(t) \leq \overline{y}_\delta(t).
\]

Because of the monotonicity of \( \tilde{g}_t \) this implies

\[
\underline{x}_\delta(t) \leq [F^{-1}(y_\delta)](t) \leq \overline{x}_\delta(t).
\]
and therefore $\|F^{-1}(y^\delta) - x_0\| \leq \max(h_1(\delta), h_2(\delta))$. Together with $\|y^\delta - y_0\| = \delta = \|y^\delta - y_0\|$ and the definition of $h_1$ and $h_2$ it holds

$$\|F^{-1}(y^\delta) - x_0\| = h_2(\delta) \quad \text{as well as} \quad \|F^{-1}(\overline{y}^\delta) - x_0\| = h_1(\delta).$$

This proves the equality (5.20).

The strict monotonicity of the function $h$ is a consequence of the strict monotonicity of $\tilde{g}$. Indeed, for $0 < \delta_1 < \delta_2 < c$ we have $\overline{\tilde{x}}^{\delta_2}(t) < \overline{\tilde{x}}^{\delta_1}(t) < \overline{x}^{\delta_1}(t) < \overline{x}^{\delta_2}(t)$, which implies $h_1(\delta_1) < h_1(\delta_2)$ and $h_2(\delta_1) < h_2(\delta_2)$. As the maximum of two strictly monotonically increasing functions is again strictly monotonically increasing this implies the strict monotonicity of $h$.

Finally it remains to remark that $y^\delta \to y_0$ and $\overline{y}^\delta \to y_0$ implies $\overline{x}^\delta \to x_0$ and $\overline{x}^\delta \to x_0$ in $C[0,1]$ (cf. Theorem 5.3.6) and therefore $\lim_{\delta \to 0} h(\delta) \to 0$.

We finish our considerations concerning the setting (A) by a sufficient condition on the generating function $f$ that ensures the Lipschitz continuity of $F^{-1}$.

**Theorem 5.3.8**

*If the function $f$ satisfies the conditions of setting (A) in Assumption 5.3.1 and is furthermore partially differentiable with respect to the second variable and this partial derivative satisfies

$$\frac{\partial f}{\partial s}(t,s) \geq c_L > 0 \quad \forall (t,s) \in [0,1] \times \mathbb{R}$$

then the generated operator $F$ has a Lipschitz continuous inverse with Lipschitz constant $L = \frac{1}{c_L}$, i.e.

$$\|F^{-1}(y_1) - F^{-1}(y_2)\| \leq L\|y_1 - y_2\| \quad \forall y_1, y_2 \in F(D).$$

*Proof:* Let $y_1 = F(x_1), y_2 = F(x_2)$. For arbitrary $t \in [0,1]$ we apply the mean value theorem to obtain

$$y_1(t) - y_2(t) = f(t,x_1(t)) - f(t,x_2(t)) = \frac{\partial f}{\partial s}(t,s(t)) (x_1(t) - x_2(t))$$

with some $s(t)$ between $x_1(t)$ and $x_2(t)$. Taking on both sides the absolute values and using (5.21) we get

$$|y_1(t) - y_2(t)| = \frac{\partial f}{\partial s}(t,s(t)) |x_1(t) - x_2(t)| \geq c_L |x_1(t) - x_2(t)|.$$

Hence,

$$|x_1(t) - x_2(t)| \leq \frac{1}{c_L} |y_1(t) - y_2(t)| \leq \frac{1}{c_L} \|y_1 - y_2\|_{C[0,1]}.$$

As $t \in [0,1]$ was arbitrary we obtain the assertion.

Now we are going to discuss the setting (B). The first question that has to be discussed is whether the conditions of setting (B) in Assumption 5.3.1 can guarantee the continuity of the inverse operator $F^{-1}$ as it was the case in the setting (A). Here the answer is negative, as the following example shows.
5. OPERATOR EQUATIONS WITH NEMYTSKII OPERATORS

Example 5.3.9 We examine the operator $F : L^2(0, 1) \to L^2(0, 1)$ generated by the function $f(t, s) = 1 - e^{-s}$. Let

$$
x_0 \equiv 1 \quad \text{and} \quad x_n(t) := \begin{cases} 1 & t < 1 - \frac{1}{n} \\
1 + \sqrt{t - 1 + \frac{1}{n} n^{3/2}} & t \geq 1 - \frac{1}{n}
\end{cases} \quad n = 1, 2, \ldots,
$$

as well as $y_0 := F(x_0)$ and $y_n := F(x_n)$. Then we have

$$
\|x_n - x_0\|_{L^2(0,1)}^2 = \int_{1-1/n}^1 (t - 1 + \frac{1}{n} t^n) n^2 dt = \frac{n}{2},
$$

and because of $y_0(t) \leq y_n(t) \leq 1$

$$
\|y_n - y_0\|_{L^2(0,1)}^2 = \int_{1-1/n}^1 (y_n(t) - y_0(t))^2 dt \leq \int_{1-1/n}^1 (1 - (1 - e^{-1}))^2 dt = \frac{1}{n e^2},
$$

which proves that $F^{-1}$ is not continuous.

In order to find sufficient conditions for the continuity of the inverse of a Nemytskii operator we remember Lemma 5.3.3 which stated that the inverse of a Nemytskii operator $F$ is again of Nemytskii type whenever the conditions of Assumptions 5.3.1 are satisfied. Moreover, the generating function $g$ of $F^{-1}$ is characterised by

$$
g(t, f(t, s)) = s \quad \forall (t, s) \in [0, 1] \times \mathbb{R}. \tag{5.22}
$$

Furthermore, we remember Proposition 5.2.3 which stated loosely spoken that a Nemytskii operator generated by a Carathéodory function satisfying certain growth conditions (cf. (5.8)) maps continuously from $L^p(0,1)$ into $L^q(0,1)$.

A direct application of these two results seems to be difficult, as the domain of the generating function $g$ of $F^{-1}$ is only defined on a subset $G$ and not on the whole set $[0, 1] \times \mathbb{R}$. Nevertheless, we can follow the ideas of the proof of Proposition 5.2.3 (cf. e.g. [1, Theorem 2.2] or [2, Satz 9.4]). We have to prove that for every sequence $\{y_n\}$ satisfying $y_n \in F(D)$ and every element $y_0 \in F(D)$ the convergence

$$
y_n \xrightarrow{L^p(0,1)} y_0 \quad \text{implies the convergence} \quad F^{-1}(y_n) \xrightarrow{L^p(0,1)} F^{-1}(y_0).
$$

As assumption we need the condition

$$
|g(t, u)| \leq \tilde{a}(t) + \tilde{b}|u|^\frac{2}{p} \quad \text{for all} \quad (t, u) \in G, \tag{5.24}
$$

with some function $\tilde{a} \in L^p(0, 1)$ and a positive constant $\tilde{b} > 0$. Remember that the set $G$ has been defined in Lemma 5.3.3. It is interesting to see that this maximal growth condition on $g$ can be translated into a minimal-growth condition on the function $f$.

To see this, we remember that every $(t, u) \in G$ can be written as $(t, u) = (t, f((t, s))$ with $s \in \mathbb{R}$. Inserting this equivalence in (5.24) and using (5.22) we obtain

$$
|s| = |g(t, f((t, s)))| \leq \tilde{a}(t) + \tilde{b}|f((t, s))|^\frac{2}{p}.
$$

or equivalently

$$
|s| - \tilde{a}(t) \leq \tilde{b}|f((t, s))|^\frac{2}{p} \quad \forall (t, s) \in H. \tag{5.25}
$$
Theorem 5.3.10
Let the function \( f : [0, 1] \times \mathbb{R} \to \mathbb{R} \) satisfy the conditions of setting (B) in Assumption 5.3.1 and (5.25) with some function \( \tilde{a} \in L^p(0, 1) \), a constant \( \tilde{b} > 0 \) and a set \( H \subset [0, 1] \times \mathbb{R} \). Let furthermore
\[
D := \{ x \in L^p(0, 1) : (t, x(t)) \in H \text{ for almost every } t \in [0, 1] \}.
\]
Then the Nemytskii operator \( F : D \subset L^p(0, 1) \to L^q(0, 1) \) generated by \( f \) has a continuous inverse
\[
F^{-1} : F(D) \subset L^q(0, 1) \to L^p(0, 1).
\]

Proof: We prove the result indirectly, i.e. in order to show (5.23) we assume that there exist \( x_n \in D \subset L^p(0, 1) \) \((n = 0, 1, 2, \ldots)\) such that
\[
F(x_n) \xrightarrow{L^q(0,1)} F(x_0) \quad \text{but} \quad x_n \not\to x_0 \text{ in } L^p(0, 1).
\]
Then there exists \( \varepsilon > 0 \) and a subsequence (without loss of generality the whole sequence) such that \( \|x_n - x_0\|_{L^p(0,1)} \geq \varepsilon \) for all \( n \in \mathbb{N} \).

Furthermore, \( \|F(x_n) - F(x_0)\|_{L^q(0,1)} \to 0 \) implies that there exists a sub-sequence \( \{x_{n_k}\} \) and an element \( h \in L^q(0, 1) \) such that
\[
F(x_{n_k})(t) \to F(x_0)(t) \quad \forall t \in [0, 1] \setminus \Omega_1
\]
\[
|F(x_{n_k})(t)| \leq h(t) \quad \forall t \in [0, 1] \setminus \Omega_1
\]
where \( \Omega_1 \subset [0, 1] \) denotes a set with Lebesgue measure zero (cf. [1, Theorem 2.3]).

Now we use the fact that there exists another set \( \Omega_2 \subset [0, 1] \) with Lebesgue measure zero such that the functions \( \tilde{f}_t \) defined by \( \tilde{f}_t(s) := f(t, s) \) are continuous and strictly monotonically increasing for all \( t \in [0, 1] \setminus \Omega_2 \) (cf. Definition 5.2.2 and Assumption 5.3.1). As every strictly monotonically increasing function has a continuous and strictly monotonically increasing inverse the functions \( \tilde{g}_t = \tilde{f}_t^{-1} \) are also continuous for all \( t \in [0, 1] \setminus \Omega_2 \). Furthermore, it holds \( x_{n_k}(t) = \tilde{g}_t(F(x_{n_k})(t)) \). Hence, we obtain
\[
x_{n_k}(t) \to x_0(t) \quad \forall t \in [0, 1] \setminus (\Omega_1 \cup \Omega_2).
\]
Moreover, from the assumption (5.25) we infer
\[
|x_{n_k}(t)| - \tilde{a}(t) \leq \tilde{b}|f(t, x_{n_k}(t))|^{\frac{p}{q}} \quad \forall t \in [0, 1].
\]
and therefore
\[
|x_{n_k}| \leq \tilde{b}|F(x_{n_k})|^{\frac{p}{q}} + \tilde{a} \quad \leq \tilde{b}|h(t)|^{\frac{p}{q}} + \tilde{a} \in L^p(0, 1).
\]
Using the Lebesgue Dominated-Convergence Theorem we conclude from (5.28) and (5.26)
\[
\|x_{n_k} - x_0\|_{L^p(0,1)} = \int_0^1 |x_{n_k}(t) - x_0(t)|^p \, dt \to 0,
\]
which is a contradiction to our assumption.
Corollary 5.3.11

Let the conditions of setting (B) in Assumption 5.3.1 be satisfied and $0 < C < \infty$. Then the Nemytskii operator $\tilde{F} : \tilde{D} \subset L^p(0, 1) \to L^q(0, 1)$ with the (restricted) domain

$$\tilde{D} := \{ x \in L^p(0, 1) : |x(t)| \leq C \text{ a. e. in } [0, 1] \}$$

has a continuous inverse.

Proof: We apply Theorem 5.3.10 with the set $H := \{(t, s) : |s| \leq C\}$, the function $\tilde{a} \equiv C$ and $b = 1$. \hfill $\blacksquare$
Chapter 6

Identification of the time-dependent volatility using option prices

In this chapter we will use option price data to identify the time-dependent volatility function \( \sigma \) or deviated quantities. In the first section we will review results of [22] and [23] concerning the identification of the squared volatility function \( a(t) = \sigma^2(t) \). Originally, these results were intended to price options in the Black-Scholes setting with time-dependent volatility. In the Black-Scholes setting it is well known that the fair option prices can be computed by the Black-Scholes formula. However, as we have seen in Lemma 3.2.3 this formula remains valid for the generalised bivariate Ornstein-Uhlenbeck model and is then functionally independent of the parameters \( \gamma, \lambda, \beta, \sigma_X \), which describe the stochastic drift. Consequently, the results of [22] and [23] can also be applied in the setting of the generalised bivariate Ornstein-Uhlenbeck model.

The second section of this chapter is concerned with the identification of the antiderivative \( S \) of \( a \) defined in (3.12). As motivation it should be pointed out that for pricing European Call and Put Options (or more generally claims for which the final payoff at time \( t_p \) depends only on \( P_{t_p} \)) the knowledge of \( S(t_p) \) is sufficient. We will therefore analyse the inverse problem of calibrating the function \( S \) from option price data and discuss differences between the \( L^2 \)-space setting and the \( C \)-space setting. Applying the results of Section 5.3 we show that the inverse problem is well-posed in the \( C \)-space setting but ill-posed in the \( L^2 \)-space setting. Moreover, even in the \( C \)-space setting there occur ill-conditioning effects for small maturities which lead to strongly oscillating solutions and a delayed convergence. Therefore, we discuss the question whether apriori information about the monotonicity of the data or of the solution can be used for stabilisation. For the \( C \)-space setting we propose a numerically effective algorithm that computes a strictly monotonically increasing approximate solution and illustrate its performance. We conclude the section with a discussion of the discrete setting together with deterministic or stochastic noise, respectively.

6.1 Tikhonov-Regularization

At time \( t_{buy} = 0 \) we denote the asset price by \( P_0 \) and the constant, risk-free interest rate by \( r \). For a fixed strike price \( K > 0 \) we assume as an idealisation the existence of an option
family with maturities $t_{\text{mat}}$ continuously varying in the interval $I := [0, T]$. The associated fair prices of the options are given by

$$u^\dagger(t_{\text{mat}}) = u_{\text{BS}}(P, K, r, t_{\text{mat}}, S^\dagger(t_{\text{mat}})),$$  \hfill (6.1)

where $S^\dagger(t) = \int_0^t a^\dagger(v) \, dv$ and $a^\dagger$ denotes the true squared volatility function.

In this section we will consider the inverse problem of finding an approximation $a^\delta$ of the function $a^\dagger$ from noisy option price data $u^\delta$ in an $L^2(I)$ setting. That is, the accuracy of the observed data $u^\delta$ is measured by $\|u^\delta - u^\dagger\|_{L^2(I)}$ and the accuracy of $a^\delta$ is measured by $\|a^\delta - a^\dagger\|_{L^2(I)}$.

According to the assumption formulated in the model (3.0.26) the essential minimum of the squared volatility function $a(t) = \sigma^2(t)$ has to be positive. Furthermore, in view of (3.12) we want to ensure that the function $a$ is integrable. We will therefore restrict the set of admissible functions $a$ to the set

$$D^+ := \{ a \in L^2(I) : \text{essinf}_{t \in I} a(t) > 0 \}.$$  \hfill (6.2)

In order to formulate this inverse problem in form of an operator equation we define the forward operator $F$ as composition

$$F = N \circ J : D^+ \subset L^2(I) \to L^2(I)$$

of an inner linear Volterra integral operator $J : L^2(I) \to L^2(I)$

$$[J(h)](t) := \int_0^t h(v) \, dv \quad (t \in I)$$  \hfill (6.3)

and an outer nonlinear Nemytskii operator $N : D(N) \subset L^2(I) \to L^2(I)$ defined as

$$[N(S)](t) := U_{\text{BS}}(P, K, r, t, S(t)) \quad (t \in I)$$  \hfill (6.4)

on the domain

$$D(N) := D^+_0 := \{ S \in C[I] : S(0) = 0 \text{ and } S(t) > 0 \; \forall t \in (0, T] \}.$$  \hfill (6.5)

Now the considered inverse problem can be formulated by the nonlinear operator equation

$$F(a) = u \quad (a \in D^+, u \in L^2(I))$$

where $L^2_+(I)$ denotes the set of all nonnegative squareintegrable functions over $I_2$. In order to analyse this operator equation properties of the forward operator such as continuity, compactness, injectivity have been proven in [22].

**Proposition 6.1.1**

The nonlinear operator $F : D^+ \subset L^2(I) \to L^2(I)$ is compact, continuous and injective. Thus, the inverse operator $F^{-1}$ defined on the range $F(D^+)$ of $F$ exists.

Concerning the range of the operator $F$ we find the following result (cf. [22, Lemma 4.2.1]).
Proposition 6.1.2
The set \( F(D^+) \) is a subset of \( C[I] \), i.e. the function \( F(a) \) is continuous for every \( a \in D^+ \). Furthermore, if \( a \) is continuous then the function \( F(a) \) is continuously differentiable.

We are now studying the differentiability of the operator \( F \). First of all we remark that the domain \( D^+ \subset L^2(I) \) has no interior points, so that the notion of Fréchet differentiability is not defined. However, assuming a lower bound \( c \) for the exact solution \( a^\dagger \) and defining the set
\[
D^+_c := \{ a \in D^+ : \mathrm{essinf}_{t \in I} a(t) \geq c \}
\]
it has been shown in [23, Theorem 5.4] that there exists an operator \( G = G(a^\dagger) : L^2(I) \to L^2(I) \) and a constant \( L > 0 \) such that
\[
\| F(a^\dagger + h) - F(a^\dagger) - Gh \|_{L^2(I)} \leq \frac{L}{2} \| h \|_{L^2(I)}^2
\]
holds for all \( h \) such that \( a^\dagger + h \in D^+_c \). The operator \( G \) can be considered as Fréchet derivative \( \tilde{F}'(a^\dagger) \) of an operator \( \tilde{F} \), for which \( F \) is the restriction to the domain \( D^+_c \) with an empty interior in the sense of [12, Remark 10.30].

Proposition 6.1.3
Let \( P_0 \neq K \). Then the linear operator \( G = G(a^\dagger) \) defined by
\[
[G(h)](t) = m(t)[J(h)](t) \quad (t \in I, h \in L^2(I))
\]
with the multiplier function
\[
m(0) := 0 \quad (6.8)
\]
\[
m(\tau) := \frac{\partial u_{BS}(P_0, K, r, \tau, [J(a^\dagger)](\tau))}{\partial s} \quad (\tau > 0) \quad (6.9)
\]
maps continuously in \( L^2(I) \) with \( m \in L^\infty(I) \). Furthermore, condition (6.6) is satisfied with a constant
\[
L = TC_2, \quad \text{where} \quad C_2 := \sup_{(t,s) \in M_c} \left| \frac{\partial^2 U_{BS}(P, K, r, t, s)}{\partial s^2} \right| < \infty
\]
is determined from the set
\[
M_c := \{(t,s) \in (0,T] \times (0,\infty) : s \geq c\}
\]

Next we address the question whether the inverse problem (SIP1) is ill-posed or well-posed (in the sense of Hadamard, cf. Definition 5.1.1). In fact we have already seen that the range of \( F \) is a subset of \( C[0,T] \), i.e. \( F \) is not surjective. In other words, the existence condition in Definition 5.1.1 is not satisfied. Furthermore, the operator \( F \) is injective, which means that the uniqueness condition is fulfilled. Finally, in [22, Satz 4.7.1] we find the following result which shows that the stability condition is not satisfied.

