Smoluchowski-Kramers Approximation for Stochastic Differential Equations and Applications in Behavioral Finance

Diplomarbeit
Studiengang Wirtschaftsmathematik

Ramona Westermann

Fakultäten für
Mathematik und Wirtschaftswissenschaften
Universität Bielefeld

07.06.2006

Erstprüfer:	Prof. Dr. Michael Röckner
Zweitprüfer:	Prof. Dr. Walter Trockel

Contents

Chapter 1

Introduction

The mathematical part of this thesis is based on the paper [11], which was written by Mark Freidlin and published in 2004. The central aim and idea of this thesis is to apply the results of [11] to financial markets. Hence in this thesis we first study [11] in detail and present the mathematical theory. Then we apply this theory to financial markets by setting up a financial model and then use the results shown in [11] in the context of our model. We find the possibilities to apply results from mathematics and physics to finance quite fascinating.

We emphasize that this thesis is addressed to mathematicians as well as to people from finance. Hence depending on the academic or practical background, on the one hand, the different parts of this thesis (mathematical vs. financial) might offer new and unknown aspects, but, on the other hand, there might be results and facts which are already well-known to the reader. This is inevitable in order to make this thesis convenient to read and well understandable to both (mathematicians and finance people).

This thesis is structured as follows: In Chapter 2 we shortly equip the reader with the mathematical basis and tools in order to be able to understand the following chapters well. This chapter is given for convenience of the reader, who might not have studied stochastic analysis extensively. Mathematicians might already be familiar with the results presented in Chapter 2. In Chapter 3 we study convergence properties of selected stochastic differential equations based on [11]. Chapter 4 represents an application in finance of the mathematical results of [11], which were as far as we know only applied to physics so far. Thus this chapter presents the main result of this thesis, since our motivation was to apply [11] to finance. We think that our considerations open up new directions for further research, hence we briefly describe possible future topics at the end of Chapter 4.

We shortly summarize the contents of the single chapters of this thesis. In Chapter 2 we present the Stratonovich stochastic integral and focus in particular on intuition, heuristics and comparison with the Ito stochastic integral. After that we give a short overview over (systems of) stochastic differential equations and finally we state (mainly technical) results of interest for this thesis. For mathematicians this chapter can be seen as a short review or reference of the needed facts in stochastic analysis; for people from finance this chapter can be seen as a short overview and conclusion of important facts from stochastic analysis, with the main attempt to give some intuition and heuristics to the needed mathematics.

Chapter 3 states and proves the main mathematical result presented in this thesis. This part of the thesis is based on [11]. We also included some background and worked out the details of Freidlin's paper [11].
We shortly give the mathematical content and physical motivation for this chapter: We consider the motion of a small particle in a force field. According to Newton's law and our assumptions the motion of this particle can be described by the following system of stochastic differential equations:

$$\begin{aligned} dX_t^{\mu,\delta} &= Y_t^{\mu,\delta} dt \\ \mu dY_t^{\mu,\delta} &= b(t, X_t^{\mu,\delta})dt + \sigma(t, X_t^{\mu,\delta})dV_t^{\delta} - dX_t^{\mu,\delta}, \\ X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2, \end{aligned} \tag{1.1}$$

where the disturbing noise V_t^{δ} satisfies certain conditions and converges to the Wiener process. We are interested in the behavior of the solution of the above system $X_t^{\mu,\delta}$ as μ and δ go to zero (i.e. the mass of the particle becomes very small and the disturbing noise converges to the Wiener process). It turns out that the limit of $X_t^{\mu,\delta}$ depends on the way μ and δ approach zero with respect to each other.
A special case of our main result will be that $X_t^{\mu,\delta}$ converges to X_t if first $\delta \searrow 0$ and then $\mu \searrow 0$ and to $\tilde{X}_t$ if it is vice versa. Here X_t $[\tilde{X}_t]$ is the solution of a first order stochastic differential equation $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ $\left[d\tilde{X}_t = b(\tilde{X}_t)dt + \sigma(\tilde{X}_t) \circ dW_t\right]$, where the stochastic integral is interpreted as an Ito [Stratonovich] stochastic integral. First of all, for people not so

familiar with such issues it might be a surprising result that we have convergence to the much simpler first order stochastic differential equation. Especially in physics, this result is of practical use: We can describe the motion of the particle by one of the much simpler first order stochastic differential equations and are still consistent with Newton's law!
Moreover, for non-mathematicians, it could - on the first glance - be surprising that the limits differ if we change the way the limits are taken. But it turns out that with a good intuition und understanding of the underlying problem it is not that surprising any more: The key is the fact that if we consider a smoother (i.e. a piecewise differentiable noise) V_t^δ, we can write the according stochastic differential equation as an ordinary (meaning non-stochastic) differential equation. Now, on the one hand, consider the first case: We first fix δ and let $\mu \searrow 0$. We will see that even in the limit for $\mu = 0$ we continue to consider an ordinary non-stochastic differential equation. Now we take into account the fact that ordinary differential equations under certain conditions converge to stochastic differential equations, where the stochastic integral is interpreted as a Stratonovich stochastic integral (first shown in [26], and well-known to people with a mathematical background) and we have established the first part at least heuristically. Now, on the other hand, consider the second case: We fix μ and let $\delta \searrow 0$, i.e. we consider the limit with our disturbing noise converging to a Wiener process. Due to the nature of the Wiener process the limiting differential equation is a stochastic one (and not an ordinary one as in the first case). If we now let additionally $\mu \searrow 0$, we still continue to have a stochastic differential equation. The important fact here is that the convergence properties of stochastic and ordinary differential equations are different and in this case it turns out that the difference takes place in the way the stochastic integral is interpreted. Keeping in mind this fact, our main result is not as surprising as it possibly seems to be to non-mathematicians on their first glance. We will even show in the next chapter that these results are applicable to more than 'only' physics (namely, behavioral finance) and identify parallels of physics and finance.

In Chapter 4 we first present an extended version of the Black-Scholes model, which we will call *momentum model*. The motivation for introducing our momentum model will be justified mainly by research and observations from the field of behavioral finance. We emphasize the fact that the idea to use the *momentum* (to be defined in the chapter) as an explanatory variable in finance is not new, since researchers already took it into account as a factor for explaining (at least partially) phenomena on financial markets. Moreover,

there are also firms using momentum trading in reality. But to the best of our knowledge a model including the momentum has not been formulated mathematically, in particular not in context with the Black-Scholes model. This thesis presents such an attempt.
In order to come up with such a mathematical model in behavioral finance we include several new factors in our model, which will be the so-called *momentum*, the *speed of spread of information* and the *liquidity of the financial product*. In order to stress the importance of the momentum we call the whole model *momentum model.* By taking into account these factors we try to adjust the Black-Scholes model better for reality, considering also the possibility of non-rationalizability of investors and market inefficiencies.
We will see that if the additional information (gained from the momentum) is executable (depending on the speed of flow of information or the liquidity of a financial product) the investor can expect an extra gain from this information and hence the Black-Scholes model is not appropriate for this situation. We calculate the resulting model and obtain a positive correction term (compared to the Black-Scholes model). But if we are in the situation that the additional information (gained from the momentum) is either not sufficiently available (e.g. due to a too slow speed in the spread of information) or not sufficiently executable (e.g. due to a too low liquidity of the considered financial product) or both, there will consequently be no extra gain for the investor and hence we obtain convergence to the Black-Scholes model in this case. Hence by the first situation we detect limits of the Black-Scholes model (in the sense of not being able to capture gains from momentum trading in the above situation), but for the second situation we obtain a newly motivated justification for the Black-Scholes model (since the Black-Scholes model seems to be - despite of the inclusion of new factors - still an appropriate model for this situation). These results will be gained from the application of our mathematical convergence results for stochastic differential equations.
Since we are interested in modeling, we will not be concerned with pricing formulae or the calculation of pricing results, as the reader might be used to. Our access to financial markets emphasizes the importance of the underlying model, not (yet) the calculations based on this model. Calculating results might be done later and would be beyond the scope of this thesis.
We conclude that we were in our opinion successful to introduce and establish a new model considering psychological factors as well as market inefficiencies such as momentum or speed of the flow of information. We were quite surprised by the parallels of physics and finance and also encountered many points where further research seems to be interesting and promising.

Chapter 2

Stochastic Calculus: Mathematical Basis and Tools

We shortly introduce the structure of Chapter 2. Section 2.1 introduces the Stratonovich stochastic integral, Section 2.2 introduces the considered stochastic differential equations and Section 2.3 gives various mathematical background needed in this thesis.
We already stated that we will be concerned with convergence properties of selected stochastic differential equations. It turns out in Chapter 3 that the stochastic integral we are concerned with will not always be a common Ito stochastic integral, but sometimes also the so-called Stratonovich stochastic integral. Hence Section 2.1 begins with stating a definition for the Stratonovich stochastic integral, which can be extended to a greater class of stochastic processes analogeous to the Ito stochastic integral (both is not done in this thesis). Rather than giving the explicit construction for the most general case of the Stratonovich stochastic integral we decided to give some intuition, heuristics and comparisons of the two stochastic integrals (Ito's and Stratonovich's), so that a reader with a financial background and possibly non-expert in stochastic calculus will get a good feeling for the different kinds of stochastic integrals, in particular the Stratonovich stochastic integral. We also state some important results concerning the stochastic integrals (e.g. pathwise Ito / Stratonovich formula).
Section 2.2 introduces the stochastic differential equations we will consider. We state desired conditions (later referred to as *regularity conditions*). After that we introduce our solution concept, which will be the *strong solution of a stochastic differential equation.* In order to make sure that our considerations make sense we devote some time to existence and uniqueness of the according stochastic differential equations (under certain conditions of course). Finally we define the needed convergence concepts for the solutions of stochastic dif-

ferential equations.
Section 2.3 on the one hand gives a short review of important facts concerning stochastic calculus and on the other hand states a few mathematical technical results, both of which will be needed later in the thesis.

2.1 The Stratonovich stochastic integral

In order to provide a good intution to a non-expert reader in mathematics, we start with the definition of the Stratonovich integral for a special case.

Definition and Remark 2.1. *(from [23] p. I.5)*
Suppose that $t \longmapsto X_t$ is a continuous and real-valued function on $[0,\infty[$, such that there exists a sequence of partitions (τ_n) with $(\tau_n) \overset{n\to\infty}{\longrightarrow} 0$ and $t^{(n)}_{N(n)} \overset{n\to\infty}{\longrightarrow} \infty$, such that the quadratic variation along (τ_n) exists for all $t \in \mathbb{R}_+$ and $t \longmapsto X_t$ is continuous on $[0,\infty[$. Let $f \in \mathcal{C}^1(\mathbb{R})$. Then

$$\oint_0^t f(X_s)dX_s := \lim_{n\to\infty} \sum_{t_i\in\tau_n, t_i\le t} f(\frac{1}{2}X_{t_{i+1}} + \frac{1}{2}X_{t_i}) \left(X_{t_{i+1}} - X_{t_i}\right).$$

This limit exists and we have that

$$\oint_0^t f(X_s) \circ dX_s = \int_0^t f(X_s)dX_s + \frac{1}{2}\int_0^t f'(X_s)d\left\langle X\right\rangle_s.$$

Under the given assumptions the Stratonovich integral exists whenever the Ito integral exists, as Stratonovich showed in [25].

If not specified, we always mean the Ito stochastic integral by just saying 'stochastic integral'.

Now we state the general definition of the Stratonovich stochastic integral (see [17] p. 156).

Definition 2.2. *Let X and Y be continuous semi-martingales with decompositions $X_t = X_0 + M_t + B_t$ and $Y_t = Y_0 + N_t + C_t$ for $0 \le t < \infty$, where M and N are continuous local martingales and B and C are adapted, continuous processes of bounded variation with $B_0 = C_0 = 0$ P-almost-surely. Then the Stratonovich stochastic integral of Y with respect to X is defined by*

$$\oint_0^t Y_s \circ dX_s := \int_0^t Y_s dM_s + \int_0^t Y_s dB_s + \frac{1}{2}\left\langle M,N\right\rangle_t \;\; \forall t \ge 0. \tag{2.1}$$

To get an intuition for the Stratonovich stochastic integral (which might be not so well-known to people with a financial background), consider the following statement (see [17], p. 156). For two continuous semi-martingales X and Y and $\Pi = \{t_0, t_1, ..., t_m\}$ a partition of $[0, T]$ with $0 = t_0 < t_1 < ... < t_m = T$ we have that

$$\sum_{i=0}^{m-1} \left(\frac{1}{2} Y_{t_{i+1}} + \frac{1}{2} Y_{t_i} \right) \left(X_{t_{i+1}} - X_{t_i} \right) \overset{\|\Pi\| \to 0}{\longrightarrow} \oint_0^t Y_s \circ dX_s$$

in probability.
So heuristically by using the Stratonovich integral, we consider the 'mean of the values of $Y_{t_{i+1}}$ and Y_{t_i}' as 'supporting points' to calculate the limiting sum, whereas by calculating the Ito integral we take the value of Y_{t_i} as 'supporting points' to calculate the limiting sum.

We will also refer to these heuristics in Chapter 4.

The next proposition states the pathwise Ito-formula (for Ito stochastic integrals) and its analogon for Stratonovich stochastic integrals.

Proposition 2.3. *Suppose that X_t is as in Definition and Remark 2.1.*

(i) (pathwise Ito formula)
Let $F \in \mathcal{C}^2$. Then for all $t \geq 0$:

$$F(X_t) - F(X_0) = \int_0^t F'(X_s) dX_s + \frac{1}{2} \int_0^t F''(X_s) d \langle X \rangle_s \,, \tag{2.2}$$

where

$$\int_0^t F'(X_s) dX_s = \lim_{n \to \infty} \sum_{t_i \in \tau_n, t_i \leq t} F'(X_{t_i}) \left(X_{t_{i+1}} - X_{t_i} \right). \tag{2.3}$$

This result can also be written in the (shorter) differential form:

$$dF(X) = F'(X) dX + \frac{1}{2} F''(X) d \langle X \rangle \tag{2.4}$$

(ii) (pathwise Stratonovich substitution rule, Fisk, Stratonovich) Suppose that $F \in \mathcal{C}^2$ and that W_t, $t \in \mathbb{R}_+$ is a standard Brownian motion. Then we have for all $t \geq 0$:

$$F(X_t) - F(X_0) = \oint_0^t F'(X_s) \circ dX_s \tag{2.5}$$

where the Stratonovich stochastic integral is defined as above. Analogeuously, this result can also be written in the (shorter) differential form:

$$dF(X) = F'(X) \circ dX \tag{2.6}$$

Proof. (i) is shown in [23] p I.4 and (ii) is shown in [23] p I.5 - I.6. $\square$

(i) of the above definition is a 'substitution rule for Ito integrals'. The difference to the substitution rule of ordinary calculus is the correction term $\frac{1}{2}f''(X)d\langle X\rangle$. Comparing with the Stratonovich substitution rule (ii), we see that we do not have this correction term in (ii), which is the 'substitution rule for Stratonovich integrals.' Thus at the first glance it seems that the Stratonovich substitution rule and hence the Stratonovich stochastic integral is more convenient, because it is usable like the one we know from ordinary calculus.
But unfortunately we have to pay the price: Stratonovich integrals do not preserve martingale properties as Ito integrals do. This statement is meant in the sense that if the integrator is a martingale, the resulting Ito integral is again a martingale (for a discussion see for example [23]), but the resulting Stratonovich integral is generally only a semi-martingale (and not a martingale).

It turns out that both stochastic integrals are useful in their according contexts. Since both integrals might be already well-known to a reader with a mathematical background, the discussion of the different concepts (Ito vs. Stratonovich) might be of more interest for a reader with a financial background. Our intention is to provide an intuition to the non-expert reader, where one integral should be used in preference to the other.

2.2 Introduction to the stochastic differential equations

The following stochastic differential equations (here given together with their initial conditions) will be in the center of this thesis:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = \xi_1, \tag{2.7}$$

and

$$d\tilde{X}_t = b(t, \tilde{X}_t)dt + \sigma(t, \tilde{X}_t) \circ dW_t, \quad X_0 = \xi_1. \tag{2.8}$$

Furthermore we consider the system of stochastic differential equations (as well with its initial conditions given):

$$\begin{gathered} dX_t^\mu = Y_t^\mu dt \\ dY_t^\mu = \frac{1}{\mu}b(t, X_t^\mu)dt + \frac{1}{\mu}\sigma(t, X_t^\mu)dW_t - \frac{1}{\mu}dX_t^\mu, \\ X_0^\mu = \xi_1,\ Y_0^\mu = \xi_2. \end{gathered} \tag{2.9}$$

In order to prove the desired convergence results, we make the following assumptions concerning the stochastic differential equations in this thesis:

(i) $\xi_1, \xi_2 \in \mathbb{R}$.
ξ_1 and ξ_2 will be called the initial conditions (of first and second order) of the (according) stochastic differential equation.

(ii) $b : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ Borel-measurable and uniformly bounded in t and x, i.e. $\sup_{t\in\mathbb{R}_+, x\in\mathbb{R}} |b(t, x)| = c_b < \infty$.
In a physical context (given in the next chapter) b will also be called 'drift' or 'trend'.

(iii) $\sigma : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ Borel-measurable, twice continuously differentiable in x; σ, $\frac{\partial\sigma(t,x)}{\partial x}$ and $\frac{\partial^2\sigma(t,x)}{\partial x^2}$ are uniformly bounded in t and x, i.e. $\sup_{t\in\mathbb{R}_+, x\in\mathbb{R}} |f(t, x)| \le c_\sigma < \infty$ for $f = \sigma$, $\frac{\partial\sigma(t,x)}{\partial x}$ and $\frac{\partial^2\sigma(t,x)}{\partial x^2}$.
In a physical context σ will also be called 'dispersion'.
b and σ are called the 'coefficients' of the according stochastic differential equation.

(iv) $W = \{W_t | t \in \mathbb{R}_+\}$ is a standard one-dimensional Brownian motion.

In the following (ii) and (iii) will be called 'regularity conditions' for b and σ. Note that the regularity conditions are not necessary for the existence and uniqueness Theorems 2.6 and 2.7 (for these we need different conditions).
Now at first we would like to formalize what is meant by saying 'a solution of the according stochastic differential equation'. To this purpose we will introduce the concept of a strong solution given by the following definition. The idea of this definition is taken from [17], p. 285, but we slightly modified that definition in order to make it more suitable for our situation.

Definition and Corollary 2.4. *(i) A strong solution of the Ito first order stochastic differential equation (2.7)*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = 0$$

on the given probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ *with respect to the fixed Brownian motion* W *is a real-valued stochastic process* $X = \{X_t | t \in \mathbb{R}_+\}$ *with continuous sample paths, adapted to the filtration* $(\mathcal{F}_t)_{t\geq 0}$*, fulfilling the initial condition* $X_0 = \xi_1$ *and for all* $t \in \mathbb{R}_+$

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$$

holds P*-a.s.*

(ii) A strong solution of the Ito Newton equation (second order Ito stochastic differential equation) (2.9):

$$dX_t^\mu = Y_t^\mu dt,$$

$$dY_t^\mu = \frac{1}{\mu}b(t, X_t^\mu)dt + \frac{1}{\mu}\sigma(t, X_t^\mu)dW_t - \frac{1}{\mu}dX_t^\mu,$$

$$X_0^\mu = \xi_1, \ Y_0^\mu = \xi_2$$

on the given probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ *with respect to the fixed Brownian motion* W *is a pair of real-valued stochastic processes* $(X, Y) = \{(X_t, Y_t) | t \in \mathbb{R}_+\}$*, each of them with continuous sample paths and adapted to the filtration* $(\mathcal{F}_t)_{t\geq 0}$*; Moreover, let* $X_0^\mu = \xi_1$ *as well as* $Y_0^\mu = \xi_2$ *hold and let the process* Y_t^μ *defined by* $Y_t^\mu dt = dX_t^\mu$ *for all* $t \in \mathbb{R}_+$ P*-a.s. be a strong solution of*

$$dY_t^\mu = \underbrace{\left(\frac{1}{\mu}b(t, X_t^\mu) - \frac{1}{\mu}Y_t^\mu\right)}_{:=b_Y(t, X_t^\mu, Y_t^\mu)} dt + \underbrace{\frac{1}{\mu}\sigma(t, X_t^\mu)}_{:=\sigma_Y(t, X_t^\mu, Y_t^\mu)} dW_t.$$

That is - analogously to (i) - for all $t \in \mathbb{R}_+$ P*-a.s.*

$$\begin{aligned} Y_t^\mu &= Y_0^\mu + \int_0^t b_Y(s, X_s^\mu, Y_s^\mu)ds + \int_0^t \sigma_Y(t, X_t^\mu, Y_t^\mu)dW_s \\ &= Y_0^\mu - \frac{1}{\mu}\int_0^t Y_s^\mu ds + \frac{1}{\mu}\int_0^t b(s, X_s^\mu)ds + \frac{1}{\mu}\int_0^t \sigma(s, X_s^\mu)dW_s \end{aligned}$$

(iii) A strong solution of the Stratonovich first order stochastic differential equation (2.8)

$$d\tilde{X}_t = b(t, \tilde{X}_t)dt + \sigma(t, \tilde{X}_s) \circ dW_s, \ \ \tilde{X}_0 = \xi_1.$$

on the given probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, P)$ with respect to the fixed Brownian motion W and initial condition ξ_1 is a real-valued stochastic process $\tilde{X} = \left\{ \tilde{X}_t | t \in \mathbb{R}_+ \right\}$ with continuous sample paths, adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, fulfilling the initial condition $\tilde{X}_0 = \xi_1$ and for all $t \in \mathbb{R}_+$

$$\tilde{X}_t = \tilde{X}_{t_0} + \int_{t_0}^{t} b(s, \tilde{X}_s)ds + \oint_{t_0}^{t} \sigma(s, \tilde{X}_s) \circ dW_s$$

holds P-a.s.

Remark 2.5. *(connecting Ito's and Stratonovich's stochastic integral in stochastic differential equations)*
We consider the stochastic differential equation

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

with $\sigma : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$, σ $\mathcal{C}^2$ in x with bounded first and second derivatives in x, and W_t is a Wiener process on $\mathbb{R}_+$. Then

$$\oint_0^t \sigma(s, X_s) \circ dW_s = \int_0^t \sigma(s, X_s)dW_s + \frac{1}{2}\int_0^t \sigma(s, X_s)\frac{\partial\sigma(s, X_s)}{\partial x}ds. \quad (2.10)$$

(2.10) can also be used to define the Stratonovich stochastic integral of the semi-martingale $\sigma(t, X_t)$ respectively to W_t (similarly done in [23] p. IV.6). From this point of view we note that 2.4(iii) is not a new definition, but rather an implication of 2.4(i), since in the given situation we can trace back the Stratonovich stochastic differential equation to the according Ito stochastic differential equation: The Ito stochastic differential equation

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

is under the given conditions equivalent to the Stratonovich stochastic differential equation

$$dX_t = b^*(t, X_t)dt + \sigma(t, X_t) \circ dW_t,$$

with

$$b^*(t, X_t) := b(t, X_t) - \frac{1}{2}\sigma(s, X_s)\frac{\partial\sigma(s, X_s)}{\partial x}.$$

Nevertheless, for completeness and in order to stress the importance of (iii) we made it part of Definition and Corollary 2.4.
For references and discussion, see for example [25], [23] (p IV.6) or [17] (p 295-296).

Next, let us discuss one of the most important fundamentals for this thesis: existence and uniqueness of the solutions of the introduced stochastic differential equations. Since we are dealing with convergence properties of the solution of several (stochastic) differential equations we have to make sure that these solutions exist and are unique.

Note that the following existence and uniqueness theorem is given for an arbitrary dimension $n \in \mathbb{N}$. We need this more general version in order to identify our situations with the following theorem.

Theorem 2.6. *Let $n \in \mathbb{N}$ and $T > 0$.*

(i) (Existence and uniqueness for ordinary differential equations)
Suppose that $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ is (Borel-) measurable and Lipschitz. Then the solution of the differential equation

$$dX_t = b(t, X_t)dt$$

exists and is unique on $[0,T]$.

(ii) (Existence and uniqueness for the Ito stochastic differential equation)
Let $W = \{W_t;\ 0 \le t < \infty\}$ be an n-dimensional Brownian motion and assume that $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ are (Borel-) measurable functions. Suppose that b and σ satisfy the growth condition

$$|b(t,x)|^2 + |\sigma(t,x)|^2 \le K^2(1 + |x|^2)\ \forall\ x \in \mathbb{R}^n,\ t \in [0,T] \tag{2.11}$$

for some constant K and the global Lipschitz condition

$$|b(t,x_0) - b(t,x_1)| + |\sigma(t,x_0) - \sigma(t,x_1)| \le D\,|x_0 - x_1| \tag{2.12}$$

$$\forall\ x_0, x_1 \in \mathbb{R}^n, t \in [0,T]$$

for some constant D, where $|b(t,x)|^2 := \sum |b_{ij}|^2$ (and for σ analogous). Then the stochastic differential equation

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

admits a unique strong solution X_t on $[0,T]$.

(iii) (Existence and uniqueness for stochastic differential equations perturbed by a certain class of stochastic processes)
Set $n = m = 1$ and let $V = \{V_t;\ 0 \le t < \infty\}$ be a stochastic process in $\mathbb{R}$ such that P-almost every path path $V(\omega)$ is continuous and has finite total variation on compact intervals of the form $[0,T]$. Suppose

that b and σ are Lipschitz and fulfill the according conditions (2.11) and (2.12) (for the case $n = 1$). Then the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dV_t$$

possesses a unique strong solution on $[0, T]$.

Proof. (i) A proof can be inferred from [22] part II, p. 10.5 - 10.8.

(ii) A proof can be found in [1] p. 119 - 124.

(iii) A proof can be found in [17] p. 297 - 298.

□

The next corollary gives existence and uniqueness for the stochastic differential equations we consider in this thesis (if not already covered directly by Theorem 2.6). We trace back our situation to the ones given above and show that the according conditions are fulfilled.

Corollary 2.7. *Let $T > 0$.*

(i) (Existence and uniqueness for the first order Stratonovich stochastic differential equation)
Suppose that the conditions of 2.6(ii) for the functions $b : [0, T] \times \mathbb{R} \to \mathbb{R}$, $\sigma : [0, T] \times \mathbb{R} \to \mathbb{R}$ (b, σ Borel-measurable) are satisfied. Furthermore let σ be $\mathcal{C}^2$ in x with bounded first and second derivative. Then the first order Stratonovich stochastic differential equation

$$d\tilde{X}_t = b(t, \tilde{X}_t)dt + \sigma(t, \tilde{X}_t) \circ dW_t$$

admits a unique, strong solution $\tilde{X}_t$ on $[0, T]$ on $[0, T]$.

