Universität Bielefeld Fakultät für Mathematik

Non-explosion criteria via Lyapunov functions for stochastic differential equations in finite dimensions

Diplomarbeit

Betreuer: Prof. Dr. Michael Röckner

Martin Dieckmann

Bielefeld, November 2013

Contents

Introduction **1**

1. Mathematical Preliminaries **7**
1.1. Basic notations and definitions . . . 7
1.2. A characterisation of convergence in probability . . . 12

2. Framework **15**
2.1. Assumptions . . . 15
2.2. Non-explosion of solutions . . . 18

3. Existence and Uniqueness **23**
3.1. Distribution of stopping times and Skorokhod's representation theorem 23
3.2. Convergence in probability of (stochastic) integrals . . . 28
3.3. Main theorem . . . 41

4. Estimates on the Transition Probability Density **57**
4.1. Properties of positive definite matrices . . . 57
4.2. Auxiliary estimates . . . 61
4.3. Estimates on the transition probability density . . . 72

5. Existence and Uniqueness (Non-degeneracy Version) **87**
5.1. Main theorem (non-degeneracy version) . . . 87

A. Appendix **89**
A.1. Basic theorems . . . 89
A.2. Itô's formula and Itô's product rule . . . 93
A.3. Convergence . . . 95
A.4. Tightness . . . 97

Bibliography **101**

Introduction

Consider the $\mathbb{R}^d$-valued stochastic differential equation

$$\begin{aligned} \mathrm{d}X(t) &= b(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t), \\ X(0) &= \xi, \end{aligned} \tag{SDE}$$

where $\big(W(t)\big)_{t\geq 0}$ is a d_1-dimensional Wiener process with respect to a normal filtration $(\mathcal{F}_t)_{t\geq 0}$ and ξ is an $\mathcal{F}_0$-measurable random vector. Let b and σ be Borel-measurable functions mapping from $\mathbb{R}_+ \times \mathbb{R}^d$ into $\mathbb{R}^d$ and $\mathbb{R}^{d\times d_1}$, respectively.

Then there exist well-known results about the existence and uniqueness of strong solutions of the equation (SDE) under certain additional assumptions on the coefficients. For example, D. W. Stroock and S. R. S. Varadhan state such a theorem in [SV79] (cf. Chapter 5.1 starting on page 124) under the further assumptions that $b(t, x)$ and $\sigma(t, x)$ are *Lipschitz continuous* in x and bounded by a constant. In [Kry99] the results (cf. Theorem 1.2 on page 2) are heavily based on the also well-known *local weak monotonicity* and *weak coercivity* assumptions on the coefficients b and σ. By assuming the continuity of $b(t, x)$ and $\sigma(t, x)$ in x as well as an integrability criterion in addition, N. V. Krylov proves existence and uniqueness in that case.

In applications, for example in mathematical biology and financial mathematics, it is often necessary to consider stochastic differential equations in a certain domain instead of the whole space $\mathbb{R}^d$. Therefore, we have to introduce so-called *non-explosion criteria* by which we can exclude that an *explosion* occurs, i.e. that a solution leaves the domain in finite time.

The aim of this thesis is to present a more elaborate version of the article "Existence of strong solutions for Itô's stochastic equations via approximations" written by I. Gyöngy and N. V. Krylov and published in the journal "Probability Theory and Related Fields" in 1996 (see [GK96]), which concentrates on the study of the equation (SDE) in a domain $D \subseteq \mathbb{R}^d$ using the concept of *Lyapunov functions* as a condition to ensure *non-explosion*.

The study of *Lyapunov functions* in the context of stochastic differential equations in finite dimensions goes, among others, back to R. Khaminskii who considered the stability of finite-dimensional stochastic differential equations in [Kha80] (in particular Chapter 5.4. and 3.4). This book had originally been published in 1969 in Russian.

Part I: Assumptions and results

In the first part of the thesis we introduce the three assumptions **A1)**, **A2)** and **A3)**, which are of main importance for the further considerations and are slightly modified in comparision to [GK96].

First of all, we assume that for some $\chi > 0$ the coefficients b and σ are bounded by non-random locally $L^{1+\chi}$-integrable functions $M_k\colon \mathbb{R}_+ \longrightarrow]0, \infty[$ on the sets $\overline{D}_k$ belonging to an *exhausting sequence* $(D_k)_{k\in\mathbb{N}}$ of bounded domains. Namely,

A1) There exists a sequence of bounded domains $(D_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^d$ such that

- $\overline{D}_k \subseteq D_{k+1}$ for all $k \in \mathbb{N}$ and $\bigcup_{k\in\mathbb{N}} D_k = D$,
- $\sup_{x\in\overline{D}_k} \|b(t,x)\|_{\mathbb{R}^d} \leq M_k(t)$ and $\sup_{x\in\overline{D}_k} \|\sigma(t,x)\|_{L_2}^2 \leq M_k(t)$ for all $k \in \mathbb{N}$, $t \in [0,k]$.

Besides, we suppose in the crucial assumption **A2)** the existence of a *Lyapunov function* V, which is the main condition to ensure that a solution of the stochastic differential equation never leaves the domain D.

A2) There exists a non-negative function $V \in C^{1,2}\big(\mathbb{R}_+ \times D; \mathbb{R}\big)$ such that

- $LV(t,x) \leq M(t)V(t,x)$ for all $t \geq 0$, $x \in D$,
- $\inf_{\substack{x\in\partial D_k\\ t\in[0,T]}} V(t,x) \xrightarrow[k\to\infty]{} \infty$ for all $T < \infty$.

Here, L is the differential operator associated with (SDE), which is given by

$$L := \frac{\partial}{\partial t} + \sum_{i=1}^{d} b_i(t,x)\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^{d} (\sigma\sigma^T)_{ij}(t,x)\frac{\partial^2}{\partial x^i \partial x^j},$$

and M is locally in $L^{1+\chi}\big(\mathbb{R}_+;]0,\infty[\big)$. In addition, the initial value of the equation (SDE) should be P-a.s. in D, i.e. we assume that

A3) $P\big[\xi \in D\big] = 1$

holds.

Except for the Borel-measurability of the coefficients b and σ we also have to make two other additional assumptions in order to prove existence and uniqueness of a strong solution of (SDE). We assume that $b(t,x)$ and $\sigma(t,x)$ are continuous in $x \in D$ as well as that *pathwise uniqueness* holds. The pathwise uniqueness, which holds e.g. under local monotonicity assumptions, will directly yield the uniqueness of a strong solution. Hence, the important part of the main theorem is the existence.

For the proof we consider the so-called *Euler "polygonal" approximations* of the equation (SDE), which are defined as processes $(X_n(t))_{t\geq 0}$, $n \in \mathbb{N}$, given by

$$X_n(t) = \xi + \int_0^t b\big(s, X_n(\kappa_n(s))\big)\,\mathrm{d}s + \int_0^t \sigma\big(s, X_n(\kappa_n(s))\big)\,\mathrm{d}W(s),$$

where $\kappa_n(s) := t_i^n$, for $s \in [t_i^n, t_{i+1}^n[$, and $\{0 = t_0^n < t_1^n < t_2^n < \cdots < t_i^n < t_{i+1}^n < \dots\}$ is a sequence of partitions of $\mathbb{R}_+$ such that the mesh tends to zero for $n \to \infty$ and $t_i^n \to \infty$ as $i \to \infty$.

Now the first main theorem (see Theorem 3.7), which is based on a theorem in [GK96] (cf. Theorem 2.4 on page 148), states that there exists a process $(X(t))_{t\geq 0}$ such that $X_n(t) \xrightarrow[n\to\infty]{} X(t)$ in probability, uniformly in t on bounded intervals, and that $(X(t))_{t\geq 0}$ is the unique solution of (SDE).

Part I: Structure of the chapters, references and own contributions

In Chapter 1 we work out the mathematical preliminaries of this thesis, which include the basic notations and definitions in the first section. Besides, we give a detailed proof for a crucial lemma from [GK96] in Section 1.2 (see Lemma 1.14), that yields an equivalent description for convergence in probability of a sequence of random variables in terms of convergence in distribution. The necessary preparations for the proof are taken from the book [Dud02] of R. M. Dudley.

The second chapter starts with the framework of the thesis. On the basis of [GK96] we state the main assumptions **A1)** to **A3)** as well as the notion of a solution and the concept of the *Euler "polygonal" approximations*. Besides, we introduce the notion of pathwise uniqueness from [GK96]. In the second section we prove that the non-explosion criterion for solutions of the stochastic differential equation holds, i.e. we consider a lemma from [GK96] (see Lemma 2.4). For the extended version of its proof we in particular need Itô's formula and Itô's product rule, which, being cited from [KS05] and [RY99], respectively, can be found in Section A.3 of the Appendix.

At the beginning of Chapter 3 we state and prove two helpful technical lemmas (see Lemma 3.1 and 3.2) that are necessary for the proof of the first main theorem. We finish the first section by stating the important *Skorokhod representation theorem*, while referring for its proof to [Bil99]. In Section 3.2 we state a crucial lemma mentioned in [GK96] about the convergence in probability of sequences of (stochastic) integrals (see Lemma 3.6). The proof is a detailed and extended version using the basic idea of a theorem from A. V. Skorokhod (see [Sko65] on page 32).

The third section contains the first main theorem and its proof (see Theorem 3.7), which is a more elaborate version of the one given in [GK96]. In particular, a significant part of the effort is the usage of tightness criteria to prove the relatively weak compactness of sequences of probability measures via Prokhorov's theorem. The applied tightness criteria from the books [Dur96] of R. Durrett and [Bil99] of P. Billingsley are gathered in Section A.4 of the Appendix.

We finish this chapter with a remark on the application of the first main theorem in the case $D = \mathbb{R}^d$ and Corollary 3.8 about the fact that local weak monotonicity implies pathwise uniqueness.

The already mentioned Appendix, which can be found at the end of the thesis, also includes plenty of basic theorems like continuous mapping theorems, a generalised Young inequality, a generalised Minkowski inequality for integrals, Itô's formula, Prokhorov's theorem and lemmas concerning the relationship between the different types of convergence of random variables. The most important references in the Appendix are [Dud02], [Dur96], [Bil99], [vdV98], [RY99] and [KS05].

Part II: Assumptions and results

In the second part of the thesis, i.e. Chapter 4 and 5, we change the assumptions from the first part slightly with the aim that we do not have to assume the continuity of the drift coefficient b anymore.

First of all, we add a fourth assumption, the so-called *non-degeneracy* condition **A4)** for the diffusion coefficient σ, to **A1)**, **A2)** and **A3)**. Namely, we assume:

A4) For every $k \in \mathbb{N}$ the domain D_k is bounded and convex and

$$\sum_{i,j=1}^{d} (\sigma\sigma^T)_{ij}(t,x)\lambda_i\lambda_j \geq \varepsilon_k\, M_k(t) \sum_{i=1}^{d} |\lambda_i|^2$$

holds for every $t \in [0,k]$, $x \in D_k$ and $\lambda_i \in \mathbb{R}$ for $i = 1, \ldots, d$, where $\varepsilon_k > 0$ are some constants.

Then it is claimed in [GK96] (cf. Theorem 2.8 on page 149) that in this case $(X_n(t)))_{t\geq 0}$ converges in probability, uniformly in t on bounded intervals, to a unique solution $(X(t))_{t\geq 0}$ of the equation (SDE) under the further assumptions that $\sigma(t,x)$ is locally Hölder continuous in x with some exponent $\alpha \in]0,1]$ and in addition, if $\alpha \neq 1$, that the pathwise uniqueness holds for (SDE). This second main theorem is stated as Theorem 5.2 in Chapter 5.

Part II: Structure of the chapters, references and own contributions

In the fourth chapter we start with the properties of positive definite matrices in Section 4.1 and follow the book [HJ85] of R. A. Horn and C. R. Johnson as a reference.

For the proof of the second main theorem we have to consider Theorem 4.8 (see also Theorem 4.2 in [GK96] on page 153) about estimates on the transition probability density, which can be found in the third section of Chapter 4. In order to show these estimates we crucially need auxiliary estimates from Lemma 4.7 in Section 4.2 (see also Lemma 4.1 in [GK96] on page 152) at first.

We tried to follow the proof of Gyöngy and Krylov given in [GK96], but could not confirm the steps of their estimates on page 153. In fact, it seems that the first step of the inequalities may not be fulfilled for any $t > 0$ with a constant N independent of t. Namely, the term

$$\exp\Big(- \langle (a + a_1)^{-1}(y - x), y - x\rangle_{\mathbb{R}^d}\Big)$$

(cf. [GK96] on page 152 for the definition of $p_a(x,y)$) is estimated from above by $\exp\Big(- \dfrac{\|y-x\|^2_{\mathbb{R}^d}}{N\,t}\Big)$ for a constant N, which should be independent of the time variable t.

This implies that

$$\langle (a+a_1)^{-1}(y-x), y-x\rangle_{\mathbb{R}^d} \geq \frac{\|y-x\|^2_{\mathbb{R}^d}}{N\,t}$$

and, therefore, $a + a_1 \leq N\,t$ Id as well would have to hold. But by assumption we only know that the inequality $\varepsilon\,t$ Id $\leq a \leq K\,t$ Id is fulfilled for the symmetric matrix a and that a_1 is the covariance matrix of an unspecified d-dimensional Gaussian vector η with zero mean. Hence, we could not manage to prove that a_1 as well as $a + a_1$ are bounded from above by $N\,t$ Id for a constant N independent of t.

Therefore, we had to modify the proof and estimate the covariance matrix a_1 by its maximal and minimal eigenvalue. In this case we could verify the assertion of the lemma and prove the same estimate but with a constant depending on these eigenvalues.

The problem with this adjusted estimate becomes clear in the proof of Lemma 4.8. There we show that we cannot avoid that the constants may depend on the time variable since the estimates for the covariance matrix depend on it in this case. The independence would be necessary for following up the idea of the proof presented in [GK96].

Hence, we are not able to finish the proof of Lemma 4.8 completely, but we clarify the occurring difficulties instead and give an extended version of the proof up to this point.

The fifth chapter contains the second main theorem (see Theorem 5.2) with the additional assumption **A4)**. In order to keep this thesis within reasonable length, we, however, do not present its proof and refer to [GK96] instead.

Acknowledgements

I would like to thank Prof. Dr. Michael Röckner for his guidance and support. Furthermore, I am grateful to Dr. Dmitri Finkelshtein for his helpful remarks and comments on this thesis. I also want to thank my parents and my brother for their full support.

1. Mathematical Preliminaries

In this first chapter we will on the one hand fix the most important basic notations used in this thesis and besides recall some definitions from probability theory. On the other hand we will also prove the crucial Lemma 1.14 concerning an equivalent description for convergence in probability of a sequence of random variables by using convergence in distribution of pairs of subsequences to an element on the diagonal.

1.1. Basic notations and definitions

For a topological space U the expression $\mathcal{B}(U)$ will always denote the Borel-σ-algebra of U. A map $f\colon U_1 \longrightarrow U_2$ between topological spaces U_1 and U_2 is said to be Borel-measurable if it is $\mathcal{B}(U_1)/\mathcal{B}(U_2)$-measurable.

Let (S,ρ) be a metric space. For $y \in S$ and $\varepsilon > 0$ we define

$$B_\varepsilon(y) := \big\{x \in S \,\big|\, \rho(x,y) < \varepsilon\big\}$$

to be the open ball of radius ε centered at point y and denote the corresponding closed ball by $\overline{B}_\varepsilon(y) := \big\{x \in S \,\big|\, \rho(x,y) \leq \varepsilon\big\}$. Furthermore, we write $\|\cdot\|_{\mathbb{R}^d}$ for the Euclidean norm (of course: $|\cdot| = \|\cdot\|_{\mathbb{R}^1}$) and $\langle\cdot,\cdot\rangle_{\mathbb{R}^d}$ for the Euclidean inner product on $\mathbb{R}^d$. From now on the interval $[0,\infty[$ will be also denoted by $\mathbb{R}_+$.

Let $T \in \mathbb{R}_+$. In the following we will consider the space $C\big([0,T];\mathbb{R}^d\big)$ of continuous functions from $[0,T]$ to $\mathbb{R}^d$. Usually this space is equipped with the supremum norm $\|\cdot\|_\infty$ defined by

$$\|f\|_\infty := \sup_{t\in[0,T]} \|f(t)\|_{\mathbb{R}^d}.$$

Then $C\big([0,T];\mathbb{R}^d\big)$ is a separable and complete normed space. The separability follows from the fact that polynomials with rational coefficients form a countable dense subset, and for the completeness we refer for example to [Bil99] on page 11, where a proof for the space $C\big([0,1];\mathbb{R}\big)$ can be found.

For $m,n,\alpha \in \mathbb{N}$ and an open set $\Lambda \subseteq \mathbb{R}^m$ we denote by $C^\alpha(\Lambda;\mathbb{R}^n)$ the space of α-times continuously differentiable functions $f\colon \Lambda \longrightarrow \mathbb{R}^n$. We also use the notation $C^{\alpha,\beta}(\Lambda_1\times\Lambda_2;\mathbb{R}^n)$ for functions mapping from a domain $\Lambda_1\times\Lambda_2 \subseteq \mathbb{R}^{m_1}\times\mathbb{R}^{m_2}$ to $\mathbb{R}^n$ that are α-times continuously differentiable in the first and β-times in the second variable, where $m_1, m_2, \beta \in \mathbb{N}$ as well. Besides, for $1 \leq p < \infty$ we denote by $L^p := L^p(\Lambda;\mathbb{R})$ the space of equivalence classes of p-th power integrable, measurable functions from $\Lambda \subseteq \mathbb{R}^m$ to $\mathbb{R}$ equipped with the L^p-norm

$$\|f\|_{L^p} := \left(\int |f(x)|^p \,\mathrm{d}x\right)^{\frac{1}{p}}.$$

In the case $p = \infty$ we consider the essentially bounded, measurable functions for the space L^∞.

Furthermore, we introduce the notation $\mathbb{R}^{n\times m}$ for the space of $n \times m$-matrices with $\mathbb{R}$-valued entries. For $A \in \mathbb{R}^{n\times m}$ we define the Hilbert-Schmidt norm $\|\cdot\|_{L_2}$ by

$$\|A\|_{L_2}^2 := \sum_{i=1}^{n}\sum_{j=1}^{m} |A_{ij}|^2,$$

where the subindex L_2 refers to the Hilbert-Schmidt operators. In addition, we write Id for the $d \times d$ unit matrix as well as A^T for the transpose, $\det A$ for the determinant and $\operatorname{tr} A$ for the trace of a matrix A as usual.

The space of all probability measures on a measurable space $(U, \mathcal{B}(U))$ will be denoted by $\mathcal{M}_1(U)$.

Given a subset $B \subseteq U$ of a topological space U, we write B^C for the complement, $\overline{B}$ for the closure and ∂B for the boundary. Besides, we define $\inf \emptyset := \infty$ as usual. Moreover, we set $\operatorname{dist}(x, B) := \inf\big\{\|x-z\|_{\mathbb{R}^n} \,\big|\, z \in B\big\}$ as the distance of a point $x \in \mathbb{R}^n$ to a set $B \subseteq \mathbb{R}^n$.

Finally, we have to fix the notation for partial derivatives. Let $m, n \in \mathbb{N}$ and $f \in C^{1,2}\big(\mathbb{R}_+ \times \mathbb{R}^m; \mathbb{R}^n\big)$. Then we write $\frac{\partial}{\partial t} f(t,x)$ for the partial derivative with respect to the time component, $\frac{\partial}{\partial x^i} f(t,x)$, $1 \le i \le m$, for the i-th spacial partial derivative and $\frac{\partial^2}{\partial x^i \partial x^j} f(t,x)$, $1 \le i, j \le m$, for the second-order mixed partial derivative with respect to the i-th and j-th spacial component.

Now we can recall some important basic definitions from probability theory, beginning with the distribution of random variables.

Definition 1.1 (Distribution of a random variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\big(U, \mathcal{B}(U)\big)$ be a measurable space and $X\colon \Omega \longrightarrow U$ be a random variable. Then define the distribution of X by $P_X := P \circ X^{-1}$.*

In particular, we write $N(m, \Sigma)$ for a normal distribution with mean vector m and covariance matrix Σ in the following. We will also need the marginal distributions of a joint random variable.

Definition 1.2 (Marginal distribution of a joint random variable). *Let $n \in \mathbb{N}$, $(\Omega, \mathcal{F}, P)$ be a probability space and let $\big(U^i, \mathcal{B}(U^i)\big)$, for $1 \le i \le n$, be measurable spaces. Suppose in addition that $X = (X^1, \dots, X^n)\colon \Omega \longrightarrow U^1 \times \cdots \times U^n$ is a joint random variable with distribution P_X. Then, for any $1 \le k \le n$ and every subset $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ with $i_l \neq i_\ell$ for $l \neq \ell$, the distribution $P_{(X^{i_1}, \dots, X^{i_k})}$ is called marginal distribution of X.*

Recall that for an index set I, a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space $\big(U, \mathcal{B}(U)\big)$ the stochastic process X is the family $\big(X(t)\big)_{t\in I}$ of random variables $X(t)\colon \Omega \longrightarrow U$. Of course, we can think of X as the map

$$X\colon \begin{aligned} I \times \Omega &\longrightarrow U \\ (t, \omega) &\longmapsto X(t, \omega) \end{aligned}$$

as well. Now we can specify the notion of finite-dimensional distributions of a stochastic process.

Definition 1.3 (Finite-dimensional distribution of a stochastic process). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\big(U, \mathcal{B}(U)\big)$ be a measurable space and I be an index set. Suppose that $X\colon I \times \Omega \longrightarrow U$ is a stochastic process. Then for any $k \in \mathbb{N}$ and every $t_1, \ldots, t_k \in I$ the distribution $P_{(X(t_1),\ldots,X(t_k))}$ is said to be a finite-dimensional distribution of X.*

A probability space $(\Omega, \mathcal{F}, P)$ is called complete if for every $N \in \mathcal{F}$ with $P[N] = 0$ and for all $N' \subseteq N$ we have $N' \in \mathcal{F}$, i.e. every subset of a P-zero set in $\mathcal{F}$ is again contained in $\mathcal{F}$. By a filtration $(\mathcal{F}_t)_{t\geq 0}$ we mean a family of sub-σ-algebras of $\mathcal{F}$ such that we have $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$.

Definition 1.4 (Stochastic basis, cf. [PR07] on page 121). *We call $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$ a stochastic basis if $(\Omega, \mathcal{F}, P)$ is a complete probability space and $(\mathcal{F}_t)_{t\geq 0}$ is a normal filtration, i.e. $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous and $\mathcal{F}_0$ contains all P-zero sets.*

Next, we clarify the concept of equality of two stochastic processes that we will use later in this thesis when it comes to the uniqueness of a solution of the considered stochastic differential equation.

Definition 1.5 (P-indistinguishable). *Let $\big(X(t)\big)_{t\geq 0}$ and $\big(Y(t)\big)_{t\geq 0}$ be two stochastic processes on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a measurable space $\big(U, \mathcal{B}(U)\big)$. Then they are called P-indistinguishable if*

$$P\Big[X(t) = Y(t),\ \forall t \geq 0\Big] = 1.$$

At this point we have to emphasise that whenever a P-a.s. continuous stochastic process is given, we can replace it by the altered and P-indistinguishable process which is continuous for every $\omega \in \Omega$. This basic idea is stated in the following remark and will be used tacitly in future.

Remark (Continuity of stochastic processes). Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and X be a P-a.s. continuous stochastic process taking values in a measurable space $\big(U, \mathcal{B}(U)\big)$. Then the set

$$\Omega_0 := \Big\{\omega \in \Omega \,\Big|\, t \longmapsto X(t, \omega) \text{ is not continuous}\Big\}$$

is a measurable P-zero set, i.e. $P[\Omega_0] = 0$, by the completeness of $(\Omega, \mathcal{F}, P)$. Therefore, we can always consider the process given by

$$\bar{X}(\cdot, \omega) := \begin{cases} X(\cdot, \omega) & \text{for } \omega \in \Omega \setminus \Omega_0, \\ 0 & \text{for } \omega \in \Omega_0, \end{cases}$$

which is in fact continuous for every $\omega \in \Omega$.

Moreover, we should recall the notion of convergence of random variables, where we can also introduce the notation of weak convergence of probability measures at first.

Definition 1.6 (Weak convergence of probability measures, cf. [Bil99], page 7). *Let U be a topological space and let μ, μ_n, for $n \in \mathbb{N}$, be probability measures on $\big(U, \mathcal{B}(U)\big)$. We say that $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ if*

$$\int f \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int f \,\mathrm{d}\mu$$

for every bounded function $f \in C(U, \mathbb{R})$, and denote the weak convergence of probability measures by $\mu_n \xrightarrow[n\to\infty]{w} \mu$.

Definition 1.7 (Convergence of random variables). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and (S, ρ) be a separable metric space. Let Z and Z_n, for $n \in \mathbb{N}$, be S-valued random variables on $(\Omega, \mathcal{F}, P)$. We say that*

i) $(Z_n)_{n\in\mathbb{N}}$ converges P-a.s. to Z if

$$P\Big[\lim_{n\to\infty} \rho(Z_n, Z) = 0\Big] = 1,$$

and denote this by $Z_n \xrightarrow[n\to\infty]{P-a.s.} Z$,

ii) $(Z_n)_{n\in\mathbb{N}}$ converges in probability to Z if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P\Big[\rho(Z_n, Z) \geq \varepsilon\Big] = 0,$$

and write $Z_n \xrightarrow[n\to\infty]{p} Z$,

iii) $(Z_n)_{n\in\mathbb{N}}$ converges in distribution (or weakly) to Z if

$$P_{Z_n} \xrightarrow[n\to\infty]{w} P_Z,$$

and denote this by $Z_n \xrightarrow[n\to\infty]{d} Z$.

For some important basic lemmas concerning the relationship between the different types of convergence of random variables we refer to Section A.3 in the Appendix.

Besides, we also mention the definition of Polish spaces, which we will use in the following Section 1.2.

Definition 1.8 (Polish space, cf. [Kle06], Definition 13.1 on page 235). *A topological space is called Polish if it is completely metrisable and separable.*

Now let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$ be a stochastic basis. First of all, we recall the notion of *bounded variation* of a process $(B(t))_{t\geq 0}$. Let $t \geq 0$ and let $\Pi = \big\{0 = s_0 < s_1 < \cdots < s_m = t\big\}$ for some $m \in \mathbb{N}$ be a partition of $[0, t]$. Then B is of bounded variation if

$$\sup_{\Pi} \sum_{i:\, s_{i+1}\in\Pi} \|B(s_{i+1}) - B(s_i)\|_{\mathbb{R}^d} < \infty$$

for every $t \geq 0$ (see e.g. [KS05] on page 32). Moreover, we recall the definition of a continuous local martingale (up to ∞).

Definition 1.9 (Continuous local martingale, [KS05], Definition 5.15 on page 36)**.** *A continuous, $(\mathcal{F}_t)$-adapted process $(Z(t))_{t\geq 0}$ with $Z(0) = 0$ P-a.s. is said to be a continuous local $(\mathcal{F}_t)$-martingale (up to ∞) if there exists a non-decreasing sequence of $(\mathcal{F}_t)$-stopping times $(\tau_n)_{n\in\mathbb{N}}$ such that $\big(Z(t\wedge\tau_n)\big)_{t\geq 0}$ is a continuous $(\mathcal{F}_t)$-martingale for every $n \in \mathbb{N}$ and*

$$P\Big[\lim_{n\to\infty} \tau_n = \infty\Big] = 1.$$

Finally, we can give the definition of a continuous semimartingale.

Definition 1.10 (Continuous semimartingale, cf. [RY99], Definition 1.17 on page 127)**.** *A continuous semimartingale $(Y(t))_{t\geq 0}$ is an $(\mathcal{F}_t)$-adapted process which has P-a.s. the decomposition*

$$Y(t) = Y_0 + Z(t) + B(t)$$

for every $t \geq 0$, where $(Z(t))_{t\geq 0}$ is an $(\mathcal{F}_t)$-adapted continuous local martingale,$(B(t))_{t\geq 0}$ is a continuous, $(\mathcal{F}_t)$-adapted process of bounded variation and Y_0 is an $\mathcal{F}_0$-measurable random vector.

1.2. A characterisation of convergence in probability

For the proof of the previously mentioned Lemma 1.14, we first have to clarify how to metrise a certain space of random variables with respect to convergence in probability. The following considerations are based on Dudley's Chapter 9.2 in [Dud02].

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(U, \mathcal{U})$ be a measurable space. Then denote by $\mathcal{E}\big(\Omega, \mathcal{F}; U, \mathcal{U}\big)$ the set of all $\mathcal{F}/\mathcal{U}$-measurable functions from Ω to U. Furthermore, let $\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; U, \mathcal{U}\big)$ be the set of all equivalence classes of elements of $\mathcal{E}\big(\Omega, \mathcal{F}; U, \mathcal{U}\big)$ with respect to P-a.s. equality.

Definition 1.11 (Ky Fan metric, cf. [Dud02] on page 289). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and (S, ρ) be a separable metric space. Then define the map $\tilde{\rho}$ by*

$$\tilde{\rho}(X, Y) := \inf \big\{ \varepsilon \geq 0 \,\big|\, P\big[\rho(X, Y) \geq \varepsilon\big] \leq \varepsilon \big\}$$

for any $X, Y \in \mathcal{E}\big(\Omega, \mathcal{F}; S, \mathcal{B}(S)\big)$.

We note at this point that $\tilde{\rho}$ is only a semimetric (or pseudometric) on $\mathcal{E}\big(\Omega, \mathcal{F}; S, \mathcal{B}(S)\big)$ because the coincidence axiom is not fulfilled due to matters of P-a.s. equality. On $\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; S, \mathcal{B}(S)\big)$ then again, $\tilde{\rho}$ is in fact a metric as it is stated in the following theorem and, therefore, said to be the Ky Fan metric.

Remark (cf. [Dud02] on page 289). Note that the definitions of P-a.s. convergence and convergence in probability are unaffected by replacing random variables by P-a.s. equal ones. Hence, these modes of convergence from Definition 1.7 are also defined on $\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; S, \mathcal{B}(S)\big)$.

Theorem 1.12 (cf. [Dud02], Theorem 9.2.2 on page 289). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and (S, ρ) be a separable metric space. Then the map $\tilde{\rho}$ is a metric on $\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}; S, \mathcal{B}(S)\big)$, which corresponds to convergence in probability, i.e. a sequence $(Z_n)_{n \in \mathbb{N}}$ of random variables converges in probability to Z if and only if $\tilde{\rho}(Z_n, Z) \xrightarrow[n \to \infty]{} 0$.*

Proof. We refer to [Dud02], Theorem 9.2.2 on page 289. □

Remark. The infimum in this definition of the Ky Fan metric is always attained. For further details we refer to [Dud02] on page 289.

Finally, we can also show the completeness of $\big(\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; E, \mathcal{B}(E)\big), \tilde{\rho}\big)$ for a Polish space $\big(E, \rho\big)$.