Proposition 6.1.4
The inverse operator \( F^{-1} : F(D^+) \subset L^2(I) \to L^2(I) \) is nowhere continuous, i.e. for every element \( \tilde{u} \in F(D^+) \) there exists a sequence \( \{ u_n \} \subset F(D^+) \) such that \( u_n \to \tilde{u} \) but \( F^{-1}(u_n) \) does not converge to \( F^{-1}(\tilde{u}) \).
As we have learned in Section 5.1 for the stable solution of an ill-posed inverse problem the application of a regularization method is necessary. Therefore, in [22], [23] and in [28] the applicability of Tikhonov-Regularization (cf. (5.4)) with the penalty functionals

\[
\Omega_{TR_0}(a) := \| a - a^* \|^2_{L^2(I)} \quad \text{and} \quad \Omega_{ent} := \int_I a(t) \ln \left( \frac{a(t)}{ea^*(t)} \right) dt
\]

has been studied. Furthermore, in [22, Subsection 4.7.4] numerical case studies with the penalty functional \( \Omega_{TR_2} = \| La \|^2_2 \) and the discretized problem have been carried out. In this context, \( L \) denotes the (discretization of) the second derivative.

Here we will review some analytical results concerning Tikhonov-Regularization with the penalty functional \( \Omega_{TR_0} \), i.e. the approximate solution of (6.5) is chosen as

\[
\| F(a) - u^\delta \|^2_{L^2(I)} + \alpha \| a - a^* \|^2_{L^2(I)} \rightarrow \min, \quad a \in D(F).
\] (6.11)

as approximation of \( a^\dagger \). We remember that \( a^* \in B_1 \) is an initial guess containing a priori information about the volatility.

We choose the domain of \( F \) as \( D(F) := D^+ \subset L^2(I) \) with the prescribed lower bound \( \zeta > 0 \) mentioned above. It is natural to assume that the constant \( \zeta \) is such that the following assumption holds.

**Assumption 6.1.5** We assume that \( a^\dagger \in D(F) \), i.e. \( \inf_{t \in I} a^\dagger(t) \geq \zeta \).

We remark that \( D(F) \) is convex and closed and hence weakly closed. Assumption 6.1.5 and Proposition 6.1.1 imply that the following three conditions are satisfied

1. \( F : D(F) \subset L^2(I) \rightarrow L^2(I) \) is continuous.
2. \( F \) is weakly closed, i.e. for every sequence \( \{a_n\} \subset D(F) \) weak convergence \( a_n \rightharpoonup a \) in \( L^2(I) \) and weak convergence \( F(a_n) \rightharpoonup u \) in \( L^2(I) \) implies \( a \in D(F) \) and \( F(a) = u \).
3. The Equation (6.5) possesses a unique solution \( a^\dagger \in D(F) \).

From the general theory of nonlinear Tikhonov regularization (cf. end of Section 5.1) we know that the minimization problem (6.11) admits a solution. In general this solution will not be unique. We denote any solution of (6.11) by \( a^\delta_n \). The general theory gives us the following results concerning continuous dependence of the regularized solutions \( a^\delta_n \) on the data \( u^\delta \) and convergence of the regularized solutions (cf. also [22, p. 69ff]).

**Proposition 6.1.6**

Let \( \alpha > 0 \) and let \( u_k \) and \( a_k \) be sequences where \( u_k \xrightarrow{L^2(I)} u^\delta \) and \( a_k \) is a minimizer of (6.5) with \( u^\delta \) replaced by \( u_k \). Then there exists a convergent subsequence of \( \{a_k\} \) and the limit of every convergent subsequence is a minimizer of (6.5).
6.1. INVERSE OPTION PRICING: TIKHONOV-REGULARIZATION

Proposition 6.1.7
Let \( u^\delta \in L^2(I) \) with \( \|u^\dagger - u^\delta\| \leq \delta \) and let \( \alpha(\delta) \) be such that

\[
\alpha(\delta) \to 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \to 0 \quad \text{as} \quad \delta \to 0.
\]

Then

\[
\lim_{\delta \to 0} x^\delta_{\alpha(\delta)} = a^\dagger.
\]

Concerning convergence rates we find the following result (cf. [23, Proposition 5.3 and Theorem 5.4]).

Proposition 6.1.8
Let \( D(F) = D_+^F \) and \( P \neq K \). Let furthermore \( u^\delta \in B_2 \) with \( \|u^\dagger - u^\delta\| \leq \delta \) and \( a^\dagger \) be the solution of (6.5). Assume the following conditions are satisfied.

1. There exists an element \( \omega \in L^2(I) \), such that \( a^\dagger - a^\ast = G(a^\dagger)^\ast \omega \). Here \( G(a^\dagger)^\ast \) denotes the adjoint of the operator \( G(a^\dagger) \), defined in (6.7).

2. It holds \( L\|\omega\|_{L^2(I)} < 1 \), where \( L \) is defined in Proposition 6.1.3.

Then for the parameter choice \( \alpha \sim \delta \) it holds

\[
\|a^\delta_{\alpha} - a^\dagger\|_{B_1} = \mathcal{O}(\sqrt{\delta}).
\]

In [22, p. 70f] we find furthermore an interpretation of the conditions 1. and 2.. In fact the adjoint operator \( G(a^\dagger)^\ast \omega \) has the form

\[
\left[G(a^\dagger)^\ast \omega\right](t) = \int_t^T m(\tau)\omega(\tau) \, d\tau \quad (t \in [0, T]),
\]

where we have used the auxiliary function \( m \) defined in (6.8).

Therefore condition 1. attains the form

\[
a^\dagger(t) - a^\ast(t) = \int_t^T m(\tau)\omega(\tau) \, d\tau.
\]

(6.12)

Setting \( t = T \) we get the necessary condition

\[
a^\dagger(T) = a^\ast(T),
\]

i.e. the exact volatility \( a^\dagger \) at the time point \( T \) has to be known in advance. Furthermore, the Equation (6.12) implies

\[
\frac{(a^\dagger - a^\ast)'}{m} \in L^2(I)
\]

and condition 2. states that the norm of this element has to be small. Furthermore, in [23] it is shown that for \( P \neq K \) the function \( \frac{1}{m(t)} \) grows exponentially to infinity as \( t \) tends to
zero. In fact, there exist positive constants $K$ and $\bar{K}$ such that with $\nu := \ln \left( \frac{K}{\bar{K}} \right) \neq 0$ and $\tau := \|a^t\|_{L^2(I)}$ the estimate

$$K \sqrt{t} \exp \left( \frac{\nu^2}{2 \tau \sqrt{t}} \right) \leq \frac{1}{m(t)} \leq \bar{K} \sqrt{t} \exp \left( \frac{\nu^2}{2 C \sqrt{t}} \right) \quad (0 < t \leq T)$$

holds. This shows that the condition (6.13) is very rigorous with respect to small $t$. Roughly speaking it implies that for small $t$ the derivative of the exact volatility $a^t$ has to be known in advance. Another necessary condition that stems from (6.12) is

$$\|a^t - a^s\|_{L^2(I)} \leq \frac{\bar{C}}{L},$$

where the constant $\bar{C}$ is defined as supremum

$$\bar{C} := \sup_{(t,s) \in \mathcal{M}_L} \left| \frac{\partial u_{BS}(P, K, r, t, s)}{\partial s} \right| < \infty$$

over the set $\mathcal{M}_L$ defined by the equation (6.10).

It remains to remark that the situation of at-the-money options ($P = K$) is a singular situation as for $P - K \to 0$ the constant $L$ tends to infinity, which makes the analysis difficult. Therefore this situation is excluded in Proposition 6.1.8 but it is addressed in the recent paper [27, Section 5.2], where sufficient conditions for convergence rates are formulated and interpreted.

### 6.2 The outer problem

In this section we discuss the identification of the antiderivative $S$ of $a$, which has been defined in (3.12). This inverse problem has been introduced in [23] as outer problem and leads to an operator equation where the forward operator is of Nemytskii type and satisfies the conditions of Assumption 5.3.1.

Hence, the theory of Section 5.3 concerning the inverse of Nemytskii operators is applicable. This enables us to supplement the analysis carried out in [23, Sections 3 and 4] by answering some open questions. We prove ill-posedness and well-posedness of the inverse problem when considered in the Hilbert space $L^2(I)$ or the Banach space $C(I)$ respectively. In a short subsection we analyse the $L^2(I)$-setting and discuss briefly the applicability of regularization methods which are based on the restriction of the set of admissible solutions.

Most of the section is devoted to the $C(I)$-setting. By proving that the inverse operator is continuous we classify the occurring instabilities as ill-conditioning phenomena. A numerical case study illustrates the slow convergence and leads to a conjecture about convergence rates. Furthermore, we study the stabilising effect of apriori information concerning the monotonicity of the data and the searched solution. Moreover, we propose a numerically effective algorithm that selects for given noiselevel $\delta$ and noisy data $u^\delta$ with $\|u^t - u^\delta\|_{C(I)} \leq \delta$ a strictly monotonically increasing approximate solution $S^\delta$ satisfying

$$\|N(S^\delta) - u^\delta\| \leq \delta.$$
Additionally, the situation of discrete stochastic noise is discussed. The described algorithms are illustrated by numerical case studies.

As in Section 6.1 let $\sigma^\dagger$ denote the exact volatility term structure, $a^\dagger$ its square, $S^\dagger$ the corresponding antiderivative, $P_0 > 0$ the current asset price, $r$ the riskless interest rate and $K > 0$ the considered strike price. Instead of the fair option price data (6.1) we observe a nonnegative data function $u^\delta(t)$ ($t \in I$). We consider two situations

(a) We assume that

$$\|u^\dagger - u^\delta\|_{C(I)} \leq \delta \quad (6.14)$$

holds and try to find an approximation $S^\delta \in C(I)$ where the accuracy of $S^\delta$ is measured in the norm of $C(I_2)$.

(b) We assume that

$$\|u^\dagger - u^\delta\|_{L^2(I)} \leq \delta \quad (6.15)$$

Our aim is to find an approximation $S^\delta \in L^2(I)$ of the function $S^\dagger$, where the accuracy of $S^\delta$ is measured in the norm of $L^2(I)$.

We will now formulate the inverse problems described by the situations (a) and (b) in form of an operator equation with a nonlinear Nemytskii operator $N$ generated by the function $f(t, s) = u^{\text{BS}}(P_0, K, r, t, s)$, i.e.

$$[N(S)](t) = f(t, S(t)) \quad t \in I.$$

For notational convenience we use in both settings (a) and (b) the same notation for the Nemytskii operator.

In the setting (a) we choose the domain as

$$\mathcal{D}_0^+ := \{ S \in C(I) : S(0) = 0, S(t) > 0 \ \forall \ t \in (0, T) \}.$$

Hence, the forward operator $N : \mathcal{D}_0^+ \subset C(I) \to C(I)$ is well defined and inverse problem (a) can be described by

$$N(S^\delta) = u^\delta \quad (S^\delta \in \mathcal{D}_0^+, u^\delta \in C(I)). \quad (6.16)$$

In the setting (b) the domain $\mathcal{D}(N)$ of $N : \mathcal{D}(N) \subset L^2(I) \to L^2(I)$ is chosen as

$$\mathcal{D}(N) := \mathcal{D}^+ := \{ S \in L^2(I) : 0 < S(t) \ \text{a.e.} \in I \}$$

and the inverse problem described by situation (b) can be formulated as the operator equation

$$N(S^\delta) = u^\delta \quad (S^\delta \in \mathcal{D}^+, u^\delta \in L^2_+(I)). \quad (6.17)$$
6. IDENTIFICATION OF THE TIME-DEPENDENT VOLATILITY USING OPTION PRICES

6.2.1 Analytical studies

Preliminaries

We start by summarising some properties of the function \( u_{BS} \) defined in (3.10), which will be useful in the following. For a proof we refer to [22].

**Lemma 6.2.1**

Let the parameters \( P, K > 0 \) and \( r \geq 0 \) be fixed. Then the function \( u_{BS}(P, K, r, t, s) \) is nonnegative and continuous for \((t, s) \in [0, \infty) \times [0, \infty)\). Moreover, for \( t \geq 0 \) and \( s > 0 \) the following properties hold.

1. The function \( u_{BS} \) is continuously differentiable with respect to \( s \) and it holds

\[
\frac{\partial u_{BS}(P, K, r, t, s)}{\partial s} = \phi(d_1) P \frac{1}{2 \sqrt{s}} > 0
\]

with \( d_1 \) defined by (3.11) and the distribution function \( \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \) of standard normal distribution.

2. The function \( u_{BS} \) is continuously differentiable with respect to \( t \), where we have

\[
\frac{\partial u_{BS}(P, K, r, t, s)}{\partial t} = r K e^{-rt} \Phi(d_2) \geq 0.
\]

with \( d_2 \) defined by (3.11).

3. Furthermore, we find the following limit conditions

\[
\lim_{s \to 0} \frac{\partial u_{BS}(P, K, r, t, s)}{\partial s} = \begin{cases} \infty & \text{if } P = K e^{-rt} \\ 0, & \text{else} \end{cases}
\]

as well as

\[
\lim_{s \to \infty} u_{BS}(P, K, r, t, s) = P. 
\]

Clearly, the properties of the function \( u_{BS} \) can also be formulated in terms of the generating function \( f \) of the Nemytskii operator \( N \). Let us formulate some important implications of Lemma 6.2.1

**Corollary 6.2.2**

For fixed parameters \( P_0, K > 0 \) and \( r \geq 0 \) the function \( f : I \times [0, \infty) \to [0, \infty) \) defined by

\[
f(t, s) = u_{BS}(P_0, K, r, t, s) \quad (t, s) \in I \times [0, \infty)
\]

is continuous. Moreover, we have the following conditions

...
1. \( f(t, 0) = \max(P_0 - Ke^{-rt}, 0) \) and \( \lim_{s \to \infty} f(t, s) = P_0 \)

2. The functions \( f(t, s) \) is strictly monotonically increasing with respect to \( s \) and monotonically non-decreasing with respect to \( t \).

Especially, the function \( f \) possesses an extension \( \tilde{f} : I \times \mathbb{R} \to \mathbb{R} \) that satisfies the condition (A) and (B) of Assumption 5.3.1 and is equal to \( f \) on the set \( I \times [0, \infty) \).

**Proof:** Properties 1 and 2 follow immediately from the definition of \( f \) and Lemma 6.2.1. The function \( \tilde{f} \) can be chosen as

\[
\tilde{f}(t, s) = \begin{cases} 
  f(t, s) & s \geq 0 \\
  f(t, 0) + s & s < 0 
\end{cases}
\]

Clearly, \( \tilde{f}(t, s) \) is continuous and strictly monotonically increasing with respect to \( s \), which proves the condition (A) of Assumption 5.3.1. Furthermore, the continuity of \( \tilde{f} \) implies that \( \tilde{f} \) is also a Carathéodory function. Besides, for every \( t \in I \) we have

\[ 0 \leq \tilde{f}(t, s) \leq P_0 \quad \text{whenever } s \geq 0 \]

and

\[ s \leq \tilde{f}(t, s) \leq \max(P_0 - Ke^{-rt}, 0) \leq P_0 \quad \text{whenever } s < 0 \]

Hence, (5.8) is satisfied with \( a = P_0 \) and \( b = 1 \). Hence also the condition (B) of Assumption 5.3.1 is fulfilled.

**The \( L^2 \)-setting**

In the following Subsubsection we are going to analyse the situation (b). Let \( P_0, K > 0 \) and \( r \geq 0 \) be fixed and \( f : I \times [0, \infty) \to [0, \infty) \) be defined by (6.22). We start by proving that in this setting the inverse problem is ill-posed. Indeed, we have a very similar situation as in Example 5.3.9. Here it holds \( 0 \leq f(t, s) \leq P_0 \) for all \( (t, s) \in I \times [0, \infty) \). We can therefore consider the functions

\[ S_n(t) := \begin{cases} 
  t & r \leq T - \frac{1}{n} \\
  T_1 + \sqrt{t - T + \frac{1}{n}t^{3/2}} & t \geq T - \frac{1}{n}
\end{cases} \quad t \in I \]

and \( S_0(t) := t \quad t \in I \). Then we have

\[ \|S_n - S_0\|^2_{L^2(I)} = \int_{T-1/n}^{T} \left( t - T + \frac{1}{n} \right) n^3 dt = \frac{n}{2} \]

and

\[ \|N(S_n) - N(S_0)\|^2_{L^2(I_2)} \leq \int_{T-1/n}^{T} P_0^2 dt \leq \frac{1}{n} P_0^2. \]
Hence,
\[ N(S_n) \to N(S_0) \quad \text{but} \quad S_n \nrightarrow S_0, \quad (6.23) \]
which proves that the inverse of \( N : N(D^+) \subset L^2(I) \to L^2(I) \) is not continuous.

Next we consider two regularization methods which are based on a restriction of the set of admissible solutions. The main point is to find a set \( \tilde{D} \) such that \( N \) restricted to the set \( \tilde{D} \) has a continuous inverse and the true function \( S^\dagger \) is an element of this set \( S^\dagger \in \tilde{D} \). Then we can choose the approximate solution \( S^\delta \) as minimizer of the extremal problem
\[ \|N(\tilde{S}) - u^\delta\|_{L^2(I)} \to \min, \quad \text{subject to} \quad \tilde{S} \in \tilde{D}. \quad (6.24) \]

In [23, Section 4] it has been concluded from the well known result formulated in Proposition 5.1.2 that the set \( \tilde{D} \) can be chosen as
\[ \tilde{D} = D^+_{\kappa,\text{mon}} := \{ S \in D^+ : 0 < S(t_1) \leq S(t_2) \leq \kappa \quad \forall t_1 \leq t_2 \} \]
if apriori information of the form
\[ S^\dagger(t) \leq \kappa \quad \text{a.e. in } I \quad (6.25) \]
is available. Indeed, we have seen in Example 5.1.3 that the set \( D^+_{\kappa,\text{mon}} \) is for every \( \kappa < \infty \) a compactum in \( L^2(I) \).

However, the application of Corollary 5.3.11 shows that the restriction of \( N \) to the set
\[ D^+_{\kappa} := \{ S \in D^+ : 0 < S(t) \leq \kappa \ \text{a.e. in } I \} \]
has also a continuous inverse. In other words, the set \( \tilde{D} \) can also be chosen as \( D^+_{\kappa} \). Combining these considerations we can formulate the following result (cf. also [23, p. Theorem 4.2]).

**Corollary 6.2.3**

Let \( \{u^\delta_n\}_{n=1}^\infty \) be as sequence of noisy data and \( S^\delta_n \) be a minimizer of (6.24) with \( u^\delta \) replaced by \( u^n \) and either \( \tilde{D} = D^+_{\kappa} \) or \( \tilde{D} = D^+_{\kappa,\text{mon}} \). Let furthermore (6.25) be satisfied. Then the convergence
\[ u^\delta_n \xrightarrow{L^2(I)} u^\dagger \quad \text{implies} \quad S^\delta_n \xrightarrow{L^2(I)} S^\dagger. \]

One can now ask which of the sets \( D^+_{\kappa} \) or \( D^+_{\kappa,\text{mon}} \) should be preferred. On the one hand one could argue that \( a^\dagger(t) > 0 \quad \text{a.e.} \) and hence \( S^\dagger \) is strictly monotonically increasing. Therefore, the additional apriori information which is incorporated in the choice \( \tilde{D} = D^+_{\kappa,\text{mon}} \) about the monotonicity of \( S^\dagger \) is objective and can therefore not lead to a misspecification of the model. On the contrary, it is preferable to include all available (and trustworthy) apriori information in the solution process as this might help to overcome the ill-conditioning effects. From the analytical point of view it seems therefore preferable to choose \( \tilde{D} = D^+_{\kappa,\text{mon}} \).