(ii) Suppose that the conditions (2.11) and (2.12) of Theorem 2.6(ii) for the functions $b : [0, T] \times \mathbb{R} \longrightarrow \mathbb{R}$ and $\sigma : [0, T] \times \mathbb{R} \longrightarrow \mathbb{R}$ (b, σ Borel-measurable) are satisfied, and that b and σ are Lipschitz continuous with Lipschitz constants K_b and K_σ, respectively. Then the second order (Ito) system of stochastic differential equations

$$dX_t^\mu = Y_t^\mu dt,$$

$$dY_t^\mu = \frac{1}{\mu} b(t, X_t^\mu)dt + \frac{1}{\mu}\sigma(t, X_t^\mu)dW_t - \frac{1}{\mu}dX_t^\mu,$$

$$X_0^\mu = \xi_1, \ Y_0^\mu = \xi_2$$

admits a unique, strong solution (X_t^μ, Y_t^μ) on $[0, T]$.

(iii) Suppose that the functions $b : [0,T] \times \mathbb{R} \longrightarrow \mathbb{R}$ and $\sigma : [0,T] \times \mathbb{R} \longrightarrow \mathbb{R}$ (b, σ Borel-measurable) are Lipschitz with Lipschitz constants K_b and K_σ, respectively. Further assume that we have an adapted and right-continuous, differentiable stochastic process V_t^δ (depending on a fixed parameter δ), which fulfills $\sup_{t\in[0,T]} \left|\dot{V}_t^\delta\right| \leq C_\delta$, where $\dot{V}_t^\delta := \frac{d}{dt}V_t^\delta$ and $C_\delta \in \mathbb{R}$. Then the second order system of (stochastic) differential equations

$$dX_t^{\mu,\delta} = Y_t^{\mu,\delta} dt,$$

$$dY_t^{\mu,\delta} = \frac{1}{\mu} b(t, X_t^{\mu,\delta})dt + \frac{1}{\mu}\sigma(t, X_t^{\mu,\delta})dV_t^\delta - \frac{1}{\mu} dX_t^{\mu,\delta}$$

$$X_0^{\mu,\delta} = \xi_1,\ Y_t^{\mu,\delta} = \xi_2$$

admits a unique, strong solution $(X_t^{\mu,\delta}, Y_t^{\mu,\delta})$ on $[0,T]$.

Proof. (i) By Remark 2.5, the considered stochastic differential equation is equivalent to

$$d\tilde{X}_t = \underbrace{b(t,\tilde{X}_t)dt + \frac{1}{2}\sigma(t,\tilde{X}_t)\frac{\partial \sigma(t,\tilde{X}_t)}{\partial x}dt}_{=:\tilde{b}(t,\tilde{X}_t)dt} + \sigma(t,\tilde{X}_t)dW_t.$$

Denote the supremum of $\left|\frac{\partial\sigma(t,x)}{\partial x}\right|$ by $c_1 \in \mathbb{R}_+$.
We show that the conditions concerning $\tilde{b}$ and σ of Theorem 2.6 (ii) are fulfilled. Then by applying Theorem 2.6 (ii) the claim follows.
The Borel-measurability of $\tilde{b}$ and σ is clear.
The growth condition is fulfilled, since σ and $\sigma(t,x)\frac{\partial\sigma(t,x)}{\partial x}$ are globally bounded by assumption.
Consider the global Lipschitz condition

$$\begin{aligned}
\left|\tilde{b}(t,x_0) - \tilde{b}(t,x_1)\right| &+ |\sigma(t,x_0) - \sigma(t,x_1)| \\
&= \left| b(t,x_0) + \frac{1}{2}\sigma(t,x_0)\frac{\partial\sigma(t,x_0)}{\partial x} \right. \\
&\quad \left. - b(t,x_1) - \frac{1}{2}\sigma(t,x_1)\frac{\partial\sigma(t,x_1)}{\partial x} \right| \\
&\quad + |\sigma(t,x_0) - \sigma(t,x_1)| \\
&\leq \underbrace{|b(t,x_0) - b(t,x_1)| + |\sigma(t,x_0) - \sigma(t,x_1)|}_{\leq D|x_0 - x_1|} \\
&\quad + \frac{1}{2}c_1 |\sigma(t,x_0) - \sigma(t,x_1)|
\end{aligned}$$

$$\begin{aligned}
&\leq D\left|x_0 - x_1\right| \\
&\quad + \frac{1}{2} c_1 \left(\left|b(t, x_0) - b(t, x_1)\right| + \left|\sigma(t, x_0) - \sigma(t, x_1)\right|\right) \\
&\leq D\left(1 + \frac{1}{2} c_1\right)\left|x_0 - x_1\right|.
\end{aligned}$$

(ii) We write the considered system of stochastic differential equations (omitting the initial conditions for the moment) as a two-dimensional stochastic differential equation and obtain

$$d\begin{bmatrix} X_t^\mu \\ Y_t^\mu \end{bmatrix} = \begin{bmatrix} Y_t^\mu \\ \frac{1}{\mu} b(t, X_t^\mu) - \frac{1}{\mu} Y_t^\mu \end{bmatrix} dt + \begin{bmatrix} 0 \\ \frac{1}{\mu}\sigma(t, X_t^\mu) \end{bmatrix} dW_t.$$

Set

$$z := \begin{bmatrix} x \\ y \end{bmatrix}, \; f(t,x,y) := \begin{bmatrix} y \\ \frac{1}{\mu} b(t,x) - \frac{1}{\mu} y \end{bmatrix}, \; g(t,x,y) := \begin{bmatrix} 0 \\ \frac{1}{\mu}\sigma(t,x) \end{bmatrix},$$

i.e. $z \in \mathbb{R}^2$, $f, g : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}^2$. Hence our second order system of stochastic differential equations can be written as

$$dZ_t = f(t, X_t^\mu, Y_t^\mu)dt + g(t, X_t^\mu, Y_t^\mu)dW_t$$

and is hence a special case of Theorem 2.6 (ii) with $n = 2$ and $m = 1$. Thus we only need to show that the growth and global Lipschitz conditions (2.11) and (2.12) concerning f and g are satisfied and we are done.
We show this with the help of the according conditions concerning b and σ (already assumed to be fulfilled).
Again, Borel-measurability of f and g is clear.
So we start with the growth condition (2.11) (for the case $n = 2$):

$$|f(t,x,y)|^2 + |g(t,x,y)|^2 = y^2 + \frac{1}{\mu^2}\left(b(t,x) - y\right)^2 + \frac{1}{\mu^2}\sigma^2(t,x)$$

$$\begin{aligned}
&\leq y^2 + \frac{2}{\mu^2}\left(b^2(t,x) + y^2\right) + \frac{1}{\mu^2}\sigma^2(t,x) \\
&\leq \frac{\mu^2 + 2}{\mu^2} y^2 + \frac{2}{\mu^2}\left(b^2(t,x) + \sigma^2(t,x)\right) \\
&\leq \frac{\mu^2 + 2}{\mu^2} y^2 + \frac{2K^2}{\mu^2}\left(1 + x^2\right) \qquad \text{(by (2.11))} \\
&\leq \left(\frac{\mu^2 + 2}{\mu^2} + 1\right)\left(\frac{2K^2}{\mu^2} + 1\right)\left(1 + x^2 + y^2\right).
\end{aligned}$$

Now consider the global Lipschitz condition:

$$
\begin{aligned}
&|f(t,x_0,y_0)-f(t,x_1,y_1)|+|g(t,x_0,y_0)-g(t,x_1,y_1)|\\
&= \sqrt{(y_0-y_1)^2+\frac{1}{\mu^2}\left[(b(t,x_0)-b(t,x_1))-(y_0-y_1)\right]^2}\\
&\quad+\sqrt{\frac{1}{\mu^2}\left(\sigma(t,x_0)-\sigma(t,x_1)\right)^2}\\
&\leq \sqrt{(y_0-y_1)^2+\frac{1}{\mu^2}\left(K_b\left|x_0-x_1\right|-(y_0-y_1)\right)^2}\\
&\quad+\sqrt{\frac{1}{\mu^2}K_\sigma^2\left|x_0-x_1\right|^2}\\
&\leq \sqrt{(y_0-y_1)^2+\frac{2}{\mu^2}\left(K_b^2\left|x_0-x_1\right|^2+(y_0-y_1)^2\right)}\\
&\quad+\sqrt{\frac{1}{\mu^2}K_\sigma^2\left|x_0-x_1\right|^2}\\
&\leq \sqrt{\frac{\mu^2+2}{\mu^2}(y_0-y_1)^2+\frac{2}{\mu^2}K_b^2\left|x_0-x_1\right|^2}\\
&\quad+\sqrt{\frac{1}{\mu^2}K_\sigma^2\left|x_0-x_1\right|^2}\\
&\leq \sqrt{\left(\frac{\mu^2+2}{\mu^2}\right)\left(\frac{2}{\mu^2}K_b^2+1\right)}\sqrt{\left|x_0-x_1\right|^2+(y_0-y_1)^2}\\
&\quad+\sqrt{\frac{1}{\mu^2}K_\sigma^2}\sqrt{\left|x_0-x_1\right|+\left|y_0-y_1\right|^2}\\
&= \left(\sqrt{\left(\frac{\mu^2+2}{\mu^2}\right)\left(\frac{2}{\mu^2}K_b^2+1\right)}+\frac{K_\sigma}{\mu}\right)\\
&\quad\cdot\sqrt{(x_0-x_1)^2+(y_0-y_1)^2},
\end{aligned}
$$

which proves the global Lipschitz property of f and g.

(iii) Fix $\delta \in]0,1[$ and $\omega \in \Omega$. We want to apply the global existence and uniqueness theorem for an ordinary system of differential equations (of second order) (see for example [22], part II, p. 10.8 - 10.9, special case $n=2$).
Note that for fixed $\omega \in \Omega$ due to the piecewise differentiability of $X^{\mu,\delta}$

and $Y_t^{\mu,\delta}$ the considered stochastic differential equation is

$$\begin{aligned} dX_t^{\mu,\delta}(\omega) &= Y_t^{\mu,\delta}(\omega)dt, \\ \frac{d}{dt}Y_t^{\mu,\delta}(\omega) &= \frac{1}{\mu}b(t, X_t^{\mu,\delta}(\omega)) + \frac{1}{\mu}\sigma(t, X_t^{\mu,\delta}(\omega))\dot{V}_t^{\delta}(\omega) - \frac{1}{\mu}dY_t^{\mu,\delta}(\omega), \end{aligned}$$

where we used the piecewise differentiability of $V_t^{\delta}(\omega)$ for all $\omega \in \Omega$ and ordinary calculus.
Due to the relationship between $X_t^{\mu,\delta}(\omega)$ and $Y_t^{\mu,\delta}(\omega)$ (i.e. $\frac{d}{dt}X_t^{\mu,\delta}(\omega) = Y_t^{\mu,\delta}(\omega)\ \forall\ \omega \in \Omega$) we have an ordinary system of differential equations of second order. It is left to show that $f : \mathbb{R}\times\mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$, $f(t,x,y) := \frac{1}{\mu}b(t,x)+\frac{1}{\mu}\sigma(t,x)\dot{V}_t^{\delta}(\omega)-\frac{1}{\mu}y$ is Lipschitz (then existence and uniqueness follows). This can be seen as follows:

$$\begin{aligned} &|f(t,x_0,y_0) - f(t,x_1,y_1)| \\ &= \left|\frac{1}{\mu}b(t,x_0) + \frac{1}{\mu}\sigma(t,x_0)\dot{V}_t^{\delta} - \frac{1}{\mu}y_0 - \frac{1}{\mu}b(t,x_1) - \frac{1}{\mu}\sigma(t,x_1)\dot{V}_t^{\delta} + \frac{1}{\mu}y_1\right| \\ &\le \frac{1}{\mu}\left|b(t,x_0) - b(t,x_1)\right| + \frac{1}{\mu}\left|\sigma(t,x_0) - \sigma(t,x_1)\right|\left|\dot{V}_t^{\delta}\right| + \frac{1}{\mu}\left|y_0 - y_1\right| \\ &\le \frac{1}{\mu}K_b\left|x_0 - x_1\right| + \frac{1}{\mu}K_\sigma\left|x_0 - x_1\right|C_\delta + \frac{1}{\mu}\left|y_0 - y_1\right| \\ &\le \frac{1}{\mu}\left(K_b - C_\delta K_\sigma\right)\left|x_0 - x_1\right| + \frac{1}{\mu}\left|y_0 - y_1\right|. \end{aligned}$$

where the global Lipschitz property of f can be obtained from.

□

Note that the above Corollary 2.7 does not state new mathematical results, but is rather an application of the well-known existence and uniqueness results stated in Theorem 2.6. Nevertheless for the sake of completeness we stated the results and identified our situations with the ones in Theorem 2.6.

Remark 2.8. *Note that it is possible to show a stronger ('stronger' in the sense of less conditions and a better result) version of the above existence-and-uniqueness result 2.7 (i). This and its proof can be found in [17] p. 295 - 297. However, taking into account our general framework we will be working with, the result presented in 2.7 (i) does not require more than we will require anyway and it gives all the results we need. Thus we are completely satisfied with the above version.*

In this thesis we will be concerned with convergence properties of stochastic processes, which are solutions to certain stochastic differential equations. For example, one can examine the behavior of the solution X_t^μ of (2.9) as μ converges to zero (this will be part of the next chapter).
Hence we need to clarify our desired convergence property, which will be 'uniformly convergence in probability on a compact interval':

Definition 2.9. *Let X_t be a stochastic process and $X_t^n, n \in \mathbb{N}$, be a sequence of stochastic processes. We say that X_t^n converges in probability uniformly on $[a,b]$ to X_t if*

$$\forall\ \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left[\sup_{a\le t\le b} |X_t^n - X_t| > \epsilon\right] = 0,$$

or, equivalently,

$$\forall\ \epsilon > 0,\ \delta > 0 \quad \exists\ n_0 \in \mathbb{N} : \mathbb{P}\left[\sup_{a\le t\le b} |X_t^n - X_t| > \delta\right] \le \epsilon \quad \forall\ n \ge n_0. \quad (2.13)$$

2.3 Technical details and needed facts

The statements in this section may seem either very technical or a bit unmotivated or both. But in fact, we will need all of them later in this thesis, so we introduce them now.

The first proposition will be needed in order to complete a proof in the next chapter. A special case ($n = 2$) has already been used in the proof of existence and uniqueness.

Proposition 2.10. *Let*
$a_i \in \mathbb{R}_+$, $1 \le i \le n$, $i, k, n \in \mathbb{N}$, $k \ge 3$, k *odd,*
or
$a_i \in \mathbb{R}$, $1 \le i \le n$, $i, k, n \in \mathbb{N}$, $k \ge 2$, k *even.*
Then

$$\left(\sum_{i=1}^n a_i\right)^k \le \sum_{i=1}^n n^{k-1} a_i^k.$$

Proof. Note that $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^k$ is a convex function on $\mathbb{R}_+$ for k odd, and $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^k$ is a convex function on $\mathbb{R}$ for k even. Hence:

$$
\begin{aligned}
\left(\sum_{i=1}^{n} a_i\right)^k &= \left(\sum_{i=1}^{n} n^k \frac{a_i}{n}\right)^k = n^k \left(\sum_{i=1}^{n} \frac{a_i}{n}\right)^k \\
&\stackrel{*}{\leq} n^k \sum_{i=1}^{n} \frac{1}{n} a_i^k = n^{k-1} \sum_{i=1}^{n} a_i^k = \sum_{i=1}^{n} n^{k-1} a_i^k,
\end{aligned}
$$

where we used the convexity of f in the inequality marked with $(*)$.

□

The next two lemmas will be needed for proofs in the next chapter as well. Lemma 2.12 - stated second - is similar to Gronwall's Lemma 2.11 - stated first -, but the conditions required in Lemma 2.12 are weaker than the conditions required in Gronwall's Lemma 2.11. The statements given by the two lemmas cannot be compared in a sense of 'weaker' or 'stronger', but their applications depend on the situation given.
Gronwall's Lemma 2.11 is well-known, Lemma 2.12 and its proof are taken from [26].

Gronwall's Lemma 2.11. *Let $g : \mathbb{R} \mapsto \mathbb{R}$ continuous, with*

$$0 \leq g(t) \leq \alpha(t) + \beta \int_0^t g(s)ds, \;\; t \in [0, T]\,,$$

with $\beta \geq 0$ and $\alpha : [0, T] \longrightarrow \mathbb{R}$ integrable. Then

$$g(t) \leq \alpha(t) + \beta \int_0^t \alpha(s) e^{\beta(t-s)} ds, \;\; t \in [0, T]\,.$$

Lemma 2.12. *Let $f : \mathbb{R} \mapsto \mathbb{R}$ continuous in $[a, b]$, $a, b \in \mathbb{R}$, $a < b$, with $f(x) \geq 0$ for all $x \in \mathbb{R}$. Let $\epsilon : \mathbb{R} \mapsto \mathbb{R}$, $\epsilon(t) \geq 0$ for all $t \in \mathbb{R}$ with $0 < \nu < \infty$, $\rho > 0$, $\nu, \rho \in \mathbb{R}$ such that:*

$$\int_a^b \epsilon(s)ds < \frac{1}{\rho\nu e^{\rho\nu(b-a)}} \tag{2.14}$$

and

$$\log\left(1 + \frac{f(t)}{\nu}\right) \leq \log\left(1 + \epsilon(t)\right) + \rho \int_a^t f(s)ds. \tag{2.15}$$

Then:

$$f(t) \leq \nu \frac{\epsilon(t) + \rho\nu e^{\rho\nu(b-a)} \int_a^t \epsilon(s)ds}{1 - \rho\nu e^{\rho\nu(b-a)} \int_a^t \epsilon(s)ds}. \tag{2.16}$$

Proof. Consider inequality (2.15):

$$\begin{aligned}
\log\left(1+\frac{f(s)}{\nu}\right) &\leq \log\left(1+\epsilon(s)\right)+\rho\int_a^s f(r)dr \\
\Rightarrow \qquad 1+\frac{f(s)}{\nu} &\leq \left(1+\epsilon(s)\right)\underbrace{exp\left(\rho\int_a^s f(r)dr\right)}_{>0}
\end{aligned}$$

$$\iff \qquad \frac{1+\frac{f(s)}{\nu}}{exp\left(\rho\int_a^s f(r)dr\right)} \leq 1+\epsilon(s) \tag{2.17}$$

$$\begin{aligned}
\iff \qquad \frac{\rho\nu+\rho f(s)}{exp\left(\rho\int_a^s f(r)dr+\rho\nu s\right)} &\leq \rho\nu\left(1+\epsilon(s)\right)e^{-\rho\nu s} \\
\Rightarrow \qquad -\frac{d}{ds}exp\left(-\rho\int_a^s f(r)dr-\rho\nu s\right) &\leq \rho\nu\left(1+\epsilon(s)\right)e^{-\rho\nu s}
\end{aligned}$$

Since the right hand side is greater than zero (because all factors are greater than zero - either by assumption or obviously -) we can take the integral of both sides and the inequality still holds. Thus:

$$\Rightarrow \qquad \int_a^t -\frac{d}{ds}exp\left(-\rho\int_a^s f(r)dr-\rho\nu s\right)ds \leq \int_a^t \rho\nu\left(1+\epsilon(s)\right)e^{-\rho\nu s}ds$$

Integrating both sides yields:

$$\begin{aligned}
e^{-\rho\nu a}-exp\left(-\rho\int_a^t f(r)dr-\rho\nu t\right) &= -exp\left(-\rho\int_a^s f(r)dr-\rho\nu s\right)\Bigg|_a^t \\
&= \int_a^t -\frac{d}{ds}exp\left(-\rho\int_a^s f(r)dr-\rho\nu s\right)ds \\
&\leq \int_a^t \rho\nu\left(1+\epsilon(s)\right)e^{-\rho\nu s}ds \\
&= \int_a^t \rho\nu e^{-\rho\nu s}ds+\int_a^t \underbrace{\rho\nu}_{>0}\,\underbrace{\epsilon(s)}_{>0}\,\underbrace{e^{-\rho\nu s}}_{\leq e^{-\rho\nu a}}ds \\
&\leq e^{-\rho\nu a}-e^{-\rho\nu t}+\rho\nu e^{-\rho\nu a}\int_a^t \epsilon(s)ds.
\end{aligned}$$

Hence it follows

$$\begin{aligned}
\frac{1}{exp\left(\rho\int_a^t f(r)dr\right) exp\left(\rho\nu t\right)} &\geq e^{-\rho\nu t} - \rho\nu e^{-\rho\nu a}\int_a^t \epsilon(s)ds \\
\Rightarrow \qquad exp\left(\rho\int_a^t f(r)dr\right) &\leq \frac{e^{-\rho\nu t}}{e^{-\rho\nu t} - \rho\nu e^{-\rho\nu a}\int_a^t \epsilon(s)ds} \\
&= \frac{1}{1-\rho\nu e^{\rho\nu(t-a)}\int_a^t \epsilon(s)ds} \\
&\leq \frac{1}{1-\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds}
\end{aligned}$$

where we used that:

$$\begin{aligned}
e^{-\rho\nu t} - \rho\nu e^{-\rho\nu a}\int_a^t \epsilon(s)ds &= e^{-\rho\nu t} - e^{-\rho\nu b}\underbrace{\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds}_{<1 \quad \text{(by assumption (2.14))}} \\
&> e^{-\rho\nu t} - e^{-\rho\nu b} \geq 0 \quad for \quad t \leq b.
\end{aligned}$$

Finally we conclude by using the above result and inequality (2.17), setting $s = t$:

$$\begin{aligned}
\frac{1+\frac{f(t)}{\nu}}{\frac{1}{1-\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds}} &\leq \frac{1+\frac{f(t)}{\nu}}{exp\left(\rho\int_a^t f(r)dr\right)} \\
&\leq 1+\epsilon(t)
\end{aligned}$$

$$\Rightarrow \qquad 1+\frac{f(t)}{\nu} - \rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds - \rho f(t) e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds \leq 1+\epsilon(t)$$

$$\Leftrightarrow \qquad f(t)\underbrace{\left(1-\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds\right)}_{>0, \quad \text{see above}} \leq \nu\left(\epsilon(t)+\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds\right)$$

$$\Leftrightarrow \qquad f(t) \leq \nu\frac{\epsilon(t)+\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds}{1-\rho\nu e^{\rho\nu(b-a)}\int_a^t \epsilon(s)ds}$$

□

In the next chapter, we will also be concerned with perturbations that are not the Brownian, but smoother processes. This processes will be a composition of Gaussian stationary processes, so let us for convenience of the non-mathematician reader finally recall the definition and a well-known property of a certain class of such processes. Here, we give a Definition taken from [17], since this definition concludes the important information in a short but still intuitive way. Nevertheless, note that there are different possibilities to define Gaussian random variables (another access for example is to define Gaussian random variables by certain properties of the characteristic function, e.g. see [24], part 2, p. III.1).

Definition 2.13. *(from [17], p 103, case $n = 1$)*
An $\mathbb{R}$-valued process $X = \{X_t; 0 \leq t < \infty\}$ is called Gaussian if, for any integer $k \geq 1$ and real numbers $0 \leq t_1 \leq t_2 < ... < t_k < \infty$, the random vector $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ has a joint normal distribution. If the distribution of $(X_{t+t_1}, X_{t+t_2}, ..., X_{t+t_k})$ does not depend on t, we say that the process is stationary.
The finite-dimensional distributions of a Gaussian process X are determined by its expectation vector $m(t) = \mathbb{E}[X_t]; t \geq 0$, and its covariance matrix $\rho(s,t) = \mathbb{E}[(X_s - m(s))(X_t - m(t))]; s, t \geq 0$. [...] If $m(t) = 0; t \geq 0$, we say that X is a mean-zero Gaussian process.

Remark 2.14. *For any stationary process in $\mathbb{R}$ there exists a function r, $r : \mathbb{R} \to \mathbb{R}$ such that $\rho(s,t) = r(s-t)$, i.e. the covariance function can be written as a function only dependent on one parameter, namely the difference in s and t. In the following, let such an r represent the covariance function for the according stochastic process.*

Finally, the last proposition for this chapter is inferred from [9]. It gives a limiting result for an upper bound for a real, normal and stationary process satisfying certain conditions. We will need this result in the next chapter as well, since we will work not only with a Wiener process (as the disturbing noise), but also with a 'smoother' Gaussian disturbing noise, to which we will apply the following argument.

Proposition 2.15. *We consider a real, normal and stationary process $\xi(t)$ with zero mean and unit variance. Let its covariance function $r(t)$ satisfy the following two conditions:*

(i) For $t \to 0$ there exist constants $\lambda_2, \lambda_4 < \infty$, $\lambda_2 \geq 0, \lambda_4 > 0$ such that:

$$r(t) = 1 - \frac{\lambda_2}{2} t^2 + \frac{\lambda_4}{24} t^4 + o(t^4),$$

and

(ii) for $t \to \infty$ there exists $\alpha > 0$ such that:

$$r(t) = O(t^{-\alpha}).$$

Then it follows that

$$\lim_{T\to\infty} \mathbb{P}\left[\max_{0\le t\le T} |\xi_t| < c\sqrt{\log T}\right] = 1.$$

Proof. Using a result of Cramer and Leadbetter (see [9] p. 257 / 272) we have that under the given conditions for all $\epsilon > 0$ there exists $K > 0$ and $\tilde{n}_0 \in \mathbb{N}$ such that for all $T > n_0 := \min\{\tilde{n}_0, 3\}$:

$$\begin{aligned}
& \mathbb{P}\left[\left|\max_{0\le t\le T} |\xi_t| - \sqrt{2\log T}\right| < \frac{K}{\sqrt{\log T}}\right] &>& \; 1-\epsilon \\
\Rightarrow \quad & \mathbb{P}\left[\max_{0\le t\le T} |\xi_t| - \sqrt{2\log T} < \frac{K}{\sqrt{\log T}}\right] &>& \; 1-\epsilon \\
\Leftrightarrow \quad & \mathbb{P}\left[\max_{0\le t\le T} |\xi_t| < \sqrt{\log T}\left(\frac{K}{\log T} + \sqrt{2}\right)\right] &>& \; 1-\epsilon \\
\Rightarrow \quad & \mathbb{P}\left[\max_{0\le t\le T} |\xi_t| < \sqrt{\log T}\left(K + \sqrt{2}\right)\right] &>& \; 1-\epsilon \quad \text{(since } T > e\text{)},
\end{aligned}$$

which proves the claim by setting $c := K + \sqrt{2}$. $\square$

Chapter 3

Smoluchowski-Kramer and further approximations: Convergence results

This Chapter represents the main mathematical part of this thesis and is based on the paper [11] written by Mark Freidlin.