Theorem 1.13 (cf. [Dud02], Theorem 9.2.3 on page 290). *If (E, ρ) is a Polish space and $(\Omega, \mathcal{F}, P)$ is a probability space, then $\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; E, \mathcal{B}(E)\big)$ is complete with respect to the Ky Fan metric $\tilde{\rho}$.*

Proof. We refer to [Dud02], Theorem 9.2.3 on page 290. □

After these preparations we can state Lemma 1.14, which is one of the fundamental ideas for the proof of the main theorem (cf. Theorem 3.7) in Section 3.3. Using this observation, which has been stated by Gyöngy and Krylov in [GK96], we can prove convergence in probability of the *Euler "polygonal" approximations* to a solution of the stochastic differential equation later.

Lemma 1.14 (cf. [GK96], Lemma 1.1 on page 144)**.** *Let (E, ρ) be a Polish space equipped with the Borel σ-algebra $\mathcal{B}(E)$ and $D := \big\{(x, y) \in E \times E \,\big|\, x = y\big\}$. Suppose in addition that $(Z_n)_{n\in\mathbb{N}}$ is a sequence of E-valued random variables on a probability space $(\Omega, \mathcal{F}, P)$. Then the following assertions are equivalent.*

i) There exists an E-valued random variable Z such that $Z_n \xrightarrow[n\to\infty]{p} Z$.

ii) For every pair $\big(Z_{n_l}, Z_{\bar{n}_l}\big)_{l\in\mathbb{N}}$ of subsequences of $(Z_n)_{n\in\mathbb{N}}$ there exists a subsequence $(z_k)_{k\in\mathbb{N}} := \big(Z_{n_{l_k}}, Z_{\bar{n}_{l_k}}\big)_{k\in\mathbb{N}}$ such that $z_k \xrightarrow[k\to\infty]{d} z$ for a D-valued random variable z.

Proof. (cf. [GK96] on page 145)
"*i) $\Rightarrow$ ii)*": Let $Z_n \xrightarrow[n\to\infty]{p} Z$ for an E-valued random variable Z. Then every subsequence of $(Z_n)_{n\in\mathbb{N}}$ converges in probability to Z. Besides, we also have the convergence in probability of pairs of subsequences (cf. Lemma A.13 in the Appendix). Hence, z is given by (Z, Z). Since convergence in probability implies convergence in distribution (cf. Lemma A.12 ii) in the Appendix), the assertion holds.

"*ii) $\Rightarrow$ i)*": Let $\tilde{Z}_n$ for $n \in \mathbb{N}$ be the equivalence class related to Z_n. Then we prove the following claim.

Claim (1)**.** $(\tilde{Z}_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\big(\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; E, \mathcal{B}(E)\big), \tilde{\rho}\big)$.

For convenience we write Z_n for the representative of the equivalence class. Since the convergence in probability and the definition of the Ky Fan metric are well-defined, the calculations are independent of the choice of the representative.

Proof of the Claim (1). Assume that $(Z_n)_{n\in\mathbb{N}}$ is not a Cauchy sequence, i.e. there exists an $\varepsilon > 0$ such that for all $l \in \mathbb{N}$ there exist $m, m' \geq l$ such that $\tilde{\rho}(Z_m, Z_{m'}) > \varepsilon$. Hence, we can find subsequences of $(Z_n)_{n\in\mathbb{N}}$ such that

$$\tilde{\rho}(Z_{n_l}, Z_{\bar{n}_l}) > \varepsilon \tag{1.1}$$

for every $l \in \mathbb{N}$. Then there exists a subsequence $(z_k)_{k\in\mathbb{N}} := \big(Z_{n_{l_k}}, Z_{\bar{n}_{l_k}}\big)_{k\in\mathbb{N}}$ such that $z_k \xrightarrow[k\to\infty]{d} z$ for a D-valued random variable z by assumption.
By the continuous mapping theorem (cf. Lemma A.1 in the Appendix) we know that $f(z_k) \xrightarrow[n\to\infty]{d} f(z)$ holds for any continuous function f between metric spaces. Since

$$\rho\colon \begin{array}{c} E \times E \longrightarrow \mathbb{R}_+, \\ (x, y) \longmapsto \rho(x, y) \end{array}$$

is continuous and $\rho(z) = 0$, we have for the sequence $\big(\rho(z_k)\big)_{k\in\mathbb{N}}$ of random variables that $\rho(z_k) \xrightarrow[n\to\infty]{d} 0$. At this point we need the separability of E to ensure that $\rho(z_k)$ for $k \in \mathbb{N}$ are random variables as desired. Details can be found in [Dud02] on page 287 and for example in [Bil99] on page 27 or in [Kle06] on page 125. Then it follows that

$$\rho(z_k) \xrightarrow[k\to\infty]{p} 0$$

holds due to the fact that convergence in distribution to a constant implies convergence in probability to this constant (cf. Lemma A.12 iii) in the Appendix).
Consequently, $\tilde{\rho}(Z_{n_{l_k}}, Z_{\bar{n}_{l_k}}) \xrightarrow[k\to\infty]{} 0$ by the definition of the Ky Fan metric. That is a contradiction to the assumed inequality (1.1). □

By the completeness of $\big(\tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; E, \mathcal{B}(E)\big), \tilde{\rho}\big)$ (see Theorem 1.13) we can conclude the convergence of the Cauchy sequence $(\tilde{Z}_n)_{n\in\mathbb{N}}$. Now Theorem 1.12 implies that $(\tilde{Z}_n)_{n\in\mathbb{N}}$ converges in probability to some $\tilde{Z} \in \tilde{\mathcal{E}}\big(\Omega, \mathcal{F}, P; E, \mathcal{B}(E)\big)$. Therefore, we certainly obtain that the sequence $(Z_n)_{n\in\mathbb{N}}$ converges in probability to an E-valued random variable Z. Hence, the assertion is fulfilled. □

2. Framework

At the beginning of this chapter we will state the three main assumptions **A1)**, **A2)** and **A3)** for the considered stochastic differential equation (shortly: SDE), which we use hereafter in the whole thesis. In addition, we will introduce the so-called *Euler "polygonal" approximations* and the notion of a solution of the SDE. Besides, we define the notion of pathwise uniqueness for the SDE in the same way as it is done in [GK96].

In the second section we will prove the *"non-explosion"* Lemma 2.4, which ensures that solutions of the SDE never leave a given domain, what makes it the crucial lemma of this chapter.

2.1. Assumptions

The following assumptions and definitions are based on the article [GK96] by Gyöngy and Krylov, but we have changed for example **A1)** and the assumed integrability of the functions M and M_k, $k \in \mathbb{N}$, slightly.

Let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$ be a stochastic basis and $d, d_1 \in \mathbb{N}$. Consider the $\mathbb{R}^d$-valued SDE

$$\begin{aligned} \mathrm{d}X(t) &= b(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t), \\ X(0) &= \xi \end{aligned} \tag{2.1}$$

in a domain $D \subseteq \mathbb{R}^d$, where $\big(W(t)\big)_{t\geq 0}$ is a d_1-dimensional Wiener process with respect to $(\mathcal{F}_t)_{t\geq 0}$ and ξ is an $\mathcal{F}_0$-measurable random vector with values in D. Furthermore,

$$\begin{aligned} b&\colon \mathbb{R}_+ \times D \longrightarrow \mathbb{R}^d, \\ \sigma&\colon \mathbb{R}_+ \times D \longrightarrow \mathbb{R}^{d\times d_1} \end{aligned}$$

are assumed to be Borel-measurable functions, and we define $b(t,x) = 0 = \sigma(t,x)$ for $x \in \mathbb{R}^d \setminus D$, $t \in \mathbb{R}_+$. Now we clarify the notion of a solution used in this thesis.

Definition 2.1 (Solution of equation (2.1))**.** *Let $(X(t))_{t\geq 0}$ be a P-a.s. continuous, $\mathbb{R}^d$-valued, $(\mathcal{F}_t)$-adapted process that satisfies P-a.s. the SDE* (2.1) *for all $t \in [0, \infty[$. Then $(X(t))_{t\geq 0}$ is called solution of equation* (2.1)*.*

Let $\chi > 0$ and let

$$M, M_k \colon \mathbb{R}_+ \longrightarrow]0, \infty[,$$

for $k \in \mathbb{N}$, be locally $L^{1+\chi}\big(\mathbb{R}_+;]0,\infty[\big)$-integrable functions. Then we can introduce the following assumptions, which are of main importance for the whole thesis.

A1) There exists a sequence of bounded domains $(D_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^d$ such that

- $\overline{D}_k \subseteq D_{k+1}$ for all $k \in \mathbb{N}$ and $\bigcup_{k\in\mathbb{N}} D_k = D$,
- $\sup_{x\in\overline{D}_k} \|b(t,x)\|_{\mathbb{R}^d} \leq M_k(t)$ and $\sup_{x\in\overline{D}_k} \|\sigma(t,x)\|_{L_2}^2 \leq M_k(t)$ for all $k \in \mathbb{N}$, $t \in [0,k]$.

A2) There exists a non-negative function $V \in C^{1,2}\big(\mathbb{R}_+ \times D; \mathbb{R}\big)$ such that

- $LV(t,x) \le M(t)V(t,x)$ for all $t \in \mathbb{R}_+$ and $x \in D$, where L is the differential operator given by

$$L := \frac{\partial}{\partial t} + \sum_{i=1}^{d} b_i(t,x)\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^{d}(\sigma\sigma^T)_{ij}(t,x)\frac{\partial^2}{\partial x^i \partial x^j},$$

- $V_k(T) := \inf_{\substack{x\in\partial D_k\\ t\in[0,T]}} V(t,x) \xrightarrow[k\to\infty]{} \infty$ for all $T < \infty$.

A3) $P\big[\xi \in D\big] = 1$.

Note that **A1)** is an assumption for the existence of an exhausting sequence for the domain D, in which we demand some kind of boundedness of the coefficients b and σ by non-random locally $L^{1+\chi}$-integrable functions. Assumption **A2)** is said to be a *Lyapunov condition* on the existence of a *Lyapunov function* V, which provides an estimate for the differential operator L associated with equation (2.1). In particular, that is central to the proof of Lemma 2.4, where we show that the occurrence of so-called *explosions* can be excluded, i.e. that solutions of the SDE (2.1) never actually leave the domain D.

Hence, we would like to emphasise that the definition of b and σ outside of D is only for convenience.

Additionally, we will consider the *Euler "polygonal" approximations* of the SDE (2.1). Therefore, we define a sequence of partitions of $\mathbb{R}_+$ given by

$$\big\{0 = t_0^n < t_1^n < t_2^n < \cdots < t_i^n < t_{i+1}^n < \ldots\big\}$$

such that $t_i^n \xrightarrow[i\to\infty]{} \infty$ and that the mesh $d_n(T)$ tends to zero for every $T > 0$, i.e.

$$d_n(T) := \sup_{i\,:\,t_{i+1}^n \le T} |t_{i+1}^n - t_i^n| \xrightarrow[n\to\infty]{} 0.$$

Now for every $n \in \mathbb{N}$ let $\kappa_n(s) := t_i^n$, for $s \in [t_i^n, t_{i+1}^n[$, and define the *Euler "polygonal" approximations* as the process $(X_n(t))_{t\ge 0}$ given by

$$X_n(t) = \xi + \int_0^t b\big(s, X_n(\kappa_n(s))\big)\,\mathrm{d}s + \int_0^t \sigma\big(s, X_n(\kappa_n(s))\big)\,\mathrm{d}W(s) \tag{2.2}$$

for $t \in [0,\infty[$.

Remark. Note that by assumption **A1)** and by the definition of b and σ outside of D both integrals on the right-hand sides of (2.2) exist for all $t \in [0,\infty[$. Indeed, we have

$$\int_0^k \sup_{x\in\overline{D}_k} \|b(s,x)\|_{\mathbb{R}^d}\,\mathrm{d}s + \int_0^k \sup_{x\in\overline{D}_k} \|\sigma(s,x)\|_{L_2}^2\,\mathrm{d}s \le 2\int_0^k M_k(s)\,\mathrm{d}s < \infty$$

for every $k \in \mathbb{N}$.

Finally, we define pathwise uniqueness for the SDE (2.1) in the same way as it is done in [GK96]. This property ensures, as an assumption in the upcoming Theorem 3.7, the uniqueness of the strong solution.

Definition 2.2 (Pathwise uniqueness for equation (2.1))**.** *Let $P_{(W,\xi)}$ be the joint distribution of (W, ξ) given by equation (2.1). We say that pathwise uniqueness holds for equation (2.1) if for any stochastic basis $(\Omega', \mathcal{F}', P', (\mathcal{F}'_t)_{t\geq 0})$ carrying a d_1-dimensional Wiener process W' and a random variable ξ', such that*

$$P'_{(W',\xi')} = P_{(W,\xi)}$$

is fulfilled, we have that equation (2.1) with W' and ξ' instead of W and ξ cannot have more than one solution (up to P'-indistinguishability).

2.2. Non-explosion of solutions

At first we have to recall a lemma that is similar to Lemma 3.1.3 from [PR07] on page 44, which is an adaption of Chebyshev's inequality including stopping times.

Lemma 2.3 (cf. [PR07], Lemma 3.1.3 on page 44)**.** *Let $(Y(t))_{t\geq 0}$ be a P-a.s. continuous, $\mathbb{R}_+$-valued, $(\mathcal{F}_t)$-adapted process on a stochastic basis $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$. Suppose that γ is an $(\mathcal{F}_t)$-stopping time, $\varepsilon \in]0,\infty[$ and $T > 0$. Define*

$$\tau_\varepsilon := \gamma \wedge \inf\big\{t \geq 0 \,\big|\, Y(t) \geq \varepsilon\big\}.$$

Then

$$P\Big[\sup_{t\in[0,\gamma]} Y(t) \geq \varepsilon, 0 < \gamma \leq T\Big] \leq \frac{1}{\varepsilon}\,\mathbb{E}\Big[Y(\tau_\varepsilon)\mathbb{1}_{\{0<\gamma\leq T\}}\Big].$$

Proof. We refer to [PR07], Lemma 3.1.3 on page 44. □

Now we can consider the previously mentioned Lemma 2.4, that is based on [GK96]. It states that a solution of the SDE (2.1) never actually leaves the domain D.

Lemma 2.4 ("Non-explosion", cf. [GK96], Lemma 2.2 on page 147)**.** *Assume that $(X(t))_{t\geq 0}$ is a P-a.s. continuous, $\mathbb{R}^d$-valued, $(\mathcal{F}_t)$-adapted process that satisfies the SDE (2.1) for $t < \tau$, where $\tau := \inf\big\{t \geq 0 \,\big|\, X(t) \notin D\big\}$. Suppose moreover that the assumptions from Section 2.1 are fulfilled. Then P-a.s. we have $\tau = \infty$.*

Proof. (cf. [GK96], Lemma 2.2 on page 147)
For $k \in \mathbb{N}$ define the stopping times

$$\tau^k := \inf\big\{t \geq 0 \,\big|\, X(t) \notin D_k\big\} \wedge k.$$

Then $\tau^k \uparrow \tau$ because of the assumptions $\overline{D}_k \subseteq D_{k+1}$ for all $k \in \mathbb{N}$ and $\bigcup_{k\in\mathbb{N}} D_k = D$ in **A1)**. Furthermore, note that for every $T \in]0,\infty[$ there exists a $K \in \mathbb{N}$ such that $V_k(T) > 0$ for all $k \geq K$ since $V_k(T) \xrightarrow[k\to\infty]{} \infty$ by **A2)**. Now we are going to prove the following claim.

Claim (1)**.** For every $T > 0$, $\delta > 0$ and $k \in \mathbb{N}$, such that $V_k(T) > 0$ and $k \geq T$, the inequality

$$\begin{aligned} P\big[\tau^k \leq T\big] \leq P\big[\xi \notin D_k\big] + P\Big[V(0,\xi) \geq \log\Big(\frac{1}{\delta}\Big)\Big] \\ + \frac{1}{\delta V_k(T)} \exp\Big(\int_0^T M(t)\,\mathrm{d}t\Big) \end{aligned} \tag{2.3}$$

holds.

Proof of the Claim (1). First of all, we mention that by **A3)** the expression $V(0,\xi)$ is defined. Now let $T > 0$, $\delta > 0$ and let $k \in \mathbb{N}$ be such that $V_k(T) > 0$ and $k \geq T$.

Then we have

$$
\begin{aligned}
P\big[\tau^k \le T\big] &\le P\big[\tau^k \le T, \xi \notin D_k\big] + P\Big[\tau^k \le T, \xi \in D_k, V(0,\xi) \ge \log\Big(\frac{1}{\delta}\Big)\Big] \\
&\qquad\qquad + P\Big[\tau^k \le T, \xi \in D_k, V(0,\xi) < \log\Big(\frac{1}{\delta}\Big)\Big] \\
&\le P\big[\xi \notin D_k\big] + P\Big[V(0,\xi) \ge \log\Big(\frac{1}{\delta}\Big)\Big] \\
&\qquad\qquad + P\Big[0 < \tau^k \le T, V(0,\xi) < \log\Big(\frac{1}{\delta}\Big)\Big]
\end{aligned}
\tag{2.4}
$$

since $\xi \in D_k$ implies $\tau^k > 0$. The latter implication holds because D_k is an open set and X is P-a.s. continuous. In order to estimate $P\big[0 < \tau^k \le T, V(0,\xi) < \log(\frac{1}{\delta})\big]$, we define

$$
\gamma(t) := \exp\Big(- \int_0^t M(s)\,\mathrm{d}s - V(0,\xi)\Big).
$$

Applying Itô's product rule for semimartingales (cf. Theorem A.11 in the Appendix) to $\gamma(t)V(t,X(t))$ yields P-a.s.

$$
\begin{aligned}
\gamma(t)V\big(t,X(t)\big) = \gamma(0)V\big(0,X(0)\big) &+ \int_0^t \gamma(s)\,\mathrm{d}V\big(s,X(s)\big) + \int_0^t V\big(s,X(s)\big)\,\mathrm{d}\gamma(s) \\
&+ \langle \gamma(\cdot), V\big(\cdot, X(\cdot)\rangle_t
\end{aligned}
$$

for every $t \in [0,\tau^k]$, where $\langle \gamma(\cdot), V\big(\cdot, X(\cdot)\big)\rangle_t = 0$ since γ is of bounded variation. Now we use Itô's formula (cf. Corollary A.10 in the Appendix) for $V(t,X(t))$, which gives us P-a.s.

$$
\begin{aligned}
V\big(t,X(t)\big) = V\big(0,X(0)\big) &+ \int_0^t \frac{\partial}{\partial t} V\big(s,X(s)\big)\,\mathrm{d}s \\
&+ \sum_{i=1}^d \int_0^t b_i\big(s,X(s)\big)\,\frac{\partial}{\partial x^i} V\big(s,X(s)\big)\,\mathrm{d}s \\
&+ \int_0^t \big\langle \nabla_x V\big(s,X(s)\big), \sigma\big(s,X(s)\big)\,\mathrm{d}W(s)\big\rangle_{\mathbb{R}^d} \\
&+ \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \int_0^t (\sigma\sigma^T)_{ij}\big(s,X(s)\big)\frac{\partial}{\partial x^i \partial x^j} V\big(s,X(s)\big)\,\mathrm{d}s
\end{aligned}
$$

for every $t \in [0,\tau^k]$. We compute

$$
\begin{aligned}
\int_0^t \gamma(s)\,\mathrm{d}V\big(s,X(s)\big) = \int_0^t \gamma(s)\,\frac{\partial}{\partial t} V\big(s,X(s)\big)\,\mathrm{d}s \\
&+ \int_0^t \gamma(s)\,\sum_{i=1}^d b_i\big(s,X(s)\big)\,\frac{\partial}{\partial x^i} V\big(s,X(s)\big)\,\mathrm{d}s \\
&+ \int_0^t \gamma(s)\,\big\langle \nabla_x V\big(s,X(s)\big), \sigma\big(s,X(s)\big)\,\mathrm{d}W(s)\big\rangle_{\mathbb{R}^d} \\
&+ \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \int_0^t \gamma(s)\,(\sigma\sigma^T)_{ij}\big(s,X(s)\big)\frac{\partial}{\partial x^i \partial x^j} V\big(s,X(s)\big)\,\mathrm{d}s
\end{aligned}
$$

and

$$\int_0^t V\big(s, X(s)\big)\,\mathrm{d}\gamma(s) = -\int_0^t V\big(s, X(s)\big)\,\gamma(s)\,M(s)\,\mathrm{d}s.$$

Then, by using the definition of the differential operator L and assumption **A2)**, we have P-a.s. for every $t \in [0, \tau^k]$

$$\begin{aligned}
&\gamma(t)V\big(t, X(t)\big)\\
&\quad= \gamma(0)V\big(0, X(0)\big) + \int_0^t \gamma(s)LV\big(s, X(s)\big)\,\mathrm{d}s - \int_0^t M(s)\gamma(s)V\big(s, X(s)\big)\,\mathrm{d}s\\
&\qquad\quad + \int_0^t \big\langle \gamma(s)\nabla_x V\big(s, X(s)\big), \sigma\big(s, X(s)\big)\,\mathrm{d}W(s)\big\rangle_{\mathbb{R}^d}\\
&\quad\le \gamma(0)V\big(0, X(0)\big) + \int_0^t M(s)\gamma(s)V\big(s, X(s)\big)\,\mathrm{d}s - \int_0^t M(s)\gamma(s)V\big(s, X(s)\big)\,\mathrm{d}s\\
&\qquad\quad + \underbrace{\int_0^t \big\langle \gamma(s)\nabla_x V\big(s, X(s)\big), \sigma\big(s, X(s)\big)\,\mathrm{d}W(s)\big\rangle_{\mathbb{R}^d}}_{=:\,m(t)}\\
&\quad= \gamma(0)V\big(0, X(0)\big) + m(t),
\end{aligned}$$

where $m(t)$, $t \in [0, \tau^k]$, is a continuous local $(\mathcal{F}_t)$-martingale with $m(0) = 0$.

Hence, for any $(\mathcal{F}_t)$-stopping time $\vartheta \le \tau^k$ and for any sequence $(\phi_s)_{s\in\mathbb{N}}$ of $(\mathcal{F}_t)$-stopping times with $\phi_s \uparrow \tau^k$ such that $m(t \wedge \phi_s)$, $t \in [0, \tau^k]$, is a martingale for every $s \in \mathbb{N}$, we have by Fatou's lemma

$$\begin{aligned}
\mathbb{E}\Big[\gamma(\vartheta)V\big(\vartheta, X(\vartheta)\big)\mathbb{1}_{\{0<\tau^k\le T\}}\Big] &\le \mathbb{E}\Big[\lim_{s\to\infty} \underbrace{\gamma(\vartheta\wedge\phi_s)V\big(\vartheta\wedge\phi_s, X(\vartheta\wedge\phi_s)\big)}_{\ge 0}\Big]\\
&\le \liminf_{s\to\infty} \mathbb{E}\Big[\gamma(\vartheta\wedge\phi_s)V(\vartheta\wedge\phi_s, X(\vartheta\wedge\phi_s))\Big]\\
&\le \liminf_{s\to\infty} \mathbb{E}\Big[\gamma(0)V(0,\xi) + m(\vartheta\wedge\phi_s)\Big]\\
&\le \mathbb{E}\Big[\gamma(0)V(0,\xi)\Big] + \liminf_{s\to\infty} \underbrace{\mathbb{E}\Big[m(\vartheta\wedge\phi_s)\Big]}_{=0,\text{ since } m(0)=0},
\end{aligned}$$

where we have used that V and γ are continuous and also that X is P-a.s. continuous in the first step. Let $R > 0$ and $\tilde{\vartheta} := \tau^k \wedge \inf\big\{t \ge 0 \,\big|\, \gamma(t)V\big(t, X(t)\big) \ge R\big\}$. Then by applying Lemma 2.3 it follows that

$$\begin{aligned}
P\Big[\sup_{t\in[0,\tau^k]} \gamma(t)V\big(t, X(t)\big) \ge R, 0 < \tau^k \le T\Big] &\le \frac{1}{R}\,\mathbb{E}\Big[\gamma(\tilde{\vartheta})V\big(\tilde{\vartheta}, X(\tilde{\vartheta})\big)\mathbb{1}_{\{0<\tau^k\le T\}}\Big]\\
&\le \frac{1}{R}\,\mathbb{E}\Big[\gamma(0)V(0,\xi)\Big]\\
&\le \frac{1}{R}
\end{aligned}\tag{2.5}$$

holds since $e^{-z}z \leq \frac{1}{e}$ for every $z \in \mathbb{R}$ and, hence,

$$\mathbb{E}\big[\gamma(0)V(0,\xi)\big] = \mathbb{E}\Big[\exp\big(-V(0,\xi)\big)V(0,\xi)\Big] \leq 1.$$

Note that for $0 < \tau^k \leq T$ we have

$$0 < V_k(T) = \inf_{\substack{x\in\partial D_k \\ r\in[0,T]}} V\big(r,x\big) \leq V\big(\tau^k, X(\tau^k)\big) \leq \sup_{r\in[0,\tau^k]} V\big(r,X(r)\big) < \infty,$$

where the finiteness is fulfilled since V is a continuous function and X is P-a.s. continuous. Now we can complete the estimate of the last summand in inequality (2.4) by calculating

$$\begin{aligned}
&P\Big[0 < \tau^k \leq T, V(0,\xi) < \log\Big(\frac{1}{\delta}\Big)\Big] \\
&\quad = P\Big[0 < \tau^k \leq T, \exp\big(-V(0,\xi)\big) > \delta\Big] \\
&\quad = P\Big[0 < \tau^k \leq T,\ \gamma(\tau^k)\,V\big(\tau^k, X(\tau^k)\big) > \delta\,V\big(\tau^k, X(\tau^k)\big)\exp\Big(-\int_0^{\tau^k} M(s)\,\mathrm{d}s\Big)\Big] \\
&\quad \leq P\Big[0 < \tau^k \leq T,\ \sup_{r\in[0,\tau^k]} \gamma(r)\,V\big(r,X(r)\big) \\
&\qquad\qquad \geq \delta \inf_{\substack{x\in\partial D_k \\ r\in[0,T]}} V\big(r,X(r)\big)\exp\Big(-\int_0^T M(s)\,\mathrm{d}s\Big)\Big] \\
&\quad \overset{(2.5)}{\leq} \frac{1}{\delta}\frac{1}{\inf_{\substack{x\in\partial D_k \\ r\in[0,T]}} V\big(r,X(r)\big)} \exp\Big(\int_0^T M(s)\,\mathrm{d}s\Big) \\
&\quad = \frac{1}{\delta\,V_k(T)}\exp\Big(\int_0^T M(t)\,\mathrm{d}t\Big)
\end{aligned}$$

such that we obtain the required term for (2.3). $\square$

Now it is left to show that inequality (2.3) implies the assertion of the lemma. Since $\tau^k \uparrow \tau$ for $k \to \infty$, we have for every $T > 0$

$$\begin{aligned}
P\big[\tau \leq T\big] &= P\big[\sup_{k\in\mathbb{N}} \tau^k \leq T\big] = P\Big[\bigcap_{k\in\mathbb{N}}\{\tau^k \leq T\}\Big] = \lim_{N\to\infty} P\Big[\bigcap_{k=1}^{N}\{\tau^k \leq T\}\Big] \\
&\leq \limsup_{k\to\infty} P\big[\tau^k \leq T\big]
\end{aligned}$$

by the continuity from above of P.

Hence, we obtain

$$
\begin{aligned}
P\big[\tau \le T\big] &\le \lim_{\delta\downarrow 0} \limsup_{k\to\infty} P\big[\tau^k \le T\big] \\
&\overset{(2.3)}{\le} \lim_{\delta\downarrow 0} \limsup_{k\to\infty} \Bigg(P\big[\xi \notin D_k\big] + P\Big[V(0,\xi) \ge \log\Big(\frac{1}{\delta}\Big)\Big] \\
&\qquad\qquad\qquad\qquad\qquad + \frac{1}{\delta V_k(T)} \exp\Big(\int_0^T M(t)\,\mathrm{d}t\Big)\Bigg) \\
&\le \underbrace{P\big[\xi \notin D\big]}_{=0} + \lim_{\delta\downarrow 0} P\Big[V(0,\xi) \ge \log\Big(\frac{1}{\delta}\Big)\Big] \\
&\qquad\qquad + \lim_{\delta\downarrow 0} \limsup_{k\to\infty} \frac{1}{\delta V_k(T)} \exp\Big(\int_0^T M(t)\,\mathrm{d}t\Big),
\end{aligned}
$$

where we have used that $D_k \uparrow D$ for $k \to \infty$ and therefore that

$$
\begin{aligned}
\limsup_{k\to\infty} P\big[\xi \notin D_k\big] &\le P\big[\limsup_{k\to\infty}\{\xi \notin D_k\}\big] = 1 - P\big[\liminf_{k\to\infty}\{\xi \in D_k\}\big] \\
&= 1 - P\Big[\bigcup_{n\in\mathbb{N}}\bigcap_{k\ge n}\{\xi \in D_k\}\Big] = 1 - P\Big[\xi \in \bigcup_{n\in\mathbb{N}}\bigcap_{k\ge n} D_k\Big] \\
&= 1 - P\big[\xi \in \liminf_{k\to\infty} D_k\big] = 1 - P\big[\xi \in D\big] = P\big[\xi \notin D\big]
\end{aligned}
$$

holds by Lemma A.8 from the Appendix. In addition, $\lim_{\delta\downarrow 0} P\big[V(0,\xi) \ge \log(\frac{1}{\delta})\big] = 0$ because for $\delta \downarrow 0$ we have $\log\big(\frac{1}{\delta}\big) \xrightarrow[\delta\downarrow 0]{} \infty$. Hence, $P\big[V(0,\xi) \ge \log(\frac{1}{\delta})\big] \xrightarrow[\delta\downarrow 0]{} 0$ since $V(0,\xi)$ is a finite number. Furthermore, we have $V_k(T) \xrightarrow[k\to\infty]{} \infty$ and, hence, $\frac{1}{\delta V_k(T)} \xrightarrow[k\to\infty]{} 0$. Therefore, $P\big[\tau \le T\big] = 0$ for every $T > 0$ and, hence, we conclude $\tau = \infty$ P-a.s. such that the assertion is proved. □

3. Existence and Uniqueness

In this chapter we will state and prove the first main theorem of this thesis (i.e. Theorem 3.7) about the existence and uniqueness of a solution for the SDE (2.1). But first of all, we have to explain all further necessary preparations for this proof in the following two sections.

3.1. Distribution of stopping times and Skorokhod's representation theorem

At the beginning of this section we will consider the distribution of certain stopping times related to random variables with equal distribution, that are defined on different probability spaces. The reason for that is the usage of Skorokhod's representation theorem (see Theorem 3.3) in the proof of Theorem 3.7, due to which we have to consider a change of the underlying probability space. Furthermore, we will also prove Lemma 3.2 about an inequality concerning important stopping times for the proof of Theorem 3.7.