On the other hand the following considerations show that the numerical realization of (6.24) with the choice \( \tilde{D} = D^+_{\kappa} \) is easier as with \( \tilde{D} = D^+_{\kappa,\text{mon}} \). Let us consider the following discretisation of the problem. Let \( V^n \) denote the subspace of \( L^2(I) \) spanned by the stepfunctions that are constant on the intervals \( \left( \frac{i-1}{n}T, \frac{i}{n}T \right] \), i.e.
\[ V^n := \left\{ \sum_{i=1}^{n} g[i] \chi_{\left( \frac{i-1}{n}T, \frac{i}{n}T \right]}(t) : g \in \mathbb{R}^n \right\}. \]
6.2. THE OUTER PROBLEM

Furthermore, let $V_n := V^n \cap L^2_+ (\mathbb{R})$ denote the set of all nonnegative functions in $V^n$ and $D^+_\kappa := V^n_+ \cap D^+_\kappa$ be the discretization of the set $D^+_\kappa$.

Defining the timepoints $t_i := \frac{i}{n} T - \frac{1}{2m} T$ we can discretized the operator $N$ as

$$
N : V^n_+ \rightarrow V^n_+
$$

$$
[N(S)](t) := \sum_{i=1}^{n} f(t_i, S(t_i)) \chi_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t) \quad \text{for } S \in V^n_+.
$$

Besides, the range $N(D^+_\kappa)$ can be described as

$$
\left\{ \sum_{i=1}^{n} g[i] \chi_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t) : g \in \mathbb{R}^n, \max \left(P_0 - Ke^{-rt_i}\right) \leq g[i] \leq f(t_i, \kappa) \forall i = 1, \ldots, n \right\}.
$$

Hence, for given $\tilde{u}^\delta \in V^n_+$ the minimizer of

$$
\|N(\tilde{S}) - u^\delta\|_{L^2(I)} \rightarrow \min \quad \text{subject to } \tilde{S} \in D^+_\kappa
$$

(6.26)

can be computed as follows.

1. Compute $\tilde{u}^\delta \in N(D^+_\kappa)$ by

$$
\tilde{u}^\delta(t) := \sum_{i=1}^{n} \tilde{u}^\delta[i] \chi_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t)
$$

with $\tilde{u}^\delta[i] := \max \left(P_0 - Ke^{-rt_i}, 0, \min \left(u^\delta(t_i), f(t_i, \kappa)\right)\right)$.

2. Compute $\tilde{S}^\delta = \sum_{i=1}^{n} \tilde{S}^\delta[i] \chi_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t)$ with $\tilde{S}^\delta$ defined by

$$
f(t_i, \tilde{S}^\delta[i]) = \tilde{u}^\delta[i] \quad (i = 1, \ldots, n).
$$

(6.27)

We remark that in this step the continuity and the strict monotonicity of $f(t, s)$ can be used. Indeed, for every $i$ (6.27) can be solved by a simple bisection algorithm.

In this way, the solution of the minimization problem (6.26) is decomposed into several independent subproblems, which reduces the complexity of the problem significantly.

As opposed to that the solution of

$$
\|N(\tilde{S}) - u^\delta\|_{L^2(I)} \rightarrow \min \quad \text{subject to } \tilde{S} \in D^+_{\kappa,\text{mon}}
$$

(6.28)

with $D^+_{\kappa,\text{mon}} := V^n_+ \cap D^+_{\kappa,\text{mon}}$ is more complicated. In this situation a decomposition of the problem into independent subproblems is not possible as the restriction $f[i] \leq f[i+1]$ has to be fulfilled. From the numerical point of view it might therefore be interesting to use descriptive regularization with the choice $\tilde{D} = D^+_\kappa$. 
We conclude our considerations with respect to the setting (b) with the remark that an extension to the situation $N : D(N) \subset L^p(I) \to L^q(I)$ with $1 \leq p, q < \infty$ is straightforward. Furthermore, Corollary 5.3.11 which shows that the restriction of $N$ to the set

$$D^+_n := \{ S \in D^+: 0 \leq S(t) \leq \kappa \ \text{a.e. in} \ I \}$$

has a continuous inverse is also valid for an Nemytskii operator acting between $L^p(I)$ and $L^q(I)$ with $1 \leq p, q < \infty$. However, extensions of the result to situations with $p = \infty$ or $q = \infty$ seem not to be obvious.

**The $C(I)$-setting**

In the rest of this section we confine our considerations to the setting (a). We start by formulating some consequences of Lemma 6.2.1 concerning properties of $u^\dagger$.

**Remark 6.2.4** Combining Lemma 6.2.1 with Definitions (3.12) and (3.10) we get immediately the following properties of $u^\dagger(t) = f(t, S^\dagger(t))$. We remark that properties 1 and 2 can also be shown on more general arbitrage-free markets (cf. [21, p. 94ff]).

1. For all $t > 0$ we have

$$\max(P_0 - Ke^{-rt}, 0) < u^\dagger(t) < P_0.$$  

2. The function $u^\dagger(t)$ is continuous and strictly monotonically increasing. Furthermore, it holds

$$u^\dagger(0) = \max(P_0 - K, 0).$$

3. For a continuous volatility function, the term structure $u^\dagger(t)$ is continuously differentiable, where we have

$$u'(t) = \frac{\partial u_{BS}}{\partial s}(P_0, K, r, t, S^\dagger(t)) S^\dagger(t) + \frac{\partial u_{BS}}{\partial t}(P_0, K, r, t, S^\dagger(t))$$

$$= P_0 \phi(d_1^\dagger(t)) \frac{a^\dagger(t)}{2\sqrt{S^\dagger(t)}} + K e^{-rt} \Phi(d_2^\dagger(t)) > 0$$

with $d_1^\dagger$ and $d_2^\dagger$ defined by

$$d_1^\dagger(t) := \frac{\ln \left( \frac{P_0}{K} \right) + rt + \frac{S^\dagger(t)}{2}}{\sqrt{S^\dagger(t)}} \quad \text{and} \quad d_2^\dagger(t) := d_1^\dagger(t) - \sqrt{S^\dagger(t)}.$$  

Especially it holds

$$u''(t) \geq K e^{-rt} \Phi(d_2^\dagger(t)). \quad (6.29)$$

In order to guarantee the existence of a solution $S^\delta \in D_0^+$ of the operator equation (6.16) it is necessary that the noisy data $u^\delta$ satisfy certain conditions. In fact, in [23, p. 1325] it is shown that for all noisy data $u^\delta$ satisfying the first and second condition of Remark 6.2.4
with \( u^\dagger \) replaced by \( u^\delta \) there exists a uniquely defined function \( S^\delta(t) \quad (t \in I) \), which satisfies the equation
\[
f(t, S^\delta(t)) = u^\delta(t) \quad t \in I.
\] (6.30)
Furthermore, the solution \( S^\delta \) is an element of \( D_0^+ \) and \( S^\delta \) is bounded by \( 0 < S^\delta \leq \overline{S} t \in I \) where \( \overline{S} \) satisfies the equation
\[
f(0, \overline{S}) = \max_{t \in I} u^\delta(t) .
\] (6.31)
Noting that the proof of this result does not make use of the strict monotonicity of \( u^\delta \) one can omit this condition and formulate the following assumption guaranteeing the existence of a solution \( S^\delta \in D_0^+ \) of the operator equation (6.16).

**Assumption 6.2.5** The data function \( u^\delta(t) \) is assumed to be continuous and to satisfy
\[
u^\delta(0) = \max(P_0 - K, 0) \quad \text{and} \quad \max(P_0 - Ke^{-rt}, 0) < u^\delta(t) < P_0 \quad \text{for all} \ t \in (0, T].
\] (6.32)
Besides, the properties of the function \( f \) imply that the conditions (6.32) are also satisfied for any function \( u \in N(D_0^+) \). Together, these considerations show that Assumption 6.2.5 is necessary and sufficient for the existence of a function \( S^\delta \) satisfying (6.30). In other words, the range of \( N \) is equal to the set
\[
C_1 := \{ u \in C(I) : u(0) = \max(P_0 - K, 0) \text{ and } \max(P_0 - Ke^{-rt}, 0) < u(t) < P_0 \ \forall \ t \in (0, T]\}.
\]
We proceed by reviewing results of [23] concerning pointwise estimates of the absolute error \( |S^\dagger(t) - S^\delta(t)| \). Denoting by \( f_s(t, s) = \frac{\partial f}{\partial s}(t, s) = \frac{\partial u}{\partial s}(P_0, K, r, t, s)(t, s) \) the partial derivative of \( f \) with respect to \( s \) and fixing \( t \in (0, T] \) we can expand the function \( f \) into a Taylor series at the point \( S^\dagger(t) \). This gives
\[
u^\dagger(t) = f(t, S^\dagger(t)) = f(t, S^\dagger(t)) + f_s(t, S_{im}(t)) (S^\delta(t) - S^\dagger(t)) ,
\]
where \( S_{im} \) denotes a positive intermediate function such that
\[
S_{im}(t) \in [\min(S^\delta(t), S^\dagger(t)), \max(S^\delta(t), S^\dagger(t))] .
\]
Because of (6.18) we can subtract \( u^\dagger(t) = f(t, S^\dagger(t)) \) on both sides and divide by \( f_s(t, S_{im}(t)) \). Taking on both sides the absolute value gives
\[
|S^\delta(t) - S^\dagger(t)| = \frac{1}{f_s(t, S_{im}(t))}|u^\delta(t) - u^\dagger(t)| .
\] (6.33)
Thus, for fixed \( S^\dagger \) and \( S^\delta \) the function \( h(t) := \frac{1}{f_s(t, S_{im}(t))} > 0 \quad 0 < t \leq T \) can be defined and interpreted as error amplification factor. From the definition of \( D_0^+ \) follows
\[
\lim_{t \to 0} S^\dagger(t) = 0 \quad \text{and} \quad \lim_{t \to 0} S^\delta(t) = 0
\]
and therefore also \( \lim_{t \to 0} S_{im}(t) = 0 \). Now, for the situation \( P_0 \neq K \) the limit condition (6.20) shows that \( \lim_{t \to 0} h(t) = \infty \). Thus, in this situation for \( t \approx 0 \) small errors in the data can lead to huge errors in the approximate solution.
These considerations show also that $F^{-1}$ is not Lipschitz continuous, i.e. there cannot exist any constant $L$, possibly depending on $S^0$, such that an estimate of the form
\[ \| S^\delta - S^0 \|_{\mathcal{C}(I)} \leq L \| u^\delta - u^0 \|_{\mathcal{C}(I)} \]
holds. If the operator $N$ were linear, we could now conclude that the inverse operator
\[ N^{-1}: \mathcal{C}_1 \subset \mathcal{C}(I) \to D(N) \subset \mathcal{C}(I) \]
is unbounded and thus not continuous. However, for the nonlinear problem the unboundedness of the error amplification factor might either be a sign of ill-conditioning or of ill-posedness effects in a neighbourhood of $t = 0$.

Thus, the question whether the convergence
\[ u_n \xrightarrow{\mathcal{C}(I)} u_0 \] (with elements $u_n \in \mathcal{C}_1 \quad n = 0, 1, 2, \ldots$) remained open. In [22] only pointwise convergence of $S_n := N^{-1}(u_n)$ to $S^0 := N^{-1}(u_0)$ could be proven. Applying Theorem 5.3.6 we are now able to answer this question.

**Theorem 6.2.6**

Let $P_0, K > 0$ and $r \geq 0$. Then the operator $N^{-1}: \mathcal{C}_1 \subset \mathcal{C}(I) \to \mathcal{C}(I)$ is continuous, i.e.
\[ \| N(S_n) - N(S^0) \|_{\mathcal{C}(I)} \to 0 \] implies \[ \| S_n - S^0 \|_{\mathcal{C}(I)} \to 0. \]

Together with the injectivity of $N$ we conclude from Theorem 6.2.6 that the inverse problem
\[ N(S^\delta) = u^\delta \quad (S^\delta \in D(N), u^\delta \in \mathcal{C}_1) \] (6.34)
is well-posed.

Nevertheless, we have seen that for $P_0 \neq K$ there occur ill-conditioning effects in the sense that small noise in the data $u^\delta$ can be strongly amplified. In terms of convergence one can conclude that for a sequence of noisy data $u^{\delta_n} \in \mathcal{C}_1$ converging to the true data $u^\dagger$ the corresponding solutions $S^{\delta_n} = N^{-1}(u^{\delta_n})$ do also converge to $S^\dagger = N^{-1}(u^\dagger)$ but the convergence can be very slow. One could therefore speak of delayed convergence. In the following subsection we will find a conjecture about the convergence rate.

### 6.2.2 Numerical case studies concerning ill-conditioning effects

Again we confine our considerations to the setting (a). In Subsubsection 6.2.1 we have shown that the inverse problem (6.34) is well-posed but it shows ill-conditioning effects in a neighbourhood of $t = 0$. In this Subsection we are going to illustrate these effects by a numerical case study with synthetic data. Furthermore, we will address the question whether apriori information about the monotonicity of the exact data $u^\dagger$ can be used to overcome these ill-conditioning effects.
In this section we use the time unit year. For a fixed asset price $P_0 = 100$, a strike $K = 85$, a riskless interest rate $r = 0.05$, discrete maturities $t_i = \frac{i}{100} (i = 0, \ldots, N = 100)$ in $I = [0, 0.2]$ and a volatility function

$$u^i(t) = 0.04 \left[0.5 + \frac{0.9}{1 + 100(2.05t - 0.2)^2}\right]$$

we computed the corresponding fair option prices $u^i(t_i)$. After that we collected these values in a vector $u^i = \left(u^i(t_0), \ldots, u^i(t_N)\right)$ and added a (random) noise vector to obtain a noisy data vector $u^\delta$, which satisfies

$$\max_{0 \leq i \leq N} \left|u^i(t_i) - u^\delta[i]\right| \leq \delta \quad \text{with} \quad \delta = 0.0126.$$

We remark that the random noise vector had been chosen in such a way that the condition

$$\max(P_0 - Ke^{-rt_i}, 0) < u^\delta[i] < P_0,$$

which can be interpreted as discrete analogue of Assumption 6.2.5, holds.

Next, we define

$$\underline{u}^\delta(t) := \max(u^i(t) - \delta, P_0 - Ke^{-rt}, 0) \quad \text{and} \quad \overline{u}^\delta(t) := u^i(t) + \delta \quad (t \in I).$$

In this context, it should be noted that $u^i(t) < P_0 - \delta$ and thus $\overline{u}^\delta(t) < P_0$ holds for all $t \in I$. Clearly, any function $u^\delta$ with noise level $\delta$ which satisfies Assumption 6.2.5 varies between $\underline{u}^\delta$ and $\overline{u}^\delta$. Figure 6.1 shows the exact data function $u^i$, the noisy data $u^\delta$ and the bounds $\underline{u}^\delta$ and $\overline{u}^\delta$. Note that we have already zoomed into the picture and displayed the functions only in the interval $[0.07, 0.13]$, otherwise the plotted functions could not be visually distinguished.

Next, we used $u^\delta$ to compute $S^\delta$ as solution of $f(t_i, S^\delta[i]) = u^\delta[i]$. From the monotonicity of $f$ follows that for $t \in I$ the inequality $\underline{u}^\delta(t) \leq u^\delta(t) \leq \overline{u}^\delta(t)$ implies $\underline{S}^\delta(t) \leq S^\delta(t) \leq \overline{S}^\delta(t)$, where $\underline{S}^\delta$ and $\overline{S}^\delta$ are defined by

$$f(t, \underline{S}^\delta(t)) = \underline{u}^\delta(t) \quad \text{and} \quad f(t, \overline{S}^\delta(t)) = \overline{u}^\delta(t) \quad (t \in I).$$

Figure 6.2 compares the exact function $S^i(t)$ with the approximation $S^\delta$ and the bounds $\underline{S}^\delta(t)$, $\overline{S}^\delta(t)$. The strong oscillations of the solution $S^\delta$ and the large distance between $\underline{S}^\delta$ and $\overline{S}^\delta$ in the interval $[0, 0.1]$ are a result of the ill-conditioning effects described above.

We will now illustrate the continuity of the inverse operator, i.e. for $\delta$ tending to zero the corresponding bounds $\underline{S}^\delta$ and $\overline{S}^\delta$ converge (slowly) to $S^i$. For $\delta_k = \frac{100}{\ell}(k = 1, \ldots, 16)$ we define $\underline{u}^{\delta_k}$, $\overline{u}^{\delta_k}$ and the corresponding $\underline{S}^{\delta_k}$, $\overline{S}^{\delta_k}$ by

$$\underline{u}^{\delta_k}(t) := \max(u^i(t) - \delta_k, P_0 - Ke^{-rt}, 0) \quad \text{and} \quad \overline{u}^{\delta_k}(t) := u^i(t) + \delta_k \quad (t \in I),$$

$$f(t, \underline{S}^{\delta_k}(t)) = \underline{u}^{\delta_k}(t) \quad \text{and} \quad f(t, \overline{S}^{\delta_k}(t)) = \overline{u}^{\delta_k}(t) \quad (t \in I).$$

See Figure 6.3 for an illustration. Each pointed line corresponds to one $\underline{S}^{\delta_k}$ or $\overline{S}^{\delta_k}$, respectively.
Although the maximal errors \( \| u^\delta_k - u^\dagger \|_{C(I)} \) and \( \| u^\delta_k - u^\dagger \|_{C(I)} \) decrease by a factor \( \frac{1}{3} \) the errors \( \| S^\delta_k - S^\dagger \|_{C(I)} \) as well as \( \| S^\delta_k - S^\dagger \|_{C(I)} \) decrease very slowly. This leads to the conjecture that the convergence rate is logarithmic, i.e. for the modulus of continuity (cf. (5.18)) it holds

\[
\omega_{N^{-1}}(\delta, S^\dagger) = C \frac{1}{\ln^\mu \left( \frac{1}{\delta} \right)}
\]

(6.37)

with appropriate constants \( C \in (0, \infty) \) and \( \mu > 0 \).

In order to see whether this conjecture is true we set \( h_k := \omega_{N^{-1}}(\delta_k, S^\dagger) \). Proposition 5.3.7 states

\[
h_k = \max \left( \| S^\delta_k - S^\dagger \|_{C(I)}, \| S^\delta_k - S^\dagger \|_{C(I)} \right).
\]

If the conjecture (6.37) is true we have \( h_k = C \frac{1}{\ln^\mu \left( \frac{1}{\delta_k} \right)} \) and therefore

\[
\ln \left( \frac{1}{h_k} \right) = \ln \left( \frac{1}{C} \ln^\mu \left( \frac{1}{\delta_k} \right) \right) = \ln \left( \frac{1}{C} \right) + \mu \ln \left( \ln \left( \frac{1}{\delta_k} \right) \right).
\]

In other words, if we plot \( \ln \left( \frac{1}{h_k} \right) \) against \( \ln \left( \frac{1}{\delta_k} \right) \) all points must be on a straight line with slope \( \mu \).

In order to be precise, one has to remark that we are not able to compute the values \( h_k \) exactly, as this would mean to compute the entire functions \( S^\delta_k \) and \( S^\delta_k \). In practice we can only compute the values of these functions at discrete timepoints \( t_i \). However, choosing \( t_i = i\Delta \) with very small \( \Delta \) this difference is neglectable.