We start with a physical motivation. The motion of a particle in a force field according to Newton's law can be described by the system of stochastic differential equations

$$\begin{aligned} dX_t^{\mu,\delta} &= Y_t^{\mu,\delta}dt \\ \mu dY_t^{\mu,\delta} &= b(X_t^{\mu,\delta})dt + \sigma(X_t^{\mu,\delta})dV_t^{\delta} - dX_t^{\mu,\delta}, \\ X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2. \end{aligned}$$

Hereby we assume that it is reasonable to consider a smoother stochastic process V_t^{δ} rather than the Wiener noise.
In this thesis we are concerned with the question: What happens to the solution of the above system of stochastic differential equations $X_t^{\mu,\delta}$ if μ and δ converge to zero? That is: What can we say about $\lim_{\delta\searrow 0,\mu\searrow 0} X_t^{\mu,\delta}$, i.e. what happens to the law of motion based on the above system of stochastic differential equations as the mass of the particle becomes very small and the disturbing noise becomes 'very near to a Wiener process'?
It turns out that the above question cannot be answered in general, but the limit of $X_t^{\mu,\delta}$ depends on the way μ and δ approach zero. Hence the aim of this chapter is to show that - depending on the way μ and δ approach zero - the solution $X_t^{\mu,\delta}$ of the above system of equation may converge to X_t or $\tilde{X}_t$. Here X_t is the solution of the stochastic differential equation

$X_t = b(X_t)dt + \sigma(X_t)dW_t$ and $\tilde{X}_t$ is the solution of the stochastic differential equation $\tilde{X}_t = b(\tilde{X}_t)dt + \sigma(\tilde{X}_t) \circ dW_t$.

This chapter is structured as follows: In Section 3.1 we introduce the (systems of) stochastic differential equations of interest and interpret them physically as well. Section 3.2 states and proves the main convergence theorem for $X_t^{\mu,\delta}$. Since the proof is quite complex, we divided the proof into smaller parts. This is done by proving several lemmas, which can directly be used in the proof of the main result.
Due to the sequential proceeding in the proof of the main result, we also gain some additional results about the nature of $X_t^{\mu,\delta}$. In other words: The lemmas being used in the main proof are also valuable considered on their own. For example the well-known Smoluchowski-Kramer approximation (stated in Subsection 3.2.1) is such a result being used in the proof, but interesting on its own as well. Moreover, this long and complex proof offers an insight in the behavior of $X_t^{\mu,\delta}$ and the underlying (stochastic) differential equations.

Another important point to note is that when we consider an at least piecewise differentiable disturbing noise V_t^δ the according stochastic differential equation can be written as an ordinary (in the sense of non-stochastic) differential equation. Since we again consider a limiting situation and state (on the way to prove the main theorem) results about this limits as well, the thesis also includes a consideration of a sequence of ordinary differential equations converging to a stochastic differential equation. These considerations are taken from [26]. We explain this connection between ordinary and stochastic differential equations in Section 3.1 in detail.

The main result at first might seem to be quite surprising to readers not so familiar with stochastic calculus, since we get different limits depending on the way μ and δ approach zero. It turns out that by taking a closer look on the nature of this problem it is not as surprising as it seems at the first glance to non-mathematicians, so we will also devote some time to explain the intuition and consider heuristics. This will be done in Subsection 3.2.2, where also the main result is stated and proved.

3.1 Introduction to the convergence problems for the Newton equations

We consider the motion of a particle of mass μ in a force field, $0 < \mu << 1$. First we assume that the deterministic force is disturbed by a Wiener process W_t, i.e. the differential of the force field is given by $b(X_s)ds + \sigma(X_s)dW_s$, where $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$. We assume that b and σ fulfill the regularity conditions stated in Section 2.2.
Let the law of motion of the particle be denoted by X_t^μ. According to Newton's law, X_t^μ is given by the following system of equations:

$$\begin{aligned} dX_t^\mu &= Y_t^\mu dt \\ \mu dY_t^\mu &= b(X_t^\mu)dt + \sigma(X_t^\mu)dW_t - dX_t^\mu, \\ X_0^\mu &= \xi_1,\ Y_0^\mu = \xi_2. \end{aligned} \tag{3.1}$$

Throughout this chapter we have $\xi_i \in \mathbb{R}$ for $i = 1, 2$.
The Newton equation above can be interpreted in the following way: The increment of the momentum (defined by momentum := mass times velocity) of the particle (left hand side) is given by the differential of a force field (consisting of a deterministic part $b(X_t^\mu)dt$, which depends on the path of the particle, and a random part, where the (stochastic) differential is represented by $\sigma(X_t^\mu)dW_t$, and we substract the differential of the friction (represented by the very right term dX_t^μ).

From a physical point of view it is reasonable to replace the Wiener process W_t by a smoother pertubation, since the Brownian motion is only an idealized model for the disturbing noise. In reality we often rather have a smoother perturbation, which might be 'very near to a Brownian motion', but is not the Brownian motion itself. Thus we consider a process V_t^δ, such that $V_t^\delta \overset{\delta\to 0}{\longrightarrow} W_t$ uniformly on $[0, T]$ P-a.s., $0 < \delta << 1$. Let V_t^δ be piecewise continuously differentiable (modeling the 'smoothness' of the disturbing noise). This leads to the following system of equations:

$$\begin{aligned} dX_t^{\mu,\delta} &= Y_t^{\mu,\delta} dt \\ \mu dY_t^{\mu,\delta} &= b(X_t^{\mu,\delta})dt + \sigma(X_t^{\mu,\delta})dV_t^\delta - dX_t^{\mu,\delta}, \\ X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2. \end{aligned} \tag{3.2}$$

Let $\dot{V}_t^\delta$ denote the first derivative of V_t^δ (exists by assumption), i.e. $\dot{V}_t^\delta = \frac{d}{dt}V_t^\delta$, or, equivalently $\dot{V}_t^\delta dt = dV_t^\delta$. Plugging that in the above equation yields

$\mu dY_t^{\mu,\delta} = b(X_t^{\mu,\delta})dt + \sigma(X_t^{\mu,\delta})\dot{V}_t^{\delta}dt - dX_t^{\mu}$, and hence the system becomes

$$
\begin{aligned}
dX_t^{\mu,\delta} &= Y_t^{\mu,\delta}dt \\
\mu dY_t^{\mu,\delta} &= \left(b(X_t^{\mu,\delta}) + \sigma(X_t^{\mu,\delta})\dot{V}_t^{\delta}\right) dt - dX_t^{\mu}, \\
X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2,
\end{aligned}
\tag{3.3}
$$

which is - because of the absence of a stochastic term - obviously an ordinary differential equation. Hence considering a differentiable noise in a stochastic differential equation is equivalent to considering an ordinary (non-stochastic) differential equation.
In order to present things even clearer, we note that for any $\omega \in \Omega$ (arbitrary but fixed) the system (3.3) becomes:

$$
\begin{aligned}
\frac{d}{dt}X_t^{\mu,\delta}(\omega) &= Y_t^{\mu,\delta}(\omega) \\
\frac{d}{dt}Y_t^{\mu,\delta}(\omega) &= \frac{1}{\mu}\left(b(X_t^{\mu,\delta}(\omega)) + \sigma(X_t^{\mu,\delta}(\omega))\dot{V}_t^{\delta} - Y_t^{\mu}(\omega)\right) \\
X_0^{\mu,\delta}(\omega) &= \xi_1,\ Y_0^{\mu,\delta}(\omega) = \xi_2.
\end{aligned}
\tag{3.4}
$$

By assuming that the according conditions for existence and uniqueness of the according differential equations are fulfilled (see Corollary 2.7 (ii) and (iii)), the solutions X_t^{μ} and $X_t^{\mu,\delta}$ of the equations of (3.1) and (3.2) (respectively) exist and are unique.

In this chapter we are interested in the convergence properties of $X_t^{\mu,\delta}$, if both μ and δ go to zero. We will show that depending on the way μ and δ approach zero, $X_t^{\mu,\delta}$ converges to the solution of one of the following two stochastic differential equations.

$$
dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad X_0 = \xi_1 \tag{3.5}
$$

Here the stochastic integral is understood as the Ito integral. Interpreting the stochastic integral in Stratonovich's way leads to the second stochastic differential equation:

$$
d\tilde{X}_t = b(\tilde{X}_t)dt + \sigma(\tilde{X}_t) \circ dW_t, \qquad \tilde{X}_0 = \xi_1 \tag{3.6}
$$

As one can see, the only difference of the equations (3.5) and (3.6) is the way in which the stochastic integral is interpreted. The main result of this chapter will be that the solution $X_t^{\mu,\delta}$ of (3.2) converges to the first order stochastic differential equation given above with the stochastic integral understood either in Ito's or Stratonovich's way (depending on the way μ and δ tend to zero).

In order to prove this main result, we will also have to work with the first order stochastic differential equation, which is disturbed not by a Wiener process W_t, but by a smoother pertubation V_t^δ. The according differential equation is the following:

$$dX_t^\delta = b(X_t^\delta)dt + \sigma(X_t^\delta)dV_t^\delta, \qquad X_0^\delta = \xi_1 \tag{3.7}$$

Remark 3.1. *(i) Consider the Newton equation with perturbation V_t^δ (3.2). Here Ito's and Stratonovich's interpretation of the stochastic integral coincide, since $X_t^{\mu,\delta}$ is continuously differentiable in t.*

(ii) Now consider the first order differential equations perturbed by a Wiener process (3.5) and (3.6) with solutions X_t and $\tilde{X}_t$, respectively. In general we will have that $X_t \neq \tilde{X}_t$, since X_t and $\tilde{X}_t$ are both continous differentiable nowwhere P-a.s.

Proof. The Stratonovich stochastic integral is given by

$$\oint_0^t \sigma(X_t^{\mu,\delta}) \circ dV_t^\delta = \int_0^t \sigma(X_t^{\mu,\delta}) dV_t^\delta + \frac{1}{2} \underbrace{\left\langle \sigma(X_t^{\mu,\delta}), V_t^\delta \right\rangle}_{=0 \ (V_t^\delta \text{ cont. diff.})}.$$

$\square$

This remark is explaining why one is concerned with the convergence problem described above. It is reasonable to think about the convergence properties of the solution $X_t^{\mu,\delta}$ of (3.2). Considering this remark the resulting question is: If $X_t^{\mu,\delta}$ converges to a solution of a first order equation, how do we have to interpret the stochastic integral of the first order equation? This depends on the way how μ and δ approach zero, as we will see in the next section.

3.2 Convergence for the modified Newton equation

The main focus of this section is still the modified Newton equation (3.2). We are interested in the convergence properties in relation to a first order stochastic differential equation. We will specify and prove the convergence results indicated in Section 3.1.

3.2.1 The Smoluchowski-Kramers approximation

For the proof of the main theorem of this thesis presented in the next subsection, we will also have to work with time-dependent functions b and σ and the according stochastic differential equations. In order to have applicable results to the situation in the next chapter, we consider the non-Markovian case, meaning that the functions b and σ exhibit an additional time dependence. The according Newton systems of stochastic differential equations for the according perturbations are then:

$$\begin{aligned} dX_t^{\mu,\delta} &= Y_t^{\mu,\delta} dt \\ \mu dY_t^{\mu,\delta} &= b(t, X_t^{\mu,\delta})dt + \sigma(t, X_t^{\mu,\delta})dV_t^\delta - dX_t^{\mu,\delta}, \\ X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2, \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} dX_t^{\mu} &= Y_t^{\mu} dt \\ \mu dY_t^{\mu} &= b(t, X_t^{\mu})dt + \sigma(t, X_t^{\mu})dW_t - dX_t^{\mu}, \\ X_0^{\mu} &= \xi_1,\ Y_0^{\mu} = \xi_2. \end{aligned} \tag{3.9}$$

Assume that b, σ and V_t^δ fulfill all conditions stated in Corollary 2.7 (ii) and (iii). Then according to Corollary 2.7 (iii) the solution $X_t^{\mu,\delta}$ of (1.1) exists and is unique and according to Corollary 2.7 (ii) the solution X_t^μ of (3.9) exists and is unique.

The following proposition and its proof are taken from [11].

Proposition 3.2. *Consider the second order system of stochastic differential equations (3.9) together with the above assumptions. Then we have that:*

$$\begin{aligned} X_t^\mu &= \xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) \\ &\quad + \int_0^t b(s, X_s^\mu)ds - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu)ds \\ &\quad + \int_0^t \sigma(s, X_s^\mu)dW_s - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^\mu)dW_s, \\ Y_t^\mu &= e^{-\frac{t}{\mu}}\xi_2 + \frac{1}{\mu}e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu)ds + \frac{1}{\mu}e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^\mu)dW_s. \end{aligned}$$

Proof. We have $dX_t^\mu = Y_t^\mu dt$. We solve for Y_t^μ first, which is now given by the following stochastic differential equation:

$$dY_t^\mu = -\frac{1}{\mu} Y_t^\mu dt + \frac{1}{\mu} b(t, X_t^\mu) dt + \frac{1}{\mu} \sigma(t, X_t^\mu) dW_t.$$

Multiply with $e^{\frac{t}{\mu}}$ to obtain:

$$\frac{1}{\mu} e^{\frac{t}{\mu}} Y_t^\mu dt + e^{\frac{t}{\mu}} dY_t^\mu = \frac{1}{\mu} e^{\frac{t}{\mu}} b(t, X_t^\mu) dt + \frac{1}{\mu} e^{\frac{t}{\mu}} \sigma(t, X_t^\mu) dW_t$$

By the pathwise Ito formula 2.3(i) we have that

$$d(e^{\frac{t}{\mu}} Y_t^\mu) = \frac{1}{\mu} e^{\frac{t}{\mu}} Y_t^\mu dt + e^{\frac{t}{\mu}} dY_t^\mu.$$

Plugging this in leads to the equation:

$$d(e^{\frac{t}{\mu}} Y_t^\mu) = \frac{1}{\mu} e^{\frac{t}{\mu}} b(t, X_t^\mu) dt + \frac{1}{\mu} e^{\frac{t}{\mu}} \sigma(t, X_t^\mu) dW_t,$$

hence

$$e^{\frac{t}{\mu}} Y_t^\mu - e^{\frac{0}{\mu}} Y_0^\mu = \int_0^t \frac{1}{\mu} e^{\frac{s}{\mu}} b(s, X_s^\mu) ds + \int_0^t \frac{1}{\mu} e^{\frac{s}{\mu}} \sigma(s, X_s^\mu) dW_s,$$

and finally

$$Y_t^\mu = e^{-\frac{t}{\mu}} Y_0^\mu + e^{-\frac{t}{\mu}} \int_0^t \frac{1}{\mu} e^{\frac{s}{\mu}} b(s, X_s^\mu) ds + e^{-\frac{t}{\mu}} \int_0^t \frac{1}{\mu} e^{\frac{s}{\mu}} \sigma(s, X_s^\mu) dW_s.$$

Now use that $dX_t^\mu = Y_t^\mu dt$. Integrate to obtain

$$\begin{aligned} X_t^\mu &= X_0^\mu + \int_0^t e^{-\frac{s}{\mu}} Y_0^\mu ds + \frac{1}{\mu} \int_0^t e^{-\frac{s}{\mu}} \left(\int_0^s e^{\frac{r}{\mu}} b(r, X_r^\mu) dr \right) ds \\ &\quad + \frac{1}{\mu} \int_0^t e^{-\frac{s}{\mu}} \left(\int_0^s e^{\frac{r}{\mu}} \sigma(r, X_r^\mu) dW_r \right) ds. \end{aligned} \tag{3.10}$$

Plug in $X_0^\mu = \xi_1$, $Y_0^\mu = \xi_2$, and integrate the first term on the right hand side. Use integration by parts for the non-stochastic integral and Ito's product rule for the stochastic integral:

$$
\begin{aligned}
X_t^\mu &= \xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) \\
&\quad + \frac{1}{\mu}\left[-\mu e^{-\frac{s}{\mu}} \cdot \left(\int_0^s e^{\frac{r}{\mu}} b(r, X_r^\mu) dr \right) \Bigg|_0^t + \int_0^t \mu e^{-\frac{s}{\mu}} e^{\frac{s}{\mu}} b(s, X_s^\mu) ds \right] \\
&\quad + \frac{1}{\mu}\left[-\mu e^{-\frac{s}{\mu}} \cdot \left(\int_0^s e^{\frac{r}{\mu}} \sigma(r, X_r^\mu) dW_r \right) \Bigg|_0^t + \int_0^t \mu e^{-\frac{s}{\mu}} e^{\frac{s}{\mu}} \sigma(s, X_s^\mu) dW_s \right] \\
&= \xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) \\
&\quad - e^{-\frac{t}{\mu}} \cdot \int_0^t e^{\frac{s}{\mu}} b(s, X_r^\mu) ds + 0 + \int_0^t b(s, X_s^\mu) ds \\
&\quad - e^{-\frac{t}{\mu}} \cdot \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_r^\mu) dW_s + 0 + \int_0^t \sigma(s, X_s^\mu) dW_s \\
&= \xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) \\
&\quad + \int_0^t b(s, X_s^\mu) ds - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu) ds \\
&\quad + \int_0^t \sigma(s, X_s^\mu) dW_s - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^\mu) dW_s
\end{aligned}
$$

□

Remark 3.3. *Note that from the above Proposition 3.2 it follows that*

$$
\begin{aligned}
X_t^{\mu,\delta} &= \xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) \\
&\quad + \int_0^t b(s, X_s^{\mu,\delta}) ds - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s, X_s^{\mu,\delta}) ds \\
&\quad + \int_0^t \sigma(s, X_s^{\mu,\delta}) dV_s^\delta - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^{\mu,\delta}) dV_s^\delta, \\
Y_t^{\mu,\delta} &= e^{-\frac{t}{\mu}}\xi_2 + \frac{1}{\mu} e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s, X_s^{\mu,\delta}) ds + \frac{1}{\mu} e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^{\mu,\delta}) dV_s^\delta,
\end{aligned}
$$

where $(X_t^{\mu,\delta}, Y_t^{\mu,\delta})$ is the solution of (1.1) (together with the above assumptions).

Analogously to the time-independent case, we also consider the according first order stochastic differential equation with time-dependent functions b and σ, together with the initial condition:

$$
dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \qquad X_0 = \xi_1. \tag{3.11}
$$

Definition 3.4. *For a mapping $a(t,x)$, $a : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$, we define the norms:*

$$\|a\|_T := \sup \left\{ |a(t,x)| \,\middle|\, 0 \le t \le T, x \in \mathbb{R} \right\},$$

and

$$\|a\| := \|a\|_\infty = \sup \left\{ |b(t,x)| \,\middle|\, 0 \le t \le \infty, x \in \mathbb{R} \right\}.$$

The following lemma is essentially taken from [11] and is concerned with properties of X_t^μ and X_t with respect to each other. The lemma states that the considered terms in (i) - (iii) are bounded with explicit bounds. Compared to Freidlin's version of this lemma, we slightly modified the bounds in (ii) and (iii) due to technical reasons and then give the slightly modified proof of Freidlin's result. The general statement of Freidlin's lemma stays the same. Note that (i) can also be inferred from (ii), but we also give the (direct) proof here.

Lemma 3.5. *Let X_t^μ and X_t be the solutions of (3.9) and (3.11), respectively.*

(i) Let $\sigma(t,X) \equiv 0$. Then

$$\sup_{0 \le t \le T} |X_t^\mu - X_t| \le \mu \left(\|b\| + |\xi_2|\right)\left(1 + TK_b e^{K_b T}\right). \tag{3.12}$$

For any $T > 0$ there exists a constant $c = c(T, K_b, K_\sigma) > 0$, such that

(ii)

$$\sup_{0 \le t \le T} \mathbb{E}\left[|X_t^\mu - X_t|^2\right] \le c\mu \left[\mu |\xi_2|^2 + \mu \|b\|^2 + \frac{1}{2}\|\sigma\|^2\right], \tag{3.13}$$

and

(iii)

$$\mathbb{P}\left[\sup_{0 \le t \le T} |X_t^\mu - X_t| > h\right] \le \frac{c\mu}{h^2}\left[\mu |\xi_2|^2 + \mu \|b\|^2 + \frac{1}{2}\|\sigma\|^2\right] \quad \forall \quad h > 0. \tag{3.14}$$

Proof. (i) Since $\sigma(t,X) \equiv 0$ by assumption, we get by the definition of a strong solution X_t of a first order differential equation (Definition

2.4(i)) and Proposition 3.2

$$
\begin{aligned}
|X_t^\mu - X_t| &= \Big| \xi_1 + \mu\xi_2(1-e^{-\frac{t}{\mu}}) + \int_0^t b(s,X_s^\mu)ds \\
&\quad - e^{-\frac{t}{\mu}} \int_0^t b(s,X_s^\mu)e^{\frac{s}{\mu}}ds - \xi_1 - \int_0^t b(s,X_s)ds \Big| \\
&\le \left|\mu\xi_2(1-e^{-\frac{t}{\mu}})\right| + \left|\int_0^t b(s,X_s^\mu) - b(s,X_s)ds\right| \\
&\quad + \left|e^{-\frac{t}{\mu}} \int_0^t b(s,X_s^\mu)e^{\frac{s}{\mu}}ds\right| \\
&\le \left|\mu\xi_2(1-e^{-\frac{t}{\mu}})\right| + K_b\int_0^t |X_s^\mu - X_s|\,ds + \mu\,\|b\| \\
&\le \mu\,(|\xi_2| + \|b\|) + K_b\int_0^t |X_s^\mu - X_s|\,ds,
\end{aligned}
$$

where we have used that due to the uniformly boundedness of b we get

$$
\begin{aligned}
\left|e^{-\frac{t}{\mu}} \int_0^t b(s,X_s^\mu)e^{\frac{s}{\mu}}ds\right| &\le \left|e^{-\frac{t}{\mu}} \int_0^t \sup_{0\le r\le T, x\in\mathbb{R}} b(r,x)e^{\frac{s}{\mu}}ds\right| \\
&= e^{-\frac{t}{\mu}} \left|\sup_{0\le r\le T, x\in\mathbb{R}} b(r,x)\right| \int_0^t e^{\frac{s}{\mu}}ds \\
&= e^{-\frac{t}{\mu}} \left|\sup_{0\le r\le T, x\in\mathbb{R}} b(r,x)\right| (\mu e^{\frac{t}{\mu}} - \mu) \\
&\le e^{-\frac{t}{\mu}} \left|\sup_{0\le r\le T, x\in\mathbb{R}} b(r,x)\right| \mu e^{\frac{t}{\mu}} \\
&= \mu\,\|b\|_T \le \mu\,\|b\|\,.
\end{aligned}
$$

Finally applying Gronwall's Lemma 2.11 leads to

$$
\begin{aligned}
|X_t^\mu - X_t| &\le \mu(\|b\| + |\xi_2)|)\left(1 + K_b\int_0^t e^{K_b t}ds\right) \\
&\le \mu\,(\|b\| + |\xi_2|)\left(1 + TK_b e^{K_b T}\right) \ \forall\ t\ \in [0,T]\,.
\end{aligned}
$$

Since this inequality is valid for all $t \in [0,T]$ (since the upper bound is not dependent on t) it follows that

$$
\sup_{0\le t\le T} |X_t^\mu - X_t| \le \mu\,(\|b\| + |\xi_2|)\left(1 + TK_b e^{K_b T}\right).
$$

Thus statement (i) is proved.

(ii) Again we use the definition of a strong solution X_t of a first order stochastic differential equation (Definition 2.4) and Proposition 3.2. In the third step, we use Proposition 2.10(i) for the case $n = 5$.

$$
\begin{aligned}
&\mathbb{E}\left[|X_t^\mu - X_t|^2\right] \\
= \;& \mathbb{E}\Bigg[\Bigg|\xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) + \int_0^t b(s, X_s^\mu)ds + \int_0^t \sigma(s, X_s^\mu)dW_s \\
& -e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu)ds - e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\sigma(s, X_s^\mu)dW_s \\
& -\xi_1 - \int_0^t b(s, X_s)ds - \int_0^t \sigma(s, X_s)dW_s\Bigg|^2\Bigg] \\
\leq \;& \mathbb{E}\Bigg[\Bigg(\mu\xi_2 + \int_0^t \left(b(s, X_s^\mu) - b(s, X_s)\right) ds \\
& + \int_0^t \left(\sigma(s, X_s^\mu) - \sigma(s, X_s)\right) dW_s \\
& -e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu)ds - e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\sigma(s, X_s^\mu)dW_s\Bigg)^2\Bigg] \\
\leq \;& 5\mathbb{E}\left[(\mu\xi_2)^2\right] + 5\mathbb{E}\left[\left(\int_0^t \left(b(s, X_s^\mu) - b(s, X_s)\right) ds\right)^2\right] \\
& +5\mathbb{E}\left[\left(\int_0^t \left(\sigma(s, X_s^\mu) - \sigma(s, X_s)\right) dW_s\right)^2\right] \\
& +5\mathbb{E}\left[\left(e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu)ds\right)^2\right] \\
& +5\mathbb{E}\left[\left(e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\sigma(s, X_s^\mu)dW_s\right)^2\right]
\end{aligned}
$$

Let us consider the terms separately.

$$5\mathbb{E}\left[(\mu\xi_2)^2\right] = 5\left(\mu\xi_2\right)^2.$$

With the help of Jensen's inequality (see [24] part I, page I.5.1) and again due to the fact that $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ is a convex function we obtain:

$$5\mathbb{E}\left[\left(\int_0^t \left(b(s, X_s^\mu) - b(s, X_s)\right) ds\right)^2\right]$$

$$= 5t^2 \; \mathbb{E}\left[\left(\int_0^t \frac{b(s,X_s^\mu)-b(s,X_s)}{t} ds\right)^2\right]$$

$$\begin{aligned}
&\leq \; 5t^2 \cdot \frac{1}{t} \; \mathbb{E}\left[\int_0^t \left(b(s,X_s^\mu)-b(s,X_s)\right)^2 ds\right] \\
&= \; 5t \; \mathbb{E}\left[\int_0^t \left(b(s,X_s^\mu)-b(s,X_s)\right)^2 ds\right] \\
&\leq \; 5tK_b^2 \; \mathbb{E}\left[\int_0^t \left(X_s^\mu - X_s\right)^2 ds\right] \\
&\leq \; 5TK_b^2 \int_0^t \mathbb{E}\left[\left(X_s^\mu - X_s\right)^2\right] ds,
\end{aligned}$$

where we used Fubini's Theorem (see [24] p. IV.2.3) in the last step. For the next term we conclude thanks to the Ito isometry (and then analogous arguments):

$$5\mathbb{E}\left[\left(\int_0^t \left(\sigma(s,X_s^\mu)-\sigma(s,X_s)\right) dW_s\right)^2\right]$$

$$\begin{aligned}
&\leq \; 5\mathbb{E}\left[\int_0^t \left(\sigma(s,X_s^\mu)-\sigma(s,X_s)\right)^2 ds\right] \\
&\leq \; 5K_\sigma^2 \int_0^t \mathbb{E}\left[\left(X_s^\mu - X_s\right)^2\right] ds
\end{aligned}$$

Next, by the argument in (i) we see:

$$\begin{aligned}
5\mathbb{E}\left[\left(e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s,X_s^\mu) ds\right)^2\right] \; &\leq \; 5\mathbb{E}\left[\left(\mu \left\|b\right\|_T\right)^2\right] \\
&= \; 5\left(\mu \left\|b\right\|_T\right)^2.
\end{aligned}$$

And the last term we estimate by using an analogous argument as in (i) after using the Ito isometry:

$$5\mathbb{E}\left[\left(e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s,X_s^\mu) dW_s\right)^2\right]$$

$$
\begin{aligned}
&= 5\mathbb{E}\left[\left(e^{-\frac{t}{\mu}}\right)^2\left(\int_0^t e^{\frac{s}{\mu}}\sigma(s,X_s^\mu)dW_s\right)^2\right] \\
&\le 5\mathbb{E}\left[e^{-\frac{2t}{\mu}}\int_0^t e^{\frac{2s}{\mu}}\left(\sigma(s,X_s^\mu)\right)^2 ds\right] \\
&\le 5\mathbb{E}\left[e^{-\frac{2t}{\mu}}\left\|\sigma^2\right\|_T\int_0^t e^{\frac{2s}{\mu}}ds\right] \\
&= 5e^{-\frac{2t}{\mu}}\left\|\sigma\right\|_T^2\left(\frac{\mu}{2}e^{\frac{2t}{\mu}}-\frac{\mu}{2}\right) \\
&\le 5\frac{\mu}{2}\left\|\sigma\right\|_T^2.
\end{aligned}
$$

With these results, continuing the original calculation yields

$$
\begin{aligned}
\mathbb{E}\left[\left|X_t^\mu - X_t\right|^2\right] &\le 5\left(\mu\xi_2\right)^2 + 5TK_b^2\int_0^t \mathbb{E}\left[\left(X_s^\mu - X_s\right)^2\right]ds \\
&\quad +5K_\sigma^2\int_0^t \mathbb{E}\left[\left(X_s^\mu - X_s\right)^2\right]ds \\
&\quad +5\left(\mu\left\|b\right\|_T\right)^2 + 5\frac{\mu}{2}\left\|\sigma\right\|_T^2 \\
&\le 5\mu\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right] \\
&\quad +5(TK_b^2 + K_\sigma^2)\int_0^t \mathbb{E}\left[\left|X_s^\mu - X_s\right|^2\right]ds,
\end{aligned}
$$

Now by Gronwall's Lemma 2.11 it follows that

$$
\begin{aligned}
\mathbb{E}\left[\left(X_t^\mu - X_t\right)^2\right] &\le 5\mu\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right] \\
&\quad \cdot\left(1 + 5(TK_b^2 + K_\sigma^2)\int_0^t e^{5(TK_b^2+K_\sigma^2)t}ds\right) \\
&\le \mu\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right] \\
&\quad \underbrace{\cdot 5\left(1 + 5T(TK_b^2 + K_\sigma^2)e^{5(TK_b^2+K_\sigma^2)T}\right)}_{=:c} \\
&= c\mu\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right] \ \forall\, t \in [0,T].
\end{aligned}
$$

Fix $T > 0$. Again, taking into account that we postulate $\|b\| < \infty$ and $\|\sigma\| < \infty$, the achieved inequality gives obviously an upper bound for

all t, $0 \leq t \leq T$, i.e. $\mathbb{E}\left[|X_t^\mu - X_t|^2\right] \leq const. < \infty$ for all $0 \leq t \leq T$.
In particular, $\mathbb{E}\left[|X_t^\mu - X_t|^2\right]$ exists for all $0 \leq t \leq T$.
Hence it finally follows that

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[|X_t^\mu - X_t|^2\right] \leq c\mu \left[\mu |\xi_2|^2 + \mu \|b\|^2 + \frac{1}{2} \|\sigma\|^2\right]$$

and (ii) is proved.