Lemma 3.1. *Let $(\Omega, \mathcal{F}, P)$, $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ be two probability spaces and let $T \geq 0$, $k \in \mathbb{N}$. Suppose that $Y \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^d\big)$ and $\tilde{Y} \colon \tilde{\Omega} \longrightarrow C\big([0,T];\mathbb{R}^d\big)$ are stochastic processes such that*

$$P_Y = \tilde{P}_{\tilde{Y}}.$$

Let $U \subseteq \mathbb{R}^d$ be an open and bounded set. Define $\tau_Y^{U^c} := \inf\big\{s \in [0,T] \,\big|\, Y(s) \notin U\big\} \wedge k$ and $\tau_{\tilde{Y}}^{U^c} := \inf\big\{s \in [0,T] \,\big|\, \tilde{Y}(s) \notin U\big\} \wedge k$. Then

$$P_{\tau_Y^{U^c}} = \tilde{P}_{\tau_{\tilde{Y}}^{U^c}}.$$

Proof. Observe that $\tau_Y^{U^c}$ and $\tau_{\tilde{Y}}^{U^c}$ are $\mathbb{R}_+$-valued random variables. Therefore, it suffices to show the equality of their distributions on generating sets of $\mathcal{B}(\mathbb{R}_+)$. Hence, we have to prove that

$$P\big[\tau_Y^{U^c} \leq t\big] = \tilde{P}\big[\tau_{\tilde{Y}}^{U^c} \leq t\big]$$

for every $t \in \mathbb{R}_+$.

Step 1: In this first step we construct an increasing sequence $(\bar{U}_n)_{n\in\mathbb{N}}$ of compact sets in order to approximate $\tau_Y^{U^c}$.

Since $U \subseteq \mathbb{R}^d$ is an open and bounded set, we can define a sequence $(\bar{U}_n)_{n\in\mathbb{N}}$ of compact sets by $\bar{U}_n := \big\{x \in \mathbb{R}^d \,\big|\, \operatorname{dist}(x, U^C) \geq \frac{1}{n}\big\}$. Then the properties $\bigcup_{n\in\mathbb{N}} \bar{U}_n = U$ and $\bar{U}_n \subseteq \bar{U}_{n+1}$ are fulfilled. Define the stopping times $\tau_Y^{\bar{U}_n^c} := \inf\big\{s \in [0,T] \,\big|\, Y(s) \notin \bar{U}_n\big\} \wedge k$ for $n \in \mathbb{N}$.

Claim (1)**.** We have P-a.s.

$$\sup_{n\in\mathbb{N}} \tau_Y^{\bar{U}_n^c} = \tau_Y^{U^c}.$$

Proof of Claim (1). First of all, we note that it suffices to prove

$$\sup_{n\in\mathbb{N}} \underbrace{\inf\left\{s\in[0,T]\,\middle|\,Y(s)\notin \bar{U}_n\right\}}_{=:\tau_n} = \underbrace{\inf\left\{s\in[0,T]\,\middle|\,Y(s)\notin U\right\}}_{=:\tau}$$

since $\sup\limits_{n\in\mathbb{N}} \tau_Y^{\bar{U}_n^c} = \Big(\sup\limits_{n\in\mathbb{N}} \tau_n\Big) \wedge k$.

"$\leq$": By construction $\bar{U}_n \subseteq U$ and, hence, $\sup\limits_{n\in\mathbb{N}} \tau_n \leq \tau$.

"$\geq$": We have to consider the inequality $\sup\limits_{n\in\mathbb{N}} \tau_n \geq \tau$ with respect to the following complementary events.

1) On $\Big\{\sup\limits_{n\in\mathbb{N}} \tau_n > T\Big\}$ we are in the trivial case.

2) Consider $\Big\{\sup\limits_{n\in\mathbb{N}} \tau_n \leq T\Big\}$. Then we have $Y\big(\tau_n\big) \in \bigcap\limits_{j=1}^{m} \bar{U}_j^C$ for every $m < n$. Hence, by letting $n\to\infty$ we obtain

$$Y\Big(\sup_{n\in\mathbb{N}} \tau_n\Big) \in \bigcap_{j=1}^{m} \bar{U}_j^C$$

for every $m \in \mathbb{N}$ since Y is continuous. Therefore, $Y\Big(\sup\limits_{n\in\mathbb{N}} \tau_n\Big) \in \bigcap\limits_{j\in\mathbb{N}} \bar{U}_j^C = U^C$ and, hence, the inequality $\sup\limits_{n\in\mathbb{N}} \tau_n \geq \tau$ follows.

$\square$

Note that the analogous property for $\tilde{Y}$ and $\tilde{P}$ in **Claim (1)** holds as well.

Step 2: In this step we will prove that $P_{\tau_Y^{\bar{U}_n^c}} = \tilde{P}_{\tau_{\tilde{Y}}^{\bar{U}_n^c}}$, where it suffices to show that $P\Big[\tau_Y^{\bar{U}_n^c} < t\Big] = \tilde{P}\Big[\tau_{\tilde{Y}}^{\bar{U}_n^c} < t\Big]$ is fulfilled for every $t \in \mathbb{R}_+$.

To do this we observe that

$$\big\{\tau_Y^{\bar{U}_n^c} < t\big\} = \bigcup_{q\in[0,t[\cap\mathbb{Q}} \big\{Y(q) \in \bar{U}_n^C\big\} \cup \{k < t\}$$

holds by using the continuity of Y. Then we write $\{q_1, q_2, q_3, \dots\}$ for the countable set $[0,t[\cap\mathbb{Q}$ and obtain

$$\bigcup_{q\in[0,t[\cap\mathbb{Q}} \big\{Y(q) \in \bar{U}_n^C\big\} = \bigcup_{N\in\mathbb{N}} \bigcup_{i=1}^{N} \big\{Y(q_i) \in \bar{U}_n^C\big\}.$$

Hence,

$$\begin{aligned} P\big[\tau_Y^{\bar{U}_n^c} < t\big] &= P\Big[\bigcup_{N\in\mathbb{N}} \bigcup_{i=1}^{N} \big\{Y(q_i) \in \bar{U}_n^C\big\} \cup \{k < t\}\Big] \\ &= \lim_{N\to\infty} P\Big[\bigcup_{i=1}^{N} \big\{Y(q_i) \in \bar{U}_n^C\big\} \cup \{k < t\}\Big] \end{aligned}$$

follows by the continuity from below of P. Now we apply the so-called *inclusion-exclusion principle* (cf. [Bil95], Equation (2.9) on page 24), which states that for arbitrary sets A_i we have

$$P\left[\bigcup_{i=1}^{N} A_i\right] = \sum_{j=1}^{N}(-1)^{j-1} \sum_{\substack{I\subseteq\{1,\dots,N\}\\ |I|=j}} P\left[\bigcap_{i\in I} A_i\right].$$

Hence, we can write

$$\begin{aligned}
&P\left[\bigcup_{i=1}^{N}\{Y(q_i)\in \bar{U}_n^C\}\right]\\
&\qquad = \sum_{j=1}^{N}(-1)^{j-1} \sum_{1\le i_1<i_2<\dots<i_j\le N} P\left[\left\{\left(Y(q_{i_1}),\dots,Y(q_{i_j})\right)\in \left(\bar{U}_n^C\right)^j\right\}\right].
\end{aligned}$$

By using the fact that the equality of the distributions P_Y and $\tilde{P}_{\tilde{Y}}$ implies the equality of their finite-dimensional distributions, we obtain that

$$P\left[\bigcup_{i=1}^{N}\left\{Y(q_i)\in \bar{U}_n^C\right\}\right] = \tilde{P}\left[\bigcup_{i=1}^{N}\left\{\tilde{Y}(q_i)\in \bar{U}_n^C\right\}\right]$$

holds. Besides, $\{k<t\}$ equals either $\emptyset$ or the whole sample space such that in fact

$$P\left[\bigcup_{i=1}^{N}\left\{Y(q_i)\in \bar{U}_n^C\right\}\cup\{k<t\}\right] = \tilde{P}\left[\bigcup_{i=1}^{N}\left\{\tilde{Y}(q_i)\in \bar{U}_n^C\right\}\cup\{k<t\}\right]$$

is fulfilled. Therefore, by also applying the calculation and arguments from above to the term $\tilde{P}\left[\bigcup_{i=1}^{N}\left\{\tilde{Y}(q_i)\in \bar{U}_n^C\right\}\cup\{k<t\}\right]$, we get that

$$P\left[\tau_Y^{\bar{U}_n^c}<t\right] = \tilde{P}\left[\tau_{\tilde{Y}}^{\bar{U}_n^c}<t\right]$$

holds.

Step 3: Finally, we will prove the assertion, i.e. that $P\left[\tau_Y^{U^c}\le t\right] = \tilde{P}\left[\tau_{\tilde{Y}}^{U^c}\le t\right]$ is fulfilled.

Since we have noticed in **Step 1** that $\tau_Y^{U^c} = \sup_{n\in\mathbb{N}}\tau_Y^{\bar{U}_n^c}$ and besides for any $n\in\mathbb{N}$ the inclusions $\left\{\tau_Y^{\bar{U}_{n+1}^c}\le t\right\}\subseteq\left\{\tau_Y^{\bar{U}_n^c}\le t\right\}$ hold, the assertion

$$\begin{aligned}
P\left[\tau_Y^{U^c}\le t\right] &= P\left[\bigcap_{n\in\mathbb{N}}\left\{\tau_Y^{\bar{U}_n^c}\le t\right\}\right] = \lim_{n\to\infty} P\left[\tau_Y^{\bar{U}_n^c}\le t\right] = \lim_{n\to\infty}\tilde{P}\left[\tau_{\tilde{Y}}^{\bar{U}_n^c}\le t\right]\\
&= \tilde{P}\left[\bigcap_{n\in\mathbb{N}}\left\{\tau_{\tilde{Y}}^{\bar{U}_n^c}\le t\right\}\right] = \tilde{P}\left[\tau_{\tilde{Y}}^{U^c}\le t\right]
\end{aligned}$$

follows by the continuity from above of P and $\tilde{P}$, where we have used the equality of the distributions from **Step 2**. □

Lemma 3.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $T \geq 0$, $k \in \mathbb{N}$. Suppose that $Y, Y_n \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^d\big)$, for $n \in \mathbb{N}$, are stochastic processes such that*

$$\big\|Y_n - Y\big\|_\infty = \sup_{t\in[0,T]} \big\|Y_n(t) - Y(t)\big\|_{\mathbb{R}^d} \xrightarrow[n\to\infty]{} 0 \tag{3.1}$$

P-a.s. holds. Define the stopping times

$$\tau_{Y_n}^k := \underbrace{\inf\big\{t \in [0,T] \,\big|\, Y_n(t) \notin D_k\big\}}_{=:\tau_{Y_n}} \wedge k \text{ and } \tau_Y^k := \underbrace{\inf\big\{t \in [0,T] \,\big|\, Y(t) \notin D_k\big\}}_{=:\tau_Y} \wedge k.$$

Then we have P-a.s.

$$\liminf_{n\to\infty} \tau_{Y_n}^k \geq \tau_Y^k.$$

Proof. First of all note that $\liminf\limits_{n\to\infty} \tau_{Y_n}^k = \big(\liminf\limits_{n\to\infty} \tau_{Y_n}\big) \wedge k$. Hence, it suffices to prove that P-a.s.

$$\liminf_{n\to\infty} \tau_{Y_n} \geq \tau_Y$$

holds. But we still have to distinguish different complementary events for this proof because we only consider the processes on $[0,T]$ and have set $\inf\emptyset := \infty$.

1) On $\Big\{\liminf\limits_{n\to\infty} \tau_{Y_n} > T\Big\}$ we are in the trivial case.

2) Now consider $\Big\{\liminf\limits_{n\to\infty} \tau_{Y_n} \leq T, \tau_Y \leq T\Big\}$. Assume that $\liminf\limits_{n\to\infty} \tau_{Y_n} < \tau_Y$, i.e. for some $\varepsilon > 0$ we have $\sup\limits_{\ell\in\mathbb{N}} \inf\limits_{n\geq\ell} \tau_{Y_n} \leq \tau_Y - \varepsilon$ and, hence,

$$\inf_{n\geq\ell} \tau_{Y_n} \leq \tau_Y - \varepsilon$$

for every $\ell \in \mathbb{N}$. Due to this boundedness we can find a subsequence $\big(\tau_{Y_{n_s}}\big)_{s\in\mathbb{N}}$ such that

$$\tau_{Y_{n_s}} \xrightarrow[s\to\infty]{} \tau_0 \tag{3.2}$$

for some $\tau_0 \leq \tau_Y - \varepsilon$. Since

$$\begin{aligned}\|Y_{n_s}(\tau_{Y_{n_s}}) - Y(\tau_0)\|_{\mathbb{R}^d} &\leq \|Y_{n_s}(\tau_{Y_{n_s}}) - Y(\tau_{Y_{n_s}})\|_{\mathbb{R}^d} + \|Y(\tau_{Y_{n_s}}) - Y(\tau_0)\|_{\mathbb{R}^d} \\ &\leq \underbrace{\sup_{t\in[0,T]} \|Y_{n_s}(t) - Y(t)\|_{\mathbb{R}^d}}_{\xrightarrow[s\to\infty]{}0 \quad \text{by (3.1)}} + \underbrace{\|Y(\tau_{Y_{n_s}}) - Y(\tau_0)\|_{\mathbb{R}^d}}_{\xrightarrow[s\to\infty]{}0 \quad \text{by (3.2)}},\end{aligned}$$

where we have used the continuity of Y for the last summand, we obtain

$$Y_{n_s}(\tau_{Y_{n_s}}) \xrightarrow[s\to\infty]{} Y(\tau_0).$$

Observe, that $Y_{n_s}(\tau_{Y_{n_s}}) \in D_k^C$ for every $s \in \mathbb{N}$ by the definition of $\tau_{Y_{n_s}}$. Hence, we have $Y(\tau_0) \in D_k^C$ since D_k^C is a closed set. Therefore, we obtain $\tau_0 \geq \tau_Y$, which contradicts the inequality $\tau_0 \leq \tau_Y - \varepsilon$.

3) Finally, we observe that by the same arguments as above we can prove that $\left\{ \liminf_{n\to\infty} \tau_{Y_n} \leq T, \tau_Y > T \right\}$ is a P-zero set.

□

Now we have a closer look at Skorokhod's representation theorem, which is of great importance for the proof of Theorem 3.7 such that it is the second main idea apart from Lemma 1.14. This version is a little less general than the one stated in [Bil99], but it will still be sufficient in our framework.

Theorem 3.3 (Skorokhod's representation theorem, cf. [Bil99], Theorem 6.7 on page 70)**.** *Let $(\mu_n)_{n\in\mathbb{N}}$ and μ be probability measures on a separable metric space (S, ρ) and suppose that $\mu_n \xrightarrow[n\to\infty]{w} \mu$. Then there exist S-valued random variables Z_n, $n \in \mathbb{N}$, and Z on a common probability space $(\Omega, \mathcal{F}, P)$ such that*

- $P_{Z_n} = \mu_n$ *for all* $n \in \mathbb{N}$,
- $P_Z = \mu$,
- $Z_n \xrightarrow[n\to\infty]{P-a.s} Z$.

Proof. We refer to [Bil99], Theorem 6.7 on page 70. □

3.2. Convergence in probability of (stochastic) integrals

In this section we state and prove Lemma 3.6 about the convergence in probability of certain sequences of (stochastic) integrals. In particular, we consider the convergence of stochastic integrals whose integrand and integrator are both a sequence. But first of all, we need the following definition and lemma about Dirac sequences that are stated in the book [Alt12] of W. Alt.

Recall that, for any $A \subseteq \mathbb{R}$, the space of all infinitely differentiable functions $f\colon A \longrightarrow \mathbb{R}$ which have a compact support $\operatorname{supp}(f) := \overline{\{x \in A \,|\, f(x) \neq 0\}}$ is denoted by $C_0^\infty(A;\mathbb{R})$.

Definition 3.4 ((General / Standard) Dirac sequence, cf. [Alt12], Definition 2.14 on page 114)**.** *Let $n \in \mathbb{N}$.*

i) A sequence $(\delta_\ell)_{\ell\in\mathbb{N}}$ in $L^1(\mathbb{R}^n;\mathbb{R})$ is called (general) Dirac sequence if

$$\delta_\ell \geq 0, \int_{\mathbb{R}^n} \delta_\ell(x)\,\mathrm{d}x = 1 \text{ and } \int_{\mathbb{R}^n\setminus B_r(0)} \delta_\ell(x)\,\mathrm{d}x \xrightarrow[\ell\to\infty]{} 0 \text{ for every } r > 0.$$

The last assumption holds for example if $\operatorname{supp}(\delta_\ell) \subseteq B_{r_\ell}(0)$ for a zero sequence $(r_\ell)_{\ell\in\mathbb{N}}$, i.e. $r_\ell \xrightarrow[\ell\to\infty]{} 0$.

ii) Let $\delta \in L^1(\mathbb{R}^n;\mathbb{R})$ be a function such that $\delta \geq 0$ and $\int_{\mathbb{R}^n} \delta(x)\,\mathrm{d}x = 1$. For $\varepsilon > 0$ define the function

$$\delta_\varepsilon(x) := \varepsilon^{-n}\delta\Big(\frac{x}{\varepsilon}\Big). \tag{3.3}$$

Then $\int_{\mathbb{R}^n} \delta_\varepsilon(x)\,\mathrm{d}x = 1$ and $\int_{\mathbb{R}^n\setminus B_r(0)} \delta_\varepsilon(x)\,\mathrm{d}x \xrightarrow[\varepsilon\to 0]{} 0$ for every $r > 0$ hold.

Hence, for every zero sequence $(\varepsilon_k)_{k\in\mathbb{N}}$, the sequence $(\delta_{\varepsilon_k})_{k\in\mathbb{N}}$ is a general Dirac sequence in the sense of i). The family of functions $(\delta_\varepsilon)_{\varepsilon\in]0,\infty[}$ is therefore called Dirac sequence of δ.

iii) Let $\delta \in C_0^\infty(B_1(0);\mathbb{R})$ be a function (extended on $\mathbb{R}^n \setminus B_1(0)$ by 0) such that

$$\delta \geq 0, \int_{\mathbb{R}^n} \delta(x)\,\mathrm{d}x = 1 \text{ and } \operatorname{supp}(\delta_\varepsilon) \subseteq B_\varepsilon(0) \text{ for every } \varepsilon > 0,$$

where δ_ε is given by (3.3). Then $(\delta_\varepsilon)_{\varepsilon\in]0,\infty[}$ is called standard Dirac sequence.

Lemma 3.5 (cf. [Alt12], Theorem 2.15 on page 115)**.** *Let Y be a Banach space, $n \in \mathbb{N}$, $1 \leq p < \infty$, $f \in L^p(\mathbb{R}^n;Y)$ and let $(\delta_\ell)_{\ell\in\mathbb{N}}$ be a Dirac sequence. Then*

i) $\big\|f(\,\cdot + h) - f\big\|_{L^p(\mathbb{R}^n;Y)} \longrightarrow 0$ for $\|h\|_{\mathbb{R}^n} \longrightarrow 0$, $h \in \mathbb{R}^n$,

ii) $\big\|\delta_\ell * f - f\big\|_{L^p(\mathbb{R}^n;Y)} \xrightarrow[\ell\to\infty]{} 0$.

Note that $\delta_\ell * f$ denotes the convolution of the functions δ_ℓ and f, i.e. $(\delta_\ell * f)(x) := \int_{\mathbb{R}^n} \delta_\ell(x-y) f(y) \, \mathrm{d}y$.

Proof. We refer to [Alt12], Theorem 2.15 on page 115. $\square$

Now we can consider the already mentioned Lemma 3.6, that we state more closely related to our application than it is done in [GK96] (see Lemma 3.1 on page 151).

Lemma 3.6. *Let $T \in [0,\infty[$ and let $(\Omega, \mathcal{F}, P)$ be a probability space. Let Y_j, for $j \in \mathbb{N}$, and Y be stochastic processes on $(\Omega, \mathcal{F}, P)$ with values in $C\big([0,T];\mathbb{R}^d\big)$ such that P-a.s.*

$$\big\|Y_j - Y\big\|_\infty = \sup_{t\in[0,T]} \big\|Y_j(t) - Y(t)\big\|_{\mathbb{R}^d} \xrightarrow[j\to\infty]{} 0 \tag{3.4}$$

holds. Assume furthermore that W_j, for $j \in \mathbb{N}$, and W are d_1-dimensional Wiener processes on $(\Omega, \mathcal{F}, P)$ with respect to normal filtrations $\mathcal{F}^{W_j}$ and $\mathcal{F}^W$ taking values in $C\big([0,T];\mathbb{R}^{d_1}\big)$ such that P-a.s.

$$\big\|W_j - W\big\|_\infty = \sup_{t\in[0,T]} \big\|W_j(t) - W(t)\big\|_{\mathbb{R}^{d_1}} \xrightarrow[j\to\infty]{} 0 \tag{3.5}$$

is fulfilled. Besides, the processes $\big(Y(t)\big)_{t\in[0,T]}$ and $\big(Y_j(t)\big)_{t\in[0,T]}$ are assumed to be adapted to $\big(\mathcal{F}_t^W\big)_{t\in[0,T]}$ and $\big(\mathcal{F}_t^{W_j}\big)_{t\in[0,T]}$, respectively. For a function M_{loc} being locally in $L^{1+\chi}\big(\mathbb{R}_+;]0,\infty[\big)$ we then have the following assertions.

i) Let $f\colon \mathbb{R}_+ \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be Borel-measurable in $s \in \mathbb{R}_+$ and continuous in $x \in \mathbb{R}^d$. Suppose furthermore that

$$\sup_{x\in\mathbb{R}^d} \|f(s,x)\|_{\mathbb{R}^d} \le M_{loc}(s)$$

for every $s \in [0,T]$. Then we have

$$\int_0^t f\big(s, Y_j(\kappa_j(s))\big) \, \mathrm{d}s \xrightarrow[j\to\infty]{p} \int_0^t f\big(s, Y(s)\big) \, \mathrm{d}s$$

uniformly in $t \in [0,T]$.

ii) Let $f\colon \mathbb{R}_+ \times \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times d_1}$ be Borel-measurable in $s \in \mathbb{R}_+$ and continuous in $x \in \mathbb{R}^d$. Suppose furthermore that

$$\sup_{x\in\mathbb{R}^d} \|f(s,x)\|_{L_2}^2 \le M_{loc}(s)$$

for every $s \in [0,T]$. Then we have

$$\int_0^t f\big(s, Y_j(\kappa_j(s))\big) \, \mathrm{d}W_j(s) \xrightarrow[j\to\infty]{p} \int_0^t f\big(s, Y(s)\big) \, \mathrm{d}W(s)$$

uniformly in $t \in [0,T]$.

Proof of Lemma 3.6.
At first we observe that $\kappa_j(s) \xrightarrow[j\to\infty]{} s$ since the mesh of the partitions $d_j(T)$ tends to zero for $j \to \infty$. Therefore, we have P-a.s.

$$\begin{aligned}\big\|Y_j(\kappa_j(s)) - Y(s)\big\|_{\mathbb{R}^d} &\le \big\|Y_j(\kappa_j(s)) - Y(\kappa_j(s))\big\|_{\mathbb{R}^d} + \big\|Y(\kappa_j(s)) - Y(s)\big\|_{\mathbb{R}^d} \\ &\le \underbrace{\sup_{r\in[0,T]} \big\|Y_j(r) - Y(r)\big\|_{\mathbb{R}^d}}_{\xrightarrow[j\to\infty]{} 0 \quad \text{by (3.4)}} + \underbrace{\big\|Y(\kappa_j(s)) - Y(s)\big\|_{\mathbb{R}^d}}_{\xrightarrow[j\to\infty]{} 0}\end{aligned} \tag{3.6}$$

for every $s \in [0,T]$, where we have used the continuity of Y for the last summand.

"$i)$": Let $\varepsilon > 0$. Then we have

$$\begin{aligned}&\limsup_{j\to\infty} P\left[\sup_{t\in[0,T]} \left\| \int_0^t f\big(s, Y_j(\kappa_j(s))\big)\,\mathrm{d}s - \int_0^t f\big(s, Y(s)\big)\,\mathrm{d}s \right\|_{\mathbb{R}^d} \ge \varepsilon\right] \\ &\le \limsup_{j\to\infty} P\left[\int_0^T \Big\| f\big(s, Y_j(\kappa_j(s))\big) - f\big(s, Y(s)\big)\Big\|_{\mathbb{R}^d}\,\mathrm{d}s \ge \varepsilon\right] \\ &\le \frac{1}{\varepsilon} \limsup_{j\to\infty} \mathbb{E}\left[\int_0^T \Big\| f\big(s, Y_j(\kappa_j(s))\big) - f\big(s, Y(s)\big)\Big\|_{\mathbb{R}^d}\,\mathrm{d}s\right]\end{aligned}$$

by the Markov inequality. We can now apply the reverse Fatou lemma (cf. Lemma A.4 in the Appendix) by using that $\sup_{x\in\mathbb{R}^d} \|f(s,x)\|_{\mathbb{R}^d} \le M_{loc}(s)$ holds for every $s \in [0,T]$. Hence, by using the continuity of f in $x \in \mathbb{R}^d$, we conclude

$$\limsup_{j\to\infty} \Big\| f\big(s, Y_j(\kappa_j(s))\big) - f\big(s, Y(s)\big)\Big\|_{\mathbb{R}^d} = 0$$

from (3.6). Therefore, the assertion

$$\lim_{j\to\infty} P\left[\sup_{t\in[0,T]} \left\| \int_0^t f\big(s, Y_j(\kappa_j(s))\big)\,\mathrm{d}s - \int_0^t f\big(s, Y(s)\big)\,\mathrm{d}s \right\|_{\mathbb{R}^d} \ge \varepsilon\right] = 0$$

follows.
"$ii)$": It suffices to prove that for every $\varepsilon > 0$

$$\limsup_{j\to\infty} P\left[\sup_{t\in[0,T]} \left\| \int_0^t f\big(s, Y_j(\kappa_j(s))\big)\,\mathrm{d}W_j(s) - \int_0^t f\big(s, Y(s)\big)\,\mathrm{d}W(s) \right\|_{\mathbb{R}^d} \ge \varepsilon\right] = 0$$

holds.

First of all, we have

$$
\begin{aligned}
&P\Bigg[\sup_{t\in[0,T]}\Big\|\int_0^t f\big(s,Y_j(\kappa_j(s))\big)\,\mathrm{d}W_j(s)-\int_0^t f\big(s,Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d}\geq\varepsilon\Bigg]\\
&\qquad\leq\underbrace{P\Bigg[\sup_{t\in[0,T]}\Big\|\int_0^t f\big(s,Y_j(\kappa_j(s))\big)-f\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s)\Big\|_{\mathbb{R}^d}\geq\frac{\varepsilon}{2}\Bigg]}_{=:I_j^1}\\
&\qquad\quad+\underbrace{P\Bigg[\sup_{t\in[0,T]}\Big\|\int_0^t f\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s)-\int_0^t f\big(s,Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d}\geq\frac{\varepsilon}{2}\Bigg]}_{=:I_j^2}.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
I_j^1&=P\Bigg[\sup_{t\in[0,T]}\Big\|\int_0^t f\big(s,Y_j(\kappa_j(s))\big)-f\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s)\Big\|_{\mathbb{R}^d}^2\geq\frac{\varepsilon^2}{4}\Bigg]\\
&\leq\frac{4}{\varepsilon^2}\,\mathbb{E}\Bigg[\sup_{t\in[0,T]}\Big\|\int_0^t f\big(s,Y_j(\kappa_j(s))\big)-f\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s)\Big\|_{\mathbb{R}^d}^2\Bigg]\\
&\leq\frac{8}{\varepsilon^2}\,\mathbb{E}\Bigg[\int_0^T\Big\|f\big(s,Y_j(\kappa_j(s))\big)-f\big(s,Y_j(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\Bigg]
\end{aligned}
$$

holds by using the Markov inequality in the second and the Burkholder-Davis-Gundy type inequality (cf. Lemma A.5 in the Appendix) in the last step. In order to prove $\limsup_{j\to\infty} I_j^1=0$, we can now apply the reverse Fatou lemma (cf. Lemma A.4 in the Appendix) using the assumption $\sup_{x\in\mathbb{R}^d}\|f(s,x)\|_{L_2}^2\leq M_{loc}(s)$ for every $s\in[0,T]$. Therefore, it suffices to conclude that

$$
\limsup_{j\to\infty}\Big\|f\big(s,Y_j(\kappa_j(s))\big)-f\big(s,Y_j(s)\big)\Big\|_{L_2}^2=0
$$

holds. By using the continuity of f in the second component and considering

$$
\big\|Y_j(\kappa_j(s))-Y_j(s)\big\|_{\mathbb{R}^d}\leq\underbrace{\big\|Y_j(\kappa_j(s))-Y(s)\big\|_{\mathbb{R}^d}}_{\xrightarrow[j\to\infty]{}0\text{ by (3.6)}}+\underbrace{\big\|Y_j(s)-Y(s)\big\|_{\mathbb{R}^d}}_{\xrightarrow[j\to\infty]{}0\text{ by (3.4)}},
$$

we obtain the convergence to 0.