Indeed, using the monotonicity of the data function \( u^\dagger \) we obtain for every \( t \in [t_i, t_{i+1}] \) the inequality

\[
\underline{u}^\delta_k(t_i) \leq u^\delta_k(t) \leq \overline{u}^\delta_k(t_{i+1}).
\]

Remembering the monotonicity of the function \( f \) (cf. Corollary 6.22) we conclude that the function \( g(t, u) \) defined in Lemma 5.3.3 is monotonically decreasing with respect to \( t \) and
Thus, defining

\[ S_{\delta}^k(t) \]

\[ g(t_i, u^k_{\delta}(t_i)) \leq g(t, u^k_{\delta}(t_i)) \leq S^k(t) \leq g(t, u^k_{\delta}(t_i)) \leq g(t, u^k_{\delta}(t_i + 1)). \]

(6.38)

In combination with \( S^i(t) \leq S^i(t) \leq S^i(t) \) we obtain for all \( t \in [t_i, t_{i+1}] \) the error estimate

\[ |S^k_{\delta}(t) - S^i(t)| \leq \max \left( g(t, u^k_{\delta}(t_i + 1)) - s^i(t_i), S^i(t_i + 1) - g(t_{i+1}, u^k_{\delta}(t_i)) \right). \]

Thus, defining

\[ h_k := \max_{i=0, \ldots, N} \max \left[ S^k_{\delta}(t_i) - S^i(t_i), S^i(t_i) - S^k_{\delta}(t_i) \right] \]

\[ \overline{h}_k := \max_{i=0, \ldots, N} \max \left[ g(t, u^k_{\delta}(t_i + 1)) - S^i(t_i), S^i(t_i + 1) - g(t_{i+1}, u^k_{\delta}(t_i)) \right] \]

we have \( h_k \leq h_k \leq \overline{h}_k \). In Figure 6.4 we have plotted \( \ln \left( \frac{h_k}{h_k} \right) \) and \( \ln \left( \frac{\overline{h}_k}{\overline{h}_k} \right) \) against \( \ln \left( \frac{1}{h_k} \right) \).

We see that the difference is really very small. Furthermore, the points are nearly on a straight line and the slope is approximately 1. Therefore in the specific situation (i.e. for the chosen parameters \( P_0, K, r, S^i \)) the modulus of continuity has the form (6.37) with \( \mu \approx 1 \) and \( C \approx e^{-4.1} \).

Clearly, the constants \( \mu \) and \( C \) may depend on \( P_0, K, r \) and on the function \( S^i \). Especially we know that the strong error amplification for \( t \approx 0 \) discussed above does not occur in the situation \( P_0 = K \). In this situation we would therefore expect a significantly better convergence rate. However, a rigorous analysis which clarifies the dependence between the parameters \( P_0, K \) and the convergence rate is beyond the scope of this thesis.

In the rest of this subsection we are going to address how apriori information about the monotonicity of the true data \( u^i \) and about the exact solution \( S^i \) can be incorporated in the solution process and which benefits we can expect from this additional information.

With respect to the strong oscillations of \( S^i \) which we have seen in Figure 6.1 one could conjecture that a smoothening of the data \( u^i \) (e.g. a monotonization) should lead to a
smoothening of the corresponding solution $S^\delta$. In order to see whether this conjecture holds true, we performed the following experiment.

Remembering that the exact data $u^\dagger$ are strictly monotonically increasing, we removed the (small) oscillations in the data by transforming the vector $u^\delta$ into a vector $u^\delta_{\text{mon}}$ satisfying

$$u^\delta_{\text{mon}}[i] - u^\delta_{\text{mon}}[i - 1] > 0 \quad \text{for } i = 1, \ldots, N.$$ 

By this monotonization the error of the data reduced slightly from

$$\max_{0 \leq i \leq N} |u^\delta[i] - u(t_i)| = 0.0126 \quad \text{to} \quad \max_{0 \leq i \leq N} |u^\delta_{\text{mon}}[i] - u(t_i)| = 0.0121. $$

Figure 6.5 compares the monotonized data $u^\delta_{\text{mon}}$ with the exact right hand side $u^\dagger$ and the original noisy data $u^\delta$. It can be seen that the monotonization has removed the oscillations in the data. However, the corresponding approximate solution $S^\delta_{\text{mon}}$ obtained from the data $u^\delta_{\text{mon}}$ is hardly better than the original solution $S^\delta$ (cf. Figure 6.6). This example shows that a monotonization of the data $u^\delta$ does in general not lead to a significant improvement of the solution.

In view of [23, Theorem 3.4] which states that the condition

$$(u^\delta)'(t) \geq K re^{-rt} \Phi \left( \frac{\ln \left( \frac{S^\delta}{K} \right) + rt - \frac{1}{2} S^\delta(t)}{\sqrt{S^\delta(t)}} \right) \quad \text{for all } t \in I \quad (6.39)$$

guarantees the monotonicity of $S^\delta = N^{-1}(u^\delta)$ one could conjecture that a further smoothening of the data $u^\delta$ by requiring a certain minimal slope improves the corresponding solution $S^\delta$.

Unfortunately, the condition (6.39) is not very helpful for practical computations. In fact, the right hand side of the inequality (6.39) contains the solution $S^\delta$. But if one has already computed $S^\delta$ one could test directly whether $S^\delta$ is monotone instead of checking whether the condition (6.39) is satisfied.
Nevertheless, we will deduce from (6.39) a bound of \((u^\dagger)'\). This can be viewed as apriori information about the exact data \(u^\dagger\). Using the numerical example introduced above we will then test whether the incorporation of this information into the solution process leads to an improvement of the obtained approximate solution.

We start by defining the function \(m\) as follows

\[
m(t) := K r e^{-rt} \Phi \left( \frac{\ln(P_0/K) + rt - 1/2 S(t)^\dagger}{\sqrt{S(t)^\dagger}} \right) \leq u^\dagger'(t) \quad (t \in I).
\] (6.40)

Assuming to be aware of certain positive bounds \(c\) and \(C\) for \(a^\dagger(t)\) we define the functions \(\underline{S}(t) = c t\) and \(\overline{S}(t) = C t\) which satisfy the inequality

\[
0 < \underline{S}(t) \leq S^\dagger(t) \leq \overline{S}(t) \quad 0 < t \leq T.
\] (6.41)

Using elementary calculations it is now easy to prove the following lemma, which gives a lower bound of \(m\) and therefore a minimal slope of \(u^\dagger\).

**Lemma 6.2.7**

Let the functions \(\underline{S} : I \to \mathbb{R}\) and \(\overline{S} : I \to \mathbb{R}\) be such that (6.41) holds. Defining the functions

\[
d_{\underline{2}}(t) := \begin{cases} \\
\frac{\ln(P_0/K) + rt - 1/2 \overline{S}(t)}{\sqrt{\overline{S}(t)}} & \text{if } \ln(P_0/K) + rt - 1/2 \overline{S}(t) \geq 0 \\
\frac{\ln(P_0/K) + rt - 1/2 \underline{S}(t)}{\sqrt{\underline{S}(t)}} & \text{else}
\end{cases}
\] (6.42)

\[
d_{\overline{2}}(t) := \begin{cases} \\
\frac{\ln(P_0/K) + rt - 1/2 \underline{S}(t)}{\sqrt{\underline{S}(t)}} & \text{if } \ln(P_0/K) + rt - 1/2 \underline{S}(t) \geq 0 \\
\frac{\ln(P_0/K) + rt - 1/2 \overline{S}(t)}{\sqrt{\overline{S}(t)}} & \text{else}
\end{cases}
\] (6.43)

the function \(d_{\underline{2}}^\dagger\) can be bounded by

\[
d_{\underline{2}}(t) \leq d_{\underline{2}}^\dagger(t) \leq d_{\overline{2}}(t) \quad (t \in I).
\]

Furthermore, using

\[
m(t) := K r e^{-rt} \Phi(d_{\underline{2}}(t)) \quad \text{and} \quad \overline{m}(t) := K r e^{-rt} \Phi(d_{\overline{2}}(t))
\] (6.44)

we have

\[
m(t) \leq m(t) \leq \overline{m}(t) \quad (t \in I).
\]

Now we come back to the question whether enforcing the minimal slope \(\overline{m}\) in the data \(u^\delta\) leads to an improvement of the corresponding solution \(S^\delta\). We consider again the numerical example introduced above. Using the bounds \(c = 0.01\) and \(C = 1\) for \(a^\dagger\) we compute \(\overline{m}\) via (6.44).

As mentioned before, in Section 6.2.3 we will propose an algorithm that transforms a function \(g\) into a function \(g_{\min}\) possessing a predefined minimal slope. Applying a discrete variant of this algorithm to the noisy data \(u^\delta\) we get data \(u^\delta_{\min}\) possessing a minimal slope \(m\).
Figure 6.7: Transformation of the monotonized data \( u_{\text{mon}}^\delta \) to data \( u_{\text{min}}^\delta \) with minimal slope.

Figure 6.8: Approximate solutions \( S_{\text{min}}^\delta \) and \( S_{\text{mon}}^\delta \) obtained from data \( u_{\text{mon}}^\delta \) and \( u_{\text{min}}^\delta \) respectively.

Figure 6.7 compares the exact data \( u^\dagger(t) \), the monotonized data \( u_{\text{mon}}^\delta \) and \( u_{\text{min}}^\delta \). We see that imposing the minimal slope \( m \) leads to a further smoothening of the data. In this situation the error of the data in the maximum norm decreased slightly from

\[
\| u_{\text{mon}}^\delta - u^\dagger \| = 0.0121 \quad \text{to} \quad \| u_{\text{min}}^\delta - u^\dagger \| = 0.0111.
\]

The corresponding solution \( S_{\text{min}}^\delta \) is shown in Figure 6.8 in comparison to the exact function \( S^\dagger(t) \) and the approximate solution \( S_{\text{mon}}^\delta \), which has been computed from the data \( u_{\text{mon}}^\delta \). We see that \( S_{\text{min}}^\delta \) possesses fewer oscillations than \( S_{\text{mon}}^\delta \) and the maximum norm of the error decreased from

\[
\| S_{\text{mon}}^\delta - S^\dagger \| = 0.00347 \quad \text{to} \quad \| S_{\text{min}}^\delta - S^\dagger \| = 0.00254.
\]

However, even for close bounds \( c \) and \( C \) the computed minimal slope \( m \) can not guarantee the monotonicity of \( S^\delta \). It is therefore desirable to introduce our apriori information about the monotonicity of \( S^\dagger \) directly in the solution process. An algorithm which performs this task will be presented in the next subsection.

### 6.2.3 Regularization by monotonization – Algorithm

In this subsection we define a numerically efficient algorithm which transforms the input (noisy data \( u^\delta \) and the corresponding noise level \( \delta \)) into an approximate solution \( \hat{S}^\delta \) of (6.16). The principal idea of this algorithm can be described as follows. Under all functions \( \hat{S} \) satisfying

\[
\| N(\hat{S}) - u^\delta \| \leq \delta \tag{6.45}
\]

choose a function which matches our apriori information, i.e. which has a predefined minimal slope \( h > 0 \) and is preferably uniformly sloped.

Obviously, (6.45) implies \( \| u^\dagger - N(\hat{S}^\delta) \| \leq 2\delta \). Thus, for a sequence of noisy data \( u_{\delta_k} \) with given noise levels \( \delta_k \to 0 \) the approximate solutions \( \hat{S}^\delta \) obtained by this algorithm converge to \( S^\dagger \) (see also Theorem 6.2.6).
The first step of the algorithm consists in computing a lower bound $u^{lb,\delta}$ and an upper bound $u^{ub,\delta}$ for $u^\dagger$, such that it holds

$$u^\dagger(t) - 2\delta \leq u^{lb,\delta}(t) \leq u^\dagger(t) \leq u^{ub,\delta}(t) \leq u^\dagger(t) + 2\delta \quad (t \in I).$$

(6.46)

Next, the solutions $S^{lb,\delta}(t)$ and $S^{ub,\delta}(t)$ of

$$f(t, S^{lb,\delta}(t)) = u^{lb,\delta}(t) \quad \text{and} \quad f(t, S^{ub,\delta}(t)) = u^{ub,\delta}(t)$$

are calculated. From the monotonicity of $f(t, s)$ in $s$ it follows that these functions are lower and upper bounds of $S^\dagger(t)$, i.e. it holds

$$0 \leq S^{lb,\delta}(t) \leq S^\dagger(t) \leq S^{ub,\delta}(t) \quad (t \in I).$$

In general $S^{lb,\delta}$ and $S^{ub,\delta}$ are not yet monotone. Therefore, the third step uses apriori information of the form

$$0 < h(t) \leq a^\dagger(t) \quad (t \in I)$$

(6.47)

to transform them into strictly monotonically increasing functions $S^{lb,\delta}_{mon}$ and $S^{ub,\delta}_{mon}$ such that it holds

$$S^{lb,\delta}_{mon}(t) \leq S^{lb,\delta}(t) \leq S^\dagger(t) \leq S^{ub,\delta}_{mon}(t) \leq S^{ub,\delta}(t) \quad (t \in I).$$

(6.48)

In this context $h$ denotes a positive function $h : I \to (0, \infty)$, which can be interpreted as minimal slope of $S^\dagger$. The usage of a time-depending minimal slope $h$ instead of a constant is motivated by the fact that it seems realistic that our apriori information of $a^\dagger(t)$ is better for small $t$ than for large values of $t$.

Finally we choose some monotone function $\tilde{S}^\dagger$ between $S^{lb,\delta}_{mon}$ and $S^{ub,\delta}_{mon}$.

The basis of the sketched algorithm are the following two lemmas. The first is concerned with a projector mapping from $C(I)$ in the set

$$\mathcal{M}_h := \left\{ g \in C(I) : g(t_2) \geq g(t_1) + \int_{t_1}^{t_2} h(\tau) \, d\tau \quad \forall \ t_1 \geq t_2 \right\}.$$

containing all continuous functions with minimal slope $h$, where $h : I \to (0, \infty)$ denotes a predefined integrable function. We remark that $\mathcal{M}_h$ is contained in the set of strictly monotonically increasing functions.

The second lemma proposes an algorithm that transforms lower and upper bounds $f^{lb}$ and $f^{ub}$ of a function $f$ having a strictly positive derivative $f' \geq h > 0$ into lower and upper bounds $f^{lb}_{mon}$, $f^{ub}_{mon} \in \mathcal{M}_h$. The proofs of the lemmas are elementary and can be found in [31].

**Lemma 6.2.8**

Let $h : I \to (0, \infty)$ be an integrable function. Then the operators $P^{1/2}_h : C(I) \to C(I)$ defined as

$$[P^{lb}_h g](t) = \max_{v \in [0, t]} \left( g(v) + \int_v^t h(\tau) \, d\tau \right)$$

(6.49a)

$$[P^{ub}_h g](t) = \min_{v \in [t, T]} \left( g(v) - \int_t^v h(\tau) \, d\tau \right)$$

(6.49b)

are well defined projectors onto the set $\mathcal{M}_h$. Furthermore, it holds

$$[P^{lb}_h g](t) \geq g(t) \quad \text{and} \quad [P^{ub}_h g](t) \leq g(t) \quad \forall \ t \in I.$$
Lemma 6.2.9
Let \( f : I \to \mathbb{R} \) be a strictly monotonically increasing function, which is continuously differentiable in \((0, T]\). Furthermore, let \( h : I \to (0, \infty) \) be an integrable function such that
\[
0 < h(t) \leq f'(t) \quad \forall t \in (0, T]
\]
(6.50)
holds. Let furthermore \( f^{lb} : I \to \mathbb{R} \) and \( f^{ub} : I \to \mathbb{R} \) be continuous lower and upper bounds of \( f \), i.e. it holds
\[
f^{lb}(t) \leq f(t) \leq f^{ub}(t) \quad \forall t \in I .
\]

Then the functions
\[
f_{mon}^{lb} := P_{h}^{lb} f^{lb} \quad \text{and} \quad f_{mon}^{ub} := P_{h}^{ub} f^{ub}
\]
(6.51)
are continuous, strictly monotonically increasing and satisfy the inequality
\[
f^{lb}(t) \leq f_{mon}^{lb}(t) \leq f(t) \leq f_{mon}^{ub}(t) \leq f^{ub}(t) \quad t \in I .
\]
(6.52)

Remark 6.2.10
Let \( f \) be as in Lemma 6.2.9. Assume that we are not aware of bounds \( f^{lb} \) and \( f^{ub} \) of \( f \) but we have some approximation \( f^\delta \in C(I) \) with noiseflevel \( \delta \), i.e. it holds \( \| f - f^\delta \|_{C(I)} \leq \delta \). Then one can show by elementary considerations that the functions \( f_{lb}^\delta \) and \( f_{ub}^\delta \) defined by
\[
f_{lb}^\delta (t) := \min_{v \in [t,T]} f^\delta (v) \quad \text{and} \quad f_{ub}^\delta (t) := \max_{v \in [0,t]} f^\delta (v) \quad \forall t \in I
\]
are continuous and it holds
\[
f(t) - \delta \leq f_{lb}^\delta (t) \leq f^\delta (t) \leq f_{ub}^\delta (t) \leq f(t) + \delta .
\]
However, it cannot be guaranteed that
\[
f_{lb}^\delta (t) \leq f(t) \leq f_{ub}^\delta (t)
\]
holds nor that there exists a function \( f_{min}^\delta \) between \( f_{lb}^\delta \) and \( f_{ub}^\delta \) possessing a minimal slope \( h \).

Lemma 6.2.9 indicated how we can compute monotone bounds \( S_{mon}^{lb,\delta}(t) \) and \( S_{mon}^{ub,\delta}(t) \) from \( S_{mon}^{lb,\delta} \) and \( S_{mon}^{ub,\delta} \). The question which has been left open is how to choose a monotone function \( S^\delta \) between these bounds. A first idea was to compute \( S^\delta(t) \) pointwise as arithmetic mean, i.e. to choose \( S^\delta(t) \) as
\[
S_{ar}^\delta (t) = \frac{1}{2} S_{mon}^{lb,\delta}(t) + \frac{1}{2} S_{mon}^{ub,\delta}(t) \quad t \in I .
\]
(6.53)

This approach works quite well in the regions where the difference between \( S_{mon}^{ub,\delta}(t) \) and \( S_{mon}^{lb,\delta}(t) \) is relatively small. Unfortunately, in the first part of the interval \( I \) this difference is in general rather large. This is due to the fact that in general \( S_{mon}^{lb,\delta}(t) \) and \( S_{mon}^{ub,\delta}(t) \) have a similar behaviour as the functions \( S^\delta \) and \( \overline{S}^\delta \) defined by (6.36). That is, for small \( t \) the function \( S_{mon}^{lb,\delta}(t) \) is (nearly) equal to \( \underline{\alpha} t \) and the function \( S_{mon}^{ub,\delta}(t) \) is close to a nearly horizontal line. Then at some timepoint the function \( S_{mon}^{lb,\delta}(t) \) grows very quickly. Consequently, the
difference between $S_{mon}^{ub,\delta}(t)$ and $S_{mon}^{lb,\delta}(t)$ decreases rapidly and after some timepoint $t^*$ this difference is sufficiently small.