(iii) We first prove the following statement:

$$\mathbb{E}\left[|X_s^\mu - X_s|^k\right] \leq c\mu^{\frac{k}{2}}, \tag{3.15}$$

with $c = c(p, |\xi_2| T, \|b\|_T, \|\sigma\|_T)$.
Analogously to the argument in (ii) we obtain by using Proposition 2.10(i)

$$\mathbb{E}\left[|X_t^\mu - X_t|^k\right]$$

$$\begin{aligned}
&= \mathbb{E}\Bigg[\Bigg|\xi_1 + \mu\xi_2\left(1 - e^{-\frac{t}{\mu}}\right) + \int_0^t b(s, X_s^\mu) ds + \int_0^t \sigma(s, X_s^\mu) dW_s \\
&\quad - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} b(s, X_s^\mu) ds - e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^\mu) dW_s \\
&\quad - \xi_1 - \int_0^t b(s, X_s) ds - \int_0^t \sigma(s, X_s) dW_s\Bigg|^k\Bigg] \\
&\leq \mathbb{E}\Bigg[\Bigg(|\mu\xi_2| + \int_0^t |b(s, X_s^\mu) - b(s, X_s)| \, ds \\
&\quad + \left|\int_0^t (\sigma(s, X_s^\mu) - \sigma(s, X_s)) \, dW_s\right| \\
&\quad + e^{-\frac{t}{\mu}} \int_0^t e^{\frac{s}{\mu}} |b(s, X_s^\mu)| \, ds + e^{-\frac{t}{\mu}} \left|\int_0^t e^{\frac{s}{\mu}} \sigma(s, X_s^\mu) dW_s\right|\Bigg)^k\Bigg]
\end{aligned}$$

$$
\begin{aligned}
\leq\ & 5^{k-1}\left(\mu\left|\xi_2\right|\right)^k + 5^{k-1}\mathbb{E}\left[\left(\int_0^t \left|b(s,X_s^\mu) - b(s,X_s)\right| ds\right)^k\right] \\
& +5^{k-1}\mathbb{E}\left[\left|\int_0^t \left(\sigma(s,X_s^\mu) - \sigma(s,X_s)\right) dW_s\right|^k\right] \\
& +5^{k-1}\mathbb{E}\left[\left(e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\left|b(s,X_s^\mu)\right| ds\right)^k\right] \\
& +5^{k-1}\mathbb{E}\left[\left|e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\sigma(s,X_s^\mu) dW_s\right|^k\right]
\end{aligned}
$$

Again, let us consider the terms separately.
With the help of Jensen's inequality (used in the second step), the Lipschitz property of b (used in the fourth step) and Fubini's theorem (used in the last step) we obtain:

$$
5^{k-1}\mathbb{E}\left[\left(\int_0^t \left|b(s,X_s^\mu) - b(s,X_s)\right| ds\right)^k\right]
$$

$$
\begin{aligned}
=\ & 5^{k-1}t^k\ \mathbb{E}\left[\left(\int_0^t \frac{\left|b(s,X_s^\mu) - b(s,X_s)\right|}{t} ds\right)^k\right] \\
\leq\ & 5^{k-1}t^k\cdot\frac{1}{t}\ \mathbb{E}\left[\int_0^t \left|b(s,X_s^\mu) - b(s,X_s)\right|^k ds\right] \\
=\ & 5^{k-1}t^{k-1}\ \mathbb{E}\left[\int_0^t \left|b(s,X_s^\mu) - b(s,X_s)\right|^k ds\right] \\
\leq\ & 5^{k-1}t^{k-1}K_b^k\ \mathbb{E}\left[\int_0^t \left|X_s^\mu - X_s\right|^k ds\right] \\
\leq\ & 5^{k-1}T^{k-1}K_b^k\int_0^t \mathbb{E}\left[\left|X_s^\mu - X_s\right|^k\right] ds.
\end{aligned}
$$

For the next term, we use the Birkholder-Davis-Gundy inequality (see [17] p. 166) in the first step (hence $L_k \in \mathbb{R}$ is a constant depending only on k). Again, the last two steps use Jensen's inequality and Fubini's theorem.

$$5^{k-1}\mathbb{E}\left[\left|\int_0^t \sigma(s,X_s^\mu)-\sigma(s,X_s)dW_s\right|^k\right]$$

$$\begin{aligned}
&\leq\; 5^{k-1}L_k\;\mathbb{E}\left[\left\langle\int_0^t\left(\sigma(s,X_s^\mu)-\sigma(s,X_s)\right)dW_s\right\rangle_t^{\frac{k}{2}}\right]\\
&\leq\; 5^{k-1}L_k\;\mathbb{E}\left[\left(\int_0^t\left|\sigma(s,X_s^\mu)-\sigma(s,X_s)\right|^2ds\right)^{\frac{k}{2}}\right]\\
&\leq\; 5^{k-1}L_kK_\sigma^k\;\mathbb{E}\left[\left(\int_0^t\left|X_s^\mu-X_s\right|^2ds\right)^{\frac{k}{2}}\right]\\
&\leq\; 5^{k-1}L_kt^{\frac{k}{2}}K_\sigma^k\;\mathbb{E}\left[\int_0^t\left|X_s^\mu-X_s\right|^kds\right]\\
&=\; 5^{k-1}L_kT^{\frac{k}{2}}K_\sigma^k\int_0^t\mathbb{E}\left[\left|X_s^\mu-X_s\right|^k\right]ds.
\end{aligned}$$

Next, analogously to the preceeding in (ii) (and using the argument of (i)), we obtain

$$\begin{aligned}
5^{k-1}\mathbb{E}\left[\left(e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\left|b(s,X_s^\mu)\right|ds\right)^k\right] &\leq\; 5^{k-1}\mathbb{E}\left[\left(\mu\left\|b\right\|_T\right)^k\right]\\
&=\; 5^{k-1}\mu^k\left\|b\right\|_T^k
\end{aligned}$$

And finally we estimate the last term by using the Birkholder-Davis-Gundy inequality again (with a constant $M_k\in\mathbb{R}$ only dependent on k):

$$5^{k-1}\mathbb{E}\left[\left|e^{-\frac{t}{\mu}}\int_0^t e^{\frac{s}{\mu}}\sigma(s,X_s^\mu)dW_s\right|^k\right]$$

$$\begin{aligned}
&=\; 5^{k-1}e^{-\frac{tk}{\mu}}\;\mathbb{E}\left[\left|\int_0^t e^{\frac{s}{\mu}}\sigma(s,X_s^\mu)dW_s\right|^k\right]\\
&\leq\; 5^{k-1}e^{-\frac{tk}{\mu}}M_k\;\mathbb{E}\left[\left(\left\langle\int_0^t e^{\frac{s}{\mu}}\sigma(s,X_s^\mu)dW_s\right\rangle_t\right)^{\frac{k}{2}}\right]
\end{aligned}$$

$$
\begin{aligned}
&= 5^{k-1}e^{-\frac{tk}{\mu}}M_k\,\mathbb{E}\left[\left(\int_0^t e^{\frac{2s}{\mu}}\sigma^2(s,X_s^\mu)ds\right)^{\frac{k}{2}}\right]\\
&\le 5^{k-1}e^{-\frac{tk}{\mu}}M_k\,\|\sigma\|_T^k\left(\int_0^t e^{\frac{2s}{\mu}}ds\right)^{\frac{k}{2}}\\
&\le 5^{k-1}e^{-\frac{tk}{\mu}}M_k\,\|\sigma\|_T^k\left(\frac{\mu}{2}e^{\frac{2t}{\mu}}\right)^{\frac{k}{2}}\\
&= \mu^{\frac{k}{2}}5^{k-1}M_k\,\|\sigma\|_T^k\,2^{-\frac{k}{2}}
\end{aligned}
$$

Hence we conclude

$$
\begin{aligned}
&\mathbb{E}\left[|X_t^\mu - X_t|^k\right]\\
&\le 5^{k-1}\mu^k|\xi_2|^k + 5^{k-1}T^{k-1}K_b^k\int_0^t \mathbb{E}\left[|X_s^\mu - X_s|^k\right]ds\\
&\quad +5^{k-1}L_kT^{\frac{k}{2}}K_\sigma^k\int_0^t \mathbb{E}\left[|X_s^\mu - X_s|^k\right]ds\\
&\quad +5^{k-1}\mu^k\|b\|_T^k + 5^{k-1}\mu^{\frac{k}{2}}M_k\|\sigma\|_T^k\,2^{-\frac{k}{2}}\\
&\le \mu^{\frac{k}{2}}\left(5^{k-1}|\xi_2|^k + 5^{k-1}\|b\|_T^k + 5^{k-1}M_k\|\sigma\|_T^k\,2^{-\frac{k}{2}}\right)\\
&\quad +\left(5^{k-1}T^{k-1}K_b^k + 5^{k-1}L_kT^{\frac{k}{2}}K_\sigma^k\right)\int_0^t \mathbb{E}\left[|X_s^\mu - X_s|^k\right]ds
\end{aligned}
$$

and thus by Gronwall's Lemma 2.11

$$
\begin{aligned}
\mathbb{E}\left[|X_t^\mu - X_t|^k\right] &\le \mu^{\frac{k}{2}}\left(5^{k-1}|\xi_2|^k + 5^{k-1}\|b\|_T^k + 5^{k-1}M_k\|\sigma\|_T^k\,2^{-\frac{k}{2}}\right)\\
&\quad\cdot\Bigg(1+\left(5^{k-1}T^{k-1}K_b^k + 5^{k-1}T^{\frac{k}{2}}K_\sigma^k\right)\\
&\quad\cdot\int_0^t \exp\left\{5^{k-1}T^{k-1}K_b^k + 5^{k-1}L_kT^{\frac{k}{2}}K_\sigma^k\right\}ds\Bigg)\\
&\le \mu^{\frac{k}{2}}\left(5^{k-1}|\xi_2|^k + 5^{k-1}\|b\|_T^k + 5^{k-1}M_k\|\sigma\|_T^k\,2^{-\frac{k}{2}}\right)\\
&\quad\cdot\Bigg(1+\left(5^{k-1}T^kK_b^k + 5^{k-1}L_kT^{\frac{k}{2}+1}K_\sigma^k\right)\\
&\quad\cdot\exp\left\{5^{k-1}T^{k-1}K_b^k + 5^{k-1}T^{\frac{k}{2}}K_\sigma^k\right\}\Bigg)\\
&= \mu^{\frac{k}{2}}c(k,|\xi_2|,T,\|b\|_T,\|\sigma\|_T,K_b,K_\sigma),
\end{aligned}
$$

and the statement is proved.

Now, let us get back to our problem. We use formula (3.10) (from the proof of Proposition 3.2) and get

$$\begin{aligned} X_t^\mu - X_t &= \int_0^t \left[e^{-\frac{s}{\mu}}\xi_2 + \frac{1}{\mu}e^{-\frac{s}{\mu}}\left(\int_0^s e^{\frac{r}{\mu}}b(r,X_r^\mu)dr\right) - b(s,X_s) \right. \\ &\quad \left. +\frac{1}{\mu}e^{-\frac{s}{\mu}}\left(\int_0^s e^{\frac{r}{\mu}}\sigma(r,X_r^\mu)dW_r\right)\right]ds - \int_0^t \sigma(s,X_s)dW_s. \end{aligned}$$

Thus we have for all $p, n \in \mathbb{N}$, $p \geq 2$, $p = 2n$ by Ito's formula:

$$\begin{aligned} (X_t^\mu - X_t)^p &= p\int_0^t (X_s^\mu - X_s)^{p-1} d\left(X_s^\mu - X_s\right) \\ &\quad +\frac{1}{2}p\left(p-1\right)\int_0^t (X_s^\mu - X_s)^{p-2} d\left\langle X_\cdot^\mu - X_\cdot\right\rangle_s \\ &= p\int_0^t (X_s^\mu - X_s)^{p-1}\left[e^{-\frac{s}{\mu}}\xi_2 + \frac{1}{\mu}e^{-\frac{s}{\mu}}\left(\int_0^s e^{\frac{r}{\mu}}b(r,X_r^\mu)dr\right)\right. \\ &\quad \left. +\frac{1}{\mu}e^{-\frac{s}{\mu}}\left(\int_0^s e^{\frac{r}{\mu}}\sigma(r,X_r^\mu)dW_r\right) - b(s,X_s)\right]ds \\ &\quad +\frac{1}{2}p\left(p-1\right)\int_0^t (X_s^\mu - X_s)^{p-2}\sigma^2(s,X_s)ds \\ &\quad -p\int_0^t (X_s^\mu - X_s)^{p-1}\sigma(s,X_s)dW_s \end{aligned}$$

Fix $T > 0$. Then we have for any stopping time $\tau \leq T$ by the above considerations (used in the first step), Hölder's inequality (see [2] p. 71) (used in the second step) and the Ito isometry (also used in the second step):

$$\begin{aligned} &\mathbb{E}\left[|X_\tau^\mu - X_\tau|^p\right] \\ &\leq p\int_0^T \mathbb{E}\Bigg[|X_s^\mu - X_s|^{p-1}\Big(e^{-\frac{s}{\mu}}|\xi_2| + \frac{1}{\mu}e^{-\frac{s}{\mu}}\int_0^s e^{\frac{r}{\mu}}\left\|b\right\|_T dr \\ &\quad +\frac{1}{\mu}e^{-\frac{s}{\mu}}\left|\int_0^s e^{\frac{r}{\mu}}\sigma(r,X_r^\mu)dW_r\right| + \left\|b\right\|_T\Big)\Bigg]ds \\ &\quad +\frac{1}{2}p\left(p-1\right)\left\|\sigma\right\|_T^2\int_0^T \mathbb{E}\left[|X_s^\mu - X_s|^{p-2}\right]ds \end{aligned}$$

$$
\begin{aligned}
&= && p\int_0^T \mathbb{E}\left[\left|X_s^\mu - X_s\right|^{p-1}\right]\left(e^{-\frac{s}{\mu}}\left|\xi_2\right| + \left(2-e^{-\frac{s}{\mu}}\right)\|b\|_T\right)ds \\
& && +\frac{1}{\mu}\int_0^T \left(\mathbb{E}\left[\left|X_s^\mu - X_s\right|^{2(p-1)}\right]\right)^{\frac{1}{2}} \\
& && \cdot\frac{1}{\mu}e^{-\frac{s}{\mu}}\left(\mathbb{E}\left[\int_0^s e^{\frac{2r}{\mu}}\sigma^2(r, X_r^\mu)dr\right]\right)^{\frac{1}{2}} ds \\
& && +\frac{1}{2}p\left(p-1\right)\|\sigma\|_T^2\int_0^T \mathbb{E}\left[\left|X_s^\mu - X_s\right|^{p-2}\right]ds \\
&\leq && \left(\sup_{0\leq s\leq T}\mathbb{E}\left[\left|X_s^\mu - X_s\right|^{2(p-1)}\right]\right)^{\frac{1}{2}}\left(pT\mu\left|\xi_2\right| + 2pT\|b\|_T\right) \\
& && +\left(\sup_{0\leq s\leq T}\mathbb{E}\left[\left|X_s^\mu - X_s\right|^{2(p-1)}\right]\right)^{\frac{1}{2}}\|\sigma\|_T\int_0^T \frac{1}{\sqrt{2\mu}}e^{-\frac{s}{\mu}}e^{\frac{s}{\mu}}ds \\
& && +\frac{1}{2}p\left(p-1\right)\|\sigma\|_T^2\left(\mathbb{E}\left[\left|X_s^\mu - X_s\right|^{2(p-2)}\right]\right)^{\frac{1}{2}} \\
&\leq && c(p,\left|\xi_2\right|, T, \|b\|_T, \|\sigma\|_T, K_b, K_\sigma)\left[\mu^{p+1} + \mu^{p-1}\mu^{-\frac{1}{2}} + \mu^{p-2}\right] \\
&\leq && c\left[3\mu^{\frac{p-2}{2}}\right] \\
&\xrightarrow{\mu\to 0} && 0 \quad \text{for } p > 2.
\end{aligned}
$$

Now we define for any $\epsilon > 0$ the stopping time

$$
\tau_\epsilon := T \wedge \left\{t \geq 0 \middle| \left|X_t^\mu - X_t\right|^p \geq \epsilon\right\}.
$$

Hence

$$
\left\{\sup_{0\leq t\leq T}\left|X_t^\mu - X_t\right|^p \geq \epsilon\right\} = \left\{\left|X_{\tau_\epsilon}^\mu - X_{\tau_\epsilon}\right|^p \geq \epsilon\right\}
$$

And thus by Chebyshev's inequality (see [24], part I, page I.6.1) we obtain for all $\epsilon \geq 0$ and all $p > 2$

$$
\begin{aligned}
\mathbb{P}\left[\sup_{0\leq t\leq T}\left|X_t^\mu - X_t\right| \geq \epsilon\right] &= \mathbb{P}\left[\sup_{0\leq t\leq T}\left|X_t^\mu - X_t\right|^p \geq \epsilon^p\right] \\
&\leq \frac{1}{\epsilon^p}\mathbb{E}\left[\left|X_{\tau_\epsilon}^\mu - X_{\tau_\epsilon}\right| \geq \epsilon\right] \\
&\leq \frac{1}{\epsilon^p}3c\mu^{\frac{p-2}{2}} \quad \xrightarrow{\mu\to 0} \; 0.
\end{aligned}
$$

Thus the proof of the Lemma is complete. $\square$

Remark 3.6. *Consider the situation of Lemma 3.5. Due to the fact that we always have a compact interval (namely, $[0,T]$) and having shown that each of the according terms of (i) - (iii) is bounded, we can conclude that each supremum equals the according maximum and thus the proved inequalities are valid for the according maxima as well.*

Corollary 3.7. *(Smoluchowski-Kramers approximation)*
The solution X_t^μ of (3.1) converges in probability uniformly on $[0,T]$ to the solution X_t of (3.5).
That is: For all $T > 0$, ξ_0, $\xi_1 \in \mathbb{R}$:

$$\lim_{\mu \searrow 0} P\left[\sup_{0\le t\le T} |X_t^\mu - X_t| > \epsilon\right] = 0 \quad \forall \quad \epsilon > 0. \tag{3.16}$$

Proof. From 3.5(iii) we have that:

$$P\left[\sup_{0\le t\le T} |X_t^\mu - X_t| > \epsilon\right] \le \frac{c\mu}{\epsilon^2}\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right].$$

Hence it follows

$$\begin{aligned}\lim_{\mu \searrow 0} P\left[\sup_{0\le t\le T} |X_t^\mu - X_t| > \epsilon\right] &\le \lim_{\mu \searrow 0}\frac{c\mu}{\epsilon^2}\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right] \\ &= 0,\end{aligned}$$

since ϵ, $c = 5e^{\left(TK_b^2 + K_\sigma^2\right)T}$, T, $\left|\xi_2\right|, \left\|b\right\|, \left\|\sigma\right\|$ are all not depending on μ and finite. □

This 'Smoluchowski-Kramers approximation' is well-known to mathematicians and physicians. It has been discussed extensively (see for example [20] for a quite early discussion in 1967, or [17] for a more recent one, or many more). Researchers came also up with other proofs than given here, but we present the idea of Freidlin in [11] for the proof of the Smoluchowski-Kramers approximation.

The Smoluchowski-Kramers approximation is the justification for using the first order equation (3.5) to describe the motion of a small particle disturbed by a Wiener process instead of using the Newton equation (3.1). Obviously it is much easier and much more convenient to use the solution of the first order stochastic differential equation X_t.
For our further consideration, our focus is motivated from a physical point of view: As we already explained in Section 3.1, it is reasonable to consider a smoother process V_t^δ than to consider a Wiener process. We are interested in

finding a first order stochastic differential equation in order to approximate the solution $X_t^{\mu,\delta}$ of the according stochastic differential system of equations, i.e. something like a 'generalized Smoluchowski-Kramers approximation'. This is the aim of this chapter.

3.2.2 The Convergence Result

Now we go back to our problem. We consider the Newton equation (3.2), which is perturbed by a smoother process V_t^δ. We take the following assumptions concerning V_t^δ :

(i) V_t^δ can be written as
$$V_t^\delta = \frac{1}{\sqrt{\delta}} \int_0^t \xi_{\frac{s}{\delta}} ds, \tag{3.17}$$
where ξ_s is a mean zero stationary Gaussian process with correlation function $r(|s|)$.

(ii) The correlation function $r(|s|)$ satisfies the conditions of Proposition 2.15, i.e. for $s \to 0$ there exist constants $\lambda_2, \lambda_4 < \infty, \lambda_2 \geq 0, \lambda_4 > 0$ such that:
$$r(s) = 1 - \frac{\lambda_2}{2}s^2 + \frac{\lambda_4}{24}s^4 + o(s^4),$$
and for $s \to \infty$ there exists $\alpha > 0$ such that:
$$r(s) = O(s^{-\alpha}).$$

(iii) $V_t^\delta(\omega)$ is piecewise continuously differentiable in t for every $\omega \in \Omega$ (in particular, $\dot{V}_t^\delta = \frac{d}{dt}V_t^\delta$ exists).

(iv) For every $\omega \in \Omega$, $k \in \mathbb{N}$ and $t \in \mathbb{R}_+$, V^δ satisfies $V_{t+k\delta}^\delta = V^\delta\left(\omega\left(t+k\delta\right)\right) = V_t^\delta\left(\omega\left(t+k\delta\right) - \omega\left(t\right)\right) + \omega\left(k\delta\right)$.

(v) V_0^δ is measurable and satisfies $\mathbb{E}\left[V_0^\delta\right] = 0$ and $\mathbb{E}\left[\left|V_0^\delta\right|^6\right] \leq c\delta^3$ for a certain constant $c \in \mathbb{R}_+$.

(vi) $\dot{V}_t^\delta$ satisfies $\mathbb{E}\left[\left(\int_0^\delta \left|\dot{V}_s^\delta\right| ds\right)^6\right] \leq c\delta^3$ for a certain constant $c \in \mathbb{R}_+$, and $\sup_{t\in[0,T]} \left|\dot{V}_t^\delta\right| \leq C_\delta$ for some constant $C_\delta \in \mathbb{R}$.

Condition (i) is stated for technical reasons, since Gaussian processes are well-studied and we use their properties in the following proofs. Heuristically, Condition (ii) imposes regularity conditions on the growth of the correlation function, given by the factors above. Condition (iii) postulates that the noise term V_t^δ is 'sufficiently smooth' (otherwise it would not be differentiable). Condition (iv) heuristically is a 'shift-invariance' of the noise, and Conditions (v) and (vi) are also stated for technical reasons.

As is stated in [17], page 393, Conditions (iii) - (vi) imply $\mathcal{L}^2-$convergence of V_t^δ to W_t, i.e. for every $T > 0$ we have that

$$\lim_{\delta \searrow 0} \mathbb{E}\left[\max_{0\leq t\leq T} \left|V_t^\delta - W_t\right|^2\right] = 0.$$

Additionally, assume that V_t^δ is such that

$$\mathbb{P}\left[\lim_{\delta\to 0} \sup_{0\leq t\leq T} \left|W_t - V_t^\delta\right| = 0\right] = 1,$$

i.e. V_t^δ converges uniformly in $t \in [0,T]$ to W_t P-a.s.

The following theorem presents the main result of this chapter. It is taken from [11].