Now we have to estimate the summand I_j^2. Therefore, let $(\delta_\ell)_{\ell\in\mathbb{N}}$ be a standard Dirac sequence in $C_0^\infty\big(\mathbb{R};\mathbb{R}\big)$ such that $\operatorname{supp}(\delta_\ell)\subseteq B_{\frac{1}{\ell}}(0)$ and define the function f_0 by

$$
f_0\big(s,Z(s)\big):=\begin{cases} f\big(s,Z(s)\big) & \text{for } s\in[0,T],\\ 0 & \text{else,}\end{cases}
$$

for $s\in\mathbb{R}$, where Z represents the stochastic processes Y_j and Y. Then define

$$
f_\ell\big(s,Z(s)\big):=\int_{\mathbb{R}}\delta_\ell(r-s)\,f_0\big(r,Z(r)\big)\,\mathrm{d}r.
$$

Note that the map $s \longmapsto f_\ell\big(s, Z(s)\big)$, $s \in [0,T]$, is now continuous since we consider this convolution (see e.g. [Alt12], Definition 2.13 on page 112). By the definition of f_0 we conclude that in fact

$$f_\ell\big(s, Z(s)\big) = \int_0^T \delta_\ell(r-s)\, f\big(r, Z(r)\big)\,\mathrm{d}r$$

holds. Then we have

$$\begin{aligned}
I_j^2 &\leq P\Bigg[\sup_{t\in[0,T]}\bigg\|\int_0^t f\big(s,Y_j(s)\big) - f_\ell\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s)\bigg\|_{\mathbb{R}^d} \geq \frac{\varepsilon}{6}\Bigg] \\
&\quad + \underbrace{P\Bigg[\sup_{t\in[0,T]}\bigg\|\int_0^t f_\ell\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s) - \int_0^t f_\ell\big(s,Y(s)\big)\,\mathrm{d}W(s)\bigg\|_{\mathbb{R}^d} \geq \frac{\varepsilon}{6}\Bigg]}_{=:\, J_{j,\ell}^2} \\
&\quad + P\Bigg[\sup_{t\in[0,T]}\bigg\|\int_0^t f_\ell\big(s,Y(s)\big) - f\big(s,Y(s)\big)\,\mathrm{d}W(s)\bigg\|_{\mathbb{R}^d} \geq \frac{\varepsilon}{6}\Bigg] \\
&\leq \frac{36}{\varepsilon^2}\,\mathbb{E}\Bigg[\sup_{t\in[0,T]}\bigg\|\int_0^t f\big(s,Y_j(s)\big) - f_\ell\big(s,Y_j(s)\big)\,\mathrm{d}W_j(s)\bigg\|_{\mathbb{R}^d}^2\Bigg] \\
&\quad + J_{j,\ell}^2 + \frac{36}{\varepsilon^2}\,\mathbb{E}\Bigg[\sup_{t\in[0,T]}\bigg\|\int_0^t f_\ell\big(s,Y(s)\big) - f\big(s,Y(s)\big)\,\mathrm{d}W(s)\bigg\|_{\mathbb{R}^d}^2\Bigg]
\end{aligned}$$

by using the Markov inequality. An application of the Burkholder-Davis-Gundy type inequality (cf. Lemma A.5 in the Appendix) yields

$$\begin{aligned}
I_j^2 &\leq \frac{72}{\varepsilon^2}\,\underbrace{\mathbb{E}\bigg[\int_0^T \Big\|f\big(s,Y_j(s)\big) - f_\ell\big(s,Y_j(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg]}_{=:\, J_{j,\ell}^1} \\
&\quad + J_{j,\ell}^2 + \frac{72}{\varepsilon^2}\,\underbrace{\mathbb{E}\bigg[\int_0^T \Big\|f_\ell\big(s,Y(s)\big) - f\big(s,Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg]}_{=:\, J_\ell^3}.
\end{aligned}$$

Observe that we can compute for the summand J_ℓ^3

$$\begin{aligned}
J_\ell^3 &= \mathbb{E}\bigg[\int_0^T \Big\|\int_{\mathbb{R}} \delta_\ell(r-s)\, f_0\big(r,Y(r)\big)\,\mathrm{d}r - f\big(s,Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg] \\
&\leq \mathbb{E}\bigg[\int_0^T \bigg(\int_{\mathbb{R}} \delta_\ell(r-s)\,\Big\|f_0\big(r,Y(r)\big) - f\big(s,Y(s)\big)\Big\|_{L_2}\,\mathrm{d}r\bigg)^2\,\mathrm{d}s\bigg]
\end{aligned}$$

by using $\int_{\mathbb{R}} \delta_\ell(x)\,\mathrm{d}x = 1$. Now we can split up the inner integral and obtain

$$\begin{aligned}
&\mathbb{E}\bigg[\int_0^T \bigg(\int_{\mathbb{R}} \delta_\ell(r-s)\,\Big\|f_0\big(r,Y(r)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}\,\mathrm{d}r\bigg)^2\,\mathrm{d}s\bigg] \\
&\qquad = \mathbb{E}\bigg[\int_0^T \bigg(\int_0^T \delta_\ell(r-s)\,\Big\|f\big(r,Y(r)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}\,\mathrm{d}r \\
&\qquad\qquad + \int_{[0,T]^C} \delta_\ell(r-s)\,\Big\|0-f\big(s,Y(s)\big)\Big\|_{L_2}\,\mathrm{d}r\bigg)^2\,\mathrm{d}s\bigg] \\
&\qquad \le 2\,\underbrace{\mathbb{E}\bigg[\int_0^T \bigg(\int_0^T \delta_\ell(r-s)\,\Big\|f\big(r,Y(r)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}\,\mathrm{d}r\bigg)^2\,\mathrm{d}s\bigg]}_{=:\,\tilde{J}_\ell^3} \\
&\qquad\qquad + 2\,\mathbb{E}\bigg[\int_0^T \big\|f\big(s,Y(s)\big)\big\|_{L_2}^2 \bigg(\int_{[0,T]^C} \delta_\ell(r-s)\,\mathrm{d}r\bigg)^2\,\mathrm{d}s\bigg]
\end{aligned}$$

by using Young's inequality in the last step. Note that for $0 < s < T$ we have

$$\int_{[0,T]^C} \delta_\ell(r-s)\,\mathrm{d}r = \int_{\mathbb{R}\setminus[-s,T-s]} \delta_\ell(r)\,\mathrm{d}r \le \int_{\mathbb{R}\setminus B_{\min\{s,T-s\}}(0)} \delta_\ell(r)\,\mathrm{d}r \xrightarrow[\ell\to\infty]{} 0$$

by definition. Hence, by using that $\sup_{x\in\mathbb{R}^d} \|f(s,x)\|_{L_2}^2 \le M_{loc}(s)$ holds for every $s \in [0,T]$ and applying Lebesgue's dominated convergence theorem, we only have to consider the first summand $\tilde{J}_\ell^3$.

Therefore, the transformation $r \longmapsto r+s$ followed by an application of the generalised Minkowski integral inequality for $p = 2$ (cf. Theorem A.6 in the Appendix) yields

$$\begin{aligned}
\tilde{J}_\ell^3 &= \mathbb{E}\bigg[\int_0^T \bigg(\int_{\mathbb{R}} \mathbb{1}_{[-s,T-s]}(r)\,\delta_\ell(r) \\
&\qquad\qquad \cdot \Big\|f\big(r+s,Y(r+s)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}\,\mathrm{d}r\bigg)^2\,\mathrm{d}s\bigg] \\
&\le \mathbb{E}\bigg[\bigg(\int_{\mathbb{R}} \bigg(\int_0^T \mathbb{1}_{[0,T]}(r+s)\,\delta_\ell(r)^2 \\
&\qquad\qquad \cdot \Big\|f\big(r+s,Y(r+s)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg)^{\frac{1}{2}}\,\mathrm{d}r\bigg)^2\bigg] \\
&= \mathbb{E}\bigg[\bigg(\int_{\mathbb{R}} \delta_\ell(r)\bigg(\int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\;\mathbb{1}_{[0,T]}(s) \\
&\qquad\qquad \cdot \Big\|f\big(r+s,Y(r+s)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg)^{\frac{1}{2}}\,\mathrm{d}r\bigg)^2\bigg].
\end{aligned}$$

Now we split up the first integral into $B_{\frac{1}{\varrho}}(0)$ and $\mathbb{R}\setminus B_{\frac{1}{\varrho}}(0)$ for $\varrho \in \mathbb{N}$. Then we have on the one hand by using Young's inequality

$$
\begin{aligned}
&\mathbb{E}\Bigg[\bigg(\int_{\mathbb{R}\setminus B_{\frac{1}{\varrho}}(0)} \delta_\ell(r) \bigg(\int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\, \mathbb{1}_{[0,T]}(s) \\
&\qquad\qquad\qquad \cdot \Big\| f\big(r+s, Y(r+s)\big) - f\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\bigg)^{\frac{1}{2}} \mathrm{d}r\bigg)^2\Bigg] \\
&\leq 2\,\mathbb{E}\Bigg[\bigg(\int_{\mathbb{R}\setminus B_{\frac{1}{\varrho}}(0)} \delta_\ell(r) \bigg(\int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\Big\| f\big(r+s, Y(r+s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s \\
&\qquad\qquad + \int_{\mathbb{R}} \mathbb{1}_{[0,T]}(s)\, \Big\| f\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\bigg)^{\frac{1}{2}} \mathrm{d}r\bigg)^2\Bigg] \\
&\leq 4\,\mathbb{E}\Bigg[\bigg(\int_{\mathbb{R}\setminus B_{\frac{1}{\varrho}}(0)} \delta_\ell(r) \bigg(\int_0^T M_{loc}(s)\,\mathrm{d}s\bigg)^{\frac{1}{2}} \mathrm{d}r\bigg)^2\Bigg] \\
&= 4 \int_0^T M_{loc}(s)\,\mathrm{d}s \bigg(\int_{\mathbb{R}\setminus B_{\frac{1}{\varrho}}(0)} \delta_\ell(r)\,\mathrm{d}r\bigg)^2 \xrightarrow[\ell\to\infty]{} 0
\end{aligned}
$$

since $\int_{\mathbb{R}\setminus B_{\frac{1}{\varrho}}(0)} \delta_\ell(r)\,\mathrm{d}r \xrightarrow[\ell\to\infty]{} 0$ by definition. On the other hand

$$
\begin{aligned}
&\mathbb{E}\Bigg[\bigg(\int_{B_{\frac{1}{\varrho}}(0)} \delta_\ell(r) \bigg(\int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\, \mathbb{1}_{[0,T]}(s) \\
&\qquad\qquad\qquad \cdot \Big\| f\big(r+s, Y(r+s)\big) - f\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\bigg)^{\frac{1}{2}} \mathrm{d}r\bigg)^2\Bigg] \\
&\leq \mathbb{E}\Bigg[\underbrace{\bigg(\int_{B_{\frac{1}{\varrho}}(0)} \delta_\ell(r)\,\mathrm{d}r\bigg)}_{\leq 1}^2 \bigg(\sup_{r\in B_{\frac{1}{\varrho}}(0)} \int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\, \mathbb{1}_{[0,T]}(s) \\
&\qquad\qquad\qquad \cdot \Big\| f\big(r+s, Y(r+s)\big) - f\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\bigg)\Bigg] \\
&\leq \mathbb{E}\Bigg[\sup_{r\in B_{\frac{1}{\varrho}}(0)} \int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\, \mathbb{1}_{[0,T]}(s) \\
&\qquad\qquad\qquad \cdot \Big\| f\big(r+s, Y(r+s)\big) - f\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\Bigg] \xrightarrow[\varrho\to\infty]{} 0
\end{aligned}
$$

holds by Lebesgue's dominated convergence theorem. The claimed convergence is fulfilled because we have

$$
\begin{aligned}
&\mathbb{1}_{[-\frac{1}{\varrho},\frac{1}{\varrho}]}(r) \int_{\mathbb{R}} \mathbb{1}_{[0,T]}(s)\, \mathbb{1}_{[0,T]}(r+s) \\
&\qquad\qquad \cdot \Big\| f\big(r+s, Y(r+s)\big) - f\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s \xrightarrow[\varrho\to\infty]{} 0
\end{aligned}
$$

by Lemma 3.5 i). In addition, an integrable dominating function is given by

$$\sup_{r\in B_{\frac{1}{\varrho}}(0)} \int_{\mathbb{R}} \mathbb{1}_{[0,T]}(r+s)\,\mathbb{1}_{[0,T]}(s)\,\underbrace{\Big\|f\big(r+s,Y(r+s)\big)-f\big(s,Y(s)\big)\Big\|_{L_2}^2}_{\leq 2\,\big(M_{loc}(r+s)+M_{loc}(s)\big)}\,\mathrm{d}s$$
$$\leq 2\int_0^T M_{loc}(s)\,\mathrm{d}s.$$

Analogously, we obtain for the summand $J^1_{j,\ell}$

$$\begin{aligned}
&\lim_{\varrho\to\infty}\lim_{\ell\to\infty}\limsup_{j\to\infty} J^1_{j,\ell}\\
&\quad= \lim_{\varrho\to\infty}\lim_{\ell\to\infty}\limsup_{j\to\infty}\mathbb{E}\Big[\int_0^T \Big\|f\big(s,Y_j(s)\big)-f_\ell\big(s,Y_j(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\Big]\\
&\quad\leq \lim_{\varrho\to\infty}\lim_{\ell\to\infty}\mathbb{E}\Big[\int_0^T \Big\|f\big(s,Y(s)\big)-f_\ell\big(s,Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\Big]\\
&\quad= 0,
\end{aligned}$$

where we have applied the reverse Fatou lemma (cf. Lemma A.4 in the Appendix) in the second step. An integrable dominating function is given by

$$\begin{aligned}
&\Big\|f\big(s,Y_j(s)\big)-f_\ell\big(s,Y_j(s)\big)\Big\|_{L_2}^2\\
&\qquad\leq \Big\|f\big(s,Y_j(s)\big)-\int_0^T \delta_\ell(r-s)\,f\big(r,Y_j(r)\big)\,\mathrm{d}r\Big\|_{L_2}^2\\
&\qquad\leq 2\,\Big\|f\big(s,Y_j(s)\big)\Big\|_{L_2}^2+2\,\Big(\int_0^T \delta_\ell(r-s)\,\Big\|f\big(r,Y_j(r)\big)\Big\|_{L_2}\,\mathrm{d}r\Big)^2\\
&\qquad\leq 2\,M_{loc}(s)+2\,T\,\sup_{\xi\in\mathbb{R}}\delta_\ell(\xi)^2\int_0^T M_{loc}(r)\,\mathrm{d}r,
\end{aligned}$$

where we have used Young's inequality in the second as well as the Cauchy-Schwarz inequality and the continuity of δ_ℓ in the last step. Hence, it remains to note that

$$\limsup_{j\to\infty}\Big\|f\big(s,Y_j(s)\big)-f_\ell\big(s,Y_j(s)\big)\Big\|_{L_2}^2=\Big\|f\big(s,Y(s)\big)-f_\ell\big(s,Y(s)\big)\Big\|_{L_2}^2$$

holds by using the continuity of f and f_ℓ in their spacial component.

For the summand $J^2_{j,\ell}$ we need a slightly different argument. We will follow the main idea of the proof of the theorem on page 32 in [Sko65] by using the continuity of the coefficients and the representation of the stochastic integral in this case.

Let $(\pi^l)_{l\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$ given by $\big\{0 = r_0^l < r_1^l < \cdots < r_{N_l}^l = T\big\}$ such that $\sup_{i\,:\,r_{i+1}^l \le T} |r_{i+1}^l - r_i^l| \xrightarrow[l\to\infty]{} 0$. Then consider

$$
\begin{aligned}
J_{j,\ell}^2 &= P\bigg[\sup_{t\in[0,T]} \Big\| \int_0^t f_\ell\big(s, Y_j(s)\big)\,\mathrm{d}W_j(s) - \int_0^t f_\ell\big(s, Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{6}\bigg] \\
&\le \underbrace{\begin{aligned}&P\bigg[\sup_{t\in[0,T]} \Big\| \int_0^t f_\ell\big(s, Y_j(s)\big)\,\mathrm{d}W_j(s) \\ &\qquad - \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y_j(r_i^l)\big)\big(W_j(r_{i+1}^l\wedge t) - W_j(r_i^l\wedge t)\big)\Big\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{18}\bigg]\end{aligned}}_{=:\,K_{j,\ell,l}^1} \\
&\quad + \underbrace{\begin{aligned}&P\bigg[\sup_{t\in[0,T]} \Big\| \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y_j(r_i^l)\big)\big(W_j(r_{i+1}^l\wedge t) - W_j(r_i^l\wedge t)\big) \\ &\qquad - \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y(r_i^l)\big)\big(W(r_{i+1}^l\wedge t) - W(r_i^l\wedge t)\big)\Big\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{18}\bigg]\end{aligned}}_{=:\,K_{j,\ell,l}^2} \\
&\quad + \underbrace{\begin{aligned}&P\bigg[\sup_{t\in[0,T]} \Big\| \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y(r_i^l)\big)\big(W(r_{i+1}^l\wedge t) - W(r_i^l\wedge t)\big) \\ &\qquad - \int_0^t f_\ell\big(s, Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{18}\bigg]\end{aligned}}_{=:\,K_{\ell,l}^3}.
\end{aligned}
$$

For the summand $K_{\ell,l}^3$ we have

$$
\begin{aligned}
K_{\ell,l}^3 &= P\bigg[\sup_{t\in[0,T]} \Big\| \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y(r_i^l)\big) \int_{r_i^l\wedge t}^{r_{i+1}^l\wedge t} \mathrm{d}W(s) - \int_0^t f_\ell\big(s, Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{18}\bigg] \\
&= P\bigg[\sup_{t\in[0,T]} \Big\| \int_0^t \Big(\sum_{r_i^l\in\pi^l} \mathbb{1}_{[r_i^l, r_{i+1}^l]}(s)\, f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) - f_\ell\big(s, Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{18}\bigg].
\end{aligned}
$$

Applying the Markov inequality yields

$$
\begin{aligned}
&P\bigg[\sup_{t\in[0,T]} \Big\| \int_0^t \Big(\sum_{r_i^l\in\pi^l} \mathbb{1}_{[r_i^l, r_{i+1}^l]}(s)\, f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) - f_\ell\big(s, Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d}^2 \ge \frac{\varepsilon^2}{324}\bigg] \\
&\quad\le \frac{324}{\varepsilon^2}\,\mathbb{E}\bigg[\sup_{t\in[0,T]} \Big\| \int_0^t \Big(\sum_{r_i^l\in\pi^l} \mathbb{1}_{[r_i^l, r_{i+1}^l]}(s)\, f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) - f_\ell\big(s, Y(s)\big)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d}^2\bigg] \\
&\quad\le \frac{648}{\varepsilon^2}\,\mathbb{E}\bigg[\int_0^T \Big\|\Big(\sum_{r_i^l\in\pi^l} \mathbb{1}_{[r_i^l, r_{i+1}^l]}(s)\, f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) - f_\ell\big(s, Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg],
\end{aligned}
$$

where we have used the Burkholder-Davis-Gundy type inequality (cf. Lemma A.5 in the Appendix) in the last step. Now we can split up the integral back again into the sum of the partition points such that we obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^T \Big\| \Big(\sum_{r_i^l \in \pi^l} \mathbb{1}_{[r_i^l, r_{i+1}^l]}(s)\, f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) - f_\ell\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\Big] \\
&\quad = \mathbb{E}\Big[\sum_{r_i^l \in \pi^l} \int_{r_i^l}^{r_{i+1}^l} \Big\| \Big(\sum_{r_i^l \in \pi^l} \mathbb{1}_{[r_i^l, r_{i+1}^l]}(s)\, f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) - f_\ell\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\Big] \qquad (3.7) \\
&\quad = \mathbb{E}\Big[\sum_{r_i^l \in \pi^l} \int_{r_i^l}^{r_{i+1}^l} \Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\Big].
\end{aligned}
$$

By using the mean value theorem for integrals, there exist $\xi_i \in [r_i^l, r_{i+1}^l]$ such that

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{r_i^l \in \pi^l} \int_{r_i^l}^{r_{i+1}^l} \Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(s, Y(s)\big)\Big\|_{L_2}^2 \,\mathrm{d}s\Big] \\
&\qquad = \mathbb{E}\Big[\sum_{r_i^l \in \pi^l} (r_{i+1}^l - r_i^l) \Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(\xi_i, Y(\xi_i)\big)\Big\|_{L_2}^2\Big]
\end{aligned}
$$

holds.

Now let $\tilde{\varepsilon} > 0$. Then by the uniform continuity of the map $s \longmapsto f_\ell\big(s, Y(s)\big)$, $s \in [0,T]$, there exists a $\delta > 0$ such that $\Big\| f_\ell\big(s_1, Y(s_1)\big) - f_\ell\big(s_2, Y(s_2)\big)\Big\|_{L_2}^2 \leq \dfrac{\tilde{\varepsilon}}{T}$ for any $s_1, s_2 \in [0,T]$ with $|s_1 - s_2| < \delta$. Using the fact that we have $\sup\limits_{i\,:\, r_{i+1}^l \leq T} |r_{i+1}^l - r_i^l| \xrightarrow[l\to\infty]{} 0$ by assumption, we can choose l large enough such that $\sup\limits_{i\,:\, r_{i+1}^l \leq T} |r_{i+1}^l - r_i^l| < \delta$. Hence,

$$
\sum_{r_i^l \in \pi^l} (r_{i+1}^l - r_i^l) \underbrace{\Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(\xi_i, Y(\xi_i)\big)\Big\|_{L_2}^2}_{\leq \frac{\tilde{\varepsilon}}{T}} \leq \sum_{r_i^l \in \pi^l} (r_{i+1}^l - r_i^l)\, \frac{\tilde{\varepsilon}}{T} = \tilde{\varepsilon}
$$

for l large enough. Therefore, we conclude that

$$
\mathbb{E}\Big[\sum_{r_i^l \in \pi^l} (r_{i+1}^l - r_i^l) \Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(\xi_i, Y(\xi_i)\big)\Big\|_{L_2}^2\Big] \xrightarrow[l\to\infty]{} 0
$$

is fulfilled because we can apply Lebesgue's dominated convergence theorem since an integrable dominating function is given by

$$
\begin{aligned}
&\sum_{r_i^l\in\pi^l} (r_{i+1}^l - r_i^l)\, \Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(\xi_i, Y(\xi_i)\big)\Big\|_{L_2}^2 \\
&\quad = \sum_{r_i^l\in\pi^l} (r_{i+1}^l - r_i^l)\, \bigg\| \int_0^T \big(\delta_\ell(r - r_i^l) - \delta_\ell(r-\xi_i)\big) f\big(r, Y(r)\big)\,\mathrm{d}r\bigg\|_{L_2}^2 \\
&\quad \le \sum_{r_i^l\in\pi^l} (r_{i+1}^l - r_i^l)\, \int_0^T \big(\delta_\ell(r - r_i^l) - \delta_\ell(r-\xi_i)\big)^2\,\mathrm{d}r \int_0^t \Big\| f\big(r, Y(r)\big)\Big\|_{L_2}^2\,\mathrm{d}r \\
&\quad \le 2\,T^2 \sup_{\xi\in\mathbb{R}} \delta_\ell(\xi)^2 \int_0^T M_{loc}(r)\,\mathrm{d}r,
\end{aligned}
$$

where we have used the Cauchy-Schwarz inequality and the continuity of the function δ_ℓ.

We obtain an analogous statement for the summand $K_{j,\ell,l}^1$, i.e. we repeat the calculation up to (3.7) and obtain

$$
\begin{aligned}
K_{j,\ell,l}^1 = P\bigg[\sup_{t\in[0,T]} \bigg\| &\int_0^t f_\ell\big(s, Y_j(s)\big)\,\mathrm{d}W_j(s) \\
&- \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y_j(r_i^l)\big)\, \big(W_j(r_{i+1}^l \wedge t) - W_j(r_i^l\wedge t)\big)\bigg\|_{\mathbb{R}^d} \ge \frac{\varepsilon}{18}\bigg] \\
\le \mathbb{E}\bigg[\sum_{r_i^l\in\pi^l} &\int_{r_i^l}^{r_{i+1}^l} \Big\| f_\ell\big(r_i^l, Y_j(r_i^l)\big) - f_\ell\big(s, Y_j(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg].
\end{aligned}
$$

Now we can apply the reverse Fatou lemma (cf. Lemma A.4 in the Appendix) by again considering $2\,T^2 \sup_{\xi\in\mathbb{R}} \delta_\ell(\xi)^2\ \int_0^T M_{loc}(r)\,\mathrm{d}r$ as an integrable dominating function. We obtain

$$
\begin{aligned}
&\lim_{\ell\to\infty}\lim_{l\to\infty}\limsup_{j\to\infty} K_{j,\ell,l}^1 \\
&\quad \le \lim_{\ell\to\infty}\lim_{l\to\infty}\limsup_{j\to\infty} \mathbb{E}\bigg[\sum_{r_i^l\in\pi^l} \int_{r_i^l}^{r_{i+1}^l} \Big\| f_\ell\big(r_i^l, Y_j(r_i^l)\big) - f_\ell\big(s, Y_j(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg] \\
&\quad \le \lim_{\ell\to\infty}\lim_{l\to\infty} \mathbb{E}\bigg[\sum_{r_i^l\in\pi^l} \int_{r_i^l}^{r_{i+1}^l} \Big\| f_\ell\big(r_i^l, Y(r_i^l)\big) - f_\ell\big(s, Y(s)\big)\Big\|_{L_2}^2\,\mathrm{d}s\bigg] \\
&\quad = 0,
\end{aligned}
$$

where we have again used the continuity of f_ℓ and (3.4) for the uniform convergence of Y_j.

Finally, we have to consider $K^2_{j,\ell,l}$ and compute

$$\begin{aligned}
K^2_{j,\ell,l} = P\Bigg[\sup_{t\in[0,T]} \Big\| \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y_j(r_i^l)\big)\big(W_j(r_{i+1}^l \wedge t) - W_j(r_i^l \wedge t)\big) \\
- \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y(r_i^l)\big)\big(W(r_{i+1}^l \wedge t) - W(r_i^l \wedge t)\big)\Big\|_{\mathbb{R}^d} \geq \frac{\varepsilon}{18}\Bigg] \\
\leq P\Bigg[\sup_{t\in[0,T]} \Big\| \sum_{r_i^l\in\pi^l} f_\ell\big(r_i^l, Y_j(r_i^l)\big)\Big(\big(W_j(r_{i+1}^l \wedge t) - W_j(r_i^l \wedge t)\big) \\
- \big(W(r_{i+1}^l \wedge t) - W(r_i^l \wedge t)\big)\Big)\Big\|_{\mathbb{R}^d} \geq \frac{\varepsilon}{36}\Bigg] \\
+ P\Bigg[\sup_{t\in[0,T]} \Big\| \sum_{r_i^l\in\pi^l} \Big(f_\ell\big(r_i^l, Y_j(r_i^l)\big) - f_\ell\big(r_i^l, Y(r_i^l)\big)\Big) \\
\cdot \big(W(r_{i+1}^l \wedge t) - W(r_i^l \wedge t)\big)\Big\|_{\mathbb{R}^d} \geq \frac{\varepsilon}{36}\Bigg].
\end{aligned}$$

By using the Markov inequality the latter can be estimated by

$$\begin{aligned}
&\frac{36}{\varepsilon}\,\mathbb{E}\Bigg[\sum_{r_i^l\in\pi^l} \Big\|f_\ell\big(r_i^l, Y_j(r_i^l)\big)\Big\|_{L_2} 2\sup_{\xi\in[0,T]} \big\|W_j(\xi) - W(\xi)\big\|_{\mathbb{R}^{d_1}}\Bigg] \\
&\qquad + \frac{36}{\varepsilon}\,\mathbb{E}\Bigg[\sup_{t\in[0,T]} \sum_{r_i^l\in\pi^l} \Big\|f_\ell\big(r_i^l, Y_j(r_i^l)\big) - f_\ell\big(r_i^l, Y(r_i^l)\big)\Big\|_{L_2} \\
&\qquad\qquad \cdot \Big\|W(r_{i+1}^l \wedge t) - W(r_i^l \wedge t)\Big\|_{\mathbb{R}^{d_1}}\Bigg] \\
&\leq \frac{72}{\varepsilon}\,\mathbb{E}\Bigg[N_l \sup_{\xi\in[0,T]} \Big\|f_\ell\big(\xi, Y_j(\xi)\big)\Big\|_{L_2} \sup_{\xi\in[0,T]} \big\|W_j(\xi) - W(\xi)\big\|_{\mathbb{R}^{d_1}}\Bigg] \\
&\qquad + \frac{72}{\varepsilon}\,\mathbb{E}\Bigg[N_l \sup_{\xi\in[0,T]} \Big\|f_\ell\big(\xi, Y_j(\xi)\big) - f_\ell\big(\xi, Y(\xi)\big)\Big\|_{L_2} \sup_{\xi\in[0,T]} \big\|W(\xi)\big\|_{\mathbb{R}^{d_1}}\Bigg].
\end{aligned}$$

Now we use (3.4) and (3.5), i.e. we have P-a.s.

$$\sup_{t\in[0,T]} \big\|Y_j(t) - Y(t)\big\|_{\mathbb{R}^d} \xrightarrow[j\to\infty]{} 0 \quad\text{and}\quad \sup_{t\in[0,T]} \big\|W_j(t) - W(t)\big\|_{\mathbb{R}^{d_1}} \xrightarrow[j\to\infty]{} 0$$

as well as the uniform continuity of the maps $s \longmapsto f_\ell\big(s, Y_j(s)\big)$ and $s \longmapsto f_\ell\big(s, Y(s)\big)$ for $s \in [0,T]$. Namely, we have

$$\begin{aligned}
&\limsup_{j\to\infty} \sup_{\xi\in[0,T]} \big\|f_\ell\big(\xi, Y_j(\xi)\big)\big\|_{L_2} \sup_{\xi\in[0,T]} \big\|W_j(\xi) - W(\xi)\big\|_{\mathbb{R}^{d_1}} \\
&\qquad \leq \limsup_{j\to\infty} \sup_{\xi\in[0,T]} \big\|f_\ell\big(\xi, Y_j(\xi)\big)\big\|_{L_2} \cdot \underbrace{\limsup_{j\to\infty} \sup_{\xi\in[0,T]} \big\|W_j(\xi) - W(\xi)\big\|_{\mathbb{R}^{d_1}}}_{=0}
\end{aligned}$$

and by the same arguments as before

$$\limsup_{j\to\infty} \sup_{\xi\in[0,T]} \big\| f_\ell\big(\xi, Y_j(\xi)\big)\big\|_{L_2} = \sup_{\xi\in[0,T]} \big\| f_\ell\big(\xi, Y(\xi)\big)\big\|_{L_2}$$

as well as an analogous statement for the second summand. Hence, we can conclude that $\limsup_{j\to\infty} K^2_{j,\ell,l} = 0$ holds by using the reverse Fatou lemma (cf. Lemma A.4 in the Appendix). $\square$

3.3. Main theorem

This section contains the first main theorem of this thesis. We will assume the continuity of the coefficients in their spacial component in addition to the assumptions from Section 2.1 and, of course, the pathwise uniqueness from Definition 2.2. The proof is an extended version of the one stated in [GK96] and uses in particular Lemma 3.6 about the convergence in probability of sequences of (stochastic) integrals as well as tightness criteria to prove the relatively weak compactness of sequences of probability measures via Prokhorov's theorem. At the end, we will also mention in Corollary 3.8 that the well-known local weak monotonicity condition on b and σ can replace some of the assumptions from the theorem.

Theorem 3.7 (cf. [GK96], Theorem 2.4 on page 148). *Let the assumptions from Section 2.1 be fulfilled. Suppose moreover that b and σ are continuous in $x \in D$ and that pathwise uniqueness holds for the equation* (2.1)*. Then we have:*

1) There exists a process $(X(t))_{t\geq 0}$ such that $X_n(t) \xrightarrow[n\to\infty]{p} X(t)$ uniformly in t on bounded intervals.

2) $(X(t))_{t\geq 0}$ is the unique solution of equation (2.1) *(up to P-indistinguishability).*

Recall that $(X_n)_{n\in\mathbb{N}}$ are the *Euler "polygonal" approximations* given by equation (2.2) in Section 2.1.

Proof. (cf. [GK96], Theorem 2.4 on page 150)
For every $T \geq 0$ and $k, n \in \mathbb{N}$ define the stopping times

$$\tau_n^k := \inf\big\{t \in [0,T] \,|\, X_n(t) \notin D_k\big\} \wedge k.$$

Since $\kappa_n(s) \leq s$ for all $s \geq 0$ by the definition of κ_n (see Section 2.1), we have for every $t \leq \tau_n^k$ (if $\tau_n^k > 0$)

$$\big\|b\big(t, X_n(\kappa_n(t))\big)\big\|_{\mathbb{R}^d} \leq M_k(t) \quad \text{and} \quad \big\|\sigma\big(t, X_n(\kappa_n(t))\big)\big\|_{L_2}^2 \leq M_k(t)$$

by assumption **A1)**.