From these considerations we can now conclude that $S_{ar}^\delta$ is in general not the best approximation of $S^\delta$ in the first part of the interval. Indeed, the large slope of $S_{mon}^{lb,\delta}(t)$ near $t^*$ leads also to a quite large slope in the convex linear combination (6.53). This would correspond to a volatility function $\tilde{\sigma}^\delta(t)$ with a very sharp and high peak. Clearly, in absence of any other information we would rather prefer a smooth or even nearly constant volatility. This would lead to a choice of $\tilde{S}^\delta$ as (6.53) in the region $t > t^*$ and as linear interpolation in the interval $[0, t^*]$. Finally, we have to be aware of the fact that it might of course happen that the underlying volatility $\sigma^\dagger$ possesses really a sharp peak. Then the data $u^\delta$ can be such that the linear interpolation intersects the bounds $S_{mon}^{ub,\delta}(t)$ and $S_{mon}^{lb,\delta}(t)$. In this situation we should therefore choose $\tilde{S}^\delta$ on the entire interval $I$ as $S_{ar}^\delta$.

Combining all these ideas we suggest the following pragmatic approach. First we compute the mean distance

$$md := \frac{1}{0.2T} \int_{0.8T}^{T} \left| S_{mon}^{ub,\delta}(t) - S_{mon}^{lb,\delta}(t) \right| dt$$

over the last part of the interval $I$ and define

$$t^* := \inf \left\{ t \in I : \left| S_{mon}^{ub,\delta}(t) - S_{mon}^{lb,\delta}(t) \right| \leq 3 \cdot md \right\}.$$  

Now, if the function

$$S_{lin}^\delta(t) := \frac{t}{t^*} S_{ar}^\delta(t^*)$$

with $S_{ar}^\delta$ defined by (6.53) satisfies

$$S_{mon}^{lb,\delta}(t) \leq S_{lin}^\delta(t) \leq S_{mon}^{ub,\delta}(t) \quad (t \leq t^*)$$

then set $\tilde{S}^\delta(t) = S_{lin}^\delta(t)$ \quad ($t \leq t^*$), otherwise

$$\tilde{S}^\delta(t) = \frac{t}{2t^*} S_{mon}^{ub,\delta}(t) + \left(1 - \frac{t}{2t^*}\right) S_{mon}^{lb,\delta}(t) \quad (t \leq t^*),$$

which is a pointwise linear combination of $S_{mon}^{lb,\delta}$ and $S_{mon}^{ub,\delta}$. In both cases it holds

$$\tilde{S}^\delta(0) = 0 \quad \text{and} \quad \tilde{S}^\delta(t^*) = S_{ar}^\delta(t^*).$$

Finally we set

$$\tilde{S}^\delta(t) := S_{ar}^\delta(t) \quad t > t^*.$$  

At the end of this subsection we combine all our considerations to formulate an algorithm that realizes the choice of a function $\tilde{S}^\delta$ possessing a minimal slope $\tilde{h}$ and satisfying (6.45).
Algorithm 6.2.11

Input:

• a noise level $\delta$ and a continuous function $u^\delta : I \to \mathbb{R}$ such that $\|u^\delta - u^1\|_{C^1(I)} \leq \delta$ and $u^\delta(t) < P - \delta$ holds

• an integrable function $h$ such that $0 < h(t) \leq a^1(t)$ holds for all $t \in I$

Output: approximate solution $\tilde{S}^\delta(t)$ such that

• $\tilde{S}^\delta(0) = 0$, $\tilde{S}^\delta \in \mathcal{M}_h$

• $\|N(\tilde{S}^\delta) - u^\delta\|_{C^1(I)} \leq \delta$

Algorithm:

1. For $t \in I$ compute lower and upper bounds of $u(t)$ as follows

$$u^{lb,\delta}(t) := \max\left(u^\delta(t) - \delta, f\left(t, \int_0^t h(\tau) \, d\tau\right)\right)$$

$$u^{ub,\delta}(t) := u^\delta(t) + \delta.$$ 

2. Compute pointwise lower bounds $S^{lb,\delta}(t)$ and upper bounds $S^{ub,\delta}(t)$ for $S^\dagger(t)$ as solution of

$$f\left(t, S^{lb,\delta}(t)\right) = u^{lb,\delta}(t) \quad (t \in I).$$

$$f\left(t, S^{ub,\delta}(t)\right) = u^{ub,\delta}(t) \quad (t \in I).$$

3. Monotonise the bounds: Set $S^{lb,\delta}_{\text{mon}} := P_h S^{lb,\delta}$ and $S^{ub,\delta}_{\text{mon}} := P_h S^{ub,\delta}$.

4. Choose $\tilde{S}^\delta(t)$ between $S^{lb,\delta}_{\text{mon}}$ and $S^{ub,\delta}_{\text{mon}}$ by the algorithm (6.54).

Proof: We show that the algorithm terminates without error and the constructed function $\tilde{S}^\delta$ possesses the asserted properties.

1. Obviously the functions $u^{lb,\delta}$ and $u^{ub,\delta}$ satisfy

$$u^\dagger(t) - \delta \leq u^{lb,\delta}(t) \leq u^\dagger(t) \leq u^{ub,\delta}(t) \leq u^\delta(t) + \delta.$$ 

Furthermore, using the properties of the function $f(t, s)$ together with the relation $\int_0^t h(\tau) \, d\tau \leq S^\dagger(t)$ one gets the chain of inequalities

$$\max(P_0 - K e^{-rt}, 0) < f\left(t, \int_0^t h(\tau) \, d\tau\right) \leq u^{lb,\delta}(t) \leq u^l(t) \leq u^{ub,\delta}(t) < P_0 \quad \forall t \in (0, T).$$

and the relation $u^{lb,\delta}(0) = f(0, 0) = \max(P_0 - K, 0)$. In addition, the continuity of $u^\delta$ and the continuity of the function $f$ implies the continuity of $u^{lb,\delta}$ and $u^{ub,\delta}$. 

2. As we have shown in 1., the function $u^{ib,\delta}$ satisfies Assumption 6.2.5 the existence and uniqueness of the pointwise defined function $S_{\text{mon}}^{ib,\delta}$ is ensured. Furthermore, this function is an element of $\mathcal{D}^+$. The existence and uniqueness of the pointwise defined, continuous function $S_{\text{mon}}^{ub,\delta}$ can be shown by analogous considerations. In fact, the only difference is that $u^{ab,\delta}(0) > \max(P - K, 0)$ which implies $S_{\text{mon}}^{ub,\delta}(0) > 0$. The strict monotonicity of $f(t, s)$ with respect to $s$ implies

$$S_{\text{mon}}^{ib,\delta}(t) \leq S_{\text{mon}}^{ub,\delta}(t) \quad \forall t \in I.$$  

3. Now, Lemma 6.2.9 guarantees that the functions $S_{\text{mon}}^{ib,\delta}(t)$ and $S_{\text{mon}}^{ub,\delta}(t)$ are continuous, possess a minimal slope $c > 0$ and satisfy the chain of inequalities

$$S_{\text{mon}}^{ib,\delta}(t) \leq S_{\text{mon}}^{ub,\delta}(t) \leq S_{\text{mon}}^{ub,\delta}(t) \leq S_{\text{mon}}^{ub,\delta}(t) \quad \forall t \in I. \quad (6.55)$$

4. The continuity and the minimal slope of $S_{\text{mon}}^{ib,\delta}$ and $S_{\text{mon}}^{ub,\delta}$ imply the continuity and the minimal slope of $S_{\text{lin}}^{ib,\delta}$. Besides, the chain of inequalities $g^* \leq S_{\text{mon}}^{ib,\delta}(t) \leq S_{\text{lin}}^{ib,\delta}(t)$ implies the minimal slope of $S_{\text{lin}}^{ib,\delta}$. The continuity of $S_{\text{lin}}^{ib,\delta}$ is obvious. Furthermore, the weight function $\frac{1}{2\pi}$ is monotonically increasing. Together with the inequality $S_{\text{lin}}^{ib,\delta} \leq S_{\text{lin}}^{ub,\delta}$ and the continuity and minimal slope of $S_{\text{mon}}^{ib,\delta}$ and $S_{\text{mon}}^{ub,\delta}$ this shows the continuity and minimal slope of $S_{\text{lin}}^{ib,\delta}(t) (t \leq t^*)$ if it is defined by (6.54).

Finally, the property $S_{\text{mon}}^{ib,\delta} \leq \tilde{S}_{\text{lin}}^{ib,\delta} \leq S_{\text{mon}}^{ub,\delta}$ together with the strict monotonicity of $f(t, s)$ with respect to $s$ implies $u^{ib,\delta} \leq N(\tilde{S}_{\text{lin}}^{ib,\delta}) \leq u^{ub,\delta}$ and thus $\|N(\tilde{S}_{\text{lin}}^{ib,\delta}) - u^{ib,\delta}\| \leq \delta$.

\textbf{Remark 6.2.12} 1. The functions $\tilde{S}_{\text{lin}}^{ib,\delta}$ and $\tilde{u}^{ib,\delta}$ are continuous. However, in general they are not continuously differentiable.

2. For the algorithm the knowledge of the noise level $\delta$ is crucial. If one is not aware of this level, one could compute $S_{\text{lin}}^{ib,\delta}$ from $N(S_{\text{lin}}^{ib,\delta}) = u^{ib,\delta}$ and use the ideas of Remark 6.2.10 and (6.54) to monotonize this function. However, this approach guarantees only monotonicity, no strict monotonicity or even a minimal slope.

3. Instead of the noisy data $u^{\delta}$ and the noise level $\delta$ one could also use bounds $u^{ib,\delta}$ and $u^{ub,\delta}$ for $u^\dagger$ as input for the algorithm. This would be motivated by the assumption that the noise in the data is solely due to the bid-ask spread.

\subsection{6.2.4 Discrete Variant}

Instead of observing the data $u^{\delta} : I \rightarrow \mathbb{R}$ on the entire interval we can rather expect to observe only discrete values $u^d(t_i)$ for several maturities $t_i (i = 0, \ldots, N)$. In order to keep notation simple, we assume that the $t_i$ form a uniform grid on $I$, i.e. we assume $t_i = i\Delta$ $(i = 0, \ldots, N)$ with a positive increment $\Delta$.

Combining the observed prices into the vector $u^d := (u^d(t_0), \ldots, u^d(t_N))$, we can formulate the following modification of Algorithm 6.2.11. For the seek of simple notation we formulate the algorithm only for the case where the function $h$ is constant.
Algorithm 6.2.13

Input:

- step width $\Delta$
- number of observations $N + 1$ and observation points $t_i = i\Delta$ ($i = 0, \ldots, N$).
- vector $u^\delta \in \mathbb{R}^{N+1}$ such that $\max_{i=0,\ldots,N} |u^\delta[i] - u^\dagger(t_i)| \leq \delta$ and $\max_{i=0,\ldots,N} u^\delta[i] < P$ is satisfied
- a constant $h$ such that $0 < h \leq a^\dagger(t)$ holds for all $t \in I$

Output: an approximate solution $\tilde{S}^\delta : I \to \mathbb{R}$ such that

- $\tilde{S}^\delta(0) = 0$
- $\tilde{S}^\delta \in \mathcal{M}_h$
- the corresponding function $\tilde{u}^\delta = N(\tilde{S}^\delta)$ satisfies $\max_{0 \leq i \leq N} |\tilde{u}^\delta(t_i) - u^\dagger(t_i)| \leq 2\delta$.

Algorithm:

0. Initialisation

\[ t := [0 : T/N : T] \]
for i from 0 step 1 to N
\[ u[i] := f(t[i], h_t[i]) \]
end %for
\[ u^{lb,\delta} := \max(u^\delta - \delta, u) \]
\[ u^{ub,\delta} := u^\delta + \delta \]

1. Using a simple bisection algorithm compute $S^{lb,\delta}$ and $S^{ub,\delta}$ from

\[ f(t[i], S^{lb,\delta}[i]) = u^{lb,\delta}[i] \]
\[ f(t[i], S^{ub,\delta}[i]) = u^{ub,\delta}[i] \] ($i = 0, \ldots, N$)

2. Compute the strictly monotonically increasing vectors $S^{lb,\delta}_{mon}$ and $S^{ub,\delta}_{mon}$ by

\[ S^{lb,\delta}_{mon}[0] := 0; \]
\[ S^{ub,\delta}_{mon}[N] := S^{ub,\delta}[N] \]
for i from 1 step 1 to N
\[ S^{lb,\delta}_{mon}[i] := \max(S^{lb,\delta}_{mon}[i - 1] + \Delta h, S^{lb,\delta}[i]); \]
end %for
for i from N-1 step -1 to 0
\[ S^{ub,\delta}_{mon}[i] := \min(S^{ub,\delta}_{mon}[i + 1] - \Delta h, S^{ub,\delta}[i]); \]
end %for
3. Compute $S_{\text{mon}}^{lb, \delta}$ and $S_{\text{mon}}^{ub, \delta}$ by linear interpolation such that

$$S_{\text{mon}}^{lb, \delta}(t_i) = S_{\text{mon}}^{lb, \delta}[i]$$
$$S_{\text{mon}}^{ub, \delta}(t_i) = S_{\text{mon}}^{ub, \delta}[i]$$

for $i = 0, \ldots, N$ holds. Choose $\tilde{S}^{\delta}$ between $S_{\text{mon}}^{lb, \delta}$ and $S_{\text{mon}}^{ub, \delta}$ by the algorithm (6.54).

**Remark 6.2.14**

1. Note that for computing $S_{\text{mon}}^{lb, \delta}[i]$ it is not necessary to compute

$$\max_{j \leq i} (S_{\text{mon}}^{lb, \delta}[j] + (i - j)\Delta h),$$

which would mean to compute the maximum of $i$ values. It suffices to compute the maximum

$$\max (S_{\text{mon}}^{lb, \delta}[i - 1] + \Delta h, S_{\text{mon}}^{lb, \delta}[i])$$

of two values.

2. Concerning the residuum the algorithm guarantees the estimate $|\tilde{u}^{\delta}(t_i) - u^\dagger(t_i)| \leq 2\delta$ at the points $t_i (i = 0, \ldots, N)$. As the functions $\tilde{u}^{\delta}$ and $u^\dagger$ are strictly monotonically increasing this implies for $t \in [t_{i-1}, t_i]$ the estimate

$$u^\dagger(t_{i-1}) - 2\delta \leq \tilde{u}^{\delta}(t_{i-1}) \leq \tilde{u}^{\delta}(t) \leq \tilde{u}^{\delta}(t_i) \leq u^\dagger(t_i) + 2\delta$$

and therefore

$$u^\dagger(t) - 2\delta - (u^\dagger(t_i) - u^\dagger(t_{i-1})) \leq \tilde{u}^{\delta}(t) \leq u^\dagger(t) + 2\delta + u^\dagger(t_i) - u^\dagger(t_{i-1}).$$

Thus, defining the constant $C := \max_{t \in I} |u''(t)|$ it holds

$$|\tilde{u}^{\delta}(t) - u^\dagger(t)| \leq 2\delta + \Delta C.$$

We will now illustrate the performance of Algorithm 6.2.13. We use again the noisy data $u^{\delta}$ which had been defined in Subsection 6.2.2 and illustrated in Figure 6.1. Furthermore, we use the constant $h = 0.01$ as input of the algorithm. Figure 6.9 compares the exact data $u^\dagger$ with the lower and upper bounds $u^{lb, \delta}$ and $u^{ub, \delta}$ which have been computed in Step 1 of the algorithm. As in Section 6.2.2 we zoomed into the picture and plotted the function only in the interval $[0.07, 0.13]$. The corresponding lower and upper bounds $S_{\text{mon}}^{lb, \delta}$ and $S_{\text{mon}}^{ub, \delta}$, which have been constructed in Step 3 are illustrated in Figure 6.10. As expected, the lower bound is in the first part (nearly) equal to $ht$. Furthermore, the noisy data were such that at $t_1 = 0.036$ and $t_2 = 0.06$ the function $S_{\text{mon}}^{ub, \delta}$ is quite close to $S^\dagger$, whereas the other points in this region are further away. Consequently, the monotonized bound $S_{\text{mon}}^{ub, \delta}$ which is illustrated together with $S_{\text{mon}}^{lb, \delta}$ and $S^\dagger$ in Figure 6.11, possesses at these points jumps. Finally, Figure 6.12 shows $\tilde{S}^{\delta}$ which has been computed in Step 4. We see that $t^* = 0.0104$ and $\tilde{S}^{\delta}(t) (t \leq t^*)$ is chosen as $S_{\text{lin}}^{\delta}$. 
Figure 6.9: Result of Step 1: Lower and upper bounds $u_{lb,\delta}$ and $u_{ub,\delta}$ in comparison with the exact data $u^\dagger(t)$

Figure 6.10: Result of Step 2: Corresponding lower and upper bounds $S_{lb,\delta}$ and $S_{ub,\delta}$

Figure 6.11: Result of Step 3: Monotone lower and upper bounds $S_{mon_{lb,\delta}}$ and $S_{mon_{ub,\delta}}$ in comparison with the exact function $S^\dagger(t)$

Figure 6.12: Result of Step 4: Approximate solution $\tilde{S}^{\delta}$ between $S_{lb,\delta}$ and $S_{ub,\delta}$
Let us now consider a sequence of stepwidths \( \Delta_k \) and corresponding grids having \( N_k := \frac{T}{\Delta_k} \) points. Let us furthermore assume that we can for each \( k \in \mathbb{N} \) observe noisy data \( (u^\delta(j \Delta_k))^N_{j=1} \) with noise level \( \delta_k \), i.e. satisfying

\[
|u^\delta(j \Delta_k) - u^\delta_i(j \Delta_k)| \leq \delta_k \quad (j = 0, \ldots, N).
\]

Then we can for each \( k \in \mathbb{N} \) construct an approximate solution \( \tilde{S}^\delta_k \in \mathcal{C}(I) \). Furthermore, if the noise levels \( \delta_k \) and the stepwidths \( \Delta_k \) converge to 0 then \( \tilde{S}^\delta_k \) converges to \( S^I \) in \( \mathcal{C}(I) \).

Before we investigate the situation of discrete data with random noise we want to remark that the algorithm might not terminate correctly if the noise level \( \delta \) in Algorithm 6.2.13 is misspecified, i.e. as input we use a level \( \delta_1 \) and noisy data \( u^\delta \) such that

\[
\max_{i=0, \ldots, N} |u^\delta[i] - u^I(t_i)| = \delta_2 > \delta_1
\]

holds. Indeed, in this situation the computed vectors \( u^{lb,\delta} \) and \( u^{ub,\delta} \) do not satisfy

\[
u^{lb,\delta}[i] \leq u^I(t_i) \leq u^{ub,\delta}[i] \quad (i = 0, \ldots, N)
\]

and hence also the corresponding vectors \( S^{lb,\delta} \) and \( S^{ub,\delta} \) do not satisfy

\[
S^{lb,\delta}[i] \leq S^I(t_i) \leq S^{ub,\delta}[i] \quad (i = 0, \ldots, N).
\]

Figure 6.13 shows a situation where this happens.

As this example shows, for wrongly specified \( \delta \) the constructed functions \( S^{lb,\delta}_{mon} \) and \( S^{ub,\delta}_{mon} \) may intersect, i.e. there exists some \( i_1 \) such that \( S^{lb,\delta}_{mon}[i_1] > S^{ub,\delta}_{mon}[i_1] \). Clearly, in this situation it is not possible to choose \( \tilde{S}^\delta \) such that

\[
S^{lb,\delta}_{mon}[i] \leq \tilde{S}^\delta(t_i) \leq S^{ub,\delta}_{mon}[i] \quad i = 0, \ldots, N
\]

holds.