Theorem 3.8. *Assume that the stochastic process V_t^δ satisfies the above assumptions. Suppose that b and σ are Lipschitz with Lipschitz constants K_b and K_σ, respectively, and that $\sigma \in \mathcal{C}^2(\mathbb{R})$, and $b(x)$, $\sigma(x)$, $\sigma'(x)$, $\sigma''(x)$ bounded. Further suppose that $b(x), \sigma(x), \sigma'(x)$ are continuous over $\mathbb{R}$, $(\sigma''(x))^2$ is also Lipschitz, and $\sigma(x) \geq \beta > 0$ or $-\sigma(x) \geq \beta > 0$ for some constant $\beta \in \mathbb{R}_+$ for all $x \in \mathbb{R}$.*
Then the solution $X_t^{\mu,\delta}$ of (3.2) converges in probability uniformly on $[0,T]$ to the solution

(i) X_t of (3.5) if first $\delta \searrow 0$ and then $\mu \searrow 0$, i.e.

$$\forall\ \epsilon > 0 : \lim_{\mu\searrow 0} \lim_{\delta\searrow 0} \mathbb{P}\left[\sup_{0\leq t\leq T} \left|X_t^{\mu,\delta} - X_t\right| > \epsilon\right] = 0.$$

More precisely: $X_t^{\mu,\delta}$ converges in probability uniformly on $[0,T]$ to X_t, if for sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \mu_n = \lim_{n\to\infty} \delta_n = 0$ we have that $\delta_n < f(\mu_n, h)$ for all $n \in \mathbb{N}$ for a certain function $f(\mu_n, h)$, $f : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, $h = 2(c_3\mu_n)^{\frac{1}{3}}$ ($c_3 \in \mathbb{R}$ constant).
That is: For any $h > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$:

$$\mathbb{P}\left[\sup_{0\leq t\leq T} \left|X_t^{\mu_n,\delta_n} - X_t\right| > h\right] \leq h, \text{ with } (\mu_n)_{n\in\mathbb{N}}, (\delta_n)_{n\in\mathbb{N}} \text{ as above.}$$

(ii) $\tilde{X}_t$ *of (3.6) if* $(\mu,\delta) \searrow 0$ *such that* $\lim_{(\mu,\delta)\searrow 0} \frac{1}{\delta^2}\mu e^{\frac{\hat{K}}{\delta}} = 0$ *for a certain constant* $\hat{K} \in \mathbb{R}_+$.
More precisely: $X_t^{\mu,\delta}$ *converges in probability uniformly on* $[0,T]$ *to* $\tilde{X}_t$, *if for sequences* $((\mu_n),(\delta_n))_{n\in\mathbb{N}}$ *with* $\lim_{n\to\infty}\mu_n = \lim_{n\to\infty}\delta_n = 0$ *we have that* $\lim_{n\to\infty}\mu_n\frac{1}{\delta_n^2}\exp\left\{\frac{\hat{K}}{\delta_n}\right\} = 0$ *for a certain constant* $\hat{K} \in \mathbb{R}_+$.
That is: For any $h > 0$ *there exists* $n_0 \in \mathbb{N}$ *such that for all* $n \geq n_0$:
$\mathbb{P}\left[\sup_{0\leq t\leq T}\left|X_t^{\mu_n,\delta_n} - X_t\right| > h\right] \leq h,$ *with* $((\mu_n),(\delta_n))_{n\in\mathbb{N}}$ *as above.*

Remark 3.9. *(i) The version of Theorem 3.8 presented here states an assumption which is stronger than the one stated in [11]: Freidlin assumes that* $\lim_{(\mu,\delta)\searrow 0}\mu e^{\frac{1}{\delta}} = 0$, *whereas we assume that* $\lim_{(\mu,\delta\searrow 0)}\frac{1}{\delta^2}\mu e^{\frac{\hat{K}}{\delta}} = 0$ *for a certain constant* $\hat{K} \in \mathbb{R}_+$. *We need this stronger assumption in the proof of Theorem 3.8(ii).*

(ii) In order to proof Theorem 3.8(i), we need to show the following:

$$\lim_{\delta\searrow 0}(X_t^{\mu,\delta}, Y_t^{\mu,\delta}) = (X_t^{\mu}, Y_t^{\mu}) \text{ in probability uniformly on } [0,T]\,. \quad (3.18)$$

Once (3.18) is shown, we are able to prove Theorem 3.8(i) with the help of the Smoluchowski-Kramers approximation 3.7.
A possible idea to prove (3.18) is to generalize the corresponding result for one-dimensional differential equations to two-dimensional differential equations, since we can interprete $(X_t^{\mu,\delta}, Y_t^{\mu,\delta})$ *as the two-dimensional solution of the two-dimensional differential equation (3.2). This one-dimensional result was first shown by Wong and Zakai in [26] (which is Lemma 3.12 in this thesis), and an alternative proof is presented by Karatzas and Shreve in [17], p. 298 - 299. The idea given in [11] is to generalize the result of Wong and Zakai in [26].*
Unfortunately, we encountered a problem with this idea: The proof of Wong and Zakai in [26] uses crucially that the noise is non-degenerated (meaning that the function $\sigma : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ *associated with the stochastic noise fulfils* $\sigma(t,x) \geq \beta > 0$ *or* $-\sigma(t,x) \geq \beta > 0$ *for all* $x \in \mathbb{R}$, $t \in [0,T]$, $\beta \in \mathbb{R}_+$, *see Lemma 3.12, condition c)). The assumption is needed several times in the proof, since the function* σ *occurs in the denominator of different terms (see the proof of Lemma 3.12) and hence the assumption cannot be omitted. But by interpreting our differential equation as a two-dimensional one, the according function (called* g *for the moment,* $g : \mathbb{R}\times\mathbb{R}^2 \longrightarrow \mathbb{R}$*) is then given by*

$$g(t,x,y) = g(t,x,y) := \begin{bmatrix} 0 \\ \frac{1}{\mu}\sigma(t,x) \end{bmatrix},$$

as we already argued in the proof of Corollary 2.7(ii). Thus obviously the according non-degeneracy condition is violated and it seems impossible to generalize the result of Wong and Zakai for our two-dimensional situation, as was agreed upon by Mark Freidlin in an email-correspondence.
Similarly, it is not possible to generalize the proof of Karatzas and Shreve given in [17] in the desired way, as they even state themselves in their 2.25 Remark on page 299 (following the proof of the one-dimensional case). The reason in this case is that the proof for one dimension uses an auxiliary function, for which there might not be a counterpart in dimensions greater than 1.
Fortunately, a far more general result for differential equations in higher dimension is shown in [13] by Ikeda und Watanabe. We were able to identify our situation with the one considered in [13] and use their result in order to prove the above statement. The result of Ikeda and Watanabe will be stated for our case in the proof of Theorem 3.8(i).

We discuss and interpret Theorem 3.8 shortly. Obviously the interesting thing (and surprising at the first glance) of this theorem is the fact that the limits differ depending on the way μ and δ approach zero.
The statements may seem a bit messy at the first glance, so let us think about intuition and heuristics. Both give a convergence property for the stochastic Newton equation if both, μ and δ go to zero. Theorem 3.8(i) states uniformly convergence in probability to the Ito first order stochastic differential equation given the convergence of μ, δ is such that δ is smaller than a certain function f depending on μ. In detail: Consider two sequences δ_n and μ_n, both converging to zero and satisfying $\delta_n < f(\mu_n, 2(c_2\mu_n)^{\frac{1}{3}})$ for all $n \in \mathbb{N}$. For such sequences we have for $n \to \infty$ uniformly convergence in probability to the Ito stochastic first order differential equation. Now, Theorem 3.8(ii) states uniform convergence in probability to the Stratonovich first order stochastic differential equation given the convergence of μ, δ is such that $\lim_{(\mu,\delta)\searrow 0} \frac{1}{\delta^2}\mu e^{\frac{\hat{K}}{\delta}} = 0$ for some constant $\hat{K} \in \mathbb{R}$. Again, in detail: Consider two sequences δ_n and μ_n, both converging to zero and satisfying $\lim_{n\to\infty} \frac{1}{\delta_n^2}\mu_n e^{\frac{\hat{K}}{\delta_n}} = 0$. For such sequences we have for $n \to \infty$ uniformly convergence in probability to the Stratonovich stochastic first order differential equation. Note that the assumption $\lim_{(\mu,\delta)\searrow 0} \frac{1}{\delta^2}\mu e^{\frac{\hat{K}}{\delta}} = 0$ is quite strong in the sense that 'μ has to converge to zero a lot faster than δ', but we will see in the proof that we need this strong assumption.

The next Corollary - which is more convenient to handle than the above statements - follows directly once the proofs are shown. Moreover, the intuition is clearer. This corollary is also taken from [11].

Corollary 3.10. *Assume that the stochastic process V_t^δ satisfies the above assumptions. Suppose that b and σ are Lipschitz with Lipschitz constants K_b and K_σ, respectively, and that $\sigma \in \mathcal{C}^2(\mathbb{R})$, and $b(x)$, $\sigma(x)$, $\sigma'(x)$, $\sigma''(x)$ bounded. Further suppose that $b(x), \sigma(x), \sigma'(x)$ are continuous over $\mathbb{R}$, $(\sigma''(x))^2$ is also Lipschitz, and $\sigma(x) \geq \beta > 0$ or $-\sigma(x) \geq \beta > 0$ for some constant $\beta \in \mathbb{R}_+$ for all $x \in \mathbb{R}$.*
Then the solution $X_t^{\mu,\delta}$ of (3.2) converges

- *(i) in probability uniformly on $[0,T]$ to the solution X_t of (3.5), if first $\delta \searrow 0$ and then $\mu \searrow 0$.*
- *(ii) in probabiliy uniformly on $[0,T]$ to the solution $\tilde{X}_t$ of (3.6), if first $\mu \searrow 0$ and then $\delta \searrow 0$.*

Proof. (i) The statement is part of Theorem 3.8(i).

(ii) Consider a sequence $(\mu_m)_{m\in\mathbb{N}}$ such that $\lim_{m\to\infty} \mu_m = 0$ and $\mu_m \geq 0$ for all m. Now choose a sequence $(\delta_n)_{n\in\mathbb{N}}$, $\delta_n \geq 0$ for all n and $\lim_{n\to\infty} \delta_n = 0$, such that $\lim_{n\to\infty} \lim_{m\to\infty} \frac{1}{\delta_n}\mu_m e^{\frac{1}{\delta_n}} = 0$. Due to the construction of the sequences Theorem 3.8(ii)is applicable and the statement is proved. □

Proof. of (3.8)(i):
In order to prove (i), we use a special case of a result shown by Ikeda and Watanabe in [13]. The following lemma states the result of Ikeda's and Watanabe's Theorem 7.2 on page 410 in [13] for the special case $r = 1$, $d = 2$, and the two-dimensional disturbing noise $\left(V_t^\delta, V_t^\delta\right)$ (then the postulated condition Assumption 7.1 on page 394 of [13] is trivially fulfilled, and, moreover, the correction term (additionally to the one caused by the Stratonovich stochastic integral) in the limiting stochastic differential equation is zero).
The proof is very technical and shall therefore not be repeated here. It can be found in [13], pages 410 - 417.

Lemma 3.11. *Fix $T > 0$ and let $f_i : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R} \in \mathcal{C}^1(\mathbb{R}^2)$, and $g_i : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R} \in \mathcal{C}^2(\mathbb{R}^2), i = 1,2$, and all bounded. Further suppose that $f := (f_1, f_2)^T$ and $g := (g_1, g_2)^T$ satisfy the the growth condition (2.11) and* *the global Lipschitz condition (2.12). Let (X_t^δ, Y_t^δ) be the solution of the two-dimensional differential equation*

$$\begin{aligned} dX_t^\delta &= f_1(X_t^\delta, Y_t^\delta)dt + g_1(X_t^\delta, Y_t^\delta)dV_t^\delta, \\ dY_t^\delta &= f_2(X_t^\delta, Y_t^\delta)dt + g_2(X_t^\delta, Y_t^\delta)dV_t^\delta, \end{aligned}$$

and let $(\tilde{X}_t, \tilde{Y}_t)$ be the solution of the two-dimensional stochastic differential equation

$$\begin{aligned} d\tilde{X}_t &= f_1(\tilde{X}_t, \tilde{Y}_t)dt + g_1(\tilde{X}_t, \tilde{Y}_t) \circ dW_t, \\ d\tilde{Y}_t &= f_2(\tilde{X}_t, \tilde{Y}_t)dt + g_2(\tilde{X}_t, \tilde{Y}_t) \circ dW_t. \end{aligned}$$

Then

$$\lim_{\delta \searrow 0} \mathbb{E}\left[\sup_{0\le t\le T} \left|X_t^\delta - \tilde{X}_t\right|^2\right] = 0 \tag{3.19}$$

and

$$\lim_{\delta \searrow 0} \mathbb{E}\left[\sup_{0\le t\le T} \left|Y_t^\delta - \tilde{Y}_t\right|^2\right] = 0 \tag{3.20}$$

Proof. Existence and uniqueness of $((X_t^\delta, Y_t^\delta))$ and $(\tilde{X}_t, \tilde{Y}_t)$ follows by Theorem 2.6(i) and (ii) (after using Remark 2.5). Then the proof of the statement can be found in [13], pages 410 - 417. $\square$

Consider a sequence $(\delta_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \delta_n = 0$ and remember that $(X_t^{\mu,\delta_n}, Y_t^{\mu,\delta_n})$ is determined by the following stochastic differential equation:

$$dX_t^{\mu,\delta_n} = Y_t^{\mu,\delta_n} dt,$$

$$\mu dY_t^{\mu,\delta_n} = b(X_t^{\mu,\delta_n})dt + \sigma(X_t^{\mu,\delta_n})dV_t^{\delta_n} - Y_t^{\mu,\delta_n} dt,$$

$$X_0^{\mu,\delta_n} = \xi_1, \; Y_0^{\mu,\delta_n} = \xi_2.$$

We identify this system of stochastic differential equations with the situation of Lemma 3.11 by fixing $\mu > 0$ and then setting

$$f_1(x, y) := y, \quad g_1(x, y) := 0, \quad f_2(x, y) := b(x) - y, \quad g_2(x, y) := \sigma(x).$$

The conditions of Lemma 3.11 are fulfilled thanks to our assumptions concerning the functions b and σ. Thus we can apply Lemma 3.11 and conclude that

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{0\le t\le T} \left|X_t^{\mu,\delta_n} - \tilde{X}_t^\mu\right|^2\right] = 0,$$

where $(\tilde{X}_t^\mu, \tilde{Y}_t^\mu)$ is the solution of the following stochastic differential equation:

$$d\tilde{X}_t^\mu = \tilde{Y}_t^\mu dt,$$
$$\mu d\tilde{Y}_t^\mu = b(\tilde{X}_t^\mu)dt + \sigma(\tilde{X}_t^\mu) \circ dW_t - \tilde{Y}_t^\mu dt,$$
$$\tilde{X}_0^\mu = \xi_1,\ \tilde{Y}_0^\mu = \xi_2.$$

Applying Markov's inequality (see [24], part I, page I.5.2) yields for any $h > 0$

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - \tilde{X}_t^\mu\right| > h\right] &\le \lim_{n\to\infty} \frac{1}{h^2}\mathbb{E}\left[\left(\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - \tilde{X}_t^\mu\right|\right)^2\right] \\
&= \frac{1}{h^2}\lim_{n\to\infty} \mathbb{E}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - \tilde{X}_t^\mu\right|^2\right] \\
&= 0,
\end{aligned}$$

which means that X_t^{μ,δ_n} converges in probability uniformly on $[0,T]$ to $\tilde{X}_t^\mu$. Now, remember that (X_t^μ, Y_t^μ) is defined to be the solution of the system (3.1):

$$dX_t^\mu = Y_t^\mu dt$$
$$\mu dY_t^\mu = b(X_t^\mu)dt + \sigma(X_t^\mu)dW_t - dX_t^\mu$$
$$X_0^\mu = \xi_1,\ Y_0^\mu = \xi_2.$$

Since X_t^μ has a continuous derivative (namely Y_t^μ), we conclude with the help of Remark 3.1 that $\int_0^T \sigma(X_t^\mu)dW_t = \oint_0^T \sigma(X_t^\mu) \circ dW_t$, i.e. the Ito and the Stratonovich integral coincide. Hence it follows that $\tilde{X}_t^\mu = X_t^\mu$ and $\tilde{Y}_t^\mu = Y_t^\mu$, and thus $X_t^{\mu,\delta}$ converges in probability uniformly on $[0,T]$ to X_t^μ (and consequently $Y_t^{\mu,\delta}$ converges in probability uniformly on $[0,T]$ to Y_t^μ).

Then

$$\begin{aligned}
\mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t\right| > h\right] &= \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t^\mu + X_t^\mu - X_t\right| > h\right] \\
&\le \mathbb{P}\left[\sup_{0\le t\le T}\left(\left|X_t^{\mu,\delta_n} - X_t^\mu\right| + |X_t^\mu - X_t|\right) > h\right].
\end{aligned}$$

Define

$$\Omega_A := \left\{\omega \in \Omega \;\middle|\; \sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t^\mu\right| > \frac{h}{2}\right\}$$

and

$$\Omega_B := \left\{\omega \in \Omega \;\middle|\; \sup_{0\le t\le T}|X_t^\mu - X_t| > \frac{h}{2}\right\}.$$

Noting that

$$\left\{\omega \in \Omega \,\middle|\, \sup_{0\le t\le T} \left(\left|X_t^{\mu,\delta_n} - X_t^{\mu}\right| + \left|X_t^{\mu} - X_t\right|\right) > h\right\} \subseteq \Omega_A \cup \Omega_B,$$

we conclude by using that $\mathbb{P}\left[\Omega_A \cup \Omega_B\right] \le \mathbb{P}\left[\Omega_A\right] + \mathbb{P}\left[\Omega_B\right]$ and continuing the above calculation

$$\begin{aligned} \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t\right| > h\right] \ &\le \ \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t^{\mu}\right| > \frac{h}{2}\right] \\ &\quad + \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu} - X_t\right| > \frac{h}{2}\right] \end{aligned} \tag{3.21}$$

Thus our task now is to estimate the two terms on the right hand side. Let us begin with the first one: Applying the above considerations, we conclude that $X_t^{\mu,\delta_n} \overset{n\to\infty}{\longrightarrow} X_t^{\mu}$ in probability uniformly in $[0,T]$, i.e.

$$\forall\ \frac{h}{2} > 0\ \ \exists\ n_0 \in \mathbb{N} : \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t^{\mu}\right| > \frac{h}{2}\right] \le \frac{h}{2} \quad \forall\ n \ge n_0.$$

Finally, because of the fact that such an n_0 exists for all $h \in \mathbb{R}$, the function $\tilde{f} = \tilde{f}(\mu, h)$ defined by the following expression exists:

$$\mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu,\delta_n} - X_t^{\mu}\right| > \frac{h}{2}\right] \le \frac{h}{2} \quad \forall\ n \ge \tilde{f}(\mu, h),$$

which is the desired estimation for the first term on the right hand side.
In order to estimate the second term (which is the 'Smoluchowski-Kramers approximation'), we use Lemma 3.5(iii) and obtain:

$$\begin{aligned} \mathbb{P}\left[\sup_{0\le t\le T}\left|X_t^{\mu} - X_t\right| > \frac{h}{2}\right] \ &\le \ \frac{c\mu}{\left(\frac{h}{2}\right)^2}\left[\mu\left|\xi_2\right|^2 + \mu\left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right] \\ &\le \ \frac{4\mu}{h^2}\underbrace{c\left[\left|\xi_2\right|^2 + \left\|b\right\|^2 + \frac{1}{2}\left\|\sigma\right\|^2\right]}_{=:c_2} \\ &= \ \frac{4c_2\mu}{h^2}. \end{aligned}$$

Plugging these results into inequality (3.21) and choosing $h = 2\sqrt[3]{\mu c_2}$ (i.e.

$\mu = \frac{h^3}{8c_2}$), it follows that for all $n \geq \tilde{f}(\mu, 2\sqrt[3]{\mu c_2})$

$$\begin{aligned}
\mathbb{P}\left[\max_{0\leq t\leq T}\left|X_t^{\mu,\delta_n} - X_t\right| > h\right] &\leq \mathbb{P}\left[\sup_{0\leq t\leq T}\left|X_t^{\mu,\delta_n} - X_t\right| > h\right] \\
&\leq \mathbb{P}\left[\sup_{0\leq t\leq T}\left|X_t^{\mu,\delta_n} - X_t^{\mu}\right| > \frac{h}{2}\right] \\
&\qquad + \mathbb{P}\left[\sup_{0\leq t\leq T}\left|X_t^{\mu} - X_t\right| > \frac{h}{2}\right] \\
&\leq \frac{h}{2} + \frac{4c_2\mu}{h^2} \\
&\leq \frac{h}{2} + \frac{4c_2}{h^2}\frac{h^3}{8c_2} = \frac{h}{2} + \frac{h}{2} = h
\end{aligned}$$

Since $\lim_{n\to\infty}\delta_n = 0$, it follows that there exists a function $f(\mu, h)$ with $h = 2\sqrt[3]{\mu c_2}$, such that for all $\delta < f(\mu, h) = f(\mu, 2\sqrt[3]{\mu c_2}) =: f(\mu)$

$$\mathbb{P}\left[\max_{0\leq t\leq T}\left|X_t^{\mu,\delta_n} - X_t\right| > h\right] \leq h,$$

hence Theorem 3.8(i) is proved. □

Now we want to prove the second part of the theorem.

Proof. of Theorem 3.8(ii).
Remember that X_t^δ is defined to be the strong solution of the first order differential equation (3.7)

$$dX_t^\delta = b(X_t^\delta)dt + \sigma(X_t^\delta)dV_t^\delta, \quad dX_0^\delta = \xi_1.$$

The solution concept for the above stochastic differential equation is analogous to Definition 2.4 (i).

The proof will be done in two steps.

Claim 1 Let $(\mu, \delta) \searrow 0$ such that $\lim_{(\mu,\delta)\searrow 0}\frac{1}{\delta^2}\mu e^{\frac{\hat{K}}{\delta}} = 0$. Then

$$\lim_{(\mu,\delta)\searrow 0}\mathbb{P}\left[\max_{0\leq t\leq T}\left|X_t^{\mu,\delta} - X_t^\delta\right| = 0\right] = 1.$$

Claim 2

$$\lim_{\delta\searrow 0}\mathbb{P}\left[\max_{0\leq t\leq T}\left|X_t^\delta - \tilde{X}_t\right| = 0\right] = 1.$$

Having shown the two claims Theorem 3.8(ii)follows.

Proof of Claim 1. Consider sequences $((\mu_n), (\delta_n))_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \frac{1}{\delta_n^2}\mu_n e^{\frac{\hat{K}}{\delta_n}} = 0$.

Let us first consider our stochastic process ξ_t. It follows by Proposition 2.15 that

$$\lim_{T\to\infty} \mathbb{P}\left[\max_{0\le t\le T} |\xi_t| < c\sqrt{\log T}\right] = 1, \tag{3.21}$$

and hence we have for any sequence $\delta_n \to 0$ that

$$\lim_{n\to\infty} \mathbb{P}\left[\max_{0\le s\le \frac{T}{\delta_n}} |\xi_s| < c\sqrt{\log \frac{T}{\delta_n}}\right] = 1.$$

We define

$$\Omega_T^{\delta_n} := \left\{\omega\in\Omega \middle| \max_{0\le s\le \frac{T}{\delta_n}} |\xi_s| < c\sqrt{\log \frac{T}{\delta_n}}\right\}.$$

We calculate for any $\epsilon > 0$, $n\in\mathbb{N}$:

$$\begin{aligned}
&\mathbb{P}\left[\max_{0\le t\le T}\left|X_t^{\mu_n,\delta_n} - X_t^{\delta_n}\right| > \epsilon\right] \\
&\le \mathbb{P}\left[\left\{\max_{0\le t\le T}\left|X_t^{\mu_n,\delta_n} - X_t^{\delta_n}\right| > \epsilon\right\}\cap\Omega_T^{\delta_n}\right] \\
&\quad +\mathbb{P}\left[\left\{\max_{0\le t\le T}\left|X_t^{\mu_n,\delta_n} - X_t^{\delta_n}\right| > \epsilon\right\}\cap\left(\Omega_T^{\delta_n}\right)^c\right] \\
&\le \mathbb{P}\left[\left\{\max_{0\le t\le T}\left|X_t^{\mu_n,\delta_n} - X_t^{\delta_n}\right| > \epsilon\right\}\cap\Omega_T^{\delta_n}\right] + \mathbb{P}\left[\left(\Omega_T^{\delta_n}\right)^c\right]
\end{aligned}$$

Now, by construction, we have for fixed $T\in\mathbb{R}_+$ that $\lim_{n\to\infty}\mathbb{P}\left[\Omega_T^{\delta_n}\right] = 1$. Hence for $\eta > 0$ arbitrary, there exists $n_1\in\mathbb{N}$ sufficiently large such that

$$\mathbb{P}\left[\left(\Omega_T^{\delta_n}\right)^c\right] < \eta \quad \forall\, n\ge n_1. \tag{3.22}$$

Note that without loss of generality we have $T > e$ and $\delta_{n_1} < \frac{1}{e}$ (will be needed later in the proof).

In order to estimate the first term, we show that there exists $n_2 \in \mathbb{N}$ such that:

$$\sup_{\omega \in \Omega_T^{\delta_{n_1}}} \max_{0 \le t \le T} \left| X_t^{\mu_n, \delta_n}(\omega) - X_t^{\delta_n}(\omega) \right| < \epsilon \quad \forall\, n \ge n_2.$$

Set for $t \ge 0$, $\omega \in \Omega_T^{\delta_{n_1}}$:

$$\tilde{b}(t,x)(\omega) := b(x) + \sigma(x) \frac{d}{dt} V_t^{\delta}(\omega) = b(x) + \frac{1}{\sqrt{\delta}} \sigma(x) \xi_{\frac{t}{\delta}}(\omega).$$

The idea is to apply Lemma 3.5(i) to the processes $X_t^{\mu,\delta}(\omega)$ and $X_t^{\delta}(\omega)$, equivalently given by the differential equations

$$dX_t^{\delta} = \tilde{b}(t, X_t^{\delta}) dt,$$

and

$$\begin{aligned} dX_t^{\mu,\delta} &= Y_t^{\mu,\delta} dt \\ \mu dY_t^{\mu,\delta} &= \tilde{b}(t, X_t^{\mu,\delta}) dt - dX_t^{\mu,\delta}, \\ X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2, \end{aligned} \tag{3.23}$$

Hence we need to show that for $\omega \in \Omega_T^{\delta_{n_1}}$ the conditions concerning $\tilde{b}$ are fulfilled (namely the boundedness of $\tilde{b}$ (as is postulated in Section 2.2 (ii)) and the Lipschitz property of $\tilde{b}$ (as is postulated in Section 3.2.1) (again, Borel-measurability is clear)).

It follows by using that $T > e$ and $\delta_{n_1} < \frac{1}{e}$:

$$\begin{aligned} \max_{0 \le s \le \frac{T}{\delta_{n_1}}} |\xi_s(\omega)| &< c \sqrt{\log \frac{T}{\delta_{n_1}}} \le c\sqrt{\log T + |\log \delta_{n_1}|} \\ &\le c\sqrt{\log T + 2\sqrt{\log T}\sqrt{\log |\delta_{n_1}|} + |\log \delta_{n_1}|} \\ &= c\sqrt{\log T} + c\sqrt{|\log \delta_{n_1}|} \\ &\le c\sqrt{\log T}\sqrt{|\log \delta_{n_1}|} + c\sqrt{\log T}\sqrt{|\log \delta_{n_1}|} \\ &= c_4(T)\sqrt{|\log \delta_{n_1}|}, \end{aligned}$$

with $c_4(T) := c\sqrt{\log T}$.