Define the family $\big\{X_n^k \,\big|\, n \in \mathbb{N}\big\}$ of stochastic processes with continuous sample paths by

$$X_n^k(t) := X_n(t \wedge \tau_n^k)$$

for $t \in [0,T]$ and $k, n \in \mathbb{N}$. Let $P_{X_n^k} := P \circ \big(X_n^k\big)^{-1}$ be the distribution of X_n^k. Then for any $T \geq 0$ we can consider $\big\{P_{X_n^k} \,\big|\, n \in \mathbb{N}\big\}$ as a family of probability measures on $C\big([0,T];\mathbb{R}^d\big)$. Since $C\big([0,T];\mathbb{R}^d\big)$ is separable, the *Prokhorov metric* metrises the space of probability measures $\mathcal{M}_1\Big(C\big([0,T];\mathbb{R}^d\big)\Big)$ with respect to weak convergence of probability measures (cf. [Bil99], Section 6 on page 72 and 73).

We will now divide this proof into three consecutive steps (**Step 1**, **Step 2** and **Step 3**) to improve the comprehensibility of the most important ideas, methods and arguments.

Step 1: At first we will show the tightness (cf. Definition A.16 in the Appendix) of the family $\{P_{X_n^k} \mid n \in \mathbb{N}\}$ as it is asserted in **Claim** (1).

Claim (1)**.** For every $k \in \mathbb{N}$ and $T \geq 0$ the family of probability measures

$$\{P_{X_n^k} \mid n \in \mathbb{N}\} \subseteq \mathcal{M}_1\Big(C\big([0,T];\mathbb{R}^d\big)\Big)$$

is tight.

Proof of Claim (1). For the proof we use Theorem A.20 and Theorem A.21 from the Appendix, which jointly add up to a tightness criterion for the distributions of stochastic processes with continuous sample paths. Hence, we have to show that for every fixed $k \in \mathbb{N}$ and $T \geq 0$ the assertions

i) For every $\varepsilon > 0$ there exists an $R > 0$ such that $P\Big[\|X_n^k(0)\|_{\mathbb{R}^d} > R\Big] \leq \varepsilon$ for all $n \in \mathbb{N}$.

ii) There exist $\alpha, \beta, K > 0$ such that $\mathbb{E}\Big[\|X_n^k(t) - X_n^k(s)\|_{\mathbb{R}^d}^\beta\Big] \leq K|t-s|^{1+\alpha}$ for all $n \in \mathbb{N}$ and $s,t \in [0,T]$.

hold.
"*i)*": We know that $\|X_n^k(0)\|_{\mathbb{R}^d} = \|\xi\|_{\mathbb{R}^d}$. Since $\bigcap\limits_{R\in\mathbb{N}} \big\{\|\xi\|_{\mathbb{R}^d} > R\big\} = \emptyset$, it follows that $P\big[\|\xi\|_{\mathbb{R}^d} > R\big] \xrightarrow[R\to\infty]{} 0$. Hence, i) is fulfilled for every $n \in \mathbb{N}$.
"*ii)*": Let $\beta := \frac{2\,(1+\chi)^2}{\chi}$ and $s,t \in [0,T]$. Without loss of generality let $t > s$ and $\tau_n^k > 0$. Since $\mathbb{E}\bigg[\Big\|X_n^k(t) - X_n^k(s)\Big\|_{\mathbb{R}^d}^\beta \mathbb{1}_{\{s>\tau_n^k\}}\bigg] = 0$, we obtain for a constant $C_1 > 0$

$$\begin{aligned}
\mathbb{E}\bigg[\Big\|X_n^k(t) - X_n^k(s)\Big\|_{\mathbb{R}^d}^\beta\bigg] &= \mathbb{E}\bigg[\Big\|X_n^k(t) - X_n^k(s)\Big\|_{\mathbb{R}^d}^\beta \mathbb{1}_{\{s\leq\tau_n^k\}}\bigg] \\
&\overset{(2.2)}{=} \mathbb{E}\Bigg[\bigg\|\int_{s\wedge\tau_n^k}^{t\wedge\tau_n^k} b\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}r + \int_{s\wedge\tau_n^k}^{t\wedge\tau_n^k} \sigma\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}W(r)\bigg\|_{\mathbb{R}^d}^\beta \mathbb{1}_{\{s\leq\tau_n^k\}}\Bigg] \\
&\leq \mathbb{E}\Bigg[\bigg\|\int_{s}^{t\wedge\tau_n^k} b\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}r + \int_{s}^{t\wedge\tau_n^k} \sigma\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}W(r)\bigg\|_{\mathbb{R}^d}^\beta\Bigg] \\
&\leq C_1\,\mathbb{E}\Bigg[\bigg\|\int_{s}^{t\wedge\tau_n^k} b\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}r\bigg\|_{\mathbb{R}^d}^\beta\Bigg] \\
&\qquad\qquad + C_1\,\mathbb{E}\Bigg[\bigg\|\int_0^T \mathbb{1}_{]s,t\wedge\tau_n^k]}(r)\,\sigma\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}W(r)\bigg\|_{\mathbb{R}^d}^\beta\Bigg],
\end{aligned}$$

where we have used a generalised Young inequality (cf. Lemma A.3 in the Appendix) in the last step. Applying the Burkholder-Davis-Gundy type inequality for $p = \beta$ (cf. Theorem A.5 in the Appendix) yields

$$\begin{aligned}
&C_1\,\mathbb{E}\left[\left\| \int_s^{t\wedge\tau_n^k} b\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}r \right\|_{\mathbb{R}^d}^{\beta}\right] \\
&\qquad\qquad + C_1\,\mathbb{E}\left[\left\| \int_0^T \mathbb{1}_{]s,t\wedge\tau_n^k]}(r)\,\sigma\big(r, X_n(\kappa_n(r))\big)\,\mathrm{d}W(r) \right\|_{\mathbb{R}^d}^{\beta}\right] \\
&\quad\le C_1\,\mathbb{E}\left[\left(\int_s^{t\wedge\tau_n^k} \underbrace{\big\|b\big(r, X_n(\kappa_n(r))\big)\big\|_{\mathbb{R}^d}}_{\le M_k(r),\text{ since } r\le\tau_n^k}\,\mathrm{d}r\right)^{\frac{2\,(1+\chi)^2}{\chi}}\right] \\
&\quad\quad + C_2\left(\int_0^T \mathbb{E}\left[\mathbb{1}_{]s,t\wedge\tau_n^k]}(r)\Big(\underbrace{\big\|\sigma\big(r, X_n(\kappa_n(r))\big)\big\|_{L_2}^2}_{\le M_k(r),\text{ since } r\le\tau_n^k}\Big)^{\frac{(1+\chi)^2}{\chi}}\right]^{\frac{\chi}{(1+\chi)^2}}\mathrm{d}r\right)^{\frac{(1+\chi)^2}{\chi}} \\
&\quad\le C_1\left(\int_s^t M_k(r)\,\mathrm{d}r\right)^{\frac{2\,(1+\chi)^2}{\chi}} + C_2\left(\int_s^t M_k(r)\,\mathrm{d}r\right)^{\frac{(1+\chi)^2}{\chi}}
\end{aligned}$$

for a constant $C_2 > 0$ by using $t\wedge\tau_n^k \le t$ as well as $M_k > 0$. By an application of Hölder's inequality we altogether obtain

$$\begin{aligned}
&\mathbb{E}\left[\left\|X_n^k(t) - X_n^k(s)\right\|_{\mathbb{R}^d}^{\beta}\right] \\
&\qquad\le C_1\left(\int_s^t M_k(r)\,\mathrm{d}r\right)^{\frac{2\,(1+\chi)^2}{\chi}} + C_2\left(\int_s^t M_k(r)\,\mathrm{d}r\right)^{\frac{(1+\chi)^2}{\chi}} \\
&\qquad\le \left(\int_s^t M_k(r)\,\mathrm{d}r\right)^{\frac{(1+\chi)^2}{\chi}} \cdot \underbrace{\left(C_1\left(\int_0^T M_k(r)\,\mathrm{d}r\right)^{\frac{(1+\chi)^2}{\chi}} + C_2\right)}_{=:C_3} \\
&\qquad\le C_3\left(\left(\int_s^t 1^{\frac{1+\chi}{\chi}}\,\mathrm{d}r\right)^{\frac{\chi}{1+\chi}}\left(\int_s^t M_k(r)^{1+\chi}\,\mathrm{d}r\right)^{\frac{1}{1+\chi}}\right)^{\frac{(1+\chi)^2}{\chi}} \\
&\qquad\le K(t-s)^{1+\chi},
\end{aligned}$$

where $C_3 > 0$ and $K := C_3\left(\int_0^T M_k(r)^{1+\chi}\,\mathrm{d}r\right)^{\frac{1+\chi}{\chi}} > 0$ are constants. Hence, assertion ii) follows with $\alpha := \chi$. □

Step 2: Now we consider P_{X_n}, i.e. the distribution of X_n, and intend to deduce the tightness of the family $\{P_{X_n} \mid n \in \mathbb{N}\}$ in $\mathcal{M}_1\big(C([0,T];\mathbb{R}^d)\big)$ from **Step 1**.

It suffices to show that the following claim is fulfilled.

Claim (2)**.** We have

$$\lim_{k\to\infty}\limsup_{n\to\infty} P\big[\tau_n^k \le T\big] = 0 \tag{3.8}$$

for every $T \in [0,\infty[$.

This claim is sufficient for the assertion because we have $X_n^k(t) = X_n(t)$ for $t \in [0,\tau_n^k]$ by definition. Hence, the processes coincide on $\{\tau_n^k > T\}$ for every $t \in [0,T]$ and the probability of the event $\{\tau_n^k \le T\}$ tends to zero by taking the limits. In fact, we have for $\varepsilon > 0$

$$\begin{aligned} P\Big[\mathcal{W}_\delta(X_n) > \varepsilon\Big] &= P\Big[\mathcal{W}_\delta(X_n) > \varepsilon, \tau_n^k > T\Big] + P\Big[\mathcal{W}_\delta(X_n) > \varepsilon, \tau_n^k \le T\Big] \\ &\le P\Big[\mathcal{W}_\delta(X_n^k) > \varepsilon\Big] + P[\tau_n^k \le T] \end{aligned}$$

and hence

$$\begin{aligned} &\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P\Big[\mathcal{W}_\delta(X_n) > \varepsilon\Big] \\ &\qquad = \lim_{k\to\infty}\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P\Big[\mathcal{W}_\delta(X_n) > \varepsilon\Big] \\ &\qquad \le \lim_{k\to\infty}\underbrace{\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P\Big[\mathcal{W}_\delta(X_n^k) > \varepsilon\Big]}_{=0 \text{ by } \textbf{Claim } (1)} + \underbrace{\lim_{k\to\infty}\limsup_{n\to\infty} P[\tau_n^k \le T]}_{=0 \text{ by } \textbf{Claim } (2)}. \end{aligned}$$

Besides, for $R > 0$

$$\begin{aligned} \lim_{R\uparrow\infty}\limsup_{n\to\infty} P\Big[\|X_n(0)\|_{\mathbb{R}^d} > R\Big] &= \lim_{R\uparrow\infty}\limsup_{n\to\infty} P\Big[\|\xi\|_{\mathbb{R}^d} > R\Big] \\ &= \underbrace{\lim_{R\uparrow\infty}\limsup_{n\to\infty} P\Big[\|X_n^k(0)\|_{\mathbb{R}^d} > R\Big]}_{=0 \text{ by } \textbf{Claim } (1)} \end{aligned}$$

holds such that we can conclude the tightness of $\{P_{X_n} \,|\, n \in \mathbb{N}\}$ from Theorem A.20.

Proof of Claim (2). Let $T \ge 0$ and let $k \in \mathbb{N}$ such that $k \ge T$. According to **Claim** (1) we know that $\{P_{X_n^k} \,|\, n \in \mathbb{N}\}$ is a tight family of probability measures on $C\big([0,T];\mathbb{R}^d\big)$. Let $\big(P_{X_{n_m}^k}\big)_{m\in\mathbb{N}} \subseteq \{P_{X_n^k} \,|\, n \in \mathbb{N}\}$ be an arbitrary subsequence, then it is again a tight sequence. Since every single probability measure on a separable and complete space is tight (cf. Lemma A.18 in the Appendix), the distribution P_W of the Wiener process W is a tight probability measure on $C\big([0,T];\mathbb{R}^{d_1}\big)$. Hence, the trivial sequence which only consists of P_W is tight. Therefore, by applying Lemma A.22 we obtain that $\big(P_{(X_{n_m}^k,W)}\big)_{m\in\mathbb{N}}$ is a tight family of probability measures on $C\big([0,T];\mathbb{R}^{d+d_1}\big)$. From Prokhorov's theorem (cf. Theorem A.17 in the Appendix) it follows that there exists a relatively weakly convergent subsequence $\big(P_{(X_{n_{m_j}}^k,W)}\big)_{j\in\mathbb{N}}$ by using the fact that in metric spaces relative compactness and relative sequential compactness are equivalent.

Now we can apply Skorokhod's representation theorem (see Theorem 3.3) to this subsequence. Then there exist a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and a sequence of continuous random variables $\big(\tilde{X}^k_{n_{m_j}}, \tilde{W}_j\big)_{j\in\mathbb{N}}$ such that

$$P_{(X^k_{n_{m_j}}, W)} = \tilde{P}_{(\tilde{X}^k_{n_{m_j}}, \tilde{W}_j)} \quad \text{for every } j \in \mathbb{N} \text{ and} \tag{3.9}$$

$$\big(\tilde{X}^k_{n_{m_j}}, \tilde{W}_j\big)_{j\in\mathbb{N}} \xrightarrow[j\to\infty]{\tilde{P}-a.s.} \big(\tilde{X}^k, \tilde{W}\big) \tag{3.10}$$

for stochastic processes $\tilde{X}^k$ and $\tilde{W}$ with values in $C\big([0,T];\mathbb{R}^d\big)$ and $C\big([0,T];\mathbb{R}^{d_1}\big)$, respectively. Therefore, we have $\tilde{P}$-a.s.

$$\begin{aligned} \big\|\tilde{X}^k_{n_{m_j}} - \tilde{X}^k\big\|_\infty &= \sup_{t\in[0,T]} \big\|\tilde{X}^k_{n_{m_j}}(t) - \tilde{X}^k(t)\big\|_{\mathbb{R}^d} \xrightarrow[j\to\infty]{} 0 \quad \text{and} \\ \big\|\tilde{W}_j - \tilde{W}\big\|_\infty &= \sup_{t\in[0,T]} \big\|\tilde{W}_j(t) - \tilde{W}(t)\big\|_{\mathbb{R}^{d_1}} \xrightarrow[j\to\infty]{} 0. \end{aligned} \tag{3.11}$$

In order to prove **Claim** (2), we now have to consider the six auxiliary claims **Claim** (2-1), **Claim** (2-2), **Claim** (2-3), **Claim** (2-4), **Claim** (2-5) and **Claim** (2-6).

We start by defining the stopping times $\tilde{\tau}^k_{n_{m_j}} := \inf\big\{t \in [0,T] \,\big|\, \tilde{X}^k_{n_{m_j}}(t) \notin D_k\big\} \wedge k$ and $\tilde{\tau}^k := \inf\big\{t \in [0,T] \,\big|\, \tilde{X}^k(t) \notin D_k\big\} \wedge k$.

Claim (2-1)**.** The inequality

$$\liminf_{j\to\infty} \tilde{\tau}^k_{n_{m_j}} \geq \tilde{\tau}^k \tag{3.12}$$

$\tilde{P}$-a.s. holds.

Proof of Claim (2-1). This is just an application of Lemma 3.2. □

Now define the σ-algebras

$$\tilde{\mathcal{F}}^j_t := \sigma\big(\tilde{X}^k_{n_{m_j}}(s), \tilde{W}_j(s) \,\big|\, s \in [0,t]\big) \text{ and } \tilde{\mathcal{F}}_t := \sigma\big(\tilde{X}^k(s), \tilde{W}(s) \,\big|\, s \in [0,t]\big).$$

Claim (2-2)**.** For every $j \in \mathbb{N}$ we have that $(\tilde{W}_j(t))_{t\in[0,T]}$ is an $(\tilde{\mathcal{F}}^j_t)$-adapted Wiener process and $(\tilde{W}(t))_{t\in[0,T]}$ is an $(\tilde{\mathcal{F}}_t)$-adapted Wiener process.

Proof of Claim (2-2). First of all, the processes $(\tilde{W}_j(t))_{t\in[0,T]}$ (for every $j \in \mathbb{N}$) and $(\tilde{W}(t))_{t\in[0,T]}$ are $\tilde{P}$-a.s. continuous as well as by definition adapted with respect to $\tilde{\mathcal{F}}^j_t$ and $\tilde{\mathcal{F}}_t$, respectively. Furthermore, we have by (3.9) and (3.10)

$$P_W = \tilde{P}_{\tilde{W}_j} \xrightarrow[j\to\infty]{w} \tilde{P}_{\tilde{W}}$$

because of the equality of the marginal distributions, Lemma A.12 and Lemma A.15 i) from the Appendix. Hence, $\tilde{W}_j$ and $\tilde{W}$ have the same distribution as the Wiener process W.

Next we note that $\tilde{W}_j(0) = 0$ holds since we can define the measurable set

$$U_1 := \Big\{ u \in C\big([0,T];\mathbb{R}^{d_1}\big) \,\Big|\, u(0) = 0 \Big\}$$

such that we obtain

$$1 = P_W[U_1] = \tilde{P}_{\tilde{W}_j}[U_1].$$

We also have $\tilde{W}(0) = 0$ $\tilde{P}$-a.s. since

$$\big\|\tilde{W}(0) - 0\big\|_{\mathbb{R}^{d_1}} = \big\|\tilde{W}(0) - \tilde{W}_j(0)\big\|_{\mathbb{R}^{d_1}} \leq \sup_{t\in[0,T]} \big\|\tilde{W}_j(t) - \tilde{W}(t)\big\|_{\mathbb{R}^{d_1}} \xrightarrow[j\to\infty]{} 0$$

follows from (3.11). For the independence of the increments we consider the following four intermediate steps (**(1)**, **(2)**, **(3)** and **(4)**).

(1): The *Euler "polygonal" approximations* $\big(X_n(t)\big)$ are $(\mathcal{F}_t)$-adapted for any $n \in \mathbb{N}$.

For $t \in [t_0^n, t_1^n]$ we have by (2.2) that

$$X_n(t) = \xi + \int_0^t b(s,\xi)\,\mathrm{d}s + \int_0^t \sigma(s,\xi)\,\mathrm{d}W(s),$$

where ξ is $\mathcal{F}_0$-measurable and b, σ are Borel-measurable by assumption. Hence, after the integration we obtain the $(\mathcal{F}_t)$-measurability of $X_n(t)$. Inductively we get for $t \in [t_i^n, t_{i+1}^n]$ that

$$X_n(t) = X_n(t_i^n) + \int_{t_i^n}^t b\big(s, X_n(t_i^n)\big)\,\mathrm{d}s + \int_{t_i^n}^t \sigma\big(s, X_n(t_i^n)\big)\,\mathrm{d}W(s)$$

holds such that the $(\mathcal{F}_{t_i^n})$-measurability of $X_n(t_i^n)$ and the Borel-measurability of b and σ again imply that $X_n(t)$ is $(\mathcal{F}_t)$-measurable.

(2): Adjustment of the σ-algebras.

The measurability of $X_{n_{m_j}}(t)$ from **(1)** yields that the stopped process $X^k_{n_{m_j}}(t)$ is $(\mathcal{F}_t)$-measurable as well, i.e. we have $\mathcal{F}_t = \sigma\big(\mathcal{F}_t, X^k_{n_{m_j}}(s) \,|\, s \leq t\big)$. Furthermore, we know that $\big(W(t)\big)_{t\geq 0}$ is $(\mathcal{F}_t)$-adapted by assumption, hence we can write $\mathcal{F}_t = \sigma\big(\mathcal{F}_t, X^k_{n_{m_j}}(s), W(s) \,|\, s \leq t\big)$. Therefore, we conclude that the increment $W(t) - W(s)$ is in fact independent of $\sigma\big(\mathcal{F}_s, X^k_{n_{m_j}}(r), W(r) \,|\, r \leq s\big)$ and certainly independent of the smaller σ-algebra given by $\sigma\big(X^k_{n_{m_j}}(r), W(r) \,|\, r \leq s\big)$.

(3): The increment $\tilde{W}_j(t) - \tilde{W}_j(s)$ is independent of $\tilde{\mathcal{F}}^j_s = \sigma\big(\tilde{X}^k_{n_{m_j}}(r), \tilde{W}_j(r) \,\big|\, r \leq s\big)$.

First of all, we know that

$$\sigma\big(\tilde{W}_j(t) - \tilde{W}_j(s)\big) = \Big\{ \big\{\tilde{W}_j(t) - \tilde{W}_j(s) \in B\big\} \,\big|\, B \in \mathcal{B}(\mathbb{R}^{d_1}) \Big\}.$$

Besides, the σ-algebra $\tilde{\mathcal{F}}^j_s$ is generated by sets of the form

$$\Big\{ \big(\tilde{X}^k_{n_{m_j}}(s_0), \tilde{W}_j(s_0), \ldots, \tilde{X}^k_{n_{m_j}}(s_N), \tilde{W}_j(s_N)\big) \in B \Big\},$$

where $N \in \mathbb{N}$, $0 = s_0 \leq s_1 \leq \cdots \leq s_N = s$ and $B \in \mathcal{B}\big(\mathbb{R}^{N(d+d_1)}\big)$.

For $N \in \mathbb{N}$, $0 = s_0 \leq s_1 \leq \cdots \leq s_N = s$, $B_1 \in \mathcal{B}(\mathbb{R}^{d_1})$, $B_2 \in \mathcal{B}(\mathbb{R}^{N(d+d_1)})$ and $\bar{B}_1 := \{(w_1, w_2) \in \mathbb{R}^{2d_1} \,|\, w_1 - w_2 \in B_1\}$ we compute

$$\begin{aligned}
\tilde{P}\Big[\tilde{W}_j(t) - \tilde{W}_j(s) \in B_1\Big] \cdot \tilde{P}\Big[\big(\tilde{X}^k_{n_{m_j}}(s_0), \tilde{W}_j(s_0), \ldots, \tilde{X}^k_{n_{m_j}}(s_N), \tilde{W}_j(s_N)\big) \in B_2\Big]& \\
= \tilde{P}_{\left(\tilde{W}_j(t), \tilde{W}_j(s)\right)}\big[\bar{B}_1\big] \cdot \tilde{P}_{\left(\tilde{X}^k_{n_{m_j}}(s_0), \tilde{W}_j(s_0), \ldots, \tilde{X}^k_{n_{m_j}}(s_N), \tilde{W}_j(s_N)\right)}\big[B_2\big]& \\
= \underbrace{P_{\left(W(t), W(s)\right)}\big[\bar{B}_1\big]}_{= P_{W(t)-W(s)}[B_1]} \cdot P_{\left(X^k_{n_{m_j}}(s_0), W(s_0), \ldots, X^k_{n_{m_j}}(s_N), W(s_N)\right)}\big[B_2\big]& \\
= P\Big[\big\{W(t) - W(s) \in B_1\big\}& \\
\cap \Big\{\big(X^k_{n_{m_j}}(s_0), W(s_0), \ldots, X^k_{n_{m_j}}(s_N), W(s_N)\big) \in B_2\Big\}\Big]& \\
= P\Big[\big\{\big(W(t), W(s)\big) \in \bar{B}_1\big\}& \\
\cap \Big\{\big(X^k_{n_{m_j}}(s_0), W(s_0), \ldots, X^k_{n_{m_j}}(s_N), W(s_N)\big) \in B_2\Big\}\Big]&,
\end{aligned}$$

where we have used in the second step that by (3.9) the distributions $P_{(X^k_{n_{m_j}}, W)}$ and $\tilde{P}_{(\tilde{X}^k_{n_{m_j}}, \tilde{W}_j)}$ and, therefore, also their finite-dimensional distributions coincide. Besides, the third step follows by the independence of $W(t) - W(s)$ from the σ-algebra $\sigma\big(X^k_{n_{m_j}}(r), W(r) \,|\, r \leq s\big)$ (see **(2)**). Hence, we obtain

$$\begin{aligned}
P\Big[\big\{\big(W(t), W(s)\big) \in \bar{B}_1\big\} \cap \big\{\big(X^k_{n_{m_j}}(s_0), W(s_0), \ldots, X^k_{n_{m_j}}(s_N), W(s_N)\big) \in B_2\big\}\Big]& \\
= P\Big[\big(W(t), W(s), X^k_{n_{m_j}}(s_0), W(s_0), \ldots, X^k_{n_{m_j}}(s_N), W(s_N)\big) \in \bar{B}_1 \times B_2\Big]& \\
= P_{\left(W(t), W(s), X^k_{n_{m_j}}(s_0), W(s_0), \ldots, X^k_{n_{m_j}}(s_N), W(s_N)\right)}\big[\bar{B}_1 \times B_2\big]& \\
= \tilde{P}_{\left(\tilde{W}_j(t), \tilde{W}_j(s), \tilde{X}^k_{n_{m_j}}(s_0), \tilde{W}_j(s_0), \ldots, \tilde{X}^k_{n_{m_j}}(s_N), \tilde{W}_j(s_N)\right)}\big[\bar{B}_1 \times B_2\big]& \\
= \tilde{P}\Big[\big\{\tilde{W}_j(t) - \tilde{W}_j(s) \in B_1\big\}& \\
\cap \big\{\big(\tilde{X}^k_{n_{m_j}}(s_0), \tilde{W}_j(s_0), \ldots, \tilde{X}^k_{n_{m_j}}(s_N), \tilde{W}_j(s_N)\big) \in B_2\big\}\Big]&,
\end{aligned}$$

where we have again used the equality of the finite-dimensional distributions in the third step.

(4): The increment $\tilde{W}(t) - \tilde{W}(s)$ is independent of $\tilde{\mathcal{F}}_s = \sigma\big(\tilde{X}^k(r), \tilde{W}(r) \,\big|\, r \leq s\big)$.

Let $N \in \mathbb{N}$. Assume that $\varphi \in C\big(\mathbb{R}^{d_1}; \mathbb{R}\big)$ and $\psi \in C\big(\mathbb{R}^{N(d+d_1)}; \mathbb{R}\big)$ are bounded functions. By using the monotone class theorem (see e.g. [Pro05], Theorem 8 on page

7) it suffices to conclude from **(3)** and (3.11) that

$$
\begin{aligned}
&\mathbb{E}\Big[\varphi\big(\tilde{W}(t)-\tilde{W}(s)\big)\,\psi\Big(\tilde{X}^k(s_0),\tilde{W}(s_0),\dots,\tilde{X}^k(s_N),\tilde{W}(s_N)\Big)\Big]\\
&\quad=\lim_{j\to\infty}\mathbb{E}\Big[\varphi\big(\tilde{W}_j(t)-\tilde{W}_j(s)\big)\,\psi\Big(\tilde{X}^k_{n_{m_j}}(s_0),\tilde{W}_j(s_0),\dots,\tilde{X}^k_{n_{m_j}}(s_N),\tilde{W}_j(s_N)\Big)\Big]\\
&\quad=\lim_{j\to\infty}\mathbb{E}\Big[\varphi\big(\tilde{W}_j(t)-\tilde{W}_j(s)\big)\Big]\\
&\qquad\qquad\qquad\cdot\mathbb{E}\Big[\psi\Big(\tilde{X}^k_{n_{m_j}}(s_0),\tilde{W}_j(s_0),\dots,\tilde{X}^k_{n_{m_j}}(s_N),\tilde{W}_j(s_N)\Big)\Big]\\
&\quad=\mathbb{E}\Big[\varphi\big(\tilde{W}(t)-\tilde{W}(s)\big)\Big]\,\mathbb{E}\Big[\psi\Big(\tilde{X}^k(s_0),\tilde{W}(s_0),\dots,\tilde{X}^k(s_N),\tilde{W}(s_N)\Big)\Big]
\end{aligned}
$$

holds for $N\in\mathbb{N}$ and $0=s_0\leq s_1\leq\cdots\leq s_N=s$ by Lebesgue's dominated convergence theorem. □

At this point it is necessary to ensure that $(\tilde{\mathcal{F}}^j_t)$ and $(\tilde{\mathcal{F}}_t)$ are normal filtrations such that they are suitable for the usual stochastic integration theory. In case they are not normal, we augment them by all $\tilde{P}$-zero sets and make them right-continuous by construction (intersection of the larger σ-algebras) as it is done for example in [PR07] on page 12. Then the processes $(\tilde{W}_j(t))_{t\in[0,T]}$ and $(\tilde{W}(t))_{t\in[0,T]}$ are still Wiener processes with respect to these new normal filtrations (cf. [PR07], Proposition 2.1.13 on page 12 for an applicable proof).

Furthermore, we have the following claim.

Claim (2-3)**.** We have $\tilde{P}$-a.s.

$$
\tilde{X}^k_{n_{m_j}}(t)=\tilde{X}^k_{n_{m_j}}(0)+\int_0^t b\big(s,\tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}s+\int_0^t\sigma\big(s,\tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}\tilde{W}_j(s)
$$

for all $t\in[0,T\wedge\tilde{\tau}^k_{n_{m_j}}]$.

Proof of Claim (2-3). To verify this equality we define the set

$$
\begin{aligned}
U_2:=\Big\{(u_{\ell_j},v_j)\in C\big([0,T];\mathbb{R}^{d+d_1}\big)\,\Big|\,&\forall t\in[0,T\wedge\tau_{u_{\ell_j}}]:\\
&u_{\ell_j}(t)=u_{\ell_j}(0)+\int_0^t b\big(s,u_{\ell_j}(\kappa_{\ell_j}(s))\big)\,\mathrm{d}s+\int_0^t\sigma\big(s,u_{\ell_j}(\kappa_{\ell_j}(s))\big)\,\mathrm{d}v_j(s)\Big\},
\end{aligned}
$$

where $\tau_{u_{\ell_j}}:=\inf\big\{t\in[0,T]\,\big|\,u_{\ell_j}(t)\notin D_k\big\}\wedge k$. Then we have by (3.9)

$$
\tilde{P}\big[(\tilde{X}^k_{n_{m_j}},\tilde{W}_j)\in U_2\big]=\tilde{P}_{(\tilde{X}^k_{n_{m_j}},\tilde{W}_j)}\big[U_2\big]=P_{(X^k_{n_{m_j}},W)}\big[U_2\big]=P\big[(X^k_{n_{m_j}},W)\in U_2\big]=1
$$

since $X^k_{n_{m_j}}$ satisfies equation (2.2) P-a.s. for every $t\in[0,T\wedge\tau^k_{n_{m_j}}]$. □

Claim (2-4)**.** We have

$$\int_0^t b\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}s \xrightarrow[j\to\infty]{p} \int_0^t b\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}s \quad \text{and}$$

$$\int_0^t \sigma\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}\tilde{W}_j(s) \xrightarrow[j\to\infty]{p} \int_0^t \sigma\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}\tilde{W}(s)$$

for $t < T \wedge \tilde{\tau}^k$.