Now we are ready for addressing the situation of discrete and independent (normally distributed) noise, i.e. we assume that we are able to observe for several discrete maturities
\(t_i = i\Delta\ (i = 0, \ldots, N)\) option prices \(u^\delta[i]\) such that the errors \(X_i := u^\delta[i] - u^\dagger(t_i)\) are i.i.d random variables with normal distribution, mean zero and variance \(\delta^2\), i.e. \(X_i \sim \mathcal{N}(0, \delta^2)\). Furthermore, we will assume to be aware of some upper bound \(\overline{S}\) such that it holds

\[
S^\dagger(T) \leq \overline{S}.
\]  

(6.57)

In order to apply the ideas of Algorithm 6.2.13 we would need an upper bound of the maximal error \(\max_{i=0,\ldots,N} |X_i|\). Clearly, we can not expect that the maximum of normally distributed random variables to be bounded almost surely. However, we can use the following coarse estimate which is well known. We remark that the estimates we use in this section are sometimes very coarse, nevertheless they lead to the desired convergence result.

**Lemma 6.2.15**

Let \(Y_i\) be i.i.d. random variables with \(Y_i \sim \mathcal{N}(0, 1)\). Then it holds

\[
P\left(\max_{i=1,\ldots,n} |Y_i| > 2\sqrt{\ln(n)}\right) \to 0 \quad \text{for } n \to \infty.
\]

Defining the constant \(\hat{\delta} := 2\sqrt{\ln(N)}\delta\), which can be interpreted as estimated noise level, Lemma 6.2.15 implies that the probability of the event

\[
\Omega_+ := \left\{ \max_{i=0,\ldots,N} |u^\delta[i] - u^\dagger(t_i)| \leq \hat{\delta}\right\}
\]

is roughly speaking high. Furthermore, in this situation the Algorithm 6.2.13 with input \(\hat{\delta}\), \(u^\delta\) terminates without error and we find \(\hat{S}^\delta\) such that the residuum \(N(\hat{S}^\delta) - u^\delta\) is bounded by \(2\hat{\delta} + \Delta C\) in the \(C(I)\) norm.

In the situation

\[
\Omega_- := \left\{ \max_{i=0,\ldots,N} |u^\delta[i] - u^\dagger(t_i)| > \hat{\delta}\right\}
\]

three cases might occur:

A The algorithm might not terminate without error.

B The algorithm terminates without error and computes the strictly monotone function \(\tilde{S}^\delta\) but it holds \(\tilde{S}^\delta(T) > \overline{S}\).

C The algorithm terminates without error and the strictly monotone function \(\tilde{S}^\delta\) satisfies \(\tilde{S}^\delta(T) \leq \overline{S}\) and therefore \(\|\tilde{S}^\delta - S^\dagger\|_{C(I)} \leq \overline{S}\).

In order to bound the error we modify the Algorithm 6.2.13 inasmuch as we set \(\tilde{S}^\delta(t) = \frac{1}{\Delta} \overline{S}(t)\) if the cases A or B occur. Thus we get the following algorithm.

**Algorithm 6.2.16** Input:

- step width \(\Delta\), number of observations \(N + 1\) and observation points \(t_i = i\Delta\ (i = 0, \ldots, N)\)
6.2. THE OUTER PROBLEM

- variance $\delta$ and a random vector $u^{\delta} \in \mathbb{R}^N$ such that the components $u^{\delta}[i]$ ($i = 1, \ldots, N$) are i. i. d. normal random variables with expectation $u^\dagger(t_i)$ and variance $\delta^2$. The first component is deterministic $u^{\delta}[0] = u^\dagger(0) = \max(P - K, 0)$.

- a constant $h$ such that $0 < h \leq a^\dagger(t)$ holds for all $t \in I$

- an upper bound $\overline{S}$ for $S^\dagger$ such that (6.57) is satisfied

Output: an approximate solution $\tilde{S}^{\delta} : I \to \mathbb{R}$ such that

- $\tilde{S}^{\delta}(0) = 0$

- $\tilde{S}^{\delta} \in \mathcal{M}_h$

- if the event $\Omega_+ := \left\{ \max_{i=1,\ldots,N} |u^{\delta}[i] - u^\dagger(t_i)| \leq \delta \right\}$ with $\delta := 2\delta \sqrt{\ln(N)}$ occurs, then the computed function $\tilde{S}^{\delta}$ is such that the corresponding function $\tilde{u}^{\delta} = N(\tilde{S}^{\delta})$ satisfies $\|\tilde{u}^{\delta}(t_i) - u^\dagger(t_i)\|_{\mathcal{C}(I)} \leq 2\delta + \Delta C$.

- if the event $\Omega_- := \left\{ \max_{i=1,\ldots,N} |u^{\delta}[i] - u^\dagger(t_i)| > \delta \right\}$ occurs, the computed function $\tilde{S}^{\delta}$ is such that $\|\tilde{S}^{\delta} - S^\dagger\|_{\mathcal{C}(I)} \leq \overline{S}$ holds.

Algorithm:

1. Set $\hat{\delta} := 2\sqrt{\ln(N)}\delta$ and perform the Steps 0.-3. of the Algorithm 6.2.13 with $\delta$ replaced by $\hat{\delta}$.

2. If $S^{lb,\delta}_*[i] \leq S^{ub,\delta}_*[i] \leq \overline{S}$ holds for all $i = 0, \ldots, N$ then choose $\tilde{S}^{\delta}$ between $S^{lb,\delta}_*$ and $S^{ub,\delta}_*$ by the algorithm (6.54). Otherwise set $\tilde{S}^{\delta}(t) := \frac{t}{T}S^\dagger$ ($t \in I$).

Now we are able to formulate and prove a convergence result for the situation of discrete and noisy errors.

Theorem 6.2.17
Let $\{N_k\}_{k=1}^{\infty}$, $N_k \in \mathbb{N}$ and $\{\delta_k\}_{k=1}^{\infty} \subset (0, \infty)$ be such that

$$\Delta_k := \frac{T}{N_k} \to 0 \quad \text{and} \quad \delta_k \sqrt{\ln(N_k)} \to 0 \quad \text{for} \quad k \to \infty \quad (6.58)$$

holds. Let furthermore exist a sequence $\{u^{\delta_k}\}_{k=1}^{\infty}$ such that for every $k$

$$u^{\delta_k}[i] - u^\dagger(i\Delta_k) \quad i = 1, \ldots, N$$

are i. i. d. normal variables with expectation zero and variance $\delta_k^2$. Given constants $0 < \underline{h} \leq a^\dagger(t)$ ($t \in I$) and $\overline{S}$ such that (6.57) is satisfied, for every $k$ the Algorithm 6.2.16 can be applied and computes a function $\tilde{S}^{\delta_k}$. Furthermore, we have

$$\mathbb{E} \left\| \tilde{S}^{\delta_k} - S^\dagger \right\|_{\mathcal{C}(I)} \to 0 \quad \text{for} \quad k \to \infty.$$
Proof: Let \( \varepsilon > 0 \). We define the events

\[
\Omega_-(k) := \left\{ \max_{i=1,\ldots,N_k} |u_i^{\delta_k} - u^\dagger(t_i)| > \hat{\delta}_k \right\}
\]

\[
\Omega_+(k) := \left\{ \max_{i=1,\ldots,N_k} |u_i^{\delta_k} - u^\dagger(t_i)| \leq \hat{\delta}_k \right\}.
\]

Using Lemma 6.2.15 we obtain

\[
\mathbb{P}(\Omega_-(k)) = \mathbb{P}\left(\left\{ \max_{i=1,\ldots,N_k} \frac{|u_i^{\delta_k} - u^\dagger(t_i)|}{\delta_k} > 2\sqrt{\ln(N_k)} \right\}\right) \to 0 \quad \text{for} \quad N_k \to \infty.
\]

Thus, there exists a value \( k_1 \) such that

\[
\mathbb{P}(\Omega_-(k)) \leq \frac{\varepsilon}{2S} \quad \forall k \geq k_1.
\]

Furthermore, on \( \Omega_+(k) \) the residuum \( \|N(\tilde{S}^{\delta_k}) - N(S^\dagger)\|_{\mathcal{C}(I)} \) is bounded by \( 2\hat{\delta}_k + C\Delta_k \), which tends to zero for \( k \to \infty \). Together with the continuity of the inverse operator this implies that there exists some value \( k_2 \) such that on the sets \( \Omega_+(k) \)

\[
\|\tilde{S}^{\delta_k} - S^\dagger\|_{\mathcal{C}(I)} \leq \frac{\varepsilon}{2} \quad \forall k \geq k_2
\]

holds.

Combining these results we obtain for \( k \geq \max(k_1, k_2) \)

\[
\mathbb{E}\left\|\tilde{S}^{\delta_k} - S^\dagger\right\|_{\mathcal{C}(I)} \leq \mathbb{P}(\Omega_-(k)) \frac{\varepsilon}{2} + \mathbb{P}(\Omega_+(k)) \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.
\]

\[\blacksquare\]
Chapter 7

Identification of the drift parameters

This section is concerned with the identification of the drift parameters in Model 3.0.26. We will concentrate mainly on the parameters $\gamma$, $\lambda$, $\beta$, $\sigma_X$ which describe the process $(q_t)_{t \geq 0}$. The deterministic drift $\mu$ and the parameter $d$, which realizes a scaling of the price process $(P_t)_{t \geq 0}$, are assumed to be known. In the first section we discuss several nonuniqueness phenomena. Especially, we show that the parameter $\sigma_X$ can be set without loss of generality to one, which corresponds to a scaling of the unobservable process $(X_t)_{t \geq 0}$.

In Section 7.2 we transform the Model 3.0.26 into state space representation and apply the Kalman filter (see [20], [36] and [39] for a general introduction). The Kalman filter does not only yield a prediction for the next time step but computes also conditional variances and covariances. These values are used in Section 7.3 to set up the likelihood function of the discrete observations. Given discrete observations we obtain maximum likelihood estimates for the searched parameters by maximising the likelihood function. Finally, the performance of this method is illustrated by means of a numerical case study.

7.1 Preliminary considerations

In this section we will see that the observation of the asset price process $(P_t)_{t \geq 0}$ or the logarithmic asset price process $(p_t)_{t \geq 0}$ does not suffice to identify all four parameters $\gamma$, $\lambda$, $\beta$, $\sigma_X$ in a unique way. Therefore, we can and will restrict the set of admissible parameter constellations in the parameter space.

A relatively evident restriction of the set of admissible parameters arises from the fact that we could replace the process $X$ by a scaled variant $\hat{X}_t := \frac{1}{\sigma_X} X_t$. (Here, it has to be taken in mind that $\sigma_X > 0$ was assumed in Model 3.0.26.) In this case the process $\hat{X}$ satisfies the stochastic differential equation

$$d\hat{X}_t = -\beta \hat{X}_t dt + 1 dW^X_t$$

with the initial condition $\hat{X}_0 = \frac{c_X}{\sigma_X}$. 

121
Furthermore, it is easy to see that the processes \((q_t)_{t \geq 0}\) and \((X_t)_{t \geq 0}\) satisfy the System (3.4) if and only if \((q_t)_{t \geq 0}\) and \((\hat{X}_t)_{t \geq 0}\) fulfill the system

\[
\begin{align*}
    dq_t &= -\left(\gamma q_t - \lambda \hat{X}_t\right)dt + \sigma(t)dW^q_t \\
    d\hat{X}_t &= -\beta \hat{X}_t dt + 1 dW^X_t
\end{align*}
\]

with initial conditions \(q_0 = c_q, \quad \hat{X}_0 = \frac{e^\gamma}{\sigma^2}\) and the parameter \(\lambda = \lambda \sigma_X\). As a result of these considerations we can assume \(\sigma_X = 1\) in the following. In other words, the parameters that have to be estimated reduces to \(\gamma, \lambda, \beta\).

In our first studies we considered the situation of a constant volatility function and tried to estimate the parameters \(\gamma, \lambda, \beta\) from the moments of the logarithmic returns \(r_\tau(t) := \frac{\log P_t}{\log P_{t-\tau}} = p_t - p_{t-\tau}\). To make life easier we assumed \(\gamma \neq \beta\) in this context. Throughout these studies we observed that for the special case of a constant volatility function there occurs another nonuniqueness phenomenon. It is interesting to see that this effect does in general not arise for a time-dependent volatility function.

In order to present these considerations we remark that for a constant volatility function \(\sigma\) the process \(\alpha = (q, X)^T\), which has been introduced in Section 3.1, is stationary, provided the initial value has the following normal distribution

\[
eq \left(\begin{array}{c} c_q \\ c_X \end{array}\right) \sim \mathcal{N} \left( \left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma^2 & \frac{2}{\beta+\gamma} \\ \frac{2}{\beta+\gamma} & 1 \end{array}\right)\right) \right).
\]

In this situation the variance and covariance of the returns \(r_\tau(t)\) satisfy \((k \geq 1)\)

\[
\mathbb{D}^2 r_\tau(t) = \frac{\lambda^2}{\beta^2 - \gamma^2} \beta - \gamma + \frac{\gamma e^{-\beta \tau} - \beta e^{-\gamma \tau}}{\beta \gamma} + \frac{\sigma^2(1 - e^{-\gamma \tau})}{\gamma} (7.3a)
\]

and

\[
\begin{align*}
\text{Cov} \left(r_\tau(t), r_\tau(t+k\tau)\right) &= \frac{\lambda^2}{\gamma^2 - \beta^2} \frac{(1 - e^{-\gamma \tau})^2 \beta e^{-\gamma(k-1)\tau} - (1 - e^{-\beta \tau})^2 \gamma e^{-\beta(k-1)\tau}}{2 \beta \gamma} \\
&\quad - \frac{1}{2} \frac{\sigma^2(1 - e^{\gamma \tau})^2}{\gamma} e^{-\gamma(k-1)\tau}, (7.3b)
\end{align*}
\]

provided \(\gamma \neq 0, \beta \neq 0\) and \(\gamma \neq \beta\). With respect to these limit cases we remark that the corresponding moments can be obtained by taking limits in (7.3). However, to make notation easier and to avoid further case differentiations, we will exclude them from the remaining considerations in this section.

Introducing the auxiliary variables

\[
\begin{align*}
    a &= \left(\sigma^2 - \frac{\lambda^2}{\gamma^2 - \beta^2}\right) \frac{1 - e^{-\gamma \tau}}{\gamma} \\
    g &= e^{-\gamma \tau} \\
    c &= \frac{\lambda^2}{\gamma^2 - \beta^2} \frac{1 - e^{-\beta \tau}}{\beta} \\
    b &= e^{-\beta \tau}
\end{align*}
\]

the variance and covariance of the returns can be written as

\[
\begin{align*}
\mathbb{D}^2 r_\tau(t) &= a + c \quad (7.5a) \\
\text{Cov} \left(r_\tau(t), r_\tau(t+k\tau)\right) &= \frac{1}{2} \left(a(g - 1)g^{k-1} + c(b - 1)b^{k-1}\right) \quad \text{for } k \geq 1. (7.5b)
\end{align*}
\]
Furthermore, the equations (7.5) show immediately that the values \(a, c, b\) and \(g\) are not uniquely determined by the moments of the returns. To be precise if \((a_1, c_1, b_1, g_1)\) solves the system (7.5), then there exists a second solution \((a_2, c_2, b_2, g_2)\) with

\[
    a_2 = c_1, \quad c_2 = a_1, \quad b_2 = g_1 \quad \text{and} \quad g_2 = b_1.
\]  

Additionally, given values \(b > 0\) and \(g > 0\) we can solve the equations (7.4b) uniquely for \(\beta\) and \(\gamma\). We obtain

\[
    \beta = -\frac{\ln b}{\tau} \quad \text{and} \quad \gamma = -\frac{\ln g}{\tau}.
\]

Together with (7.4a) we obtain the unique solutions

\[
    \lambda = \sqrt{c(\gamma^2 - \beta^2)} \quad \text{and} \quad \sigma = \sqrt{\frac{\gamma}{1 - e^{-\gamma \tau}} a + \frac{\beta}{1 - e^{-\beta \tau}} c}
\]

for \(\lambda\) and \(\sigma\), provided the terms under the square roots are nonnegative.

Combining these considerations we see that for every parameter constellation

**Par1**: \(\gamma_1 > 0, \quad \beta_1 > 0, \quad \lambda_1 \geq 0, \quad \sigma_1 > 0\) with \(\gamma \neq \beta\) as well as \(\lambda_1^2 - \sigma_1^2(\gamma_1^2 - \beta_1^2) \geq 0\)

there exists a second parameter constellation

**Par2**: \(\gamma_2 = \beta_1, \quad \beta_2 = \gamma_1, \quad \lambda_2 = \sqrt{\lambda_1^2 - \sigma_1^2(\gamma_1^2 - \beta_1^2)}\), \(\sigma_2 = \sigma_1\),

such that the corresponding moments of the returns (7.3) are equal, provided the initial values of the process \(\alpha\) are given by (7.2).

A concrete example for this phenomenon is

**Par1\***: \(\gamma_1 = 0.3748, \quad \beta_1 = 0.0106, \quad \lambda_1 = 0.004797, \quad \sigma_1 = 0.0074\)

**Par2\***: \(\gamma_2 = 0.0106, \quad \beta_2 = 0.3748, \quad \lambda_2 = 0.003914709, \quad \sigma_2 = 0.0074\).

This parameter situation corresponds to the empirical moments

\[
    D^2 r_\tau(t) = 7.225 \cdot 10^{-5}, \quad \rho_\tau(1) = 0.1838, \quad \rho_\tau(5) = 0.0323, \quad \text{and} \quad \rho_\tau(25) = -0.0092
\]

which have been estimated in [37, p. 107] from historical daily returns of the Center for Research in Security Prices (CRSP) value weighted market index from 1962 to 1990. We will use the parameter constellation Par1\* with a time-depending volatility varying around 0.0074 for our numerical case studies in Section 7.4.

In view of these observations it is interesting to consider the processes \(q^1\) and \(q^2\) which are defined by substituting the parameter situations Par1 and Par2 into Equation (3.8) with initial values (7.2). In [33] we have shown that these processes are indeed stochastically equivalent in the broad sense. This means that the considered model (with constant volatility and initial conditions (7.2)) is in general not identifiable.

Moreover, even for initial conditions which differ from (7.2) one should expect problems when trying to identify the parameters. Indeed, as it can be seen from the explicit solutions (3.8) and (3.9) the influence of the initial conditions decreases rapidly for increasing \(t\), provided \(\gamma > 0, \beta > 0\).
Finally, it should be remarked that the four equations (7.5a) and (7.5b) with $k = 1, \ldots, 3$ have no more than two solutions, provided the additional conditions $a \neq 0$ and $c \neq 0$ are imposed. Besides, these two solutions can be transformed into each other using (7.6). The case $c = 0$ corresponds to the situation $\lambda = 0$. Clearly, in this situation the state process $X$ does not influence the drift of $q$. Therefore, it is not possible to estimate the parameter $\beta$ in this situation. However, in this situation the parameter $\beta$ is uninteresting.

The case $a = 0$ corresponds to a special situation in which the parameters $\beta$ and $\sigma$ are uniquely identified by the equations (7.5a) and (7.5b). However, $\gamma$ and $\lambda$ can not be uniquely identified as only the relation $\lambda^2 \gamma^2 - \beta^2 = \sigma^2$ is known.