On the basis of the above considerations, we now derive the Lipschitz property of $\tilde{b}$ for all $\omega \in \Omega_T^{\delta_{n_1}}$ by using the Lipschitz property of σ and the argument above. Let $\omega \in \Omega_T^{\delta_{n_1}}$. Then for $t \in [0,T]$

$$\begin{aligned}
\left|\tilde{b}(t,x_1)(\omega) - \tilde{b}(t,x_2)(\omega)\right| &\leq \left|\frac{1}{\sqrt{\delta_{n_1}}}\sigma(x_1)\xi_{\frac{t}{\delta_{n_1}}}(\omega) - \frac{1}{\sqrt{\delta_{n_1}}}\sigma(x_2)\xi_{\frac{t}{\delta_{n_1}}}(\omega)\right| \\
&\quad + |b(x_1) - b(x_2)| \\
&= \frac{1}{\sqrt{\delta_{n_1}}}\left|\xi_{\frac{t}{\delta_{n_1}}}(\omega)\right| |\sigma(x_1) - \sigma(x_2)| \\
&\quad + |b(x_1) - b(x_2)| \\
&\leq \frac{1}{\sqrt{\delta_{n_1}}} c_4(T)\sqrt{|\log \delta_{n_1}|} K_\sigma |x_1 - x_2| + K_b |x_1 - x_2| \\
&\leq \frac{\sqrt{|\log \delta_{n_1}|}}{\sqrt{\delta_{n_1}}} c_4(T) K_\sigma |x_1 - x_2| + K_b |x_1 - x_2| \\
&\leq \frac{\sqrt{|\log \delta_{n_1}|}}{\sqrt{\delta_{n_1}}} \tilde{K} |x_1 - x_2|,
\end{aligned}$$

with $\tilde{K} = c_4(T) K_\sigma + K_b$ (since $\frac{\sqrt{|\log \delta_{n_1}|}}{\sqrt{\delta_{n_1}}} \geq 1$, because $\delta_{n_1} > e^{-1}$ without loss of generality). Thus the desired Lipschitz property is shown for all $\omega \in \Omega_T^{\delta_{n_1}}$.

Next, we show that $\tilde{b}$ is uniformly bounded in t and x for all $\omega \in \Omega_T^{\delta_{n_1}}$. We estimate for $\omega \in \Omega_T^{\delta_{n_1}}$:

$$\begin{aligned}
\left|\tilde{b}(t,x)(\omega)\right| &= \left|b(x) + \frac{1}{\sqrt{\delta_{n_1}}}\sigma(x)\xi_{\frac{t}{\delta_{n_1}}}(\omega)\right| \leq \|b\| + \frac{1}{\sqrt{\delta_{n_1}}}\|\sigma\| \left|\xi_{\frac{t}{\delta_{n_1}}}(\omega)\right| \\
&\leq \|b\| + \|\sigma\| c_4(T) \frac{\sqrt{|\log \delta_{n_1}|}}{\sqrt{\delta_{n_1}}} \\
&\leq \|b\| + c_4(T) \frac{\sqrt{|\log \delta_{n_1}|}}{\sqrt{\delta_{n_1}}} \|\sigma\|
\end{aligned}$$

The right side does not depend on x any longer, hence we have the desired uniformly boundedness of $\tilde{b}$ for all $\omega \in \Omega_T^{\delta_{n_1}}$:

$$\sup_{0 \leq t \leq T, x \in \mathbb{R}_+} \tilde{b}(t,x)(\omega) \leq \|b\| + c_4(T) \frac{\sqrt{|\log \delta_{n_1}|}}{\sqrt{\delta_{n_1}}} \|\sigma\|$$

Thus it follows by Lemma 3.5(i) and Remark 3.6 that for we have for all $n \in \mathbb{N}$ for $\omega \in \Omega_T^{\delta_{n_1}}$:

$$
\begin{aligned}
& \max_{0 \le t \le T} \left| X_t^{\mu_n, \delta_n}(\omega) - X_t^{\delta_n}(\omega) \right| \\
\le \; & \mu_n \left(\|b\| + |\xi_2| + \|\sigma\| \, c_4 \, (T) \, \frac{\sqrt{|\log \delta_n|}}{\sqrt{\delta_n}} \right) \\
& \cdot \left(1 + \tilde{K} T \frac{\sqrt{|\log \delta_n|}}{\sqrt{\delta_n}} \exp \left\{ \tilde{K} T \frac{\sqrt{|\log \delta_n|}}{\sqrt{\delta_n}} \right\} \right) \\
\le \; & \mu_n \left(\|b\| + |\xi_2| + \|\sigma\| \, c_4 \, (T) \, \frac{1}{\delta_n} \right) \left(1 + \underbrace{\tilde{K} T}_{:= \hat{K}(T)} \frac{1}{\delta_n} \exp \left\{ \tilde{K} T \frac{1}{\delta_n} \right\} \right) \\
\le \; & \mu_n \left(\|b\| + |\xi_2| + \|\sigma\| \, c_4 \, (T) \, \frac{1}{\delta_n} \right) \\
& + \hat{K} \, (T) \left(\|b\| + |\xi_2| \right) \mu_n \frac{1}{\delta_n} \exp \left\{ \frac{\hat{K} \, (T)}{\delta_n} \right\} \\
& + \hat{K} \, (T) \, \|\sigma\| \, c_4 \, (T) \, \mu_n \frac{1}{\delta_n^2} \exp \left\{ \frac{\hat{K} \, (T)}{\delta_n} \right\},
\end{aligned}
$$

where we used in the second step that $\log \delta_n \ge 1 - \frac{1}{\delta_n}$ and hence for $0 < \delta_n \le 1$:

$$
\frac{\sqrt{|\log \delta_n|}}{\sqrt{\delta_n}} \le \sqrt{\frac{\frac{1}{\delta_n} - 1}{\delta_n}} = \sqrt{\frac{1}{\delta_n^2} - \frac{1}{\delta_n}} \le \frac{1}{\delta_n}.
$$

Note that the above bound is independent of of $\omega \in \Omega_T^{\delta_{n_1}}$.
Now due to our assumption that $\lim_{n \to \infty} \mu_n \frac{1}{\delta_n^2} \exp \left\{ \hat{K} \frac{1}{\delta_n} \right\} = 0$ it follows by the above considerations that for any $\epsilon > 0$ there exists $n_2 \in \mathbb{N}$ such that

$$
\sup_{\omega \in \Omega_T^{\delta_{n_1}}} \max_{0 \le t \le T} \left| X_t^{\mu_n, \delta_n}(\omega) - X_t^{\delta_n}(\omega) \right| \le \epsilon \quad \forall \; n \ge n_2.
$$

Define $n_0 := \max \{n_1, n_2\}$. Hence for any $\eta > 0$ and any $\epsilon > 0$ (both arbitrary) there exist $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$:

$$
\mathbb{P} \left[\max_{0 \le t \le T} \left| X_t^{\mu_n, \delta_n} - X_t^{\delta_n} \right| > \epsilon \right]
$$

$$
\begin{aligned}
&= \mathbb{P}\left[\left\{\max_{0\le t\le T}\left|X_t^{\mu_n,\delta_n}-X_t^{\delta_n}\right|>\epsilon\right\}\cap\left(\Omega_T^{\delta_n}\right)^c\right]\\
&\le \mathbb{P}\left[\left(\Omega_T^{\delta_n}\right)^c\right]\le\eta \quad \forall\, n\ge n_0,
\end{aligned}
$$

which is: $X_t^{\mu,\delta}$ converges in probability to X_t^{δ} uniformly on $[0,T]$, as $\lim_{n\to\infty}\mu_n\frac{1}{\delta_n}\exp\left\{\hat{K}\left(T\right)\frac{1}{\delta_n^2}\right\}=0$.

Proof of Claim 2.

We proof this result with the help of the following lemma. The lemma and its proof are taken from [26]. Note that we even prove a more general case than the one we really need here, in the sense that in the following lemma our functions b and σ exhibit an additional time-dependency.

Lemma 3.12. *Fix $T\in\mathbb{R}_+$ and let the following conditions be satisfied:*

a) $b(t,x),\sigma(t,x),\frac{\partial\sigma(t,x)}{\partial t},\frac{\partial\sigma(t,x)}{\partial x}$ continuous in $-\infty<x<\infty,\ 0\le t\le T$,

b) $b(t,x),\sigma(t,x)$ and $\frac{\partial\sigma^2(t,x)}{\partial x}$ satisfy a Lipschitz property:

$$
|f(t,x_1)-f(t,x_2)|\le K_f\,|x_1-x_2|
$$

for $\quad f=b,\sigma,\frac{\partial\sigma^2(t,x)}{\partial x}$,

c) $\sigma(t,x)\ge\beta>0$ or $-\sigma(t,x)\ge\beta>0$ for all $x\in\mathbb{R}$, $t\in[0,T]$, $\beta\in\mathbb{R}_+$, and $\left|\frac{\partial\sigma(t,x)}{\partial t}\right|\le K_3\sigma^2(t,x)$ for all $x\in\mathbb{R}$, $t\in[0,T]$, $\beta\in\mathbb{R}$, where K_3 is a constant, $K_3<\infty$, and

d) V_t^δ satisfies

d1) V_t^δ is piecewise continuously differentiable, and

d2) $V_t^\delta(\omega)\overset{\delta\to0}{\longrightarrow}W_t(\omega)$ uniformly on $[0,T]$ P-a.s.

Let for all $t\in[0,T]$ $\tilde{X}_t^\delta$ and X_t satisfy

$$
dX_t^\delta=b(t,X_t^\delta)dt+\sigma(t,X_t^\delta)dV_t^\delta,\quad X_0^\delta=\xi_1,
$$

and

$$
d\tilde{X}_t=b(t,\tilde{X}_t)dt+\sigma(t,\tilde{X}_t)\circ dW_t,\qquad \tilde{X}_0=\xi_1,
$$

respectively.
Then $X_t^\delta\overset{\delta\to0}{\longrightarrow}\tilde{X}_t$ in probability uniformly on $[0,T]$.

Proof. Due to Remark 2.5, the (Stratonovich) stochastic differential equation

$$d\tilde{X}_t = b(t, \tilde{X}_t)dt + \sigma(t, \tilde{X}_t) \circ dW_t, \qquad \tilde{X}_0 = \xi_1$$

is equivalent to the following (Ito) stochastic differential equation:

$$d\tilde{X}_t = b(t, \tilde{X}_t)dt + \frac{1}{2}\sigma(t, \tilde{X}_t)\frac{\partial\sigma(t, \tilde{X}_t)}{\partial\tilde{X}}dt + \sigma(t, \tilde{X}_t)dW_t, \qquad \tilde{X}_0 = \xi_1.$$

We consider a fixed, but arbitrary sequence $\delta_n = (\delta(n))_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty}\delta_n = 0$. The idea of the proof is to apply Lemma 2.12 to the function

$$f_{\delta_n}(t) = \left|X_t^{\delta_n} - \tilde{X}_t\right|.$$

We have for fixed $\omega \in \Omega$ that $f_{\delta_n} : \mathbb{R} \longrightarrow \mathbb{R}$, $f_{\delta_n}(t) \geq 0$, and continuous in $t \in [0, T]$ for all δ_n for P-almost surely (by Theorems 2.6(iii), 2.7(i) (existence and uniqueness) and Definition 2.4 (i), (iii) (strong solution)).
Let $K_6 \in \mathbb{R}_+$. We set

$$\epsilon_{\delta_n}(t) := exp\left\{K_6\left(\left|V_t^{\delta_n} - W_t\right| + \left|V_0^{\delta_n} - W_0\right|\right)\right\} - 1 \quad \geq 0,$$

so $\epsilon : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$. We will show that the conditions (2.14) and (2.15) of Lemma 2.12 are fulfilled.
First consider condition (2.14). By assumption we have that $V_t^{\delta_n} \overset{n\to\infty}{\longrightarrow} W_t$ uniformly P-a.s., which is for P-almost all $\omega \in \Omega$:

$$\forall\ \tilde{\gamma} > 0 \quad \exists\ \tilde{n}_0(\omega) \in \mathbb{N} : \sup_{0\leq t\leq T}\left|V_t^{\delta_n} - W_t\right| < \tilde{\gamma} \quad \forall\ n \geq \tilde{n}_0(\omega).$$

Hence for P-almost all $\omega \in \Omega$:

$$\forall\ \gamma > 0 \quad \exists\ n_0(\omega) \in \mathbb{N} : \sup_{0\leq t\leq T}\epsilon_{\delta_n}(t) < \gamma \quad \forall\ n \geq n_0(\omega).$$

Set $\gamma := \frac{1}{T\rho\nu\exp\{\rho\nu T\}}$ with $\rho, \nu > 0$ arbitrary (to be specified later in the proof), and we have that for P-almost all $\omega \in \Omega$:

$$\int_0^T \epsilon_{\delta_n}(s)ds < \int_0^T \gamma ds = \gamma T = \frac{1}{\rho\nu e^{\rho\nu T}} \quad \forall n \geq n_0(\omega)$$

Thus condition (2.14) is fulfilled for P-almost all $\omega \in \Omega$ for sufficiently large n.

Now consider condition (2.15). We show: For all $\omega \in \Omega$ there exists $\nu := \nu(\omega)$ such that:

$$\log\left(1 + \frac{\left|X_t^{\delta_n} - \tilde{X}_t\right|}{\nu}\right) \leq K_6\left(\left|V_t^{\delta_n} - W_t\right| + \left|V_0^{\delta_n} - W_0\right|\right) + K_6\nu \int_0^t \left|X_s^{\delta_n} - \tilde{X}_s\right| ds. \tag{3.24}$$

This is equivalent to: For all $\omega \in \Omega$ there exists $\nu := \nu(\omega)$ such that:

$$\log\left(1 + \frac{\left|X_t^{\delta_n} - \tilde{X}_t\right|}{\nu}\right) \leq \log\left(\epsilon_{\delta_n}(t) + 1\right) + \underbrace{K_6\nu}_{=:\rho>0} \int_0^t \underbrace{\left|X_s^{\delta_n} - \tilde{X}_s\right|}_{=f_{\delta_n}(s)} ds,$$

which is condition (2.15) in Lemma 2.12.
So, once (3.24) is shown, applying Lemma 2.12 to P-almost all $\omega \in \Omega$ yields for all $n \geq n_0(\omega)$:

$$\left|X_t^{\delta_n} - \tilde{X}_t\right| = f_{\delta_n}(t) \leq \nu \frac{\epsilon_{\delta_n}(t) + \rho\nu e^{\rho\nu(T-0)} \int_0^T \epsilon_{\delta_n}(s)ds}{1 - \rho\nu e^{\rho\nu(T-0)} \int_0^T \epsilon_{\delta_n}(s)ds} \quad \text{P-a.s.}$$

Condition d2) states that $V_t^{\delta_n}(\omega) \overset{n\to\infty}{\longrightarrow} W_t(\omega)$ uniformly in $[0,T]$ P-almost surely. Hence it follows that $\epsilon_{\delta_n}(t) \overset{n\to\infty}{\longrightarrow} 0$ P-a.s. and then by Lebesgue's dominated convergence theorem (see [24], p. I. 3.9) that $\int_0^T \epsilon_{\delta_n}(s)ds \overset{n\to\infty}{\longrightarrow} 0$ P-a.s. Thus for P-almost all $\omega \in \Omega$:

$$\begin{aligned}
&\sup_{t\in[0,T]} \nu \frac{\epsilon_{\delta_n}(t) + \rho\nu e^{\rho\nu(T)} \int_0^T \epsilon_{\delta_n}(s)ds}{1 - \rho\nu e^{\rho\nu(T)} \int_0^T \epsilon_{\delta_n}(s)ds} \\
\leq \quad & \nu \frac{\overbrace{\sup_{t\in[0,T]} \epsilon_{\delta_n(t)}}^{\overset{n\to\infty}{\longrightarrow} 0 \text{ P-a.s.}} + \rho\nu e^{\rho\nu(T)} \overbrace{\int_0^T \epsilon_n(s)ds}^{\overset{n\to\infty}{\longrightarrow} 0 \text{ P-a.s. (Lebesgue)}}}{1 - \rho\nu e^{\rho\nu(T)} \underbrace{\int_0^T \epsilon_n(s)ds}_{\overset{n\to\infty}{\longrightarrow} 0 \text{ P-a.s. (Lebesgue)}}} \\
\overset{n\to\infty}{\longrightarrow} \quad & \nu \cdot \frac{0}{1} = 0
\end{aligned}$$

Thus for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{t\in[0,T]} \left|X_t^{\delta_n} - \tilde{X}_t\right| > \epsilon\right] = 0,$$

which is $X_t^{\delta_n} \stackrel{\delta\to 0}{\longrightarrow} \tilde{X}_t$ uniformly in probability in $[0,T]$, since we chose δ_n arbitrary as long as $\lim_{n\to\infty} \delta_n = 0$. Hence the lemma is proved.

Left to show: Claim (3.24).
Set $\Phi(t,x) := \int_0^x \frac{1}{\sigma(t,u)} du$.
Claim 1:

$$\left|\Phi(t, X_t^{\delta_n}) - \Phi(t, \tilde{X}_t)\right| \stackrel{!}{\le} K_5 \nu \int_0^t \left|X_s^{\delta n} - \tilde{X}_s\right| ds + K_5 \left(\left|V_t^{\delta n} - W_t\right| + \left|V_0^{\delta_n} - W_0\right|\right),$$

for some constant $K_5 \in \mathbb{R}$ and $\nu < \infty$ a.s.
We calculate with the help of ordinary and stochastic calculus:

$$\begin{aligned}
\Phi(t, X_t^{\delta_n}) &= \int_0^t \frac{\partial \Phi(s, X_s^{\delta_n})}{\partial s} ds + \int_0^t \frac{\partial \Phi(s, X_s^{\delta_n})}{\partial X_t^{\delta_n}} dX_t^{\delta_n} + \Phi(0, X_0^{\delta_n}) \\
&= \int_0^t \frac{\partial \Phi(s, X_s^{\delta_n})}{\partial s} ds + \int_0^t \frac{b(t, X_t^{\delta_n})}{\sigma(s, X_s^{\delta_n})} ds + \int_0^t 1 dV_t^{\delta_n} + \Phi(0, X_0^{\delta_n}) \\
&= \int_0^t \frac{\partial \Phi(s, X_s^{\delta_n})}{\partial s} ds + \int_0^t \frac{b(s, X_s^{\delta_n})}{\sigma(s, X_s^{\delta_n})} ds + V_t^{\delta_n} - V_0^{\delta_n} + \Phi(0, X_0^{\delta_n}).
\end{aligned}$$

Now recall that

$$d\tilde{X}_t = \left(b(t, \tilde{X}_t) + \frac{1}{2}\sigma(t, \tilde{X}_t)\frac{\partial \sigma(t, \tilde{X}_t)}{\partial \tilde{X}}\right) dt + \sigma(t, \tilde{X}_t) dW_t.$$

We fix $\omega \in \Omega$. Then for fixed ω (omitted for the sake of shortness) we apply Ito's formula and obtain

$$\begin{aligned}
\Phi(t, \tilde{X}_t) &= \Phi(0, \tilde{X}_0) + \int_0^t \frac{\partial \Phi(s, \tilde{X}_s)}{\partial s} ds + \frac{1}{2}\int_0^t \frac{\partial^2 \Phi(s, \tilde{X}_s)}{\partial x^2} \sigma^2(s, \tilde{X}_s) ds \\
&\quad + \int_0^t \frac{\partial \Phi(s, \tilde{X}_s)}{\partial x} \left(b(s, \tilde{X}_s) + \frac{1}{2}\sigma(s, \tilde{X}_s)\frac{\partial \sigma(s, \tilde{X}_s)}{\partial x}\right) ds \\
&\quad + \int_0^t \frac{\partial \Phi(s, \tilde{X}_s)}{\partial x} \sigma(s, \tilde{X}_s) dW_s
\end{aligned}$$

$$
\begin{aligned}
&= \Phi(0,\tilde{X}_0) + \int_0^t \frac{\partial \Phi(s,\tilde{X}_s)}{\partial s} ds + \frac{1}{2}\int_0^t \frac{\partial^2 \Phi(s,\tilde{X}_s)}{\partial x^2}\sigma^2(s,\tilde{X}_s) ds \\
&\quad + \int_0^t \frac{\partial \Phi(s,\tilde{X}_s)}{\partial x} b(s,\tilde{X}_s) ds + \frac{1}{2}\int_0^t \frac{\partial \Phi(s,\tilde{X}_s)}{\partial x}\sigma(s,\tilde{X}_s)\frac{\partial \sigma(s,\tilde{X}_s)}{\partial x} ds \\
&\quad + \int_0^t \frac{\partial \Phi(s,\tilde{X}_s)}{\partial x}\sigma(s,\tilde{X}_s) dW_s \\
&= \Phi(0,\tilde{X}_0) + \int_0^t \frac{\partial \Phi(s,\tilde{X}_s)}{\partial s} ds - \frac{1}{2}\int_0^t \frac{\partial \sigma(s,\tilde{X}_s)}{\partial x} ds \\
&\quad + \int_0^t \frac{b(s,\tilde{X}_s)}{\sigma(s,\tilde{X}_s)} ds + \frac{1}{2}\int_0^t \frac{\partial \sigma(s,\tilde{X}_s)}{\partial x} ds + \int_0^t 1 dW_s \\
&= \Phi(0,\tilde{X}_0) + \int_0^t \frac{\partial \Phi(s,\tilde{X}_s)}{\partial s} ds + \int_0^t \frac{b(s,\tilde{X}_s)}{\sigma(s,\tilde{X}_s)} ds + W_t - W_0,
\end{aligned}
$$

where we used that

$$
\frac{\partial \Phi(t,x)}{\partial x} = \frac{1}{\sigma(t,x)} \quad \text{and} \quad \frac{\partial^2 \Phi(t,x)}{\partial x^2} = -\frac{1}{\sigma^2(t,x)}\frac{\partial \sigma(t,x)}{\partial x}.
$$

Combining these two results gives

$$
\begin{aligned}
\left|\Phi(t,X_t^{\delta_n}) - \Phi(t,\tilde{X}_t)\right| &= \Bigg| \int_0^t \frac{\partial \Phi(s,X_s^{\delta_n})}{\partial s} ds - \int_0^t \frac{b(s,X_s)}{\sigma(s,X_s)} ds - W_t + W_0 \\
&\quad + \Phi(0,\tilde{X}_0) - \int_0^t \frac{\partial \Phi(s,X_s)}{\partial s} ds + \int_0^t \frac{b(s,X_s^{\delta_n})}{\sigma(s,X_s^{\delta_n})} ds \\
&\quad + V_t^{\delta_n} - V_0^{\delta_n} - \Phi(0,X_0^{\delta_n}) \Bigg| \\
&\le \int_0^t \left| \frac{\partial \Phi(s,X_s^{\delta_n})}{\partial s} - \frac{\partial \Phi(s,X_s)}{\partial s} \right| ds \\
&\quad + \int_0^t \left| \frac{b(s,X_s^{\delta_n})}{\sigma(s,X_s^{\delta_n})} - \frac{b(s,X_s)}{\sigma(s,X_s)} \right| ds \\
&\quad + \left|V_t^{\delta_n} - W_t\right| + \left|V_0^{\delta_n} - W_0\right|
\end{aligned}
$$

Estimate the terms $\left|\frac{\partial \Phi(s,x)}{\partial s} - \frac{\partial \Phi(s,\xi)}{\partial s}\right|$ and $\left|\frac{b(s,x)}{\sigma(s,x)} - \frac{b(s,\xi)}{\sigma(s,\xi)}\right|$ first, where we assume w.l.o.g. $\xi \le x$. With the use of these results, we will proceed estimating the whole expression above.

$$
\begin{aligned}
\left|\frac{\partial\Phi(s,x)}{\partial s}-\frac{\partial\Phi(\xi,s)}{\partial s}\right| &= \left|\frac{\partial\left(\int_0^x \frac{1}{\sigma(s,u)}du\right)}{\partial s}-\frac{\partial\left(\int_0^\xi \frac{1}{\sigma(s,u)}du\right)}{\partial s}\right| \\
&= \left|\int_0^x \frac{\partial\left(\frac{1}{\sigma(s,u)}\right)}{\partial s}du-\int_0^\xi \frac{\partial\left(\frac{1}{\sigma(s,u)}\right)}{\partial s}du\right| \\
&= \left|\int_0^x \frac{1}{\sigma^2(s,u)}\cdot\frac{\partial\sigma(s,u)}{\partial s}du-\int_0^\xi \frac{1}{\sigma^2(s,u)}\cdot\frac{\partial\sigma(s,u)}{\partial s}du\right| \\
&= \left|\int_\xi^x \frac{1}{\sigma^2(s,u)}\cdot\frac{\partial\sigma(s,u)}{\partial s}du\right| \overset{(*)}{\leq} \left|\int_\xi^x K_3 du\right| \\
&= \left|K_3x-K_3\xi\right| = K_3\left|x-\xi\right| \leq K_3\left(1+|\xi|\right)\left|x-\xi\right|,
\end{aligned}
$$

where we used part of the assumption c) $\left(\frac{1}{\sigma^2(s,u)}\cdot\left|\frac{\partial\sigma(s,u)}{\partial s}\right|\leq K_3\right)$ in the inequality marked with $(*)$.
Furthermore

$$
\begin{aligned}
\left|\frac{b(s,x)}{\sigma(s,x)}-\frac{b(s,\xi)}{\sigma(s,\xi)}\right| &\leq \left|\frac{b(s,x)}{\sigma(s,x)}-\frac{b(s,\xi)}{\sigma(s,x)}\right|+\left|\frac{b(s,\xi)}{\sigma(s,x)}-\frac{b(s,\xi)}{\sigma(s,\xi)}\right| \\
&= \frac{\overbrace{|b(s,x)-b(s,\xi)|}^{\leq K_b|x-\xi|(\text{b Lipschitz})}}{\underbrace{|\sigma(s,x)|}_{\geq\beta(\text{assumption c}))}}+\frac{\overbrace{|b(s,\xi)|}^{\leq K_1(1+|\xi|)}\ \overbrace{|\sigma(s,\xi)-\sigma(s,x)|}^{\leq K_\sigma|x-\xi|(\sigma\text{ Lipschitz})}}{\underbrace{|\sigma(s,x)\sigma(s,\xi)|}_{\geq\beta^2}} \\
&\leq \frac{K_b\left|x-\xi\right|}{\beta}+\frac{K_1\left(1+|\xi|\right)K_\sigma\left|x-\xi\right|}{\beta^2} \\
&\leq \frac{1}{\beta}\left|x-\xi\right|\left(1+|\xi|\right)\underbrace{\left(K_b+\frac{K_1K_\sigma}{\beta}\right)}_{=:K_2} \\
&= \frac{K_2}{\beta}\left|x-\xi\right|\left(1+|\xi|\right)
\end{aligned}
$$

We used that due to the uniformly boundedness and Lipschitz property of b we have that $|b(s,x)|\leq K_1\left(1+|\xi|\right)$ for some constant $K_1\in\mathbb{R}$, since $|b(t,x)|\leq|b(t,1)|+K_b\left(1+|x|\right)\leq\left(c_b+K_b\right)\left(1+|x|\right)$.