Proof of Claim (2-4). With the help of Lemma A.8 from the Appendix we calculate

$$\begin{aligned}
\limsup_{j\to\infty} \tilde{P}\Big[\tilde{\tau}^k_j \le t,\, t < \tilde{\tau}^k \wedge T\Big] &\le \tilde{P}\Big[\limsup_{j\to\infty}\big\{\tilde{\tau}^k_j \le t\big\},\, t < \tilde{\tau}^k \wedge T\Big] \\
&= \tilde{P}\Big[\bigcap_{\varrho\in\mathbb{N}}\bigcup_{j\ge\varrho}\big\{\tilde{\tau}^k_j \le t\big\},\, t < \tilde{\tau}^k \wedge T\Big] \\
&\le \tilde{P}\Big[\bigcap_{\varrho\in\mathbb{N}}\Big\{\inf_{j\ge\varrho}\tilde{\tau}^k_j \le t\Big\},\, t < \tilde{\tau}^k \wedge T\Big] \\
&= \tilde{P}\Big[\Big\{\sup_{\varrho\in\mathbb{N}}\inf_{j\ge\varrho}\tilde{\tau}^k_j \le t\Big\},\, t < \tilde{\tau}^k \wedge T\Big] \\
&= \tilde{P}\Big[\liminf_{j\to\infty}\tilde{\tau}^k_j \le t,\, t < \tilde{\tau}^k \wedge T\Big].
\end{aligned}$$

Therefore, since $\tilde{P}\big[\liminf_{n\to\infty}\tilde{\tau}^k_j \ge \tilde{\tau}^k\big] = 1$ holds by **Claim** (2-1), we conclude

$$\limsup_{j\to\infty} \tilde{P}\Big[\tilde{\tau}^k_j \le t,\, t < \tilde{\tau}^k \wedge T\Big] = 0. \tag{3.13}$$

The latter implies that we only have to consider the convergence in probability with respect to the event $\{t < \tilde{\tau}^k_j\} \cap \{t < \tilde{\tau}^k \wedge T\}$.

Now, in order to apply Lemma 3.6, we have to localise the functions b and σ. Therefore, define the set A as the complement of D_{k+1} in D, i.e. $A := D \setminus D_{k+1}$. Let $\gamma\colon D \to [0,1]$ be the function given by

$$\gamma(x) := \frac{\operatorname{dist}\big(x, A\big)}{\operatorname{dist}\big(x, \overline{D}_k\big) + \operatorname{dist}\big(x, A\big)}$$

for $x \in D$.

Then for σ we can consider the function $\bar{\sigma}\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times d_1}$ given by

$$\bar{\sigma}(s,x) := \begin{cases} \sigma(s,x)\,\gamma(x) & \text{for } x \in D_{k+1}, s \in [0,T], \\ 0 & \text{else,} \end{cases}$$

which is continuous in $x \in \mathbb{R}^d$ and Borel-measurable in $s \in \mathbb{R}_+$. By this definition we also know that $\bar{\sigma}(s,x) = \sigma(s,x)$ for $x \in \overline{D}_k, s \in [0,T]$ and

$$\sup_{x\in\mathbb{R}^d} \|\bar{\sigma}(s,x)\|_{\mathbb{R}^d} = \sup_{x\in D_{k+1}} \|\bar{\sigma}(s,x)\|_{\mathbb{R}^d} \le \sup_{x\in D_{k+1}} \|\sigma(s,x)\|_{\mathbb{R}^d} \le M_{k+1}(s)$$

for every $s \in [0,T]$ since $k \geq T$ by assumption. Let $\varepsilon > 0$. Then we have

$$
\begin{aligned}
&\limsup_{j\to\infty} \tilde{P}\Bigg[\bigg\| \int_0^t \sigma\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}\tilde{W}_j(s) \\
&\qquad\qquad - \int_0^t \sigma\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}\tilde{W}(s)\bigg\|_{\mathbb{R}^d} \geq \varepsilon,\, t < \tau_j^k,\, t < \tau^k \wedge T\Bigg] \\
&= \limsup_{j\to\infty} \tilde{P}\Bigg[\bigg\| \int_0^{t\wedge\tilde{\tau}_j^k\wedge T} \sigma\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}\tilde{W}_j(s) \\
&\qquad\qquad - \int_0^{t\wedge\tilde{\tau}^k\wedge T} \sigma\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}\tilde{W}(s)\bigg\|_{\mathbb{R}^d} \geq \varepsilon,\, t < \tau_j^k,\, t < \tau^k \wedge T\Bigg] \\
&\leq \limsup_{j\to\infty} \tilde{P}\Bigg[\bigg\| \int_0^{t\wedge T} \bar{\sigma}\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}\tilde{W}_j(s) \\
&\qquad\qquad - \int_0^{t\wedge T} \bar{\sigma}\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}\tilde{W}(s)\bigg\|_{\mathbb{R}^d} \geq \varepsilon\Bigg] \\
&= 0
\end{aligned}
$$

by applying Lemma 3.6. Analogously, for a function $\bar{b}\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ the estimate

$$
\begin{aligned}
&\limsup_{j\to\infty} \tilde{P}\Bigg[\bigg\| \int_0^t b\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}s \\
&\qquad\qquad - \int_0^t b\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}s\bigg\|_{\mathbb{R}^d} \geq \varepsilon,\, t < \tau_j^k,\, t < \tau^k \wedge T\Bigg] \\
&= \limsup_{j\to\infty} \tilde{P}\Bigg[\bigg\| \int_0^{t\wedge\tilde{\tau}_j^k\wedge T} b\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}s \\
&\qquad\qquad - \int_0^{t\wedge\tilde{\tau}^k\wedge T} b\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}s\bigg\|_{\mathbb{R}^d} \geq \varepsilon,\, t < \tau_j^k,\, t < \tau^k \wedge T\Bigg] \\
&\leq \limsup_{j\to\infty} \tilde{P}\Bigg[\bigg\| \int_0^{t\wedge T} \bar{b}\big(s, \tilde{X}^k_{n_{m_j}}(\kappa_{n_{m_j}}(s))\big)\,\mathrm{d}s - \int_0^{t\wedge T} \bar{b}\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}s\bigg\|_{\mathbb{R}^d} \geq \varepsilon\Bigg] \\
&= 0
\end{aligned}
$$

holds. This directly yields the assertion of the claim by using (3.13). □

From **Claim** (2-3) we can now conclude that the process $\tilde{X}^k$ satisfies the SDE for every $t \in [0, T \wedge \tilde{\tau}^k]$.

Claim (2-5)**.** The process $\tilde{X}^k$ satisfies $\tilde{P}$-a.s. the equation

$$
\tilde{X}^k(t) = \tilde{X}^k(0) + \int_0^t b\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}s + \int_0^t \sigma\big(s, \tilde{X}^k(s)\big)\,\mathrm{d}\tilde{W}(s) \tag{3.14}
$$

for $t \in [0, T \wedge \tilde{\tau}^k]$.

Proof of Claim (2-5). By (3.11) we know that $\tilde{X}^k_{n_{m_j}}(t) \xrightarrow[j\to\infty]{\tilde{P}-a.s.} \tilde{X}^k(t)$ uniformly in $t \in [0,T]$. In particular, we have $\tilde{X}^k_{n_{m_j}}(0) \xrightarrow[j\to\infty]{p} \tilde{X}^k(0)$ (cf. Lemma A.12 i), which states that a.s. convergence implies convergence in probability). Furthermore, we have the convergence in probability of the integrals for $t \in [0, T \wedge \tilde{\tau}^k[$ by **Claim** (2-4). Therefore, **Claim** (2-3) implies that equation (3.14) $\tilde{P}$-a.s. holds for every $t \in [0, T \wedge \tilde{\tau}^k[$ because the $\tilde{P}$-a.s. limit and the limit in probability have to be $\tilde{P}$-a.s. equal.

Note that both integrals in (3.14) are continuous in t and exist for $t \in [0, T \wedge \tilde{\tau}^k]$. Since $\tilde{X}^k(t)$ is continuous for $t \in [0,T]$, we conclude that the equation actually holds for $t \in [0, T \wedge \tilde{\tau}^k]$. □

Furthermore, we have to prove that the following claim holds.

Claim (2-6)**.** We have $\lim\limits_{k\to\infty} \tilde{P}\big[\tilde{\tau}^k \leq T\big] = 0$.

Proof of Claim (2-6). In the proof of inequality (2.3) we have only used that the process considered in Lemma 2.4 is P-a.s. continuous and satisfies the equation (2.1) until it hits ∂D_k as well as that its initial value is in D. Therefore, we work with the same kind of inequality as in the lemma for $\tilde{\tau}^k$ and $\tilde{X}^k$ in this case.

Namely, we obtain analogously to (2.4) the estimate

$$\begin{aligned}\tilde{P}\big[\tilde{\tau}^k \leq T\big] \leq \tilde{P}\big[\tilde{X}^k(0) \notin D_k\big] &+ \tilde{P}\Big[V\big(0, \tilde{X}^k(0)\big) \geq \log\Big(\frac{1}{\delta}\Big), \tilde{X}^k(0) \in D_k\Big] \\ &+ \tilde{P}\Big[\tilde{\tau}^k \leq T, \tilde{X}^k(0) \in D_k, V\big(0, \tilde{X}^k(0)\big) < \log\Big(\frac{1}{\delta}\Big)\Big]\end{aligned}$$

because $\tilde{X}^k$ has continuous sample paths and satisfies the equation (2.1) with $\tilde{W}$ instead of W until it hits ∂D_k by **Claim** (2-5). Besides, we have the condition $\tilde{X}^k(0) \in D_k$ in the second and third summand such that $V\big(0, \tilde{X}^k(0)\big)$ is still defined. Hence, repeating the arguments from the proof of Lemma 2.4 yields

$$\begin{aligned}\tilde{P}\big[\tilde{\tau}^k \leq T\big] \leq \tilde{P}\big[\tilde{X}^k(0) \notin D_k\big] &+ \tilde{P}\Big[V\big(0, \tilde{X}^k(0)\big) \geq \log\Big(\frac{1}{\delta}\Big), \tilde{X}^k(0) \in D_k\Big] \\ &+ \frac{1}{\delta V_k(T)} \exp\Big(\int_0^T M(t)\,\mathrm{d}t\Big).\end{aligned} \tag{3.15}$$

Now note that we also have $P_\xi = \tilde{P}_{\tilde{X}^k(0)}$ since (3.9), (3.10) and Lemma A.15 i) from the Appendix imply

$$P_\xi = P_{X^k_{n_{m_j}}(0)} = \tilde{P}_{\tilde{X}^k_{n_{m_j}}(0)} \xrightarrow[j\to\infty]{w} \tilde{P}_{\tilde{X}^k(0)}.$$

Therefore, we can transform estimate (3.15) into

$$\tilde{P}\big[\tilde{\tau}^k \leq T\big] \leq P\big[\xi \notin D_k\big] + P\Big[V\big(0, \xi\big) \geq \log\Big(\frac{1}{\delta}\Big)\Big] + \frac{1}{\delta V_k(T)} \exp\Big(\int_0^T M(t)\,\mathrm{d}t\Big),$$

which coincides with the estimate (2.3). Hence, we conclude $\lim\limits_{k\to\infty} \tilde{P}\big[\tilde{\tau}^k \leq T\big] = 0$. □

Since the inequality

$$\tilde{P}\Big[\limsup_{j\to\infty}\big\{\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big] = \tilde{P}\Big[\bigcap_{\varrho\in\mathbb{N}}\bigcup_{j\ge\varrho}\big\{\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big] \le \tilde{P}\Big[\bigcap_{\varrho\in\mathbb{N}}\big\{\inf_{j\ge\varrho}\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big]$$

$$= \tilde{P}\Big[\big\{\sup_{\varrho\in\mathbb{N}}\inf_{j\ge\varrho}\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big] = P\Big[\big\{\liminf_{j\to\infty}\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big]$$

is fulfilled and the stopping times $\tau^k_{n_{m_j}}$ and $\tilde{\tau}^k_{n_{m_j}}$ have the same distribution (cf. Lemma 3.1), it follows from **Claim** (2-6) that

$$\lim_{k\to\infty}\limsup_{j\to\infty} P\big[\tau^k_{n_{m_j}} \le T\big] = \lim_{k\to\infty}\limsup_{j\to\infty}\tilde{P}\big[\tilde{\tau}^k_{n_{m_j}} \le T\big] \le \lim_{k\to\infty}\tilde{P}\Big[\limsup_{j\to\infty}\big\{\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big]$$

$$\le \lim_{k\to\infty}\tilde{P}\Big[\big\{\liminf_{j\to\infty}\tilde{\tau}^k_{n_{m_j}} \le T\big\}\Big] \le \lim_{k\to\infty}\tilde{P}\big[\tilde{\tau}^k \le T\big] = 0$$

holds, where we have used Lemma A.8 from the Appendix and $\tilde{\tau}^k \le \liminf_{j\to\infty}\tilde{\tau}^k_{n_{m_j}}$ $\tilde{P}$-a.s. from **Claim** (2-1).

Hence, we have proved that equation (3.8) holds for the subsequence $(\tau^k_{n_{m_j}})_{j\in\mathbb{N}}$ of an arbitrary subsequence $(\tau^k_{n_m})_{m\in\mathbb{N}}$. Now we obtain the assertion for the whole sequence $(\tau^k_n)_{n\in\mathbb{N}}$ as follows. Since $\limsup_{j\to\infty} P\big[\tau^k_{n_{m_j}} \le T\big] \le \tilde{P}\big[\tilde{\tau}^k \le T\big]$, we can apply Lemma A.14 such that $\limsup_{n\to\infty} P\big[\tau^k_n \le T\big] \le \tilde{P}\big[\tilde{\tau}^k \le T\big]$ and, therefore, $\lim_{k\to\infty}\limsup_{n\to\infty} P[\tau^k_n \le T] = 0$ hold. □

Step 3: In this last step we apply Lemma 1.14 to conclude convergence in probability of the *Euler "polygonal" approximations* by proving convergence in distribution of certain subsequences.

Therefore, we take two arbitrary subsequences $(X_{n_l})_{l\in\mathbb{N}}$ and $(X_{\bar{n}_l})_{l\in\mathbb{N}}$ of the *Euler "polygonal" approximations* $(X_n)_{n\in\mathbb{N}}$ and additionally the Wiener process W. Let $P_{(X_{n_l},W,X_{\bar{n}_l},W)}$ be the joint distribution of the stochastic process $(X_{n_l},W,X_{\bar{n}_l},W)$. Then as in **Step 2**, by using Lemma A.22, we obtain for any $T\ge 0$ that $\big\{P_{(X_{n_l},W,X_{\bar{n}_l},W)}\,|\,l\in\mathbb{N}\big\} \subseteq \mathcal{M}_1\Big(C\Big([0,T];\mathbb{R}^{2(d+d_1)}\Big)\Big)$ is a tight family of probability measures and, hence, a relatively weakly compact set (cf. Prokhorov's theorem, Theorem A.17). Consequently, there exists a relatively weakly convergent subsequence $\big(P_{(X_{n_{l_j}},W,X_{\bar{n}_{l_j}},W)}\big)_{j\in\mathbb{N}}$ with a limit that we first of all label as μ. Again by applying Skorokhod's representation theorem (see Theorem 3.3) to this sequence, there exists a probability space $(\hat{\Omega},\hat{\mathcal{F}},\hat{P})$ and a sequence of continuous random processes $\big(\hat{X}_{n_{l_j}},\hat{W}_j,\check{X}_{\bar{n}_{l_j}},\check{W}_j\big)_{j\in\mathbb{N}}$ such that

$$P_{(X_{n_{l_j}},W,X_{\bar{n}_{l_j}},W)} = \hat{P}_{(\hat{X}_{n_{l_j}},\hat{W}_j,\check{X}_{\bar{n}_{l_j}},\check{W}_j)} \quad \text{for every } j\in\mathbb{N}, \tag{3.16}$$

$$\mu = \hat{P}_{(\hat{X},\hat{W},\check{X},\check{W})}, \tag{3.17}$$

$$\big(\hat{X}_{n_{l_j}},\hat{W}_j,\check{X}_{\bar{n}_{l_j}},\check{W}_j\big) \xrightarrow[j\to\infty]{\hat{P}-a.s.} \big(\hat{X},\hat{W},\check{X},\check{W}\big) \tag{3.18}$$

for stochastic processes $\hat{X}, \check{X}, \hat{W}$ and $\check{W}$ taking values in $C\big([0,T];\mathbb{R}^d\big)$ and $C\big([0,T];\mathbb{R}^{d_1}\big)$, respectively. Similarly to **Step 2** we have $\hat{P}$-a.s.

$$\begin{aligned} \big\|\hat{X}_{n_{l_j}} - \hat{X}\big\|_\infty &= \sup_{t\in[0,T]} \big\|\hat{X}_{n_{l_j}}(t) - \hat{X}(t)\big\|_{\mathbb{R}^d} \xrightarrow[j\to\infty]{} 0, \\ \big\|\check{X}_{\bar{n}_{l_j}} - \check{X}\big\|_\infty &= \sup_{t\in[0,T]} \big\|\check{X}_{\bar{n}_{l_j}}(t) - \check{X}(t)\big\|_{\mathbb{R}^d} \xrightarrow[j\to\infty]{} 0, \\ \big\|\hat{W}_j - \hat{W}\big\|_\infty &= \sup_{t\in[0,T]} \big\|\hat{W}_j(t) - \hat{W}(t)\big\|_{\mathbb{R}^{d_1}} \xrightarrow[j\to\infty]{} 0, \\ \big\|\check{W}_j - \check{W}\big\|_\infty &= \sup_{t\in[0,T]} \big\|\check{W}_j(t) - \check{W}(t)\big\|_{\mathbb{R}^{d_1}} \xrightarrow[j\to\infty]{} 0 \end{aligned} \tag{3.19}$$

for any $T \geq 0$.

Now we can define the set

$$U_1 := \Big\{(u,v,w,x) \in C\Big([0,T];\mathbb{R}^{2(d+d_1)}\Big) \,\Big|\, v(t) = x(t) \text{ for every } t \in [0,T]\Big\}$$

for which we have

$$1 = P_{(X_{n_{l_j}},W,X_{\bar{n}_{l_j}},W)}\big[U_1\big] = \hat{P}_{(\hat{X}_{n_{l_j}},\hat{W}_j,\check{X}_{\bar{n}_{l_j}},\check{W}_j)}\big[U_1\big]$$

for every $j \in \mathbb{N}$ by (3.16). Hence, for any $T \geq 0$, we obtain that $\hat{P}$-a.s. $\hat{W}_j(t) = \check{W}_j(t)$ for every $t \in [0,T]$. By using the $\hat{P}$-a.s. convergence from (3.18) and Lemma A.15 iii) from the Appendix, we can conclude that $\hat{P}$-a.s.

$$\hat{W}(t) = \check{W}(t)$$

for every $t \in [0,\infty[$.

In the same way as in the proof of **Claim** (2) we obtain that for every $k \in \mathbb{N}$ and $T \geq 0$ the processes $\hat{X}$ and $\check{X}$ satisfy equation (2.1) with $\hat{W}$ instead of W on the time intervals $[0, T \wedge \hat{\tau}^k]$ and $[0, T \wedge \check{\tau}^k]$, respectively, where $\hat{\tau}^k := \inf\big\{t \in [0,T] \,\big|\, \hat{X}(t) \notin D_k\big\} \wedge k$ and $\check{\tau}^k := \inf\big\{t \in [0,T] \,\big|\, \check{X}(t) \notin D_k\big\} \wedge k$ are stopping times.

Again as in **Step 2**, by using the inequality (2.3) from Lemma 2.4, we can proof for every $T \geq 0$ that $\lim\limits_{k\to\infty} \hat{P}\big[\hat{\tau}^k \leq T\big] = 0$ and $\lim\limits_{k\to\infty} \hat{P}\big[\check{\tau}^k \leq T\big] = 0$ hold such that the processes $(\hat{X}(t))_{t\geq 0}$ and $(\check{X}(t))_{t\geq 0}$ actually satisfy equation (2.1) with $\hat{W}$ instead of W on $[0,\infty[$.

Note that $\hat{X}_{n_{l_j}}(0) = \check{X}_{n_{l_j}}(0)$ for every $j \in \mathbb{N}$. This equality is fulfilled because $X_{n_{l_j}}(0) = \xi = X_{\bar{n}_{l_j}}(0)$ and the equality of the distributions (3.16) hold. Consequently, we can define for any $T \geq 0$ the measurable set

$$U_2 := \Big\{(u,v,w,x) \in C\Big([0,T];\mathbb{R}^{2(d+d_1)}\Big) \,\Big|\, u(0) = w(0)\Big\}$$

and calculate

$$1 = P_{(X_{n_{l_j}},W,X_{\bar{n}_{l_j}},W)}\big[U_2\big] = \hat{P}_{(\hat{X}_{n_{l_j}},\hat{W}_j,\check{X}_{\bar{n}_{l_j}},\check{W}_j)}\big[U_2\big].$$

Hence, we can conclude from (3.18) that the initial values $\hat{X}(0)$ and $\check{X}(0)$ also have to be $\hat{P}$-a.s. equal.

Then $P_{(\xi,W)} = \hat{P}_{(\check{X}(0),\hat{W})}$ holds because we have (3.16) as well as (3.18) and, therefore, in particular

$$P_{(\xi,W,\xi,W)} = P_{(X_{n_{l_j}}(0),W,X_{\bar{n}_{l_j}}(0),W)} = \hat{P}_{(\hat{X}_{n_{l_j}}(0),\hat{W}_j,\check{X}_{\bar{n}_{l_j}}(0),\hat{W}_j)} \xrightarrow[j\to\infty]{w} \hat{P}_{(\hat{X}(0),\hat{W},\check{X}(0),\hat{W})}$$

such that an application of Lemma A.15 i) from the Appendix yields the equality of these distributions.

Hence, by pathwise uniqueness we conclude that $\hat{P}$-a.s.

$$\hat{X}(t) = \check{X}(t)$$

for all $t \in [0,\infty[$.

Therefore, we finish the proof as follows. We have proved that

$$P_{(X_{n_{l_j}},W,X_{\bar{n}_{l_j}},W)} = \hat{P}_{(\hat{X}_{n_{l_j}},\hat{W}_j,\check{X}_{\bar{n}_{l_j}},\hat{W}_j)} \xrightarrow[j\to\infty]{w} \hat{P}_{(\hat{X},\hat{W},\hat{X},\hat{W})}$$

such that μ is supported on the diagonal since $\mu = \hat{P}_{(\hat{X},\hat{W},\hat{X},\hat{W})}$ according to (3.17). The application of Lemma 1.14 yields

$$(X_n, W) \xrightarrow[n\to\infty]{p} (X, Z)$$

for stochastic processes X and Z taking values in $C\big([0,T];\mathbb{R}^d\big)$ and $C\big([0,T];\mathbb{R}^{d_1}\big)$, respectively. Note that by Lemma A.15 ii) from the Appendix we can conclude that $Z = W$ $\hat{P}$-a.s. and obtain furthermore $X_n \xrightarrow[j\to\infty]{p} X$. Hence, we have proved assertion 1) of the theorem.

Finally, we will show that the process X satisfies the SDE (2.1). Since we know that every subsequence of $(X_n)_{n\in\mathbb{N}}$ converges in probability to X, we can conclude that the joint process $(X_{n_{l_j}}, W, X_{\bar{n}_{l_j}}, W)$ also converges in probability to (X,W,X,W) (cf. Lemma A.13 in the Appendix). Hence, $(X_{n_{l_j}}, W, X_{\bar{n}_{l_j}}, W)$ converges in distribution (cf. Lemma A.12 ii) in the Appendix) such that we can identify the limit μ as the distribution of (X,W,X,W). Therefore, for any $T \geq 0$ we can define the set

$$U_3 := \Bigg\{ (u,v,w,x) \in C\big([0,T];\mathbb{R}^{2(d+d_1)}\big) \,\Bigg|\, \forall t \in [0,T] : \\ u(t) = u(0) + \int_0^t b\big(s,u(s)\big)\,\mathrm{d}s + \int_0^t \sigma\big(s,u(s)\big)\,\mathrm{d}v(s) \Bigg\}$$

for which we have by (3.17) that

$$P_{(X,W,X,W)}\big[U_3\big] = \mu\big[U_3\big] = \hat{P}_{(\hat{X},\hat{W},\hat{X},\hat{W})}\big[U_3\big] = 1$$

holds since $(\hat{X}(t))_{t\geq 0}$ satisfies $\hat{P}$-a.s. the equation (2.1) with $\hat{W}$ instead of W on $[0,\infty[$. Hence, we conclude that the stochastic process $(X(t))_{t\geq 0}$ satisfies P-a.s. the SDE (2.1) on $[0, \infty[$. The $(\mathcal{F}_t)$-adaptedness of X follows by construction of the *Euler "polygonal" approximations* because the processes $(X_n(t))_{t\geq 0}$ are $(\mathcal{F}_t)$-adapted (cf. **Claim** (2-2), **(1)**) for every $n \in \mathbb{N}$.

The uniqueness asserted in 2) follows directly from the assumed pathwise uniqueness. $\square$

We finish this chapter with a remark on the application of Theorem 3.7 in the case $D = \mathbb{R}^d$ and, as mentioned before, Corollary 3.8 about assuming the local weak monotonicity due to which we can drop the pathwise uniqueness and the continuity of σ in this case.

Remark (cf. [GK96], Remark 2.5 on page 148)**.** In the case $D := \mathbb{R}^d$ and $D_k := \big\{x \in \mathbb{R}^d \,\big|\, \|x\|_{\mathbb{R}^d} < k\big\}$ we can restate the assumptions **A1)** and **A2)** by taking

$$V(t,x) := (1 + \|x\|^2_{\mathbb{R}^d}) \exp\Big(-\int_0^t M(s)\,\mathrm{d}s\Big)$$

as

A1') $\sup\limits_{x\in \overline{D}_k} \|b(t,x)\|_{\mathbb{R}^d} + \|\sigma(t,x)\|^2_{L_2} \leq M_k(t)$ for every $t \geq 0$ and $k \in \mathbb{N}$,

A2') $2x\,b(t,x) + \|\sigma(t,x)\|^2_{L_2} \leq M(t)(1 + \|x\|^2_{\mathbb{R}^d})$ for every $t \geq 0$ and $x \in \mathbb{R}^d$.

Corollary 3.8 (cf. [GK96], Corollary 2.6 on page 148)**.** *Let the assumptions from Section 2.1 be fulfilled. Suppose moreover that b and σ satisfy the local weak monotonicity condition on D, i.e.*

$$2(x-y)\big(b(t,x) - b(t,y)\big) + \|\sigma(t,x) - \sigma(t,y)\|^2_{L_2} \leq M_k(t)\|x - y\|_{\mathbb{R}^d}$$

*for every $k \in \mathbb{N}$ and $t \geq 0$, $x, y \in D_k$. Or in the case $D = \mathbb{R}^d$ we may assume that **A1')** and **A2')** are fulfilled and that the local weak monotonicity condition is satisfied for $D_k = \big\{x \in \mathbb{R}^d \,\big|\, \|x\|_{\mathbb{R}^d} < k\big\}$. Assume moreover that b is continuous in $x \in D$. Then the conclusions of Theorem 3.7 hold.*

Proof. We refer to [GK96], Corollary 2.6 on page 148. $\square$

4. Estimates on the Transition Probability Density

In this fourth chapter we will work on the necessary preparations and estimates for the proof of Theorem 5.2 in Chapter 5. First of all, we have to consider some properties of positive definite matrices in the first section. These properties are important for the very technical proof of Lemma 4.7 in the second section.

We will apply the crucially needed auxiliary estimates from Lemma 4.7 in the third section in the proof of Lemma 4.8, but unfortunately we have to omit certain parts of that proof. The reason for this restriction will be discussed later in Section 4.3 in more detail. However, it is claimed that Lemma 4.8 provides estimates for the transition probability density, which are required in the proof of Theorem 5.2. Both lemmas 4.7 and 4.8 are based on the work of Krylov and Gyöngy in [GK96].

4.1. Properties of positive definite matrices

In this first section we will gather the essential framework concerning positive definite matrices and those of their basic properties that are necessary for the proof of Lemma 4.7 in Section 4.2. We will follow the book [HJ85] of R. Horn and C. Johnson called "Matrix Analysis" and refer to it as the main reference such that we can omit most of the proofs here.

First of all, we recall the notion of self-adjoint and positive (semi-)definite matrices.

Definition 4.1 (adjoint / self-adjoint matrix). *Let $A \in \mathbb{R}^{n\times n}$. Then the adjoint matrix A^* of A is given by $A^* = A^T$. A is called self-adjoint if $A = A^*$.*

By this definition the adjoint matrix A^* has the property $\langle Ax, y\rangle_{\mathbb{R}^n} = \langle x, A^*y\rangle_{\mathbb{R}^n}$ for every $x, y \in \mathbb{R}^n$. Furthermore, note that self-adjoint matrices with $\mathbb{R}$-valued entries are in fact symmetric matrices.

Definition 4.2 (positive (semi-)definite matrix). *Let $A \in \mathbb{R}^{n\times n}$ be a self-adjoint matrix. Then A is called positive definite if*

$$\langle Ax, x\rangle_{\mathbb{R}^n} > 0$$

for every $x \in \mathbb{R}^n$. It is called positive semi-definite if the strict inequality is weakened to

$$\langle Ax, x\rangle_{\mathbb{R}^n} \geq 0$$

for every $x \in \mathbb{R}^n$.

In particular, every eigenvalue of a positive semi-definite matrix is non-negative. Moreover, every eigenvalue of a positive definite matrix $A \in \mathbb{R}^{n\times n}$ is positive such that this also ensures the existence of its inverse matrix A^{-1}.

There is a common approach to define a partial order on the set of all self-adjoint matrices.

Definition 4.3. *Let $n \in \mathbb{N}$ and $\mathbb{R}^{n\times n}_*$ be the space of all self-adjoint matrices. Then define a partial order "$\leq$" on $\mathbb{R}^{n\times n}_*$ by*

$$A \leq B \text{ if and only if } \big\langle (B-A)x, x \big\rangle_{\mathbb{R}^n} \geq 0 \text{ for every } x \in \mathbb{R}^n,$$

where $A, B \in \mathbb{R}^{n\times n}_*$.

Now we can state three lemmas about properties of self-adjoint and positive definite matrices that are important for the upcoming estimates in the proof of Lemma 4.7.

Lemma 4.4 (cf. [HJ85], Corollary 7.7.4 on page 471)**.** *Let $A, B \in \mathbb{R}^{n\times n}$ be self-adjoint and positive definite matrices. Then:*

i) $A \leq B$ if and only if $A^{-1} \geq B^{-1}$,

ii) If $A \leq B$, then $\det A \leq \det B$ and $\operatorname{tr} A \leq \operatorname{tr} B$.

Proof. We refer to [HJ85], Corollary 7.7.4 on page 471. □

Lemma 4.5 (cf. [HJ85], Theorem 4.2.2 on page 176)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a self-adjoint matrix with eigenvalues λ_i, $1 \leq i \leq n$. Then we have*

$$\lambda_A^{min}\, Id \leq A \leq \lambda_A^{max}\, Id,$$

where $\lambda_A^{min} := \min\limits_{1\leq i\leq n} \lambda_i$ and $\lambda_A^{max} := \max\limits_{1\leq i\leq n} \lambda_i$.