In the numerical case studies carried out in [33] it is illustrated that these nonuniqueness phenomena need not occur if the volatility is time-dependent. Here we give just a short motivation for this effect by considering a piecewise constant volatility function

$$
\sigma^*(t) := \begin{cases} 
\sigma_1 & t \leq T/2 \\
\sigma_2 & t > T/2 
\end{cases}
$$

with two positive values $\sigma_1 \neq \sigma_2$. Provided $T$ is sufficiently large we have for $t_1 = T/2$ and $t_2 = T$

$$
\mathbb{D}^2 r_\tau(t_1) \approx \frac{\lambda^2}{\beta - \gamma} - \frac{\beta - \gamma e^{-\beta \tau} - \beta e^{-\gamma \tau}}{\beta (\beta + \gamma) \gamma} + \frac{\sigma_i^2 (1 - e^{-\gamma \tau})}{\gamma} \quad (i = 1, 2)
$$

and therefore

$$
\mathbb{D}^2 r_\tau(t_1) - \mathbb{D}^2 r_\tau(t_2) \approx \frac{(\sigma_1^2 - \sigma_2^2)(1 - e^{-\gamma \tau})}{\gamma}.
$$

Note that the auxiliary function $h : \mathbb{R}^+ \to [0, \tau]$ with $h(\gamma) := \frac{1 - e^{-\gamma \tau}}{\gamma}$ is strictly monotonically increasing in $\gamma$, which implies that the inverse function $h^{-1}$ exists. Therefore, one can expect that in this situation also the parameter $\gamma$ can be identified uniquely. Substituting this value in (7.3) we obtain also $\beta$ and $\lambda$ in a unique way.

The last consideration shows especially that the discussed nonuniqueness can not be overcome by a restriction of the parameter space. In fact, let us assume that we have observed the process $(P_t)_{t \in [0, T/2]}$ which is described by Model 3.0.26 with the volatility $\sigma^*$ and the parameters Par2*. Assuming a restriction of the parameter space (e.g. $\gamma \geq \beta$) we would (as an idealisation) determine the parameters Par1*. As a further idealisation we could obtain $\sigma^*$ on the interval $[T/2, T]$ from option price data (cf. Chapter 6). However, these data together would not describe the process $(P_t)_{t \in [T/2, T]}$.

We finish this section by a short discussion about the parameters $\mu$ and $d$ in Model 3.0.26. Let us start by the special situation $\gamma = \lambda = 0$ for which Model 3.0.26 reduces to the Black-Scholes model, i.e. the logarithmic asset price process satisfies

$$
p_t = q_0 + d + \mu t + \int_0^t \sigma dW^q.
$$

This representation shows that in this specific situation the parameter $d$ is redundant as $(p_t)_{t \geq 0}$ could also be described by

$$
p_t = \tilde{q}_0 + \mu t + \int_0^t \sigma dW^q.
$$
with $\tilde{q}_0 := q_0 + d$. Note that even the restriction $E q_0 = 0$ would not suffice to guarantee that the parameter $d$ can be estimated from observation of one trajectory of $(p_t)_{t \geq 0}$. With respect to the deterministic drift $\mu$ it is well known that

$$\mu_N = \frac{1}{t_N} \sum_{i=1}^{t_N} p_{t_i} - p_{t_{i-1}} = \frac{p_{t_N} - p_0}{t_N} \tag{7.7}$$

is a consistent estimator for $\mu$ when considering the asymptotic $t_N \to \infty$.

Similar phenomena occur for the special situations $\gamma = 0$ or $\beta = 0$. Moreover, the situation $\gamma = \beta = 0$ and $\lambda > 0$ is even more difficult. In this situation the logarithmic asset price process $(p_t)$ satisfies (cf. (3.9))

$$p_t = c_q + d + \lambda t c_X + \mu t + \lambda \int_0^t (t - s) dW^X_s + \int_0^t \sigma(s) dW^q_s,$$

which shows that both, the parameter $d$ and $\mu$ are redundant as they can be incorporated in the distribution of $c_q$ and $c_X$, respectively.

On the other hand, it is clear that in general (for positive mean reversion parameters $\gamma > 0$ and $\beta > 0$) both parameters $d$ and $\mu$ are important to describe the process $(p_t)_{t \geq 0}$, see also Chapter 3 for an interpretation. To avoid the discussed difficulties with the estimation of $\mu$ and $d$ we will assume in Sections 7.2 and 7.3 that both parameters are known. However, in our numerical case studies we will also illustrate (for known $d$ and deterministic initial values $c_\alpha = 0$ of the process $\alpha$) the estimation of the parameter $\mu$ by

$$\hat{\mu} := \frac{2}{T(N+1)} \sum_{k=1}^{N} (p_{t_k} - d), \tag{7.8}$$

where we assume to have observations $p_{t_k}$ on an equispaced timegrid $t_k = k \tau$ with some stepwidth $\tau > 0$. Note that we have shown in [33] that this estimator has in general a better performance than the simple estimator $\mu_N$, provided $\gamma > 0$, $\beta > 0$ and the initial conditions are deterministic equal zero, i.e. $c_\alpha = 0$.

### 7.2 State space representation and Kalman filter

Using the representation (3.7) one gets immediately the following recursive representation of the values $\alpha_{t_k}$

$$\alpha_{t_k} = e^{t_k A} \alpha_{t_{k-1}} + \int_{t_{k-1}}^{t_k} e^{(t_k-s)A} \Sigma(s) dW_s. \tag{7.9}$$

In order to avoid case differentiations we will do the following computations only for the situation $\gamma \neq \beta$. We remark that the case $\gamma = \beta$ can be handled analogously. Introducing the notation $T := e^{t_k A}$ as well as $R := \begin{pmatrix} 1 & \lambda \\ \frac{\gamma}{\gamma - \beta} & 1 \end{pmatrix}$ and

$$\eta_k := \begin{pmatrix} \int_{t_{k-1}}^{t_k} \sigma(s) e^{-\gamma(t_k-s)} dW^q_s - \frac{\lambda}{\gamma - \beta} \int_{t_{k-1}}^{t_k} e^{-\gamma(t_k-s)} dW^X_s \\ \int_{t_{k-1}}^{t_k} e^{-\beta(t_k-s)} dW^X_s \end{pmatrix}, \quad k = 1, \ldots, N$$
we can rewrite (7.9) in form of
\[ \alpha_{t_k} = T \alpha_{t_{k-1}} + R \eta_{t_k}, \quad k = 1, \ldots, N. \]

The vectors \( \eta_{t_k} \) are independent, normally distributed random vectors. It holds
\[ \eta_{t_k} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, Q_{t_k} \right), \]
where \( Q_{t_k} := \mathbb{E} \eta_{t_k} \eta_{t_k}^T \) denotes the covariance matrix. Introducing the notation
\[ H(t_1, t_2) := \int_{t_1}^{t_2} \sigma^2(s) e^{-2\gamma(t_2-s)} \, ds \]
and the auxiliary function \( h_{\tau}(c) : [0, \infty) \to \mathbb{R} \), defined by
\[ h_{\tau}(c) := \int_0^\tau e^{-c(\tau-s)} \, ds = \begin{cases} \frac{1}{\tau} (1 - e^{-c\tau}) & c > 0 \\ 0 & c = 0 \end{cases}, \quad (7.10) \]
we can compute the matrices \( Q_{t_k} \) as follows
\[ Q_{t_k} = \begin{pmatrix} H(t_{k-1}, t_k) + \left( \frac{\lambda}{\gamma-\beta} \right)^2 h_{\tau}(2\gamma) & -\frac{\lambda}{\gamma-\beta} h_{\tau}(\beta + \gamma) \\ -\frac{\lambda}{\gamma-\beta} h_{\tau}(\beta + \gamma) & h_{\tau}(2\beta) \end{pmatrix}, \quad k = 1, \ldots, N. \quad (7.11) \]

To keep the notation consistent with the one introduced in [39, p. 625 ff.] and [33] we will denote the process \((p_t)_{t \geq 0}\) in the following by \((y_t)_{t \geq 0}\), i.e. we set \( y_t := p_t = q_t + \mu t + d \). Combining the above considerations and introducing the notation \( z := (1 0)^T \) as well as \( d_{t_k} := \mu t_k + d \) we can describe \( y_{t_k} \) as follows
\[ y_{t_k} = z^T \alpha_{t_k} + d_{t_k}, \quad k = 1, \ldots, N, \quad (7.12a) \]
where \( \alpha_{t_k} \) satisfies the equation
\[ \alpha_{t_k} = T \alpha_{t_{k-1}} + R \eta_{t_k}, \quad k = 1, \ldots, N. \quad (7.12b) \]

This representation is also referred to as state space model. As the process \((y_t)_{t \geq 0}\) can be observed (measured), the equation (7.12a) is called measurement equation. The values \( y_{t_k} \) depend on the states \( \alpha_{t_k} \) which can not be observed directly. The vector \( \alpha_{t_k} \) is called state vector and the equation (7.12b) transition equation.

To complete the state space model for the logarithmic price process \( p_t \) we need some additional assumptions on the initial conditions (cf. [20]):

**Assumption 7.2.1** The initial value \( \alpha_{t_0} = c_0 \) is normally distributed with expectation \( \tilde{a} \) and covariance matrix \( \tilde{P} \), i.e. it holds
\[ c_0 \sim N \left( \tilde{a}, \tilde{P} \right). \quad (7.13) \]
7.2. STATE SPACE REPRESENTATION AND KALMAN FILTER

**Definition 7.2.2**

Let \( y_0, y_1, \ldots, y_k \) be the values of the logarithmic asset price process \( y_t = p_t \) at the points \( t_k = k\tau \; k = 0, 1, \ldots, N \) for some positive step width \( \tau \). The smallest \( \sigma \)-algebra that is created by these random variables is denoted by \( \mathcal{F}_{t_k} \), i.e. it holds

\[
\mathcal{F}_{t_k} := \mathcal{A} \left( y_0, y_1, \ldots, y_k \right), \quad k = 0, 1, \ldots, N. \tag{7.14}
\]

For \( k, l = 0, 1, \ldots, N \) and \( l \leq k \) we define the conditional expectations and covariance matrices of \( \alpha_{t_k} \) by

\[
a_{t_k|t_l} := \mathbb{E} \left( \alpha_{t_k} \mid \mathcal{F}_{t_l} \right) \quad \text{and} \quad P_{t_k|t_l} := \mathbb{D}^2 \left( \alpha_{t_k} \mid \mathcal{F}_{t_l} \right) \tag{7.15a}
\]

\[
= \mathbb{E} \left[ \left( \alpha_{t_k} - \mathbb{E} \left( \alpha_{t_k} \mid \mathcal{F}_{t_l} \right) \right) \left( \alpha_{t_k} - \mathbb{E} \left( \alpha_{t_k} \mid \mathcal{F}_{t_l} \right) \right)^T \right] \tag{7.15b}
\]

For \( k = l \) we write simply \( a_{t_k} \) and \( P_{t_k} \).

Remember that the vectors \( \eta_k, \; k = 1, \ldots, N \) and the vector \( \alpha_0 = \alpha_\alpha \) are mutually independent. Using (7.12b) it is easy to show the following relations (see also [33])

\[
a_{t_k|t_{k-1}} = T \alpha_{t_{k-1}} \quad \text{and} \quad P_{t_k|t_{k-1}} = TP_{t_{k-1}}T^T + RQ_{t_k}R^T \quad \text{for} \; k = 1, \ldots, N. \tag{7.16}
\]

The equations (7.16) enable us to compute a prediction of the state \( \alpha_{t_k} \) at the basis of the data \( y_0, y_1, \ldots, y_{t_k-1} \). Therefore, they are called prediction equations. Clearly, the prediction \( a_{t_k|t_{k-1}} \) of the state \( \alpha_{t_k} \) gives the following prediction of the value \( y_{t_k} \) (\( k = 1, \ldots, N \))

\[
y_{t_k|t_{k-1}} := \mathbb{E} \left( y_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) \nonumber
\]

\[
= \mathbb{E} \left( z^T \alpha_{t_k} + \eta_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) = z^T \mathbb{E} \left( \alpha_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) + \eta_{t_k} \nonumber
\]

\[
= z^T a_{t_k|t_{k-1}} + \eta_{t_k}. \tag{7.17}
\]

The prediction error is defined by (\( k = 1, \ldots, N \))

\[
v_{t_k} := y_{t_k} - y_{t_k|t_{k-1}} 
onumber
\]

\[
= z^T \left( \alpha_{t_k} - a_{t_k|t_{k-1}} \right). \tag{7.18}
\]

Furthermore, for \( k = 1, \ldots, N \) the conditional expectation and the conditional variance of \( v_{t_k} \) under \( \mathcal{F}_{t_{k-1}} \) are obtained by

\[
\mathbb{E} \left( v_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) = \mathbb{E} \left( y_{t_k} - y_{t_k|t_{k-1}} \mid \mathcal{F}_{t_{k-1}} \right) \nonumber
\]

\[
= \mathbb{E} \left( y_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) - y_{t_k|t_{k-1}} = 0 \quad \text{and} \quad \tag{7.19}
\]

\[
f_{t_k} := \mathbb{D}^2 \left( v_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) \nonumber
\]

\[
= z^T \mathbb{E} \left[ \left( \alpha_{t_k} - a_{t_k|t_{k-1}} \right) \left( \alpha_{t_k} - a_{t_k|t_{k-1}} \right)^T \right] \mathbb{D} \left( \alpha_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) = z^T P_{t_k|t_{k-1}} z. \tag{7.21}
\]
We remark that we will use the values $v_{t_k}$ and $f_{t_k}$ later to set up the likelihood function.

As soon as a new value $y_{t_k}$ is observed we can update the estimate for $\alpha_{t_k}$. This is done by the so-called updating equations (cf. [20], [36])

$$a_{t_k} = a_{t_k|t_{k-1}} + P_{t_k|t_{k-1}} z \int_{t_{k-1}}^{t_k} (y_{t_k} - z^T a_{t_k|t_{k-1}} - d_{t_k})$$

$$= a_{t_k|t_{k-1}} + P_{t_k|t_{k-1}} z \int_{t_{k-1}}^{t_k} v_{t_k}$$

for $k = 1, \ldots, N$ (7.22)

and

$$P_{t_k} = P_{t_k|t_{k-1}} - P_{t_k|t_{k-1}} z \int_{t_{k-1}}^{t_k} z^T P_{t_k|t_{k-1}}$$

for $k = 1, \ldots, N$. (7.23)

The initialisation of the Kalman filter is done as follows. As input we take the expectation $\bar{a}$ and the covariance matrix $\bar{P}$ of the initial value $c_\alpha$ cf. (7.13). After the first measurement at the time point $t_0$ we compute $a_{t_0}$ and $P_{t_0}$ by

$$a_{t_0} := \mathbb{E} (\alpha_{t_0} | \mathcal{F}_{t_0}) = \bar{a} + \bar{P} z \int_0^{t_0} (y_{t_0} - z^T \bar{a} - d)$$

(7.24a)

$$P_{t_0} := \mathbb{D}^2 (\alpha_{t_0} | \mathcal{F}_{t_0}) = \bar{P} - \bar{P} z \int_0^{t_0} z^T \bar{P}$$

(7.24b)

with $\int := z^T \bar{P} z$ (cf. also [36, p. 73]). After the initialisation the prediction and updating steps are performed in turn for $k = 1, \ldots, N$.

Remark 7.2.3 In the computation of $P_{t_k|t_{k-1}}$ ($k = 1, \ldots, N$) the product $RQ_{t_k} R^T$ occurs frequently. For the numerical evaluation it should be taken in mind that the term $\frac{1}{\gamma - \beta}$ in $R$ gets quite large for $\gamma \approx \beta$. As this can lead to numerical difficulties it is advisable to set up the matrix $\tilde{Q} := RQ_{t_k} R^T$ immediately. The entries of $\tilde{Q}_{t_k}$ are given by

$$\tilde{Q}_{t_k}(1,1) = H(t_{k-1}, t_k) + \left( \frac{\lambda}{\gamma - \beta} \right)^2 \int_0^\tau (e^{-\gamma(s-t)} - e^{-\beta(s-t)})^2 \, ds$$

$$\tilde{Q}_{t_k}(2,1) = \tilde{Q}_{t_k}(1,2) = \frac{\lambda}{\gamma - \beta} \int_0^\tau e^{-\beta(s-t)} (e^{-\beta(s-t)} - e^{-\gamma(s-t)}) \, ds$$

$$\tilde{Q}_{t_k}(2,2) = \int_0^\tau e^{-2\beta(s-t)} \, ds.$$ 

Because of $\lim_{\gamma \to \beta} \frac{e^{-\gamma(s-t)} - e^{-\beta(s-t)}}{\gamma - \beta} = (s - \tau)e^{-\beta(s-t)}$ these terms are bounded.

Furthermore, we remark that the matrix $Q_{t_k}$ is symmetric and positive definite (cf. [33]). Therefore, it allows a Cholesky decomposition $Q_{t_k} = L_{t_k} L_{t_k}^T$, which shows that (7.12b) could also be written as

$$\alpha_{t_k} = T \alpha_{t_{k-1}} + \tilde{R}_{t_k} \tilde{\eta}_{t_k}$$

with $\tilde{R}_{t_k} := RL_{t_k}$ and independent random vectors $\tilde{\eta}_{t_k} \sim \mathcal{N}(0,I)$. 

128 7. IDENTIFICATION OF THE DRIFT PARAMETERS
7.3 Likelihood function

We assumed that the initial value \( \alpha_{t_0} = c_\alpha \) of the process \( \alpha \) is normally distributed with expectation \( \tilde{\alpha} \) and covariance matrix \( \tilde{P} \). Assuming these values to be known and applying the theorem of normal correlation (cf. [36, p. 61]) one can set up the density function for the observation \( \tilde{y} := (y_{t_0}, \ldots, y_{t_N}) \) under the parameter vector \( \tilde{\varphi} := (\gamma, \lambda, \beta)^T \). We will sketch this process in the following. For further details we refer to [33].

For simplicity we assume the covariance matrix of the initial values \( \tilde{P} \) to be nonsingular. For a discussion of the situation with constant initial conditions we refer to [33].

Let
g_{y_{t_0},\ldots,y_{t_k}}(\cdot,\tilde{\varphi}), \quad k = 0, \ldots, N
denote the density of the vector \((y_{t_0},\ldots,y_{t_k})^T\) under the parameter vector \(\tilde{\varphi}\) and
g_{y_{t_k}}(\cdot,\tilde{\varphi} | y_{t_0} = x_0, \ldots, y_{t_{k-1}} = x_{k-1}), \quad k = 1, \ldots, N.
the conditional density of \(y_{t_k}\) given \(y_{t_0} = x_0, \ldots, y_{t_{k-1}} = x_{k-1}\) under the parameter vector \(\tilde{\varphi}\).