Applying the above results to our expression yields

$$\begin{aligned}
\left|\Phi(t,X_t^{\delta_n})-\Phi(t,\tilde{X}_t)\right| &\leq \int_0^t K_3\left(1+\left|\tilde{X}_s\right|\right)\left|X_s^{\delta_n}-\tilde{X}_s\right|ds \\
&\quad+\int_0^t \frac{K_2}{\beta}\left(1+\left|\tilde{X}_s\right|\right)\left|X_s^{\delta_n}-\tilde{X}_s\right|ds+\left|V_t^{\delta_n}-W_t\right| \\
&\quad+\left|V_0^{\delta_n}-W_0\right| \\
&\leq K_3\nu\int_0^t\left|X_s^{\delta_n}-\tilde{X}_s\right|ds+\frac{K_2\nu}{\beta}\int_0^t\left|X_s^{\delta_n}-\tilde{X}_s\right|ds \\
&\quad+\left|V_t^{\delta_n}-W_t\right|+\left|V_0^{\delta_n}-W_0\right|,
\end{aligned}$$

where we set $\nu := 1+\max\left\{\left|\tilde{X}_t\right| \middle| 0\leq t\leq T\right\}$. Note that ν is dependent on ω, i.e. $\nu=\nu(\omega)$ and for almost all $\omega\in\Omega$ we have that $\nu<\infty$.

$$\begin{aligned}
&= \left(K_3+\frac{K_2}{\beta}\right)\nu\int_0^t\left|X_s^{\delta_n}-\tilde{X}_s\right|ds+\left|V_t^{\delta_n}-W_t\right| \\
&\quad+\left|V_0^{\delta_n}-W_0\right| \\
&\leq K_5\nu\int_0^t\left|X_s^{\delta_n}-\tilde{X}_s\right|ds+K_5\left|V_t^{\delta_n}-W_t\right| \\
&\quad+K_5\left|V_0^{\delta_n}-W_0\right|,
\end{aligned}$$

with $K_5 := max\left\{1; K_3+\frac{K_2}{\beta}\right\}$.

Claim 2:

$$\left|\Phi(t,X_t^{\delta_n})-\Phi(t,\tilde{X}_t)\right| \geq K_4\log\left(1+\frac{\left|X_t^{\delta_n}-\tilde{X}_t\right|}{\nu}\right),$$

for some constant $K_4\in\mathbb{R}$ and $\nu := 1+\max\left\{|X_t| \middle| 0\leq t\leq T\right\}$ as in the proof of Claim 1.
Let $x,\xi\in\mathbb{R}$. Set $y := min(x,\xi)$ and $z := max(x,\xi)$. Then:

$$\begin{aligned}
|\Phi(t,x)-\Phi(t,\xi)| &= \left|\int_0^x\frac{1}{\sigma(s,u)}du-\int_0^\xi\frac{1}{\sigma(s,u)}du\right| \\
&= \left|\int_{min(x,\xi)}^{max(x,\xi)}\frac{1}{\sigma(s,u)}du\right|
\end{aligned}$$

$$
\begin{aligned}
&= \int_y^z \frac{1}{|\sigma(s,u)|}du && \text{(due to assumption c))}\\
&\geq \int_y^z \frac{1}{C\,(1+|u|)}du && \text{(analogous to above)}\\
&= \frac{1}{C}\int_y^z \frac{1}{1+|u|}du
\end{aligned}
$$

Case 1: $sign(x) = sign(\xi)$.

W.l.o.g.: $x, \xi \geq 0$.

(Due to the symmetry of $\frac{1}{1+|u|}$ we have

$$
\int_{min(x,\xi)}^{max(x,\xi)} \frac{1}{1+|u|}du = \int_{-max(x,\xi)}^{-min(x,\xi)} \frac{1}{1+|u|}du = \int_{min(-x,-\xi)}^{max(-x,-\xi)} \frac{1}{1+|u|}du.
$$

So if $x, \xi < 0$, set $x' = -x$ and $\xi' = -\xi$ and continue with x', ξ'.)

Then the above calculation yields

$$
\begin{aligned}
|\Phi(t,x) - \Phi(t,\xi)| &\geq \frac{1}{C}\int_y^z \frac{1}{1+u}du = \frac{1}{C}\left[\log\left(1+u\right)\Big|_y^z\right]\\
&= \frac{1}{C}\left[\log\left(1+z\right) - \log\left(1+y\right)\right] = \frac{1}{C}\log\left(\frac{1+z}{1+y}\right)\\
&= \frac{1}{C}\log\left(\frac{1+y+z-y}{1+y}\right) = \frac{1}{C}\log\left(1+\frac{z-y}{1+y}\right)\\
&\geq \frac{1}{C}\log\left(1+\frac{|x-\xi|}{1+\xi}\right)
\end{aligned}
$$

Case 2: $sign(x) \neq sign(\xi)$.

W.l.o.g. $z \geq |y|$.

(This is again due to the symmetry of $\frac{1}{1+|u|}$ and the argument given above: If $z < |y|$, set $x' = -x$ and $\xi' = -\xi$ and continue with x', ξ'.)

Hence

$$
\begin{aligned}
|\Phi(t,x) - \Phi(t,\xi)| &\geq \frac{1}{C}\int_y^z \frac{1}{1+|u|}du\\
&= \frac{1}{C}\left(\int_y^0 \frac{1}{1+|u|}du + \int_0^z \frac{1}{1+|u|}du\right)\\
&= \frac{1}{C}\left(\int_y^0 \frac{1}{1-u}du + \int_0^z \frac{1}{1+u}du\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{C}\left(-\log\left(|1-u|\right)\Big|_y^0 + \log\left(1+u\right)\Big|_0^z\right)\\
&= \frac{1}{C}\left(-0+\log\left(|1-y|\right)+\log\left(1+z\right)-0\right)\\
&= \frac{1}{C}\left(\log\left(1+|y|\right)+\log\left(1+z\right)\right)\\
&= \frac{1}{C}\log\left(1+|y|+z+|y|\,z\right)\\
&\geq \frac{1}{C}\log\left(1+|y|+z\right)\\
&\geq \frac{1}{C}\log\left(1+|x-\xi|\right)\\
&\geq \frac{1}{C}\log\left(1+\frac{|x-\xi|}{1+|\xi|}\right).
\end{aligned}
$$

Set $K_4 = \frac{1}{C}$. Then it follows from the above calculation

$$
\begin{aligned}
\left|\Phi(t, X_t^{\delta_n}) - \Phi(t, \tilde{X}_t)\right| &\geq K_4 \log\left(1 + \frac{\left|X_t^{\delta_n} - \tilde{X}_t\right|}{1+\left|\tilde{X}_t\right|}\right)\\
&\geq K_4 \log\left(1 + \frac{\left|X_t^{\delta_n} - \tilde{X}_t\right|}{\nu}\right),
\end{aligned}
$$

and hence Claim 2 is proved.
Bringing Claims 1 and 2 together yields

$$
\begin{aligned}
K_4 \log\left(1 + \frac{\left|X_t^{\delta_n} - \tilde{X}_t\right|}{\nu}\right) &\leq \left|\Phi(t, X_t^{\delta_n}) - \Phi(t, \tilde{X}_t)\right|\\
\leq K_5\nu \int_0^t \left|X_s^{\delta_n} - \tilde{X}_s\right| ds \qquad &+K_5\left(\left|V_s^{\delta_n} - W_s\right| + \left|V_0^{\delta_n} - W_0\right|\right),
\end{aligned}
$$

which completes the proof of Claim (3.24) setting $K_6 = \frac{K_5}{K_4}$.
Thus the statement is completely proved. $\square$

The proof is now immediate. We consider the differential equation

$$dX_t^\delta = b(X_t^\delta)dt + \sigma(X_t^\delta)dV_t^\delta, \quad dX_0^\delta = \xi_1. \tag{3.25}$$

Now, due to our assumptions it can easily be seen that conditions a) - d) of Lemma 3.12 are fulfilled (by using the fact that our functions b and σ do not exhibit a time-dependency). Lemma 3.12 states that $X_t^\delta \stackrel{\delta\to 0}{\longrightarrow} X_t$ in probability uniformly on $[0,T]$.

Concluding Claim 1 and 2, the desired uniform convergence in probability follows. $\square$

Chapter 4

Applications in Finance: The Momentum Model

The aim of this chapter is to extend the Black-Scholes model and study this extended model on the basis of the results presented in Chapter 3 (mainly with the help of the main theorem (Theorem 3.8)).

This chapter is structured as follows: First, we shortly recall the well-known Black-Scholes model. We consider the Black-Scholes model for a non-dividend paying risky asset (Section 4.1). In Section 4.2 we construct and interpret a momentum model. This is done by first introducing new variables and parameters (Subsection 4.2.1) and then defining the momentum model (Subsection 4.2.2). Afterwards - in Section 4.3 - we present the actual applications of the convergence results of Chapter 3 on the basis of the constructed model. Section 4.4 concludes this chapter.

Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ adapted to a right-continuous and complete filtration $(\mathcal{F}_t)_{t\geq 0}$. We assume that in the following all processes are adapted to $(\mathcal{F}_t)_{t\geq 0}$. In order to describe the behavior of price processes we consider a continuous-time model with one risky asset - for example a stock - with price process X_t (in the Black-Scholes model) or $X_t^{\mu,\delta}$ (in the momentum model) and a risk free asset with price process B_t. Let T be the maximal time horizon we are interested in (for example the maturity of some derivative we want to price). Hence the models are valid on the interval $[0, T]$.

4.1 The Black-Scholes Model

The purpose of this section is to introduce the Black-Scholes model for a non-dividend paying risky asset. The ideas of this section are essentially taken from [18].
We first state the assumptions that are made in the Black-Scholes model. Some of the ideas are taken from [28], p. 30. A justification and discussion of the assumptions can also be found in [28], p. 27 - 30 and shall not be repeated here.
The assumptions of the Black-Scholes model are:

(i) The short-term interest rate r is known and is constant over time.

(ii) The underlying asset pays no dividend.

(iii) The price process of the risky asset exhibits a lognormal distribution (or, equivalently: The returns on the risky asset exhibit a normal distribution).

(iv) There are no transaction costs in buying or selling the financial product.

(v) Any fraction of any financial product can immediately be traded.

(vi) It is possible to borrow any amount of cash at the risk free interest rate r.

(vii) Trading can be carried on continuously.

(viii) Information on the financial market spreads arbitrarily fast. No time does elapse beween the release of new information and the reaction to this information (i.e. no time delay for analyzing or interpreting new information).

(ix) There are no penalties to short selling.

(x) The driving noise of the underlying stock is a Wiener process.

Lateron we will see that assumption (x) is - in our setup - a consequence of assumptions (i) - (ix). Assumption (x) is essentially influenced by assumptions (iii) (the distribution of the process), (v), (vii) (trading is carried out continuously and immediately) and (viii) (information spreads arbitrarily fast). The P-a.s. non-differentiability of the driving noise (as does the Wiener noise exhibit) is due to the absence of time delays due to trading or spread of information in the Black-Scholes model. These interdependences will be discussed in detail in the following, but for the sake of completeness we shortly mention the relationships between the assumptions already at this point.

On the basis of these assumptions the behavior of the risk free and risky assets can be described by differential equations.
First, consider the risk free asset. It's price process is determined by the ordinary differential equation

$$dB_t = rB_t dt, \quad B_0 = 1, \tag{4.1}$$

with $r \in \mathbb{R}_+$. B_0 represents the value of the risk free asset at time $t = 0$, hence we can set $B_0 = 1$ without loss of generality. r is the instantaneous interest rate. It is immediate to check that for $0 \leq t \leq T$ equation (4.1) is solved by

$$B_t = e^{rt}. \tag{4.2}$$

Furthermore due to the lognormal assumption the incremental return on the risky asset $\frac{dX_t}{X_t}$ can be described by the following stochastic differential equation

$$\frac{dX_t}{X_t} = h_1 dt + h_2 dW_t, \quad X_0 = \xi_1, \tag{4.3}$$

with $\xi_1 \in \mathbb{R}_+$ and $h_1, h_2 \in \mathbb{R}$.
A proof for the lognormal property of X_t can be found in [12] p. 237 and many other. Moreover, as is shown in [18], p. 47-48, the unique strong solution of (4.3) is given by

$$X_t = \xi_1 \exp\left\{ h_1 t - \frac{h_2^2}{2} t + h_2 W_t \right\}. \tag{4.4}$$

This so-called 'Black-Scholes Model' has been studied extensively. There have been a whole bunch of results, such as the famous 'Black-Scholes formula', which prices a European call option (see for example [4] p. 101 or many other) as well as a pricing formula for a European put (see for example [18] p. 70).

4.2 Construction and Interpretation of the Momentum Model

Our aim is to extend the introduced Black-Scholes model.

4.2.1 Momentum, Liquidity and Flow of Information

We will implement a new variable (the *momentum*) and two new parameteres (measures for *liquidity* (μ) and *speed of flow of information* (δ)) which will be considered in the momentum model.

Let the parameter μ represent the *liquidity* of the considered stock in the way that μ is an inverse and standardized measure for the liquidity of the considered financial product. Given a financial market, μ should reflect elements like the daily volumes traded and the number of transactions. Let μ be standardized in such a way that $0 \leq \mu < 1$, where $\mu = 0$ represents the case that the stock is 'infinitely liquid' and $\mu = 1$ the case that the stock is 'totally illiquid'. Very highly liquid stocks (near to 'infinitely liquid') are for example most DAX or Dow Jones stocks; highly illiquid stocks are for example stocks of firms with only very few stockholders, who might even hold stocks for strategic reasons.
The actual calculation of μ for the model should be done with the help of empirical data.

Now we are ready to define the *momentum:*

Definition 4.1. *Let $X_t = X_t(\omega)$ be a stochastic process modeling a price process of a financial product and μ be an inverse measure for the liquidity of the considered asset, as introduced above. If $Y_t := \frac{d}{dt}X_t$ exists, the value μY_t is called the momentum of X_t. That is: For fixed $\omega \in \Omega$ the momentum evaluated at time t of $X_t(\omega)$ is given by $\mu Y_t(\omega)$.*

Since the momentum is so important for our model, let us devote a few words to the interpretation and intuition of the momentum:
Consider a financial market. First of all, the momentum is simply the time derivative of the price process multiplied with the liquidity μ. That means it reflects the in- and decreases in the price process weighted with the liquidity of the product. We emphasize the fact that on the market the momentum does always exist for any given time t. (Then the momentum is defined as the difference in the price process from time $t_i - t_{i-1}$ divided by the length of the (small) interval $t_i - t_{i-1}$.)

Clearly, if we want to model a continuous price process this issue gets a bit more complicated: Obviously the existence of the momentum depends on the driving process if we model price processes with the help of stochastic differential equations. If we consider a Brownian motion as the driving process we know due to the properties of the Brownian motion that the price process in the Black-Scholes model will be P-almost surely nowwhere differentiable and thus the momentum will P-almost surely not exist at all times t. But if we consider a differentiable driving process, the price process will be P-almost surely differentiable as well and thus the momentum will P-almost surely exist at P-almost all times t. We conclude that depending on the driving process and the model the momentum might or might not exist (P-a.s.), whereas in the financial market the momentum does always exist for any time t (due to the fact that quotes are just released at discrete time points, where the length of the time interval depends on the financial market).
Comparing the financial interpretation of the momentum with the physical interpretation yields a common intuition: Consider the motion of a particle modelled as in Chapter 3. Then the *momentum* of the particle is defined as the mass of the particle times velocity (see Section 3.1). Now it becomes clear that the momentum of a financial product is simply the analogon of the momentum of the motion of a particle (assuming that the analogon to the 'mass of a particle' in a phyiscal context is the 'liquidity of the financial product' in a financial context). Heuristically: Interpreting the price process of a financial product simply as a motion, the momentum is just the momentum (in the physical sense) of the motion.
Why are we interested in the momentum on the basis of considering the price process of a financial product? Empirical studies show that there exists a whole class of traders and trading strategies making profits using momentum effects (see for example [7], [8], [10], [15], and many more). Thus it is reasonable to include the momentum in our financial model.

Now some of the arising questions are: What is the effect of the momentum to the price process like and how should this effect be implemented in the model? What kind of properties do the according price and momentum processes have? How are these considerations related to the Black-Scholes model? We will construct our momentum model and try to answer these questions as best as possible.

Now we introduce the parameter δ, which reflects the market inefficiencies concerning the flow of information on the given financial market (let δ be standardized and inverse as well). Similarly to the modeling of μ, we suppose that $0 \leq \delta < 1$, where $\delta = 0$ represents the case that the considered financial market exhibits an 'infinitely fast spread of information'. The greater δ is, the slower does information spread on the financial market. For our model δ will (in general) be small (near zero), meaning that the flow of information is measured in small (absolute) values.

We shortly consider some examples for different values of δ. A very high speed of spread of information (corresponding to a small δ) will in general be observed at the stocks of firms or sectors, which are in the center of public interest: New information will spread very fast, is available for everybody and even important background information might be presented by the media. If we in opposite consider stocks of firms which are of lesser public interest, it might be very difficult to analyze and interpret new information. Banks employ experts for this 'equity research', who are only concerned with the classification and analysis of new information. Roughly speaking, the more complex and difficult this analysis is, the greater will δ - as an inverse indicator for the speed of spread of information - be. Note that in the 'traditional models' (in particular in the Black-Scholes model) it is assumed to have an arbitrary high speed of information, i.e. $\delta = 0$ (see assumption (viii) of the Black-Scholes model). So assuming that the speed of information is not arbitrarily high is a new feature in our model compared to e.g. the Black-Scholes model.

The justification for choosing $\delta > 0$ (in opposite to most other models) is as follows. First of all, it is clear that information spreads fast because of the technology we use today, but still not arbitrarily fast. Moreover, even professional investors need time to interpret new information once they are known resulting in the fact that investors do not act and react arbitrarily fast to new information. Furthermore private investors might not be able to access all information at any time, but of course professional investors are of more interest for a financial market.

More precise, we let δ represent not only the speed of the flow of information, but also the speed of the reaction to new information (as already indicated above). That means we will include δ in such a way that it also 'smoothes out' new information which contradicts the 'overall impression' or 'long-term expectation' of an investor in regard of a certain financial product.

We conclude that hence δ is not only a technical variable but include psychological factors as well. Due to the above reasons we think it does make sense to include such a variable δ in our model and to assume that δ is not zero.

In our model, δ determines the driving noise V_t^δ. This driving noise V_t^δ is such that it satisfies the condition stated in Subsection 3.2.2, in particular

we have that the sample paths of V_t^δ are piecewise differentiable, i.e. we have a 'smoother noise' compared to the Wiener process. Moreover, for $\delta \searrow 0$, V_t^δ converges to the Wiener process P-a.s., i.e. $V_t^\delta \overset{\delta\searrow 0}{\longrightarrow} W_t$ P-a.s. Hence the important point in the modeling of V_t^δ is that for every $\delta > 0$ the sample paths are piecewise differentiable (for fixed $\omega \in \Omega$), but for $\delta \searrow 0$ converging to a process which is P-a.s. nowwhere differentiable, i.e. 'arbitrarily serrated'. The economic motivation for this kind of modeling is as follows: Because of 'time delays' due to the speed of spread of information or analysis of information investors are not able to react to new information instantaneously. Hence reaction of investors to new information is 'smoother' and not arbitrarily irregular. We think it is appropriate to model this 'smoother reaction to information' by a 'smoother' (in the sense of piecewise differentiable) driving noise. The higher the speed of spread of information is, the 'lesser smooth' is the driving noise, since the reaction of investors becomes more 'serrated'. This is modeled by the P-a.s. convergence of V_t^δ to W_t as $\delta \searrow 0$. Finally, the case $\delta = 0$ describes a market where information is exchanged instantaneously and reaction to it is immediate. Hence in this case we take the Wiener process as the driving noise (as is done for example in the Black-Scholes model), representing the immediate reaction to information by having P-a.s. nowwhere differentiable sample paths.
In detail, we model the above introduced V_t^δ by taking a smooth Gaussian field and suppose that V_t^δ satisfies certain regularity conditions (stated in Subsection 3.2.2). This will be specified in Section 4.2.2, where we discuss the underlying assumptions of the momentum model (see (x') in Section 4.2.2 for the assumptions concerning V_t^δ).

Similarly to μ, our δ has to be chosen on the basis of empirical data with respect to the best fit of the following model.

4.2.2 Introduction to the Momentum Model

Since we have introduced all the variables and parameters we need, we are now able to present the momentum model. The proceeding will be analogously to the one for the Black-Scholes model.

The assumptions of the momentum model are:

(i') The short-term interest rate r is known and is constant over time.

(ii') The underlying asset pays no dividend.

(iii') The price process of the risky asset can be described by a stochastic differential equation of a specific form including the momentum (to be specified later).

(iv') There are no transaction costs in buying or selling the financial product.

(v') The quantity of the asset or option which can be traded depends on the liquidity of the financial product and might not be arbitrarily high. Depending on the liquidity there might be limits for trading or delays for the trading of high quantities of the underlying financial product.

(vi') It is possible to borrow any amount of cash at the risk free interest rate r.

(vii') Trading can be carried on continuously, as long as this is consistent with assumption (v').

(viii') In general, information on the financial market does not spread arbitrarily fast. Depending on the financial market and the financial product there might be time delays due to a slow spread of information, interpreted as needed time to analyze or interpret new information.

(ix') There are no penalties to short selling.

(x') The driving noise of the underlying stock is a stochastic process V_t^δ, defined by

$$V_t^\delta = \frac{1}{\sqrt{\delta}} \int_0^t \xi_{\frac{s}{\delta}} ds, \tag{4.5}$$

where ξ_s is a mean zero stationary Gaussian process. We suppose that V_t^δ fulfills the regularity conditions stated in the beginning of the Section 3.2.2.

Again, assumption (x') is - in our setup - a consequence of assumptions (i') - (ix'). Assumption (x') is essentially influenced by assumptions (iii') (the price process is assumed to be the solution of a system of stochastic differential equations), (v'), (vii') (problems arising because of illiqudity) and (viii') (information does not spread arbitrarily fast). Hence due to the arising time delays and liquidity problems in the model (due to trading or speed of spread of information) our driving process is now assumed to be piecewise continuously differentiable, since the above properties 'smooth out' the arbitrarily serrated and thus P-a.s. nowhere differentiable Wiener noise. For convenience we furthermore assume that the process V_t^δ can be written in the above way and fulfills the above conditions.

Note that assumptions (i'), (ii'), (iv'), (vi') and (ix') are the same as in the Black-Scholes model. The remaining assumptions have been changed or modified due to our extensions.

Now the momentum model states on the basis of these assumptions that the behavior of the risk free and risky assets can be described by differential equations.
First, consider the risk free asset. It's price process is determined by the following ordinary differential equation

$$dB_t = rB_t dt, \quad B_0 = 1, \tag{4.6}$$

with $r \in \mathbb{R}_+$.
Note that the behavior of the risk free asset is modeled exactly as in the Black-Scholes model. Hence again B_0 represents the value of the risk free asset at time $t = 0$, so we can set $B_0 = 1$ without loss of generality. r is the instantaneous interest rate again, and equation (4.6) (the same as equation (4.1)) is for $0 \le t \le T$ solved by

$$B_t = e^{rt}. \tag{4.7}$$

Now consider the price process of the risky asset $X_t^{\mu,\delta}$ and its momentum $\mu Y_t^{\mu,\delta}$ ((as long as it exists) given a financial market. The momentum model states that $X_t^{\mu,\delta}$ and $Y_t^{\mu,\delta}$ solve the following system of stochastic differential equations:

$$\begin{aligned} dX_t^{\mu,\delta} &= Y_t^{\mu,\delta} dt \\ \frac{dX_t^{\mu,\delta}}{X_t^{\mu,\delta}} &= h_1 dt + h_2 dV_t^\delta - \frac{\mu dY^{\mu,\delta}}{X_t^{\mu,\delta}} \\ X_0^{\mu,\delta} &= \xi_1,\ Y_0^{\mu,\delta} = \xi_2. \end{aligned} \tag{4.8}$$

Note that the name 'momentum model' is a bit imprecise, since we extended the Black-Scholes model not 'only' by the momentum: We rather include liquidity and, moreover, a measure for the speed of the spread of information as well resulting in a smoother driving noise. But due to convenience and in order to stress the most important extension we decided to take the name 'momentum model'.

Now we want to discuss the economic ideas of the introduced momentum model.

Equation (4.6) describes the behavior of the risk free asset, which is exactly as in the Black-Scholes model.

The first equation of system (4.8) states the relationship between the price process and the momentum given the liquidity μ according to Definition 4.1. The second equation of system (4.8) models the 'law of motion' for the financial product. First, let us compare the stochastic differential equations with the Black-Scholes model represented by the stochastic differential equation (4.3). In both equations we consider the increment of the price process divided by the price process itself. Now, in both models, this can be decomposed in a deterministic part represented by h_1 (the *trend* of the considered product) and a stochastic part represented by h_2 (the *volatility* of the product).
The first aberration from the Black-Scholes model is now that we do not consider a Wiener process as the underlying noise, but a smoother stochastic process named V_t^δ, which converges uniformly P-a.s. to the Wiener process as δ converges to zero.
The economic interpretation is as follows: We assume that because of the fact that the flow of information is not arbitrarily fast, we do not have an arbitrary flexibel and 'serrated' noise, but rather a smoother one. In other words: By choosing V_t^δ to model the disturbing noise we take into account the fact that in reality the market cannot react arbitrarily fast (due to market inefficiencies and human nature). We argued above that δ should also represent a slow and hence smoother reaction to information. By considering a smoother stochastic process we model the 'smoothing property' of the driving process. In particular, consider the special case that newly released information (for example 'bad news') contradicts the 'overall impression' and 'long-term expectation' (for example a 'good overall impression') of an investor concerning a financial product: Then, the 'good overall impression' will not be destroyed because of the newly released 'bad' information and thus the 'positive overall impression' might be able to 'smooth out' the effect of the 'bad news'.
By letting δ converging to zero, we consider a financial market where the 'spread of information becomes better and better', until the financial market finally exhibits an 'infinitely fast spread of information' and an immediate response to it. For such a financial market the Wiener process W_t seems to be a good choice to model the disturbing noise (as is done in the Black-Scholes

model), since we do not have any 'delays' because of market inefficiencies. Consequently we assume that V_t^δ converges uniformly to W_t P-a.s. as δ converges to zero.
For convenience we finally suppose that V_t^δ can be written as in equation (4.5) and in order to apply our convergence result we postulate the regularity conditions concerning V_t^δ.