Proof. We refer to [HJ85], Theorem 4.2.2 on page 176. □

Lemma 4.6. *Let $A \in \mathbb{R}^{n\times n}$ be a self-adjoint matrix and $c_1, c_2 > 0$ such that*

$$c_1\, Id \leq A \leq c_2\, Id. \tag{4.1}$$

Then the following assertions are fulfilled:

i) The inequality $(c_1)^n \leq \det A \leq (c_2)^n$ holds. In particular, A is invertible.

ii) There exists a self-adjoint invertible matrix $A^{\frac{1}{2}} \in \mathbb{R}^{n\times n}$ such that $A = A^{\frac{1}{2}} A^{\frac{1}{2}}$.

iii) We have $|A_{ij}| \leq c_2$ for every $1 \leq i, j \leq d$.

iv) Consider the map $x \longmapsto \langle A(x-y), x-y\rangle_{\mathbb{R}^n}$. Then for any $1 \leq j \leq n$ we have

$$\frac{\partial}{\partial x^j} \langle A(x-y), x-y\rangle_{\mathbb{R}^n} = 2\big(A(x-y)\big)_j$$

for every $x, y \in \mathbb{R}^n$.

Recall that an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ satisfies the property $QQ^T = \mathrm{Id} = Q^TQ$ by definition.

Proof. "$i)$": This is just an application of Lemma 4.4 ii).
"$ii)$": Since A is symmetric, there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that $A = QDQ^T$ for a diagonal matrix $D \in \mathbb{R}^{n\times n}$. For the diagonal matrix D, which has only positive elements on the main diagonal because of the assumed inequality (4.1), there exists $D^{\frac{1}{2}} \in \mathbb{R}^{n\times n}$ by taking the square root. Hence $A = QD^{\frac{1}{2}}Q^TQD^{\frac{1}{2}}Q^T$ and we can define $A^{\frac{1}{2}} := QD^{\frac{1}{2}}Q^T$. Note that since $\det\left(A^{\frac{1}{2}}\right) = \det(Q)\det\left(D^{\frac{1}{2}}\right)\det(Q)^{-1} = \det\left(D^{\frac{1}{2}}\right) > 0$, the matrix $A^{\frac{1}{2}}$ is invertible. Furthermore, it is also self-adjoint because the symmetry follows from

$$\left(A^{\frac{1}{2}}\right)^T = \left(QD^{\frac{1}{2}}Q^T\right)^T = \left(Q^T\right)^T\left(D^{\frac{1}{2}}\right)^TQ^T = QD^{\frac{1}{2}}Q^T = A^{\frac{1}{2}}.$$

"$iii)$": For the self-adjoint matrix A we compute

$$\begin{aligned}\left\langle Ax, y\right\rangle_{\mathbb{R}^n} &= \frac{1}{2}\Big(\left\langle Ax, x\right\rangle_{\mathbb{R}^n} + \left\langle Ay, y\right\rangle_{\mathbb{R}^n} - \left\langle A(x-y), x-y\right\rangle_{\mathbb{R}^n}\Big)\\ &\leq \frac{c_2}{2}\Big(\|x\|_{\mathbb{R}^n}^2 + \|y\|_{\mathbb{R}^n}^2\Big)\end{aligned}$$

since $0 \leq c_1\|z\|_{\mathbb{R}^n}^2 \leq \left\langle Az, z\right\rangle_{\mathbb{R}^n} \leq c_2\|z\|_{\mathbb{R}^n}^2$ for every $z \in \mathbb{R}^n$ is fulfilled by assumption (4.1). Besides, we have

$$\begin{aligned}\left\langle Ax, y\right\rangle_{\mathbb{R}^n} &= \frac{1}{2}\Big(\left\langle A(x+y), x+y\right\rangle_{\mathbb{R}^n} - \left\langle Ax, x\right\rangle_{\mathbb{R}^n} - \left\langle Ay, y\right\rangle_{\mathbb{R}^n}\Big)\\ &\geq -\frac{c_2}{2}\Big(\|x\|_{\mathbb{R}^n}^2 + \|y\|_{\mathbb{R}^n}^2\Big)\end{aligned}$$

such that altogether the inequality $\left|\left\langle Ax, y\right\rangle_{\mathbb{R}^n}\right| \leq \frac{c_2}{2}\left(\|x\|_{\mathbb{R}^n}^2 + \|y\|_{\mathbb{R}^n}^2\right)$ holds for every $x, y \in \mathbb{R}^n$. Hence, for every $1 \leq i, j \leq n$ we obtain

$$|A_{ij}| = \left|\left\langle Ae_j, e_i\right\rangle_{\mathbb{R}^n}\right| \leq \frac{c_2}{2}\left(\|e_j\|_{\mathbb{R}^n}^2 + \|e_i\|_{\mathbb{R}^n}^2\right) = c_2,$$

where e_i, for $1 \leq i \leq n$, are the canonical basis vectors of $\mathbb{R}^n$.
"$iv)$": We compute

$$\begin{aligned}\frac{\partial}{\partial x^j}\langle A(x-y), x-y\rangle_{\mathbb{R}^n} &= \frac{\partial}{\partial x^j}\sum_{\ell=1}^{n}\big(A(x-y)\big)_\ell\big(x-y\big)_\ell\\ &= \frac{\partial}{\partial x^j}\left(\sum_{\ell=1}^{n}\Big(\sum_{k=1}^{n}A_{\ell k}(x-y)_k\Big)\big(x-y\big)_\ell\right)\\ &= \frac{\partial}{\partial x^j}\left(\sum_{\substack{\ell=1\\ \ell\neq j}}^{n}A_{\ell j}(x-y)_j\big(x-y\big)_\ell\right) + \frac{\partial}{\partial x^j}\left(\sum_{k=1}^{n}A_{jk}(x-y)_k\big(x-y\big)_j\right)\\ &= 2\big(A(x-y)\big)_j\end{aligned}$$

for $1 \leq j \leq n$ and $x, y \in \mathbb{R}^n$ by using the symmetry of A. □

Remark. Let $A \in \mathbb{R}^{n\times n}$ be a self-adjoint and positive definite matrix. Since

$$\det(A^{-1}) = \det(A)^{-1} \text{ and } \det A = \det\left(A^{\frac{1}{2}}A^{\frac{1}{2}}\right) = \det\left(A^{\frac{1}{2}}\right)^2$$

hold, we also have $\det\left(A^{-\frac{1}{2}}\right) = \left(\det A\right)^{-\frac{1}{2}}$.

4.2. Auxiliary estimates

This section contains the auxiliary Lemma 4.7 that we crucially need for the estimates on the transition probability density in Section 4.3. The proof is an extended and modified version of the one given by Gyöngy and Krylov in [GK96]. We laid particular emphasis on the development of the dependencies of every constant by labeling them explicitly with individual numbers in the following. This helps us to observe and ensure that the occurring constants are really independent of t as it is asserted in the lemma.

Lemma 4.7 (cf. [GK96], Lemma 4.1 on page 152)**.** *Let $K, t, \varepsilon > 0$ and $\alpha \in]0,1[$. Furthermore, let $\Lambda\colon\ \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times d}$ be a map such that for every $y \in \mathbb{R}^d$*

$$\varepsilon\, t\ \mathrm{Id} \leq \Lambda(y) = \Lambda(y)^* \leq K\, t\ \mathrm{Id} \tag{4.2}$$

holds and assume that $g\colon\ \mathbb{R}^d \longrightarrow \mathbb{R}$ is an α-Hölder continuous function, i.e.

$$|g(x) - g(y)| \leq K \|x - y\|^{\alpha}_{\mathbb{R}^d}$$

for all $x, y \in \mathbb{R}^d$. Suppose that ξ and η are independent d-dimensional Gaussian vectors on $(\Omega, \mathcal{F}, P)$ with $\xi \sim N(0, \mathrm{Id})$ and $\eta \sim N(0, \Lambda_\eta)$, respectively, where Λ_η is a positive definite covariance matrix. Besides, set λ^{max}_η as the largest and λ^{min}_η as the smallest eigenvalue of the matrix Λ_η. For bounded Borel-measurable functions $f\colon \mathbb{R}^d \longrightarrow \mathbb{R}$ consider the operator T^*, which is defined by

$$T^* f(y) := \mathbb{E}\Big[f\big(y + \Lambda(y)^{\frac{1}{2}}\xi\big)\Big],$$

and let T be the adjoint operator of T^* in $L^2\big(\mathbb{R}^d;\mathbb{R}\big)$-sense.
Then for any $i, j = 1, \dots d$, $x \in \mathbb{R}^d$, $p \in [1,\infty]$ and for any bounded Borel-measurable function $f\colon \mathbb{R}^d \longrightarrow \mathbb{R}$ the inequalities

$$\begin{aligned} &\left| g(x)\, \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T f\Big)(x+\eta)\Big] - \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T(g\, f)\Big)(x+\eta)\Big]\right| \\ &\qquad \leq C_{(4.3)}\, t^{-\frac{d}{2p} - 1 + \frac{\alpha}{2}} \|f\|_{L^p}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} &\left\| g(\cdot)\, \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T f\Big)(\cdot+\eta)\Big] - \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T(g\, f)\Big)(\cdot+\eta)\Big]\right\|_{L^p} \\ &\qquad \leq C_{(4.4)}\, t^{-1 + \frac{\alpha}{2}} \|f\|_{L^p} \end{aligned} \tag{4.4}$$

hold, where the constants $C_{(4.3)} = C_{(4.3)}\big(K, \varepsilon, d, p, \lambda^{min}_\eta, \lambda^{max}_\eta\big)$ and $C_{(4.4)} = C_{(4.4)}\big(K, \varepsilon, d, \lambda^{min}_\eta, \lambda^{max}_\eta\big)$ are independent of t.

Proof. (cf. [GK96], Lemma 4.1 on page 152)
Step 1: At first we calculate the adjoint operator T in $L^2\big(\mathbb{R}^d;\mathbb{R}\big)$-sense. Let $h_1, h_2 \in L^2\big(\mathbb{R}^d;\mathbb{R}\big)$, then we have

$$\begin{aligned} \langle T^* h_1, h_2\rangle_{L^2} &= \int_{\mathbb{R}^d} T^* h_1(y)\, h_2(y)\,\mathrm{d}y = \int_{\mathbb{R}^d}\int_\Omega h_1\big(y + \Lambda(y)^{\frac{1}{2}}\xi\big)\,\mathrm{d}P\, h_2(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_1\big(y + \Lambda(y)^{\frac{1}{2}}x\big)\, h_2(y)\,(2\pi)^{-\frac{d}{2}} \exp\Big(-\frac{1}{2}\langle x, x\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

Now we use the transformation $x \longmapsto \Lambda(y)^{-\frac{1}{2}}(x-y)$ with the Jacobian determinant given by $\det\big(\Lambda(y)^{-\frac{1}{2}}\big)$ and thus

$$\begin{aligned}\langle T^* h_1, h_2\rangle_{L^2} &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_1(x)\, h_2(y)\,(2\pi)^{-\frac{d}{2}}\,\big|\det\big(\Lambda(y)^{-\frac{1}{2}}\big)\big| \\ &\qquad\qquad \cdot \exp\Big(-\frac{1}{2}\langle \Lambda(y)^{-\frac{1}{2}}(x-y), \Lambda(y)^{-\frac{1}{2}}(x-y)\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_1(x)\, h_2(y)\,(2\pi)^{-\frac{d}{2}}\,\big(\det\Lambda(y)\big)^{-\frac{1}{2}} \\ &\qquad\qquad \cdot \exp\Big(-\frac{1}{2}\langle \Lambda(y)^{-1}(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

where we have used that $\det\Lambda(y) \geq 0$, that Lemma 4.6 ii) provides the existence of $\Lambda(y)^{-\frac{1}{2}}$ and that the inverse of a self-adjoint matrix is again self-adjoint.

Note that by assumption (4.2) we have the inequality $\varepsilon\, t \,\mathrm{Id} \leq \Lambda(y) \leq K\, t\,\mathrm{Id}$ and, hence, $\frac{1}{\varepsilon\, t}\,\mathrm{Id} \geq \Lambda(y)^{-1} \geq \frac{1}{K\,t}\,\mathrm{Id}$ by Lemma 4.4 i). We can also apply Lemma 4.6 i) and obtain the estimate

$$(\varepsilon\, t)^d \leq \det\Lambda(y) \leq (K\, t)^d \tag{4.5}$$

for every $y \in \mathbb{R}^d$. Hence, we conclude that

$$\begin{aligned}&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big| h_1(x)\, h_2(y)\,(2\pi)^{-\frac{d}{2}}\,\big(\det\Lambda(y)\big)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\langle \Lambda(y)^{-1}(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\Big|\,\mathrm{d}x\,\mathrm{d}y \\ &\quad \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |h_1(x)|\,|h_2(y)|\,(2\pi)^{-\frac{d}{2}}\,(\varepsilon\, t)^{-\frac{d}{2}} \exp\Big(-\frac{1}{2}\langle \frac{1}{K\,t}\,\mathrm{Id}\,(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad = (2\pi\,\varepsilon\, t)^{-\frac{d}{2}} \int_{\mathbb{R}^d} |h_2(y)| \int_{\mathbb{R}^d} |h_1(x)|\,\exp\Big(-\frac{1}{2\,K\,t}\,\|y-x\|^2_{\mathbb{R}^d}\Big)\,\mathrm{d}x\,\mathrm{d}y\end{aligned}$$

is finite by using the fact that the heat kernel in $\mathbb{R}^d$ is an operator from $L^2\big(\mathbb{R}^d;\mathbb{R}\big)$ to $L^2\big(\mathbb{R}^d;\mathbb{R}\big)$ (cf. [Gri09], Lemma 2.18 on page 41). Therefore, we can apply Fubini's theorem and obtain

$$T h_2(x) = \int_{\mathbb{R}^d} h_2(y)\,(2\pi)^{-\frac{d}{2}}\big(\det\Lambda(y)\big)^{-\frac{1}{2}}\,\exp\Big(-\frac{1}{2}\langle \Lambda(y)^{-1}(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}y.$$

Step 2: In this step we compute the partial derivatives of Tf and $T(g\,f)$.

Claim (1)**.** We have

$$\begin{aligned}&\frac{\partial}{\partial x^j} Tf(x) \\ &\quad = \int_{\mathbb{R}^d} (2\pi)^{-\frac{d}{2}}\big(\det\Lambda(y)\big)^{-\frac{1}{2}}\, f(y)\,\frac{\partial}{\partial x^j}\exp\Big(-\frac{1}{2}\langle \Lambda(y)^{-1}(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}y\end{aligned}$$

for every $x \in \mathbb{R}^d$.

Proof of Claim (1). Following the theory about partial differentiation of a Lebesgue integral with respect to a parameter (see e.g. [AE08], Theorem 3.18 on page 111), we have to verify that there exists a function $h \in L^1(\mathbb{R}^d;\mathbb{R})$ such that the estimate

$$\left| \frac{\partial}{\partial x^j} (2\pi)^{-\frac{d}{2}} \big(\det \Lambda(y)\big)^{-\frac{1}{2}} f(y) \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big)\right| \leq h(y) \tag{4.6}$$

holds for every $y \in \mathbb{R}^d$ and $\xi \in B_\varrho(x) \subseteq \mathbb{R}^d$ for some $\varrho > 0$. Note that

$$\begin{aligned}
&\frac{\partial}{\partial x^j} \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big) \\
&\quad = -\frac{1}{2} \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big) \frac{\partial}{\partial x^j} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d} \\
&\quad = -\frac{1}{2} \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big) 2\big(\Lambda(y)^{-1}(\xi - y)\big)_j
\end{aligned}$$

is fulfilled, where we have used Lemma 4.6 iv) in the last step. Since by assumption (4.2) and Lemma 4.4 i) the inequality $\frac{1}{K\,t}\,\mathrm{Id} \leq \Lambda(y)^{-1} \leq \frac{1}{\varepsilon\, t}\,\mathrm{Id}$ holds, we can apply Lemma 4.6 iii) and obtain

$$\big\|\Lambda(y)^{-1}\big\|_{L_2}^2 = \sum_{i,j=1}^{d} \big|\Lambda_{ij}(y)^{-1}\big|^2 \leq \Big(\frac{d}{\varepsilon\, t}\Big)^2.$$

Hence, we compute

$$\begin{aligned}
&\left|\frac{\partial}{\partial x^j} \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big)\right| \\
&\qquad = \left| \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big) \big(\Lambda(y)^{-1}(\xi - y)\big)_j \right| \\
&\qquad \leq \exp\Big(-\frac{1}{2} \langle \Lambda(y)^{-1}(\xi - y), \xi - y\rangle_{\mathbb{R}^d}\Big) \|\Lambda(y)^{-1}(\xi - y)\|_{\mathbb{R}^d} \\
&\qquad \leq \exp\Big(-\frac{1}{2\,K\,t} \|\xi - y\|_{\mathbb{R}^d}^2\Big) \|\Lambda(y)^{-1}\|_{L_2} \|\xi - y\|_{\mathbb{R}^d}.
\end{aligned} \tag{4.7}$$

Since the triangle inequality yields $-\|\xi - y\|_{\mathbb{R}^d}^2 \leq -\|y\|_{\mathbb{R}^d}^2 + 2\,\|y\|_{\mathbb{R}^d}\|\xi\|_{\mathbb{R}^d} - \|\xi\|_{\mathbb{R}^d}^2$, we have

$$\begin{aligned}
\exp\Big(-\frac{1}{2\,K\,t} \|\xi - y\|_{\mathbb{R}^d}^2\Big) &\leq \exp\left(\frac{1}{2\,K\,t} \Big(-\|y\|_{\mathbb{R}^d}^2 + 2\,\|y\|_{\mathbb{R}^d} \sup_{\xi \in B_\varrho(x)} \|\xi\|_{\mathbb{R}^d}\Big)\right) \\
&= \exp\left(\frac{C_{(4.8)}^2}{2\,K\,t}\right) \exp\left(-\frac{\big(\|y\|_{\mathbb{R}^d} - C_{(4.8)}\big)^2}{2\,K\,t}\right),
\end{aligned} \tag{4.8}$$

where $C_{(4.8)} := \sup_{\xi \in B_\varrho(x)} \|\xi\|_{\mathbb{R}^d}$ is a finite constant.

Then by using (4.7), the integrability of the function $z \longmapsto z\exp(-z^2)$ for $z \geq 0$ and the assumed boundedness of f, we only have to conclude from (4.5) that $\big(\det \Lambda(y)\big)^{-\frac{1}{2}}$ is bounded in order to estimate the term in (4.6). Hence, we can find an integrable dominating function h. $\square$

Moreover, we have to consider the case where $g\,f$ instead of f is given. Therefore, we need the α-Hölder continuity of g which implies

$$|g(y)| \leq K\,\|x-y\|_{\mathbb{R}^d}^{\alpha} + |g(x)|$$

because we have $|g(y)| - |g(x)| \leq |g(y) - g(x)| \leq K\,\|x-y\|_{\mathbb{R}^d}^{\alpha}$ for every $x, y \in \mathbb{R}^d$. Hence, we can use the integrability of the function $z \longmapsto z^{1+\alpha}\exp(-z^2)$ for $z \geq 0$ in addition to the arguments from above.

For the second partial derivative we apply the analogous argumentation from the proof of **Claim** (1). Hence, we obtain

$$\begin{aligned}
&\frac{\partial^2}{\partial x^i \partial x^j} Tf(x) \\
&= \int_{\mathbb{R}^d} (2\pi)^{-\frac{d}{2}} \big(\det \Lambda(y)\big)^{-\frac{1}{2}} f(y)\, \frac{\partial^2}{\partial x^i \partial x^j} \exp\Big(-\frac{1}{2}\langle \Lambda(y)^{-1}(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^d} (2\pi)^{-\frac{d}{2}} \big(\det \Lambda(y)\big)^{-\frac{1}{2}} f(y)\, \frac{\partial^2}{\partial y^i \partial y^j} \exp\Big(-\frac{1}{2}\langle \Lambda^{-1}(y-x), y-x\rangle_{\mathbb{R}^d}\Big)\Big|_{\Lambda=\Lambda(y)}\,\mathrm{d}y,
\end{aligned}$$

where considering the second partial derivative with respect to y turns out to be helpful in **Step 3**. For $\frac{\partial^2}{\partial x^i \partial x^j} T(g\,f)(x)$ we can also repeat the arguments from above.

Step 3: In this step we compute $\mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} Tf\Big)(x+\eta)\Big]$.

Using the calculation from **Step 2** yields

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} Tf\Big)(x+\eta)\Big] \\
&\quad= \mathbb{E}\Bigg[\int_{\mathbb{R}^d} (2\pi)^{-\frac{d}{2}} \big(\det \Lambda(y)\big)^{-\frac{1}{2}} f(y) \\
&\qquad\qquad \cdot \frac{\partial^2}{\partial y^i \partial y^j} \exp\Big(-\frac{1}{2}\langle \Lambda^{-1}(y-(x+\eta)), y-(x+\eta)\rangle_{\mathbb{R}^d}\Big)\Big|_{\Lambda=\Lambda(y)}\,\mathrm{d}y\Bigg] \qquad (4.9)\\
&\quad= \int_{\mathbb{R}^d} f(y)\, \frac{\partial^2}{\partial y^i \partial y^j}\, \mathbb{E}\Big[(2\pi)^{-\frac{d}{2}} \big(\det \Lambda\big)^{-\frac{1}{2}} \\
&\qquad\qquad \cdot \exp\Big(-\frac{1}{2}\langle \Lambda^{-1}(y-(x+\eta)), y-(x+\eta)\rangle_{\mathbb{R}^d}\Big)\Big]\Big|_{\Lambda=\Lambda(y)}\,\mathrm{d}y,
\end{aligned}$$

where the interchange of the expectation and the integral follows from Fubini's theorem and the one with the partial derivative results from the same arguments as in **Step 2**.

Since $\Lambda^{\frac{1}{2}}\xi \sim N(0,\Lambda)$ and $\eta \sim N(0,\Lambda_\eta)$ are independent Gaussian vectors, we have

$$\begin{aligned}
&\mathbb{E}\Big[(2\pi)^{-\frac{d}{2}}\big(\det\Lambda\big)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\langle\Lambda^{-1}(y-x-\eta),y-x-\eta\rangle_{\mathbb{R}^d}\Big)\Big]\\
&\qquad=\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}\big(\det\Lambda\big)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\langle\Lambda^{-1}(y-x-z),y-x-z\rangle_{\mathbb{R}^d}\Big)\\
&\qquad\qquad\cdot(2\pi)^{-\frac{d}{2}}\,(\det\Lambda_\eta)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\langle\Lambda_\eta^{-1}z,z\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}z\\
&\qquad=(2\pi)^{-\frac{d}{2}}\,\big(\det(\Lambda+\Lambda_\eta)\big)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\langle(\Lambda+\Lambda_\eta)^{-1}(y-x),y-x\rangle_{\mathbb{R}^d}\Big),
\end{aligned}$$

where we have used the representation of the convolution of two probability density functions belonging to independent normal-distributed random variables in the last step (cf. [Bau02], Theorem 8.4 on page 55). Define

$$\begin{aligned}
p_A(x,y):=(2\pi)^{-\frac{d}{2}}\,&\big(\det(A+\Lambda_\eta)\big)^{-\frac{1}{2}}\\
&\cdot\exp\Big(-\frac{1}{2}\langle(A+\Lambda_\eta)^{-1}(y-x),y-x\rangle_{\mathbb{R}^d}\Big)
\end{aligned}\tag{4.10}$$

for any positive definite matrix $A\in\mathbb{R}^{d\times d}$. Hence, plugging the definition (4.10) into equation (4.9) yields

$$\mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i\partial x^j}Tf\Big)(x+\eta)\Big]=\int_{\mathbb{R}^d}f(y)\,\frac{\partial^2}{\partial y^i\partial y^j}\,p_\Lambda(x,y)\Big|_{\Lambda=\Lambda(y)}\mathrm{d}y.$$

Let $\Theta^{(y)}:=\big(\Lambda(y)+\Lambda_\eta\big)^{-1}$. Now we compute by using Lemma 4.6 iv)

$$\begin{aligned}
\frac{\partial}{\partial y^j}\,p_\Lambda(x,y)\Big|_{\Lambda=\Lambda(y)}&=p_{\Lambda(y)}(x,y)\cdot\frac{\partial}{\partial y^j}\Big(-\frac{1}{2}\langle\Theta(y-x),y-x\rangle_{\mathbb{R}^d}\Big)\Big|_{\Theta=\Theta^{(y)}}\\
&=-\frac{1}{2}\,p_{\Lambda(y)}(x,y)\cdot\Big(2\,\big(\Theta(y-x)\big)_j\Big)\Big|_{\Theta=\Theta^{(y)}}
\end{aligned}$$

since $\Theta^{(y)}$ is symmetric. We obtain

$$\begin{aligned}
&\frac{\partial^2}{\partial y^i\partial y^j}\,p_\Lambda(x,y)\Big|_{\Lambda=\Lambda(y)}\\
&\qquad=\frac{\partial}{\partial y^i}\Big(-p_\Lambda(x,y)\,\big(\Theta(y-x)\big)_j\Big)\Big|_{\substack{\Lambda=\Lambda(y)\\ \Theta=\Theta^{(y)}}}\\
&\qquad=-p_{\Lambda(y)}(x,y)\cdot\Theta^{(y)}_{ij}+\big(-\Theta^{(y)}(y-x)\big)_j\cdot\Big(-p_{\Lambda(y)}(x,y)\cdot\big(\Theta^{(y)}(y-x)\big)_i\Big)\\
&\qquad=p_{\Lambda(y)}(x,y)\cdot\Big(\big(\Theta^{(y)}(y-x)\big)_i\big(\Theta^{(y)}(y-x)\big)_j-\Theta^{(y)}_{ij}\Big).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i\partial x^j}Tf\Big)(x+\eta)\Big]&=\int_{\mathbb{R}^d}f(y)\,\frac{\partial^2}{\partial y^i\partial y^j}\,p_\Lambda(x,y)\Big|_{\Lambda=\Lambda(y)}\mathrm{d}y\\
&=\int_{\mathbb{R}^d}f(y)\,p_{\Lambda(y)}(x,y)\,\Big(\big(\Theta^{(y)}(y-x)\big)_i\big(\Theta^{(y)}(y-x)\big)_j-\Theta^{(y)}_{ij}\Big)\,\mathrm{d}y.
\end{aligned}$$

Step 4: Next we show that

$$|p_{\Lambda(y)}(x,y)| \le (2\pi)^{-\frac{d}{2}} \exp\Big(- \frac{1}{2(Kt+\lambda_\eta^{max})}\|y-x\|^2_{\mathbb{R}^d}\Big) \Big(\frac{1}{\varepsilon\, t + \lambda_\eta^{min}}\Big)^{\frac{d}{2}}$$

holds, where λ_η^{min} and λ_η^{max} are the smallest and biggest, respectively, eigenvalue of the matrix Λ_η. To do this, we will prove the following two claims.

Claim (2)**.** We have

$$0 < \lambda_\eta^{min}\,\mathrm{Id} \le \Lambda_\eta \le \lambda_\eta^{max}\,\mathrm{Id}\,. \tag{4.11}$$

Proof of Claim (2). By Lemma 4.5 we obtain $\lambda_\eta^{min}\,\mathrm{Id} \le \Lambda_\eta \le \lambda_\eta^{max}\,\mathrm{Id}$. Since the covariance matrix Λ_η is positive definite by assumption, we also know that $\lambda_\eta^{min} > 0$. □

Claim (3)**.** We have

$$-\frac{1}{2}\langle \Theta^{(y)}(y-x), y-x\rangle_{\mathbb{R}^d} \le -\frac{1}{2(Kt+\lambda_\eta^{max})}\,\|y-x\|^2_{\mathbb{R}^d}$$

for every $x, y \in \mathbb{R}^d$.

Proof of Claim (3). By (4.2) the inequality $\varepsilon\, t\,\mathrm{Id} \le \Lambda(y) \le K\, t\,\mathrm{Id}$ holds. Hence, by using the estimate (4.11) for Λ_η from **Claim** (2), we obtain

$$\big(\varepsilon\, t + \lambda_\eta^{min}\big)\,\mathrm{Id} \le \Lambda(y) + \Lambda_\eta \le \big(K\,t + \lambda_\eta^{max}\big)\,\mathrm{Id}\,.$$

Therefore, according to Lemma 4.4 i)

$$\frac{1}{\varepsilon\, t + \lambda_\eta^{min}}\,\mathrm{Id} \ge \underbrace{\big(\Lambda(y)+\Lambda_\eta\big)^{-1}}_{=\Theta^{(y)}} \ge \frac{1}{K\,t+\lambda_\eta^{max}}\,\mathrm{Id} \tag{4.12}$$

is fulfilled and, hence,

$$-\frac{1}{2}\langle \Theta^{(y)}(y-x), y-x\rangle_{\mathbb{R}^d} \le -\frac{1}{2}\Big\langle \frac{1}{K\,t+\lambda_\eta^{max}}\,\mathrm{Id}\,(y-x), y-x\Big\rangle_{\mathbb{R}^d} = -\frac{\|y-x\|^2_{\mathbb{R}^d}}{2\big(K\,t+\lambda_\eta^{max}\big)}$$

follows. □

Thus we conclude

$$\begin{aligned}
&|p_{\Lambda(y)}(x,y)|\\
&\quad= \Big|(2\pi)^{-\frac{d}{2}}\,\big(\det(\Lambda(y)+\Lambda_\eta)\big)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\langle\big(\Lambda(y)+\Lambda_\eta\big)^{-1}(y-x), y-x\rangle_{\mathbb{R}^d}\Big)\Big|\\
&\quad= (2\pi)^{-\frac{d}{2}}\,\exp\Big(-\frac{1}{2}\langle\Theta^{(y)}(y-x), y-x\rangle_{\mathbb{R}^d}\Big)\,\Big|\big(\det\Theta^{(y)}\big)^{\frac{1}{2}}\Big|\\
&\quad\le (2\pi)^{-\frac{d}{2}}\,\exp\Big(-\frac{1}{2(Kt+\lambda_\eta^{max})}\|y-x\|^2_{\mathbb{R}^d}\Big)\,\big(\det\Theta^{(y)}\big)^{\frac{1}{2}}\\
&\quad\le (2\pi)^{-\frac{d}{2}}\,\exp\Big(-\frac{1}{2(Kt+\lambda_\eta^{max})}\|y-x\|^2_{\mathbb{R}^d}\Big)\,\Big(\frac{1}{\varepsilon\, t+\lambda_\eta^{min}}\Big)^{\frac{d}{2}}
\end{aligned}$$

since $0 \le \big(\det\Theta^{(y)}\big)^{\frac{1}{2}} \le \Big(\frac{1}{\varepsilon\, t+\lambda_\eta^{min}}\Big)^{\frac{d}{2}}$ holds by estimate (4.12) and Lemma 4.6 i).