Using the notation
g_{y_{t_k}}(x_k,\tilde{\varphi} | x_0, \ldots, x_{k-1}) := g_{y_{t_k}}(x_k,\tilde{\varphi} | y_{t_0} = x_0, \ldots, y_{t_{k-1}} = x_{k-1}), \quad k = 1, \ldots, N
as well as
\[ m(x_0, \ldots, x_{k-1}) := \mathbb{E} \left( y_{t_k} \mid y_{t_0} = x_0, \ldots, y_{t_{k-1}} = x_{k-1} \right) \]
\[ s^2(x_0, \ldots, x_{k-1}) := \mathbb{D}^2 \left( y_{t_k} \mid y_{t_0} = x_0, \ldots, y_{t_{k-1}} = x_{k-1} \right) \]
for \(k = 1, \ldots, N\) we can write the densities \(g_{y_{t_k}}(x_k,\tilde{\varphi} | x_0, \ldots, x_{k-1})\) for \(k = 1, \ldots, N\) as follows
\[ g_{y_{t_k}}(x_k,\tilde{\varphi} | x_0, \ldots, x_{k-1}) = \frac{1}{\sqrt{2\pi s(x_0, \ldots, x_{k-1})}} \cdot \exp \left( -\frac{(x_k - m(x_0, \ldots, x_{k-1}))^2}{2s^2(x_0, \ldots, x_{k-1})} \right). \quad (7.25) \]

Furthermore, it holds with \( \tilde{f} = z^T \tilde{P} z = \mathbb{D}^2 y_0 \) and \( \tilde{a}(1) = z^T \tilde{a} = \mathbb{E} y_0 - d \)
\[ g_{y_{t_0}}(x_0,\tilde{\varphi}) = \frac{1}{\sqrt{2\pi \tilde{f}}} \cdot \exp \left( -\frac{(x_0 - \tilde{a}(1) - d)^2}{2\tilde{f}} \right). \]

Using the theorem of normal correlation (cf. [36, p. 61]) and the relation
\[ g_{y_{t_k}}(x_k,\tilde{\varphi} | x_0, \ldots, x_{k-1}) = \frac{g_{y_{t_0},\ldots,y_{t_k}}(x_0, \ldots, x_k,\tilde{\varphi})}{g_{y_{t_0},\ldots,y_{t_{k-1}}}(x_0, \ldots, x_{k-1},\tilde{\varphi})}, \quad k = 1, \ldots, N \quad (7.26) \]
we obtain the density \(g_{y_{t_0},\ldots,y_{t_N}}(x_0, \ldots, x_N,\tilde{\varphi})\) as
\[ g_{y_{t_0},\ldots,y_{t_N}}(x_0, \ldots, x_N,\tilde{\varphi}) = g_{y_{t_0}}(x_0,\tilde{\varphi}) \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi s(x_0, \ldots, x_{k-1})}} \cdot \exp \left( -\frac{(x_k - m(x_0, \ldots, x_{k-1}))^2}{2s^2(x_0, \ldots, x_{k-1})} \right). \]
In general, conditional expectations and conditional covariances are stochastic variables. For the fixed elementary event \( \omega_0 \) which produced an observed trajectory \( y^* \) the conditional expectations and conditional covariances have a fixed value. Furthermore, in our situation these values \( m(y^*_{t_0}, y^*_{t_1}, \ldots, y^*_{t_k}) \) and \( s^2(y^*_{t_0}, y^*_{t_1}, \ldots, y^*_{t_k}) \) \( (k = 1, \ldots, N) \) can be computed as follows (cf. Section 7.2)

\[
m(y^*_{t_0}, y^*_{t_1}, \ldots, y^*_{t_k}) = \mathbb{E} \left( y_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) (\omega_0) = y_{t_k \mid t_{k-1}}(\omega_0)
\]

\[
=: y^*_{t_k \mid t_{k-1}}
\]  

and

\[
s^2(y^*_{t_0}, y^*_{t_1}, \ldots, y^*_{t_k}) = \mathbb{E} \left[ (y_{t_k} - y_{t_k \mid t_{k-1}})^2 \mid \mathcal{F}_{t_{k-1}} \right] (\omega_0)
\]

\[
= \mathbb{E} \left( v^2_{t_k} \mid \mathcal{F}_{t_{k-1}} \right) (\omega_0) = f^*_{t_k}(\omega_0)
\]

\[
=: f^*_{t_k}.
\]  

Here we have denoted the values of \( y^* \) at the time points \( t_k \) by \( y^*_{t_0}, y^*_{t_1}, \ldots, y^*_{t_N} \).

**Remark 7.3.1** In the setting considered here (with gaussian random vectors) the conditional covariance matrices \( P_{t_k} \) and \( P_{t_k \mid t_{k-1}} \) (and therefore also \( f_{t_k} \)) are deterministic. Indeed, Formulas (7.24b), (7.16) and (7.23) show that these matrices depend only on the timepoint \( t_k \), the parameters \( \gamma, \lambda, \beta \), the covariance matrix \( \bar{P} \) of the initial condition \( c_\alpha \), and the volatility function \( \sigma \), but not on the concrete trajectory \( \omega_0 \). In our situation we could therefore omit the * in \( f^*_{t_k} \).

Together with the notation \( v^*_{t_k} = y^*_{t_k} - y^*_{t_k \mid t_{k-1}} \) we obtain the following representation of the likelihood function.

**Corollary 7.3.2**

The likelihood function of the observed values \( \bar{y}^* = (y^*_{t_0}, y^*_{t_1}, \ldots, y^*_{t_N})^T \) of the trajectory \( y^* \) under the parameter vector \( \bar{p} \) satisfies

\[
L(\bar{y}^* \mid \bar{p}) = g_{y|\bar{p}}(y^*_{t_0}, \bar{p}) \prod_{k=1}^N \frac{1}{\sqrt{2\pi f^*_{t_k}}} \cdot \exp \left( \frac{-v^2_{t_k}}{2f^*_{t_k}} \right).
\]  

The values \( v^*_{t_k} \) and \( f^*_{t_k} \) are obtained directly from the Kalman filter recursions. As \( v^*_{t_k} \) denotes the prediction error of \( y^*_{t_k} \) and \( f^*_{t_k} \) the variance of \( v^*_{t_k} \), this decomposition is also called prediction error decomposition form.

Taking on both sides the logarithm we obtain the log-likelihood function \( \ln L(\bar{y}^* \mid \bar{p}) \). As the logarithm is a strictly monotonically increasing function maximising the likelihood function is equivalent to the following problem

\[
\ln L(\bar{y}^* \mid \bar{p}) = \ln g_{y|\bar{p}}(y^*_{t_0}, \bar{p}) - \frac{N}{2} \ln (2\pi) - \frac{1}{2} \sum_{k=1}^N \ln f^*_{t_k} - \frac{1}{2} \sum_{k=1}^N f^*_{t_k}^{-1} v^2_{t_k} \rightarrow \max_{\bar{p} \in \mathcal{D}}.
\]  

(7.30)
Assuming that the distribution function \( g_{y_0}(\cdot, \vec{p}) \) does not contain unknown parameters we can neglect the constant term \( \ln \left( g_{y_{t_0}}(y_{t_0}, \vec{p}) \right) - \frac{N}{2} \ln (2\pi) \) and multiply both sides by \((-2)\). This yields the following optimisation problem which is equivalent to maximising the likelihood function

\[
\sum_{k=1}^{N} \ln f^*_{t_k} + \sum_{k=1}^{N} f^*_{t_k} v^*_{t_k}^2 \rightarrow \min_{\vec{p} \in D}.
\]  
(7.31)

In the next section we will use the procedure \texttt{fmisearchbnd} in MATLAB to solve this minimization problem numerically.

### 7.4 Numerical Case Studies

In this section we want to give a short illustration of the performance of maximum likelihood estimation. Here, we are especially interested in the following questions.

1. Which observation period is necessary to obtain acceptable results?
2. Do small errors in the specified volatility (which is used as input of the procedure) cause large errors in the estimates?

For further investigations we refer to [33].

For our numerical case studies we set the initial values \( q_0 \) and \( X_0 \) to zero, set \( d := 0 \) and used a piecewise constant volatility function \( \sigma \) which is given by

\[
\sigma(t) = \begin{cases} 
0.0050 & t \in [0, 0.2T) \\
0.0098 & t \in [0.2T, 0.4T) \\
0.0074 & t \in [0.4T, 0.6T) \\
0.0084 & t \in [0.6T, 0.8T) \\
0.0064 & t \in [0.8T, T].
\end{cases}
\]  
(7.32)

See also Figure 7.1 for an illustration.

The value \( d \) is assumed to be known, the parameter \( \mu \) has been estimated in a first step out of the data using the estimator (7.8). In a second step we estimate the remaining parameters by maximum likelihood estimation.

Using the parameters \( \mu^\dagger = 1.6483 \cdot 10^{-4} \) and \( \vec{p}^\dagger = (0.3748, 0.004797, 0.0106)^T \) we simulated 100 trajectories on a daily time grid, i.e. \( \tau = 1 \) for several time intervals \([0, T]\). After that the parameter \( \mu \) has been estimated using the estimator \( \hat{\mu} \) defined in (7.8). Finally, we used the estimate \( \hat{\mu} \) to estimate \( \vec{p} \) via (7.31), where we used the iterative minimization procedure \texttt{fmisearchbnd} in MATLAB to solve this optimisation problem numerically. The iterations started at the point

\[
\vec{p}_0 := (\gamma_0, \lambda_0, \beta_0)^T = (0.5, 0.002, 0.03)^T.
\]
To answer the first question we computed the percentage of trajectories for which the relative error of the maximum likelihood estimate was less or equal than 20%. The results are given in Table 7.1. Although a time interval of 1000 days (approximately three years) seems already to be quite long we see that the results were very poor. Clearly, for an increasing observation period the results get better.

One could now argue whether the aim of achieving a relative error of 20% with a high probability is too ambiguous. It is therefore also interesting to consider charts as they are given in Figure 7.2 and Figure 7.3. Every blue line corresponds to one estimate. The exact parameters and the upper and lower 5% quantiles are indicated by a bold red line and dashed red lines, respectively.

In Figure 7.2, which corresponds to an observation period of $T = 1000$ days, the estimates are rather spread out, whereas they are clustered around the exact parameters in Figure 7.3. Therefore, these plots confirm the thesis that an observation period of 1000 days is too short to get reliable estimates. In this context it should be noted that for an observation period of 1000 days also the estimates for the parameter $\mu$ are not reliable as can be seen from the fourth box in Figure 7.2.

We are now addressing the second question. As the volatility $\sigma$ which is used as input for the maximum likelihood estimation is itself obtained by statistical methods (cf. Chapter 4) it is interesting to see which effects are caused by an error in the volatility input.
In our study we used a noisy function
\[ a^\delta(t) := a(t)(1 + \delta). \]  
(7.33)

for the maximum likelihood estimation. In this context \( \delta \in (-1, 1) \) denotes a real-valued number which needs not necessarily be positive. In fact a positive value \( \delta \) means that the input is too large, whereas a negative \( \delta \) corresponds to the situation where the volatility input is too small.

Analogously, we define the noisy volatility
\[ \sigma^\delta(t) := \sigma(t) \sqrt{1 + \delta} = \sqrt{a^\delta(t)}. \]  
(7.34)

In other words, \( \delta \) can be interpreted as relative error of the squared volatility function \( a \). It should be noted that we studied the cases \( \delta \geq 0 \) and \( \delta \leq 0 \) separately. With respect to independent errors on certain intervals one can assume that the errors are not larger than those in the situations considered here.

Again, we simulated 100 trajectories over a time interval of 10000 days and compared the maximum likelihood estimates for several volatility inputs \( \sigma^\delta \). The Figures 7.4-7.7 present the results for different \( \delta \leq 0 \). As the Figures 7.4 and 7.5 are nearly identical we can conclude that for a small relative noise level \( |\delta| \) the effects on the precision of the estimates are neglectable. However, for \( \delta = -0.05 \) (cf. Figure 7.6) we see that the estimates \( \hat{\gamma} \) and \( \hat{\lambda} \) obtained with the noisy volatility input \( \sigma^\delta \) tend to be too large. For a larger perturbation of the volatility this effect is even more pronounced (cf. Figure 7.7). The parameter \( \beta \) seems to be relatively stable with respect to errors in the volatility input. However, it should be taken in mind that the estimates \( \hat{\beta} \) are rather spread out so that the slight trend \( \hat{\beta} < \beta^\dag \) does not strike the eye.

With respect to the parameter \( \mu \) it should be remarked that the estimator \( \hat{\mu} \) does not require any specification of the volatility function, therefore a misspecification has no influence on the performance of \( \hat{\mu} \).

Finally, we remark that we observed an inverse effect in the case \( \delta > 0 \), inasmuch as the estimates \( \hat{\gamma} \) and \( \hat{\lambda} \) tended to be too small in this situation. As expected, the errors increased for increasing \( |\delta| \). However, small noise in the volatility input had little effect on the precision of the maximum likelihood estimates.
7. IDENTIFICATION OF THE DRIFT PARAMETERS

Figure 7.4: Results for the situation $\delta = 0$

Figure 7.5: Results for the situation $\delta = -0.01$

Figure 7.6: Results for the situation $\delta = -0.05$

Figure 7.7: Results for the situation $\delta = -0.1$
Bibliography


Theses for the dissertation

“Identification in Financial Models with Time-Dependent Volatility and Stochastic Drift Components”,
Dipl.-Math. Romy Krämer,
Chemnitz University of Technology, Faculty of Mathematics

1. This dissertation is concerned with parameter estimation in financial models with time-dependent volatility and stochastic drift components. The intention is not to calibrate the considered models to real market data but to perform a theoretical analysis of the chances and limitations of several calibration techniques. The focus is on market models with incomplete information in which the stochastic drift of the logarithmic asset price process depends on some unobservable state process.

As a toy example we consider a slightly modified version of the Bivariate Trending Ornstein-Uhlenbeck model introduced by Lo and Wang. In this example the logarithmic asset price process \( p \) is characterised by a time-dependent volatility function and a stochastic drift component. The stochastic drift depends on the logarithmic price process \( p \) itself and on an unobservable state process \( X \). The aim is to calibrate both, the time-dependent volatility function (or deviated quantities) and the remaining real-valued parameters which characterise the process \( X \) and the stochastic drift.

2. The literature proposes several methods for the identification of a time-dependent volatility function. On the one hand, methods of nonparametric statistics, such as wavelet projection or kernel estimates, are used to estimate the volatility from high-frequency asset price data. This approach is applicable for the identification of the volatility function over a time-interval in the past.

On the other hand recent papers have been concerned with the so-called inverse problem of option pricing, i.e. the identification of the squared volatility function from observed option prices. In contrast to the first approach this method is especially suitable for the identification of the volatility function on a future time interval. However, the corresponding inverse problem is ill-posed, which means that the solution does not depend stably on the data. Therefore, the application of a regularization approach is necessary.

With regard to the identification of a finite number of real-valued parameters the ansatz of maximum-likelihood estimation is well-known. However, this method needs the time-dependent volatility as input. Unfortunately, in our situation this function is not known exactly but can only be estimated by the above-mentioned methods. Therefore, we have to address the question which effects are caused by small errors in the volatility input. To estimate all parameters in the considered model we combine the three discussed approaches.

3. The generalised Ornstein-Uhlenbeck model, which is used as toy example for our numerical case studies, is formulated in form of a stochastic differential equation describing the logarithmic price process. Furthermore, the unique solution of this stochastic differential equation is computed. Besides, we review results stating that the option
price formula derived in the Black Scholes setting (extended to the situation of a time-dependent volatility) also holds true if the price of the underlying asset follows the generalised Ornstein-Uhlenbeck model.

4. With respect to the identification of the time-dependent volatility function from high-frequency asset price data we consider an estimator that is based on a projection on an orthonormal wavelet basis. To handle models with a stochastic drift term that depends on an unobservable state process it was necessary to generalise the existent theory. Under appropriate conditions on the stochastic drift, the wavelet basis and the volatility function we prove weak convergence of the estimator. Moreover, a convergence rate $2^{-\frac{2m}{m+1}} n$ for the mean integrated square error is proven. In this context, $n$ specifies the number of observations and $m \geq 1$ denotes the Sobolev index of the squared volatility function. To be precise, the distance between two subsequent observations is $\Delta_n = 2^{-n}$ and a sufficient condition on the wavelet basis is that the scaling function and the mother wavelet are both continuously supported and $r$-times continuously differentiable with $1 \leq m \leq r$.

Furthermore, the proven convergence rate is illustrated by means of a numerical case study. Besides, the data-driven choice of the resolution level by the L-method is discussed. Although this method yields often very good results we give an example of a volatility function for which the method fails in the sense that the corresponding estimator does not converge to the exact volatility function.

5. For the analysis of inverse problems arising in option pricing the concept of Nemytskii operators is of great importance. Therefore, this concept is defined and results concerning acting conditions and continuity are reviewed. Although Nemytskii operators have a relatively simple structure many questions about the existence and properties of the corresponding inverse operators are still open. A part of this thesis contributes to close this gap. In this context, the focus is on Nemytskii operators generated by a monotone function. We consider two settings. In the first one the operator maps between $C$-spaces of continuous functions and in the second it is considered as operator between $L^p$-spaces of Lebesgue-integrable functions.

Assuming these conditions we show that the inverse of a Nemytskii operator exists and is again of Nemytskii type. Furthermore, for the $C$-space setting it is proven that the inverse operator is continuous and a sufficient condition for Lipschitz continuity is formulated. With regard to the $L^p$-setting a simple example shows that in this situation the inverse operator is not necessarily continuous. Therefore, we derive a sufficient condition which characterises the growth of the generating function and guarantees the continuity of the inverse operator.

6. The inverse problem of option pricing can be formulated as an operator equation

$$F(a) = u \quad (a \in D(F) \subset L^2_+(0, T), u \in L^2_+(0, T)).$$

Here, the forward operator $F$ describes the mapping of the squared volatility function $a(t) = \sigma^2(t) \ (t \in [0, T])$ to the fair option prices $u$. $D(F)$ denotes the domain of $F$ and $L^2_+(0, T)$ is the set of all nonnegative square integrable functions over the interval $[0, T]$. The forward operator $F$ can be decomposed as $F = N \circ J$ with an inner integral operator.
$J$ and an outer Nemytskii operator $N$. After reviewing results about the ill-posedness of the problem and the applicability of Tikhonov regularization we concentrate on the outer problem

$$N(S) = u \quad (S \in D(N) \subset B_1, u \in B_2).$$

Applying the results concerning Nemytskii operators discussed above we prove that this outer problem is well-posed in the setting (a) $B_1 = B_2 = C[0,T]$, whereas it is ill-posed in the setting (b) $B_1 = B_2 = L^2(0,T)$. For the setting (b) the problem can be regularized by restricting the domain of $F$ to the set $D^+_\kappa$ containing nonnegative uniformly bounded functions.

For the setting (a) we illustrate occurring ill-conditioning effects which lead to highly oscillating solutions and a delayed convergence. In order to stabilise the solution process we propose a numerically effective algorithm which computes a strictly monotonically increasing solution and prove convergence of the method. Extensions to the situation of discrete data and discrete data with independent stochastic noise as well as a numerical case study illustrating the performance of the algorithm are presented.

7. With respect to the identification of the real-valued parameters we discuss several nonuniqueness phenomena. Replacing the underlying state process $X$ by a scaled version the constant volatility $\sigma_X$ of the (scaled) state process can be set without loss of generality to $\sigma_X := 1$. Besides, for a constant volatility of the price process there occur further nonuniqueness phenomena. However, these phenomena do in general not occur in the extended model with a time-dependent volatility.

For the estimation of the real-valued parameters we suggest maximum likelihood estimation. For deriving the likelihood function we transform the considered model into state space formulation. Then the conditional expectations which enter the likelihood function are obtained by Kalman filter recursions. Some numerical case studies illustrate the performance of the method. On the one hand quite long observation periods are necessary to obtain reliable estimates. On the other hand small errors in the volatility input had little effects on the precision of the maximum likelihood estimates.
Erklärung

Ich erkläre an Eides Statt, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angeschafft habe.

Chemnitz, den 21.02.2007