Finally we have included a new term compared to the Black-Scholes model in the stochastic differential equation of the momentum model: We substract $\frac{\mu dY_t^{\mu,\delta}}{X_t^{\mu,\delta}}$ representing the incremental momentum divided by the price of the financial product.
The economic interpretation is motivated by considerations from behavioral finance and as follows: Take the liquidity μ and the market with its speed of spread of information measured by δ as given and fixed. The investor observes the price process and hence the momentum up to a certain time t (the 'present'). We assume that the investor considers the rescaled momentum (as given above) as an indicator for the price process of a financial product for the near future. Hence the investor will be interested in buying financial products if the rescaled momentum is large or at least positive, and be interested in selling a financial product if the rescaled momentum is negative (or he might even be interested in short-selling the financial product if the rescaled momentum is negative and large in the absolute value). We stress again the key fact that the investor takes the rescaled momentum as an indicator for the development in the price process of the near future.
We note that this kind of trading - which we will call 'momentum trading' in the following - exhibits the more problems in reality, the more illiquid the considered financial product is. On the one hand, if the financial product is very illiquid, it will in general not be possible to trade the financial product in the quantity a momentum trader might desire (see assumption (v')). Hence depending on the illiquidity of the financial product the order might be delayed and thus take more time, in which the price of the product might change considerably. On the other hand, if the financial product is very liquid, large quantities can be traded relatively fast and thus momentum trading is better realizable. We conclude that the liquidity of the considered financial product influences momentum trading gravely.
We model this phenomena by including the new parameters μ and δ and the new variable momentum as in (4.8).

In conclusion: Motivated by the above economic interpretations, the momentum model states that the incremental price process $dX_t^{\mu,\delta}$ divided by the price process $X_t^{\mu,\delta}$ is given by the increment of a trend $h_1 dt$ added to the incremental volatility measured with a (smooth) stochastic process given by $h_2 dV_t^{\delta}$ and then we have to substract the incremental scaled momentum $\frac{\mu dY_t^{\mu,\delta}}{X_t^{\mu,\delta}}$ (scaled with respect to the price process).

The difference of the momentum model to the Black-Scholes model is the considered disturbing noise, which is now smoother, and the additional term with the rescaled momentum. Since the momentum is calculated by the first derivative of the price process (if it exists), the momentum model is given by a system of stochastic differential equations consisting of the two given above together with an ordinary differential equation describing the behavior of the price process of the risk free asset.

4.3 Application and Interpretation of the Convergence results

We consider the momentum model represented by (4.6) and (4.8), as was introduced in the previous section.

4.3.1 Identification of the stochastic differential equations of the momentum model

We want to identify the stochastic differential equation describing the price process of the risky asset of our momentum model (equation (4.8)) with the system of stochastic differential equations considered in Chapter 3.
We set $b(X_t^{\mu,\delta}) := h_1 X_t^{\mu,\delta}$ and $\sigma(X_t^{\mu,\delta}) := h_2 X_t^{\mu,\delta}$. Plugging this definition in the second equation of (4.8) yields:

$$\begin{array}{rrcl} & \frac{dX_t^{\mu,\delta}}{X_t^{\mu,\delta}} & = & h_1 dt + h_2 dV_t^{\delta} - \frac{\mu dY_t^{\mu,\delta}}{X_t^{\mu,\delta}} \\ \iff & dX_t^{\mu,\delta} & = & X_t^{\mu,\delta} h_1 dt + X_t^{\mu,\delta} h_2 dV_t^{\delta} - \mu dY_t^{\mu,\delta} \\ \iff & dX_t^{\mu,\delta} & = & b(X^{\mu,\delta}) dt + \sigma(X^{\mu,\delta}) dV_t^{\delta} - \mu dY_t^{\mu,\delta} \\ \iff & \mu dY_t^{\mu,\delta} & = & b(X^{\mu,\delta}) dt + \sigma(X^{\mu,\delta}) dV_t^{\delta} - dX_t^{\mu,\delta}, \end{array}$$

which is exactly the second equation of the system (2.7). Thus the stochastic differential equation for the risky asset in the momentum model can be identified with the system (3.2).

Since we are concerned with the convergence results consider also the system for $\delta = 0$, i.e.

$$dX_t^\mu = Y_t^\mu dt$$
$$\frac{dX_t^\mu}{X_t^\mu} = h_1 dt + h_2 dW_t - \frac{\mu dY_t^\mu}{X_t^\mu} \tag{4.9}$$
$$X_0^\mu = \xi_1,\ Y_0^\mu = \xi_2.$$

Economically, this system reflects a momentum model with an financial market which exhibits an 'infinitely fast spread of information' and an immediate response to it.
By setting $b(X_t^\mu) := h_1 X_t^\mu$ and $\sigma(X_t^\mu) := h_2 X_t^\mu$ analogous arguments yield that the above system can be fully identified with the system (3.1).

4.3.2 Application of the previous convergence results and their economic interpretations

In this section we will apply the results from the previous chapter mathematically and then interpret the applications to the momentum model economically.

Note that nearly all conditions are obviously fulfilled with the above choice of b and σ (measurability; $\in \mathcal{C}^2$; Lipschitz; global growth condition (2.11); $\sigma'(x)$, $\sigma''(x)$ bounded; $b(x)$, $\sigma(x)$, $\sigma'(x)$ continous; $\sigma''(x)$ Lipschitz). Due to the fact that $\sup_{0\leq t\leq T} Z_t$ for $Z_t = X_t, X_t^\mu, X_t^\delta, X_t^{\mu,\delta}$ we have that b and σ are bounded for P-almost all considered x. Finally, since economically only a strict positive price makes sense, we also have that σ is non-degenerated as desired. Hence the results from the previous chapter (existence and uniqueness, convergence) are applicable to the momentum model.

Let us start with a basic (but very important) corollary concerning existence and uniqueness:

Corollary 4.2. *(i) The system of stochastic differential equations for the risky asset in the general momentum model (4.8) exhibits a unique strong solution* $(X_t^{\mu,\delta}, Y_t^{\mu,\delta})$.

(ii) The system of stochastic differential equations for the risky asset in the momentum model for a perfect financial market (4.9) exhibits a unique strong solution (X_t^μ, Y_t^μ).

Next consider convergence of the momentum model (4.8) as μ and δ approach zero. Therefore we additionally introduce the two following stochastic (first order) differential equation:

$$\frac{dX_t^\delta}{X_t^\delta} = h_1 dt + h_2 dV_t^\delta, \quad X_0^\delta = \xi_1, \tag{4.10}$$

and

$$\frac{d\tilde{X}_t}{\tilde{X}_t} = h_1 dt + h_2 \circ dW_t, \quad \tilde{X}_0 = \xi_1. \tag{4.11}$$

Note that both equations are similar to the equation describing the behavior of the risky asset in the Black-Scholes model. Equation (4.10) differs from the one of the Black-Scholes model by the disturbing noise (V_t^δ instead of W_t) and equation (4.11) differs in the way the stochastic integral is interpreted (Stratonovich instead Ito).

Corollary 4.3. *For $\delta > 0$ fixed and $\mu \searrow 0$, the solution $X_t^{\mu,\delta}$ of (4.8) converges in probability uniformly on $[0,T]$ to the solution X_t^δ of (4.10).*

Proof. The proof is a direct implication of Lemma 3.12 (and its proof) by setting again $b(x,t) := h_1 x$ and $\sigma(t,x) := h_2 x$. $\square$

At the first glance this corollary might seem a bit abstract. But in fact, we have a very meaningful interpretation and result:

First of all remember the definitions of μ and δ in our financial context: μ measures the liquidity of the considered firm and δ measures the market speed of the spread of information.
Therefore one economic interpretation can be as follows: We consider a fixed financial market with a given speed of spread of information δ. Now we consider a highly liquid financial product, meaning that problems arising because of illiquidity of a financial products are irrelevant and can be neglected. We model this by choosing a sequence μ_n such that $\lim_{n\to\infty} \mu_n = 0$.
Then the above corollary states that the solution $X_t^{\mu,\delta}$ of the system of stochastic differential equations for the risky asset of the momentum model converges to the solution X_t^δ of equation (4.10) on the given interval uniformly in probability as $\mu \searrow 0$ and δ fixed. This means that by considering financial products with a very high liquidity (which are in general very interesting to investors) the price process in this model can be described by an equation of 'Black-Scholes type' but with our V_t^δ as the disturbing noise!

We have argued the case δ fixed and $\mu \searrow 0$. Now we consider the consecutive case, i.e. we have already $\mu \searrow 0$ and now let $\delta \searrow 0$.

Corollary 4.4. *For $\delta \searrow 0$, the solution X_t^δ of (4.10) converges in probability uniformly on $[0,T]$ to the solution $\tilde{X}_t$ of (4.11).*

Proof. The corollary is shown by Claim 2 in the proof of Theorem 3.8. $\square$

We now consider a financial product which is very highly liquid, modeled here by setting $\mu = 0$. We now 'let the speed of spread of information become infinitely fast', i.e. we let $\delta \searrow 0$. So we ask the following question: 'How can the price process of a sufficiently highly liquid financial product be modeled in the case that the underlying financial market exhibits a fast spread of information?' The above corollary answers this question and the answer is: The according price process of the financial product can be modeled by a 'Black-Scholes type of equation', but with the stochastic integral interpreted in Stratonovich's way. By using the properties of the Stratonovich stochastic integral we see that the according system of stochastic differential equation (4.11) is equivalent to:

$$\frac{d\tilde{X}_t}{\tilde{X}_t} = \left(h_1 + \frac{1}{2}h_2^2\right) dt + h_2 dW_t, \quad \tilde{X}_0 = \xi_1. \tag{4.12}$$

Compared to the Black-Scholes model we obtain the additional 'correction term' $\frac{1}{2}h_2^2 dt$. Economically this can be interpreted as the extra gain from the additional information extracted from the momentum. This is only possible if the considered financial product is sufficiently liquid, such that momentum trading is realizable (represented by the way the limits are taken).
Thus by this model the Black-Scholes model might be a bad choice when we have a relatively high liquidity of the considered financial product with respect to the spread of information on the market! We conclude that for the above situation the 'Black-Scholes equation with Stratonovich stochastic integral' might be a better choice to model the price process of the according financial product. Our first important result is that the extra gain from the additional information extracted from the momentum is represented in the limiting stochastic differential equations as well in the above way.

We still want to go deeper with our analysis and are now concerned with the question 'Where does the 'correction term' come from mathematically and how is it linked with the economic explanation?' Now we consider a financial product with a very high liquidity. We used V_t^δ to model a not arbitrary high speed of information, resulting in a still smooth driving noise. Due to this smoothness we can extract information about the very near future with the help of the rescaled momentum (determined by the first derivative of the price process). But in order to know the first derivative of the price

process, the investor has to know the price process in a whole interval around time t (otherwise he would not be able to calculate the derivative!). We assume that the investor does have this knowledge about the very near future by including the momentum in our model. Again, the justification is that the investor 'infers the price process of the very near future' from the observed price process until time t. Due to 'smoothing delays' in the spread of information the investor has a real advantage from this knowledge about the arbitrary near future (since momentum trading is realizable because of the sufficiently high liquidity), which leads to the above (positive) correction term.
The interesting and fascinating thing now is that the Stratonovich integral in the stochastic differential equation our process converges to does exactly reflect this underlying 'knowledge about the future'! As we argued in Section 2.1, the Stratonovich integral takes the 'mean of the arbitrary small intervals' to compute the limiting sum resulting in the Stratonovich stochastic integral, whereas the Ito integral takes the 'left side of the arbitrary small intervals' to compute the limiting sum resulting in the Ito stochastic integral. So by taking into account this intuition for the two different stochastic integrals, the Stratonovich integral uses this 'knowledge about the arbitrary near future', which we exactly have in our convergence result above.

We conclude that if the liquidity is sufficiently high with respect to the speed of spread of information (but still both high), we have an extra gain due to the additional information included in the momentum, where the 'sufficiently high liquidity' ensures that we are able to execute our information about the future (available because the driving process is not arbitrarily 'serrated') before it is worthless. Note that this result is also supported by the mathematical explanation concerning the different stochastic integrals as given above.

Now we also want to consider the second convergence case.

Corollary 4.5. *For $\mu > 0$ fixed and $\delta \searrow 0$, the solution $X_t^{\mu,\delta}$ of (4.8) converges in probability uniformly on $[0, T]$ to the solution X_t^{μ} of (4.9).*

Proof. The corollary follows directly from the first part of the proof of Theorem 3.8(i). $\square$

We give the economic interpretation for this case as well. Keeping in mind the definitions of μ and δ we consider a financial product with a certain liquidity μ. We consider the case that we have a financial market with

almost arbitrary high speed of spread of information. We let this very high speed of information be represented by an according sequence of stochastic processes V_t^δ and we let our piecewise continuously differentiable driving noise V_t^δ converge to the Wiener noise W_t uniformly as $\delta \searrow 0$ P-a.s.
Then the above corollary states that the solution $X_t^{\mu,\delta}$ of the momentum model converges to the solution X_t^μ of equation (4.9) uniformly in probability on the given interval as $\delta \searrow 0$ and μ fixed.
What do we see from this result? One observation is that by 'letting the speed of spread of information become arbitrarily high' our price process of the momentum model converges to the price process of the 'momentum model supplied with a disturbing noise given by W_t' (not suprising so far).
So we conclude that for modeling markets that are 'very near to exhibit an arbitrary fast spread of information' it might be a good choice to take the momentum model with a Wiener process as the disturbing noise and with letting the stochastic integral be of Ito's sense. But keep in mind that this idea can only be applied for a financial product which is 'not too highly liquid' compared to the speed of spread of information: As we have seen in the proof in the previous section, it is essential that δ (as a measure for the inverse speed of flow of information) is sufficiently small relative to μ (as a measure for the inverse liquidity).

Analogously to our proceeding in the previous case, we will now continue the analysis of this case:

Corollary 4.6. *For $\mu \searrow 0$, the solution X_t^μ of (4.9) converges in probability uniformly on $[0,T]$ to the solution X_t of (4.3).*

Proof. This result follows from the Smoluchowski-Kramers approximation 3.7. □

How do we economically interpret this result? In this situation we consider a financial market with arbitrarily fast spread of information, represented by the Wiener process as the disturbing noise ($\delta = 0$). By letting $\mu \searrow 0$, we represent the case that the liquidity of the considered financial product is very high, i.e. problems arising due to illiquidity can be neglected. Then the above corollary states that with considering financial products with very high liquidity the price process of the risky asset in the momentum model for a perfect financial market converges to the price process described by the according stochastic differential equation of the Black-Scholes model for a risky asset.

So the result from this corollary is that if the market 'exhibits a sufficiently high enough speed of the spread of information with respect to the liquidity of the considered product' the Black-Scholes model might be a good choice to model the price process even if we want to include the momentum in our model!
Intuitively, this result is not surprising: We consider a financial market with arbitrarily high speed of spread of information and a financial product with a high liquidity, but we still assume that the speed of the flow of information is sufficiently high relative to the liquidity of the considered financial product. First of all this means that we do not have any information about the future, since the underlying process is a Wiener process and thus not smooth at all, and, secondly, due to the not infinitely high liquidity, the investor would not even be able to execute the additional information. Thus in this case (sufficiently fast spread of information and high liquidity) the investor does not have any extra gain, since there is no additional information available.
Consequently, the result for this case is that the investor does not have any extra gain and we end up with our traditional Black-Scholes model even though we took our new parameters and variables into account.
Furthermore note that also the assumptions of the momentum model (i') - (x') 'converge' to the assumptions of the Black-Scholes model (i) - (x). This observation supports the consistency of the 'limiting momentum model' with the Black-Scholes model.

In order to understand and study the structure of our model as best as possible, we considered the according limits - so far - separately (first $\mu \searrow 0$ and then $\delta \searrow 0$ or vice versa). But of course, due to Theorem 3.8, we also have a much stronger result: We even have knowledge about the limiting solution X_t or $\tilde{X}_t$ if we take the limits in a certain way (and not only 'one after the other', which is a far stronger condition). Thus keeping in mind that we analyzed the above cases in order to understand the financial structure of our model, but mathematically, we even have a stronger result, which has as well a financial interpretation. Note that even the above intuition is not completely rigorous with respect to the given cases, since we sometimes interpreted the situation as 'sufficiently high liquidity with respect to the flow of information' or vice versa. Nevertheless, since we do have this mathematical results we anticipated a bit in the above interpretation.

So let us finally conclude our mathematical results. Corollary 4.7 is the analogon of the main convergence theorem of the previous chapter. As said above, it is here given in order to specify what we mean by saying 'the liquidity is sufficiently high with respect to the flow of information' or vice versa. The stated corollary gives exactly the cases in which the two cases are applicable.

Corollary 4.7. *Assume that the stochastic process V_t^δ satisfies the usual assumptions.*
Then the solution for the momentum model $X_t^{\mu,\delta}$ of (4.8) (where V_t^δ is given by the above definition) converges in probability uniformly on $[0,T]$

(i) to the solution X_t of (4.3) if first $\delta \searrow 0$ and then $\mu \searrow 0$, i.e.

$$\forall\ \epsilon > 0 : \lim_{\mu \searrow 0} \lim_{\delta \searrow 0} \mathbb{P}\left[\sup_{0 \le t \le T} \left| X_t^{\mu,\delta} - X_t \right| > \epsilon\right] = 0.$$

More precisely: $X_t^{\mu,\delta}$ converges in probability uniformly on $[0,T]$ to X_t, if for sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\mu_n = \lim_{n\to\infty}\delta_n = 0$ we have that $\delta_n < f(\mu_n, h)$ for all $n \in \mathbb{N}$ for a certain function $f(\mu_n, h)$, $f : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, $h = 2(c_3\mu_n)^{\frac{1}{3}}$, ($c_3 \in \mathbb{R}$ constant).
That is: For any $h > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$:

$$\mathbb{P}\left[\sup_{0 \le t \le T} \left| X_t^{\mu_n,\delta_n} - X_t \right| > h\right] \le h, \text{ with } (\mu_n)_{n\in\mathbb{N}}, (\delta_n)_{n\in\mathbb{N}} \text{ as above.}$$

(ii) to the solution $\tilde{X}_t$ of (4.11) if $(\mu,\delta) \searrow 0$ such that $\lim_{(\mu,\delta)\searrow 0} \frac{1}{\delta^2}\mu e^{\frac{\hat{K}}{\delta}} = 0$ for a certain constant $\hat{K} \in \mathbb{R}_+$.
More precisely: $X_t^{\mu,\delta}$ converges in probability uniformly on $[0,T]$ to $\tilde{X}_t$, if for sequences $((\mu_n),(\delta_n))_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\mu_n = \lim_{n\to\infty}\delta_n = 0$ we have that $\lim_{n\to\infty}\mu_n \frac{1}{\delta_n^2}\exp\left\{\frac{\hat{K}}{\delta_n}\right\} = 0$ for a certain constant $\hat{K} \in \mathbb{R}_+$.
That is: For any $h > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$:

$$\mathbb{P}\left[\sup_{0 \le t \le T} \left| X_t^{\mu_n,\delta_n} - X_t \right| > h\right] \le h, \text{ with } ((\mu_n),(\delta_n))_{n\in\mathbb{N}} \text{ as above.}$$

Proof. Application of Theorem 3.8 proves the corollary. □

We have already interpreted - economically and mathematically - all cases appearing above. The only additional information here is that we specify the conditions for μ and δ, since we only did that heuristically and intuitively before.

We sum up our convergence results in words as well:
We assume that both the liquidity of the considered financial product and the speed of spread of information are high.

Case (i): **The speed of spread of information is sufficiently high with respect to the liquidity of the considered financial product (Corollary 4.7(i)).**
In this case the driving process is not smooth enough such that the rescaled momentum would be an indicator for the price process in the near future and / or additional information resulting from the momentum cannot be executed before it becomes worthless (due to the not arbitrary high liquidity). Thus in this case the investor is not able to achieve any additional gain, which is represented by convergence of the momentum model to the Black-Scholes model.

Case (ii): **The liquidity of the considered financial product is sufficiently high with respect to the speed of spread of information (Corollary 4.7(ii)).**
In this case the driving process is sufficiently smooth such that the rescaled momentum can be seen as an indicator for the price process in the very near future and due to the sufficiently high liquidity of the considered financial product the investor is able to execute this additional information before it becomes worthless. This is represented by a 'Black-Scholes type of equation' but with a Stratonovich integral, which exhibits a positive correction term compared with the Black-Scholes equation (4.3).

We conclude that the important thing here is which of the effects is more dominant (resulting from the not arbitrarily high speed of spread of information and the not arbitrarily high liquidity). Depending on the dominance of one of the two effects it is either possible to do momentum trading (information available and executable) (Case (ii)) or not (Case (i)).

4.4 Conclusion and Future Prospects

We presented a new mathematical model - our *momentum model* - taking into account a rescaled momentum, the speed of the spread of information on a given financial market and implicitly also the liquidity of the considered financial product.

Let us summarize the results: To establish a relationship to the famous Black-Scholes model, we found that for a financial market with a fast spread of information relative to the liquidity of the considered financial product the Black-Scholes model might be a good choice even with taking into account the considered factors, but we also saw that for a financial market with a higher liquidity of the considered financial product relative to the spread of information the Black-Scholes model might be a bad choice to model reality. For this case we discovered the Black-Scholes model to be limited in the sense of not being able to capture the factors above as well, but we also presented an alternative proposal for this case (a 'Black-Scholes model with Stratonovich Stochastic integral'). We found that this model leads to an extra gain represented in the stochastic differential equation describing the model, and traced it back to the additional information due to the above factors, which are in this case available and executable. We proved this results mathematically with the help of the results presented in Chapter 3.

We think we touched a still relatively unexplored field of research here with possibly enormous potential for the future. Practioners are already familiar with ideas and concepts like the momentum, but to the best of our knowledge it has not been modeled mathematically. Thus we are convinced that there are a lot of open points for further research such as:

- What are pricing formulae for our momentum model?
- How should the model be calibrated? - Meaning: On which decision rules should our parameters be chosen and what is a suitable value for a certain financial market and a certain financial product?
- Which further economical and mathematical interpretations and heuristics are possible or reasonable?
- Can this model be evaluated with the help of empirical research?
- We find it amazing that even on the level of classical mechanics there are parallels to finance. What consequences do those parallels between finance and Newtonian dynamics in physics exhibit and are there more parallels between physics and finance?

Of course, there are a lot more arising questions and topics to be discussed. Further research in very different areas (for example mathematical, empirical, economical...) seems to be promising.

Bibliography

[1] **Arnold, L. (1973).** Stochastische Differentialgleichungen. *R. Oldenbourg Verlag, München - Wien.*

[2] **Bauer, H. (1974).** Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, 2. Auflage. *De Gruyter, Berlin - New York.*

[3] **Ball, R., Kothari, S.P. (1989)** Nonstationary expected returns: Implications for tests of market efficiency and serial correlations in returns. *Journal of Financial Economics, Vol. 25, 1989, p. 51-74.*

[4] **Björk, T. (2004).** Arbitrage Theory in Continuous Time, Second Edition. *Oxford University Press, Oxford - New York.*

[5] **Black, F., Scholes, M. (1973).** The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy, Vol. 81, No. 3, pp. 637 - 654.*

[6] **Cerrai, S., Freidlin, M. (2004).** On the Smoluchowski-Kramers approximation for a system with an infinite number of degrees of freedom. *Preprints di Matematica - n.16, Scuola Normale Superiore, Pisa.*

[7] **Chan, L.K.C., Jegadeesh, N., Lakonishok, J. (1996).** Momentum Strategies. *The Journal of Finance, Vol. 51, No 5, December 1996, pp. 1681-1713.*

[8] **Chopra, N., Lakonishok, J. Ritter, J.R. (1992)** Measuring abnormal performance: Do stocks overreact? *Journal of Financial Economics, Vol. 31, 1992, p. 235-268.*

[9] **Cramér, H., Leadbetter, M.R. (1967).** Stochastic Differential Equations. Stationary and Related Stochastic Processes. *Wiley Series in Probability and Mathematical Statistics, New York.*

[10] **De Bondt, W.F.M., Thaler, R. (1985).** Does the Stock Market overreact? *The Journal of Finance, Vol. 40, No 3, July 1985, p. 793-805.*

[11] **Freidlin, M. (2004).** Some Remarks on the Smoluchowski-Kramers Approximation. *Journal of Statistical Physics, Vol. 117, No 3/4, November 2004, pp. 617-634.*

[12] **Hull, J.C. (2000).** Options, Futures, & other Derivatives, Fourth Edition, International Edition. *Prentice-Hall, Inc., New Jersey.*

[13] **Ikeda, N., Watanabe, S. (1981).** Stochastic Differential Equations and Diffusion Processes. *North-Holland Publishing Company, Amsterdam - New York - Oxford.*

[14] **Ito, K. (1951).** On a formula concerning stochastic differentials. *Nagoya Mathematical Journal, No 3, 1951, pp. 55 - 65.*

[15] **Jegadeesh, N., Titman, S. (1993).** Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance, Vol. 48, No 1, March 1993, pp. 65-91.*

[16] **Kallenberg, O. (2002).** Foundations of Modern Probability, Second Edition. *Springer, Berlin.*

[17] **Karatzas, I., Shreve, S.E. (1991).** Brownian Motion and Stochastic Calculus, Second Edition. *Springer, Berlin.*

[18] **Lamberton, D., Lapeyre, B. (1996).** Introduction to Stochastic Calculus Applied to Finance. *Chapman & Hall, London.*

[19] **Levy, R.A. (1967).** Relative Strength as a Criterion for Investment Selection. *The Journal of Finance, Vol. 22, No 4, December 1967, pp. 595-610.*

[20] **Nelson, E. (1967).** Dynamical Theories of Brownian Motion. *Princeton University Press, Princeton, New Jersey.*

[21] **Øksendal, B. (1995).** Stochastic Differential Equations: An Introduction with Applications, Fourth Edition. *Springer, Berlin.*

[22] **Röckner, M. (Winter Term 2000/2001, Summer Term 2002)** Vorlesungsskript Analysis.

[23] **Röckner, M. (Winter Term 2003/2004)** Vorlesungsskript Stochastische Analysis.

[24] **Röckner, M. (Winter Term 2002/2003, Summer Term 2003)** Vorlesungsskript Wahrscheinlichkeitstheorie.

[25] **Stratonovich, R.L. (1966).** A new representation for stochastic integrals and equations. *Siam Journal on Control, Vol. 4, pp. 362 - 371.*

[26] **Wong, E., Zakai, M. (1965).** On the Convergence of Ordinary Integrals to Stochastic Integrals. *The Annals of Mathematical Statistics, Vol. 36, No. 5, pp. 1560-1564.*

[27] **Zarowin, P. (1990)** Size, seasonality and stock market overreaction. *Journal of Financial and Quantitative Analysis. Vol. 25, 1990, p. 113-125.*

[28] **Zhang, P.G. (1998).** Exotic Options. A Guide to Second Generation Options, Second Edition. *World Scientific, Singapore - New Jersey - London - Hong Kong.*