Step 5: In this step we prove inequality (4.3).

Observe that

$$\begin{aligned}
&\Big|g(x)\,\mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T f\Big)(x+\eta)\Big] - \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T(g\,f)\Big)(x+\eta)\Big]\Big| \\
&\quad = \Big|\int_{\mathbb{R}^d} \big(g(x)-g(y)\big)\, f(y)\, p_{\Lambda(y)}(x,y)\, \Big(\big(\Theta^{(y)}(y-x)\big)_i \big(\Theta^{(y)}(y-x)\big)_j - \Theta^{(y)}_{ij}\Big)\, \mathrm{d}y\Big| \\
&\quad \le \int_{\mathbb{R}^d} \underbrace{|g(x)-g(y)|}_{\le K\, \|x-y\|^{\alpha}_{\mathbb{R}^d}}\, |f(y)|\, |p_{\Lambda(y)}(x,y)|\, \Big|\big(\Theta^{(y)}(y-x)\big)_i \big(\Theta^{(y)}(y-x)\big)_j - \Theta^{(y)}_{ij}\Big|\, \mathrm{d}y
\end{aligned}$$

is fulfilled. Since we have (4.12) and Lemma 4.6 iii), it follows that the inequalities

$$\big|\Theta^{(y)}_{ij}\big| \le \frac{1}{\varepsilon\, t + \lambda^{min}_{\eta}}$$

and

$$\begin{aligned}
\Big|\big(\Theta^{(y)}(y-x)\big)_i \big(\Theta^{(y)}(y-x)\big)_j\Big| &= \Big|\Big(\sum_{l=1}^{d} \Theta^{(y)}_{il}(y-x)_l\Big)\Big(\sum_{\ell=1}^{d} \Theta^{(y)}_{j\ell}(y-x)_\ell\Big)\Big| \\
&\le \Big(\sum_{l=1}^{d} \big|\Theta^{(y)}_{il}\big|\big|(y-x)_l\big|\Big)\Big(\sum_{\ell=1}^{d} \big|\Theta^{(y)}_{j\ell}\big|\big|(y-x)_\ell\big|\Big) \\
&\le \frac{1}{(\varepsilon\, t + \lambda^{min}_{\eta})^2}\Big(\sum_{l=1}^{d} \big|(y-x)_l\big|\Big)^2 \\
&\le \frac{1}{(\varepsilon\, t + \lambda^{min}_{\eta})^2}\Big(\sum_{l=1}^{d} (y-x)_l^2\Big)\Big(\sum_{l=1}^{d} 1^2\Big) \\
&= \frac{d}{(\varepsilon\, t + \lambda^{min}_{\eta})^2}\|y-x\|^2_{\mathbb{R}^d}
\end{aligned}$$

hold for every $1 \le i,j \le d$, where we have used the Cauchy-Schwarz inequality in the second last step. Hence, we get

$$\Big|\big(\Theta^{(y)}(y-x)\big)_i \big(\Theta^{(y)}(y-x)\big)_j - \Theta^{(y)}_{ij}\Big| \le \frac{d}{(\varepsilon\, t + \lambda^{min}_{\eta})^2}\|y-x\|^2_{\mathbb{R}^d} + \frac{1}{\varepsilon\, t + \lambda^{min}_{\eta}}.$$

With this calculation and the estimate for $|p_{\Lambda(y)}(x,y)|$ from **Step 4** we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^d} K\,\|y-x\|_{\mathbb{R}^d}^{\alpha}\,|f(y)|\,|p_{\Lambda(y)}(x,y)| \\
&\qquad\qquad\qquad \cdot \Big|\big(\Theta^{(y)}(y-x)\big)_i\big(\Theta^{(y)}(y-x)\big)_j - \Theta^{(y)}_{ij}\Big|\,\mathrm{d}y \\
&\le \frac{K}{(2\pi)^{\frac{d}{2}}(\varepsilon\, t+\lambda_\eta^{min})^{\frac{d}{2}}}\int_{\mathbb{R}^d}|f(y)|\,\|y-x\|_{\mathbb{R}^d}^{\alpha}\bigg(\frac{d\,\|y-x\|_{\mathbb{R}^d}^2}{(\varepsilon\, t+\lambda_\eta^{min})^2}+\frac{1}{\varepsilon\, t+\lambda_\eta^{min}}\bigg) \\
&\qquad\qquad\qquad \cdot \exp\Big(-\frac{1}{2(Kt+\lambda_\eta^{max})}\|y-x\|_{\mathbb{R}^d}^2\Big)\,\mathrm{d}y \qquad (4.13)\\
&= C_{(4.13)}\,(\varepsilon\, t+\lambda_\eta^{min})^{-\frac{d}{2}-1}\int_{\mathbb{R}^d}|f(y)|\,\|y-x\|_{\mathbb{R}^d}^{\alpha}\bigg(\frac{d\,\|y-x\|_{\mathbb{R}^d}^2}{\varepsilon\, t+\lambda_\eta^{min}}+1\bigg) \\
&\qquad\qquad\qquad \cdot \exp\Big(-\frac{\|y-x\|_{\mathbb{R}^d}^2}{2(Kt+\lambda_\eta^{max})}\Big)\,\mathrm{d}y,
\end{aligned}
$$

where $C_{(4.13)} := K\,(2\pi)^{-\frac{d}{2}}$. Now, in order to simplify the occurring constants, we estimate the latter by

$$
\begin{aligned}
&C_{(4.13)}\,\big(C_{(4.14)}^{min}\,(t+1)\big)^{-\frac{d}{2}-1}\int_{\mathbb{R}^d}|f(y)|\,\|y-x\|_{\mathbb{R}^d}^{\alpha}\bigg(\frac{d\,\|y-x\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{min}\,(t+1)}+1\bigg) \\
&\qquad\qquad\qquad \cdot \exp\Big(-\frac{\|y-x\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{max}\,(t+1)}\Big)\,\mathrm{d}y, \qquad (4.14)
\end{aligned}
$$

where $C_{(4.14)}^{min} := \min\big\{\varepsilon, \lambda_\eta^{min}\big\}$ and $C_{(4.14)}^{max} := 2\,\max\big\{K, \lambda_\eta^{max}\big\}$ are constants.

Applying Hölder's inequality for $p \in\,]1,\infty]$ and its conjugate $q := \frac{p}{p-1} \in [1,\infty[$ yields

$$
\begin{aligned}
&C_{(4.13)}\,\big(C_{(4.14)}^{min}\,(t+1)\big)^{-\frac{d}{2}-1}\int_{\mathbb{R}^d}|f(y)|\,\|y-x\|_{\mathbb{R}^d}^{\alpha}\bigg(\frac{d\,\|y-x\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{min}\,(t+1)}+1\bigg) \\
&\qquad\qquad\qquad \cdot \exp\Big(-\frac{\|y-x\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{max}\,(t+1)}\Big)\,\mathrm{d}y \\
&\le C_{(4.13)}\,\big(C_{(4.14)}^{min}\,(t+1)\big)^{-\frac{d}{2}-1}\,\|f\|_{L^p}\Bigg(\int_{\mathbb{R}^d}\|y-x\|_{\mathbb{R}^d}^{q\alpha}\bigg(\frac{d\,\|y-x\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{min}\,(t+1)}+1\bigg)^q \\
&\qquad\qquad\qquad \cdot \exp\Big(-\frac{q\,\|y-x\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{max}\,(t+1)}\Big)\,\mathrm{d}y\Bigg)^{\frac{1}{q}} \\
&= C_{(4.13)}\,\big(C_{(4.14)}^{min}\,(t+1)\big)^{-\frac{d}{2}-1}\,\|f\|_{L^p}\Bigg(\int_{\mathbb{R}^d}\|y\|_{\mathbb{R}^d}^{q\alpha}\bigg(\frac{d\,\|y\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{min}\,(t+1)}+1\bigg)^q \\
&\qquad\qquad\qquad \cdot \exp\Big(-\frac{q\|y\|_{\mathbb{R}^d}^2}{C_{(4.14)}^{max}\,(t+1)}\Big)\,\mathrm{d}y\Bigg)^{\frac{1}{q}}
\end{aligned}
$$

by shifting the integral via $y \longmapsto y+x$ in the last step. Next we use the transformation

$y \longmapsto (t+1)^{\frac{1}{2}} y$ with the Jacobian determinant given by $(t+1)^{\frac{d}{2}}$ and obtain

$$
\begin{aligned}
& C_{(4.13)} \left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1} (t+1)^{-\frac{d}{2}-1} \left\| f \right\|_{L^p} \\
& \qquad \cdot \left(\int_{\mathbb{R}^d} (t+1)^{\frac{d}{2}} \left\| y \right\|^{q\alpha}_{\mathbb{R}^d} (t+1)^{\frac{q\alpha}{2}} \left(\frac{d \left\| y \right\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}} + 1 \right)^q \exp\left(- \frac{q \left\| y \right\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}} \right) \mathrm{d}y \right)^{\frac{1}{q}} \\
& \quad = \underbrace{(t+1)^{-\frac{d}{2}-1+\frac{\alpha}{2}+\frac{d}{2q}}}_{=(t+1)^{-\frac{d}{2p}-1+\frac{\alpha}{2}}} \left\| f \right\|_{L^p} C_{(4.3)},
\end{aligned}
$$

where the constant

$$
C_{(4.3)} := C_{(4.13)} \left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1} \left(\int_{\mathbb{R}^d} \left\| y \right\|^{q\alpha}_{\mathbb{R}^d} \left(\frac{d \left\| y \right\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}} + 1 \right)^q \exp\left(- \frac{q \left\| y \right\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}} \right) \mathrm{d}y \right)^{\frac{1}{q}}
$$

depends on $d, \varepsilon, K, \lambda^{min}_{\eta}, \lambda^{max}_{\eta}$ and p. Since the power $-\frac{d}{2p} - 1 + \frac{\alpha}{2}$ of $(t+1)$ is negative, we have

$$
(t+1)^{-\frac{d}{2p}-1+\frac{\alpha}{2}} \leq t^{-\frac{d}{2p}-1+\frac{\alpha}{2}}
$$

and, therefore, we obtain the inequality (4.3).

Moreover, in the case $p = 1$ we obtain from (4.14) by using the transformation $y \longmapsto (t+1)^{\frac{1}{2}} y + x$ with the Jacobian determinant given by $(t+1)^{\frac{d}{2}}$

$$
\begin{aligned}
& C_{(4.13)} \left(C^{min}_{(4.14)} (t+1)\right)^{-\frac{d}{2}-1} \int_{\mathbb{R}^d} \left| f(y) \right| \left\| y - x \right\|^{\alpha}_{\mathbb{R}^d} \left(\frac{d \left\| y - x \right\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)} (t+1)} + 1 \right) \\
& \qquad \qquad \cdot \exp\left(- \frac{\left\| y - x \right\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)} (t+1)} \right) \mathrm{d}y \\
& \quad = C_{(4.13)} \left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1} (t+1)^{-1+\frac{\alpha}{2}} \int_{\mathbb{R}^d} \left| f\left((t+1)^{\frac{1}{2}} y + x\right) \right| \\
& \qquad \qquad \cdot \left\| y \right\|^{\alpha}_{\mathbb{R}^d} \left(\frac{d \left\| y \right\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}} + 1 \right) \exp\left(- \frac{\left\| y \right\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}} \right) \mathrm{d}y.
\end{aligned}
$$

An application of Hölder's inequality yields

$$
\begin{aligned}
& C_{(4.13)} \left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1} (t+1)^{-1+\frac{\alpha}{2}} \int_{\mathbb{R}^d} \left| f\left((t+1)^{\frac{1}{2}} y + x\right) \right| \\
& \qquad \qquad \cdot \left\| y \right\|^{\alpha}_{\mathbb{R}^d} \left(\frac{d \left\| y \right\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}} + 1 \right) \exp\left(- \frac{\left\| y \right\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}} \right) \mathrm{d}y \\
& \quad \leq C_{(4.13)} \left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1} (t+1)^{-1+\frac{\alpha}{2}} \int_{\mathbb{R}^d} \left| f\left((t+1)^{\frac{1}{2}} y + x\right) \right| \mathrm{d}y \\
& \qquad \qquad \cdot \left\| \left\| \cdot \right\|^{\alpha}_{\mathbb{R}^d} \left(\frac{d \left\| \cdot \right\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}} + 1 \right) \exp\left(- \frac{\left\| \cdot \right\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}} \right) \right\|_{L^\infty}.
\end{aligned}
$$

Hence, we just have to consider the reversing transformation $y \longmapsto (t+1)^{-\frac{1}{2}}(y-x)$ with the Jacobian determinant given by $(t+1)^{-\frac{d}{2}}$ for the integral

$$\int_{\mathbb{R}^d} \Big| f\big((t+1)^{\frac{1}{2}} y + x\big)\Big| \,\mathrm{d}y$$

to obtain the term

$$\begin{aligned} &C_{(4.13)} \left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1} (t+1)^{-1+\frac{\alpha}{2}-\frac{d}{2}} \, \|f\|_{L^1} \\ &\quad \cdot \left\| \|\cdot\|^{\alpha}_{\mathbb{R}^d} \left(\frac{d \, \|\cdot\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}} + 1 \right) \exp\Big(- \frac{\|\cdot\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}} \Big) \right\|_{L^\infty} \end{aligned}$$

as an estimate such that inequality (4.3) follows.
Step 6: Finally, inequality (4.4) results from the calculation in **Step 5** if we use the generalised Minkowski inequality (cf. Theorem A.6 in the Appendix) for integrals instead of Hölder's inequality.

Therefore, we have for $p \in [1, \infty[$

$$\begin{aligned} &\left\| g(\cdot)\, \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T f \Big)(\cdot + \eta) \Big] - \mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x^i \partial x^j} T(g\, f) \Big)(\cdot + \eta) \Big] \right\|_{L^p} \\ &\leq C_{(4.13)} \big(C^{min}_{(4.14)}\,(t+1)\big)^{-\frac{d}{2}-1} \Bigg(\int_{\mathbb{R}^d} \bigg(\int_{\mathbb{R}^d} |f(y)| \, \|y-x\|^{\alpha}_{\mathbb{R}^d} \Big(\frac{d\, \|y-x\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}\,(t+1)} + 1 \Big) \\ &\qquad \cdot \exp\Big(- \frac{\|y-x\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}\,(t+1)} \Big) \,\mathrm{d}y \bigg)^p \,\mathrm{d}x \Bigg)^{\frac{1}{p}} \\ &\leq C_{(4.13)} \big(C^{min}_{(4.14)}\,(t+1)\big)^{-\frac{d}{2}-1} \int_{\mathbb{R}^d} \bigg(\int_{\mathbb{R}^d} \bigg(\|y-x\|^{\alpha}_{\mathbb{R}^d} \Big(\frac{d\, \|y-x\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}\,(t+1)} + 1 \Big) \\ &\qquad \cdot \exp\Big(- \frac{\|y-x\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}\,(t+1)} \Big) \bigg)^p \,\mathrm{d}x \bigg)^{\frac{1}{p}} \,\mathrm{d}y \end{aligned}$$

by using (4.14) in the first and the generalised Minkowski inequality for integrals in the second step. Shifting the integral via $y \longmapsto y + x$, where the Jacobian determinant is given by 1, yields

$$\begin{aligned} &C_{(4.13)} \big(C^{min}_{(4.14)}\,(t+1)\big)^{-\frac{d}{2}-1} \|f\|_{L^p} \int_{\mathbb{R}^d} \|y\|^{\alpha}_{\mathbb{R}^d} \Big(\frac{d\, \|y\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}\,(t+1)} + 1 \Big) \\ &\qquad \cdot \exp\Big(- \frac{\|y\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}\,(t+1)} \Big) \,\mathrm{d}y \end{aligned}$$

since $\|f\|_{L^p} = \int_{\mathbb{R}^d} |f(y+x)| \,\mathrm{d}x$. Again by using the transformation $y \longmapsto (t+1)^{\frac{1}{2}} y$

with the Jacobian determinant given by $(t+1)^{\frac{d}{2}}$, we obtain

$$C_{(4.13)}\left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1}(t+1)^{-\frac{d}{2}-1}\|f\|_{L^p}\int_{\mathbb{R}^d}\|y\|^{\alpha}_{\mathbb{R}^d}\left(\frac{d\,\|y\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}(t+1)}+1\right)\cdot\exp\left(-\frac{\|y\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}(t+1)}\right)\mathrm{d}y$$

$$\begin{aligned}&=(t+1)^{-1+\frac{\alpha}{2}}\|f\|_{L^p}\\&\qquad\cdot\underbrace{C_{(4.13)}\left(C^{min}_{(4.14)}\right)^{-\frac{d}{2}-1}\int_{\mathbb{R}^d}\|y\|^{\alpha}_{\mathbb{R}^d}\left(\frac{d\,\|y\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}}+1\right)\exp\left(-\frac{\|y\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}}\right)\mathrm{d}y}_{=:C_{(4.4)}=C_{(4.4)}\left(K,\varepsilon,d,\lambda^{min}_{\eta},\lambda^{max}_{\eta}\right)}\\&\le t^{-1+\frac{\alpha}{2}}\|f\|_{L^p}\,C_{(4.4)}.\end{aligned}$$

For $p=\infty$ we have again by (4.14) and Theorem A.6 the estimate

$$\begin{aligned}&\left\|g(\cdot)\,\mathbb{E}\left[\left(\frac{\partial^2}{\partial x^i\partial x^j}Tf\right)(\cdot+\eta)\right]-\mathbb{E}\left[\left(\frac{\partial^2}{\partial x^i\partial x^j}T(g\cdot f)\right)(\cdot+\eta)\right]\right\|_{L^\infty}\\&\le\left\|C_{(4.13)}\left(C^{min}_{(4.14)}(t+1)\right)^{-\frac{d}{2}-1}\int_{\mathbb{R}^d}|f(\cdot)|\,\|\cdot-x\|^{\alpha}_{\mathbb{R}^d}\left(\frac{d\,\|\cdot-x\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}(t+1)}+1\right)\right.\\&\qquad\qquad\left.\cdot\exp\left(-\frac{\|\cdot-x\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}(t+1)}\right)\mathrm{d}y\right\|_{L^\infty}\\&\le C_{(4.13)}\left(C^{min}_{(4.14)}(t+1)\right)^{-\frac{d}{2}-1}\operatorname*{ess\,sup}_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d}|f(y)|\,\|y-x\|^{\alpha}_{\mathbb{R}^d}\left(\frac{d\,\|y-x\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}(t+1)}+1\right)\\&\qquad\qquad\cdot\exp\left(-\frac{\|y-x\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}(t+1)}\right)\mathrm{d}y\\&\le C_{(4.13)}\left(C^{min}_{(4.14)}(t+1)\right)^{-\frac{d}{2}-1}\int_{\mathbb{R}^d}\operatorname*{ess\,sup}_{x\in\mathbb{R}^d}|f(x+y)|\,\|y\|^{\alpha}_{\mathbb{R}^d}\left(\frac{d\,\|y\|^2_{\mathbb{R}^d}}{C^{min}_{(4.14)}(t+1)}+1\right)\\&\qquad\qquad\cdot\exp\left(-\frac{\|y\|^2_{\mathbb{R}^d}}{C^{max}_{(4.14)}(t+1)}\right)\mathrm{d}y\end{aligned}$$

since we can consider the shift $y\longmapsto y+x$ as before. Then we have $\|f\|_{L^\infty}=\operatorname*{ess\,sup}_{x\in\mathbb{R}^d}|f(x+y)|$ such that we can repeat the calculation from above at this point to obtain inequality (4.4). $\square$

4.3. Estimates on the transition probability density

On the basis of Lemma 4.7 we will state and prove parts of Theorem 4.8 about estimates on the transition probability density belonging to the considered stochastic process in the L^q-norm. We will describe later in this section to what extent we have to omit parts of the proof because of occurring issues with the dependence of the constants as well as in order to keep the thesis within reasonable length.

Anyway, for every $n \in \mathbb{N}$ we start over with the simplified process $\big(Y_n(t)\big)_{t\geq 0}$ defined by

$$Y_n(t)\Big(:= Y_n(t, Y_0)\Big) := \; Y_0 + \int_0^t \sigma\big(s, Y_n(\kappa_n(s))\big)\, \mathrm{d}W(s), \tag{4.15}$$

where $Y_0 \in \mathbb{R}^d$ is non-random and $\sigma \colon \mathbb{R}_+ \times \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times d_1}$ is a Borel-measurable map satisfying the conditions

$$\varepsilon \text{ Id} \leq (\sigma\sigma^T)(s,x) \leq K \text{ Id} \tag{4.16}$$

and

$$\|\sigma(s,x) - \sigma(s,y)\|_{L_2} \leq K\|x-y\|_{\mathbb{R}^d}^{\alpha} \tag{4.17}$$

for some constants $\alpha \in]0,1[$, $K, \varepsilon > 0$ and every $x, y \in \mathbb{R}^d$, $s > 0$. For fixed $n \in \mathbb{N}$ and $t > 0$ consider the corresponding transition semigroup π_t^n associated with (4.15), which is given by

$$\pi_t^n(Y_0, \mathrm{d}y) := P \circ Y_n(t, Y_0)^{-1}(\mathrm{d}y).$$

Then as usual we set

$$\pi_t^n f(Y_0) := \mathbb{E}\Big[f\big(Y_n(t,Y_0)\big)\Big] = \int_\Omega f\big(Y_n(t,Y_0)\big)\, \mathrm{d}P = \int_{\mathbb{R}^d} f(y)\, \pi_t^n(Y_0, \mathrm{d}y)$$

for bounded Borel-measurable maps $f \colon \mathbb{R}^d \longrightarrow \mathbb{R}$.
Let $p_n(t,y)$ be the density of $\pi_t^n(Y_0, \mathrm{d}y)$ with respect to the Lebesgue measure. Denote its supremum by $m_n(t) := \sup\limits_{y\in\mathbb{R}^d} p_n(t,y)$.

Theorem 4.8 (cf. [GK96], Theorem 4.2 on page 153)**.** *For every $n \in \mathbb{N}$ let $\big(Y_n(t)\big)_{t\geq 0}$ be the process given by* (4.15) *such that* (4.16) *and* (4.17) *are fulfilled for some constants $\alpha \in]0,1[$ and $K, \varepsilon > 0$. Suppose furthermore that there exists a density $p_n(t,y)$ of $\pi_t^n(Y_0, \mathrm{d}y)$ with respect to the Lebesgue measure. Then the following assertions hold.*

a) Let $1 \leq q < \frac{d}{d-\alpha}$. Then there exists a constant $C_{(4.18)} = C_{(4.18)}(d, \alpha, K, \varepsilon, p)$ such that for every $t > 0$ and $n \in \mathbb{N}$ we have

$$\|p_n(t,\cdot)\|_{L^q} = \bigg(\int_{\mathbb{R}^d} p_n(t,x)^q \,\mathrm{d}x\bigg)^{\frac{1}{q}} \leq C_{(4.18)}\big(t^{-\frac{d}{2p}} + 1\big), \tag{4.18}$$

where $p := \frac{q}{q-1}$ is the conjugate of q.

b) If the partitions $\{0 = t_0^n < t_1^n < \dots\}$ satisfy the additional conditions $\kappa_n(s) \geq \varepsilon s$ for every $n \in \mathbb{N}$ and $s > t_1^n$, then there exists a constant $C_{(4.19)} = C_{(4.19)}(d, \alpha, K, \varepsilon)$ such that

$$m_n(t) \leq C_{(4.19)}\big(t^{-\frac{d}{2}} + 1\big) \tag{4.19}$$

for every $t > 0$ and $n \in \mathbb{N}$. In this case (4.18) *holds as well for any $q \in [1, \infty]$, $t > 0$ and $n \in \mathbb{N}$.*

The following proof is based on the one given by Gyöngy and Krylov in [GK96]. But we will only prove assertion a) of the theorem and only for a constant $C_{(4.31)}$ instead of $C_{(4.18)}$ which also depends on an upper time bound $T \in [0, \infty[$. We have to omit the proof of assertion b) as well as the method to obtain a constant independent of T because both considerations would exceed the extent of this thesis. Hence, we refer to [GK96] at this point.

But more importantly, we will describe in a remark, that is stated before **Step 4** within this proof, in which way the application of Lemma 4.7 turns out to be problematic. In particular, we will see that we cannot exclude that the occurring constants may not be independent of the time variable.

Proof of Theorem 4.8 a). (cf. [GK96], Theorem 4.2 on page 153)
Observe that $\|p_n(t, \cdot)\|_{L^1} = \int_{\mathbb{R}^d} p_n(t, x)\,\mathrm{d}x = 1$ holds since $p_n(t, y)$ is a probability density. Hence, we only have to consider $1 < q < \frac{d}{d-\alpha}$ in the following.

Step 1: First of all, we define the operator $T_{s,t}^*$ as well as its adjoint $T_{s,t}$ and prove some of their properties that are necessary for the application of Lemma 4.7 later.

For $0 \leq s \leq t < \infty$ and bounded Borel-measurable functions $f \colon \mathbb{R}^d \longrightarrow \mathbb{R}$ consider the operator $T_{s,t}^*$ defined by

$$T_{s,t}^* f(y) := \mathbb{E}\left[f\left(y + \int_s^t \sigma(r, y)\,\mathrm{d}W(r)\right)\right],$$

and let $T_{s,t}$ be the adjoint operator of $T_{s,t}^*$ in $L^2\big(\mathbb{R}^d; \mathbb{R}\big)$-sense.

In order to apply Lemma 4.7 later, we have to show the following claim at first.

Claim (1)**.** If $s < t$ and $y \in \mathbb{R}^d$, then the stochastic integral

$$\xi(s, t, y) := \int_s^t \sigma(r, y)\,\mathrm{d}W(r)$$

is a d-dimensional Gaussian vector with distribution $N\big(0, \Lambda(s, t, y)\big)$, where $\Lambda(s, t, y) := \int_s^t (\sigma\sigma^T)(r, y)\,\mathrm{d}r$ is a covariance matrix such that

$$\varepsilon\,(t - s)\ \mathrm{Id} \leq \Lambda(s, t, y) = \Lambda(s, t, y)^* \leq K\,(t - s)\ \mathrm{Id} \tag{4.20}$$

holds for every $y \in \mathbb{R}^d$.

Proof of Claim (1). In order to prove that for fixed $s < t$ and $y \in \mathbb{R}^d$ the stochastic integral $\xi(s,t,y) = \big(\xi^1, \ldots, \xi^d\big)(s,t,y)$ is a d-dimensional Gaussian vector, we will show that for any arbitrary $\psi \in \mathbb{R}^d$ the linear combination $\sum\limits_{i=1}^{d} \psi_i \, \xi^i(s,t,y)$ has a one-dimensional normal distribution (cf. [Bau02], §30 on page 260). In this case the vector of means is given by

$$\mathbb{E}\big[\xi(s,t,y)\big] = \Big(\mathbb{E}\big[\xi^1(s,t,y)\big], \ldots, \mathbb{E}\big[\xi^d(s,t,y)\big]\Big)$$

and the covariance matrix $\Lambda(s,t,y) = \big(\Lambda_{ij}(s,t,y)\big)_{ij}$ consists of elements

$$\Lambda_{ij}(s,t,y) = \operatorname{cov}\big(\xi^i(s,t,y), \xi^j(s,t,y)\big).$$

For the linear combination $\sum\limits_{i=1}^{d} \psi_i \, \xi^i(s,t,y)$ we have

$$\sum_{i=1}^{d} \psi_i \, \xi^i(s,t,y) = \sum_{i=1}^{d} \psi_i \sum_{k=1}^{d_1} \int_s^t \sigma_{ik}(r,y) \, \mathrm{d}W^k(r) = \sum_{k=1}^{d_1} \int_s^t \sum_{i=1}^{d} \psi_i \, \sigma_{ik}(r,y) \, \mathrm{d}W^k(r),$$

where $\int_s^t \sum\limits_{i=1}^{d} \psi_i \, \sigma_{ik}(r,y) \, \mathrm{d}W^k(r)$, for $k = 1, \cdots, d_1$, are $\mathbb{R}$-valued integrals with deterministic integrands. By [Shr04], Example 4.7.3 on page 223 we know that these integrals have a normal distribution. Since linear combinations of independent normal distributed random variables are again normal distributed (cf. [Bau02], Theorem 8.4 on page 55 and Example 3 on page 56), the assertion follows by the independence of the components of the Wiener process W.

Furthermore, we have $\mathbb{E}\big[\xi^i(s,t,y)\big] = 0$ for every $1 \leq i \leq d$. Observe that Lemma A.7 from the Appendix and (4.16) imply that

$$\big\|\sigma(r,y)\big\|_{L_2}^2 \leq d\,K$$

holds. This yields by using the Itô isometry

$$\begin{aligned}
\mathbb{E}\Big[\langle \xi^i(s,\cdot,y)\rangle_t\Big] &= \mathbb{E}\bigg[\Big\langle \sum_{k=1}^{d_1} \int_s^{\cdot} \sigma_{ik}(r,y) \, \mathrm{d}W^k(r) \Big\rangle_t\bigg] \\
&= \mathbb{E}\bigg[\sum_{k=1}^{d_1} \Big\langle \int_s^{\cdot} \sigma_{ik}(r,y) \, \mathrm{d}W^k(r) \Big\rangle_t\bigg] \\
&= \mathbb{E}\bigg[\int_s^t \sum_{k=1}^{d_1} \big|\sigma_{ik}(r,y)\big|^2 \, \mathrm{d}r\bigg] \leq \mathbb{E}\bigg[\int_s^t \big\|\sigma(r,y)\big\|_{L_2}^2 \, \mathrm{d}r\bigg] < \infty,
\end{aligned}$$

where the second step holds since we have $\langle Z_1 + Z_2\rangle_t = \langle Z_1\rangle_t + 2\,\langle Z_1, Z_2\rangle_t + \langle Z_2\rangle_t$ for continuous local martingales Z_1 and Z_2, the bilinearity of the covariation $\langle\cdot,\cdot\rangle_t$ and

$$\begin{aligned}\mathbb{E}\Big[\Big\langle \int_s^{\cdot} \sigma_{ik}(r,y)\,\mathrm{d}W^k(r), \int_s^{\cdot} \sigma_{i\tilde{k}}(r,y)\,\mathrm{d}W^{\tilde{k}}(r)\Big\rangle_t\Big] &\\ &= \mathbb{E}\Big[\int_s^t \sigma_{ik}(r,y)\,\sigma_{i\tilde{k}}(r,y)\,\mathrm{d}\langle W^k, W^{\tilde{k}}\rangle_t\Big]\\ &= \mathbb{E}\Big[\int_s^t \sigma_{ik}(r,y)\,\sigma_{i\tilde{k}}(r,y)\,\delta_{k\tilde{k}}\,\mathrm{d}r\Big]\\ &= 0\end{aligned}$$

for $k\neq\tilde{k}$ because $\delta_{k\tilde{k}}$ means the Kronecker delta. Therefore, $\xi^i(s,\tilde{t},y)$, $\tilde{t}\in[s,t]$, is in fact a martingale (cf. [RY99], Corollary 1.25 on page 130) and hence $\mathbb{E}\big[\xi^i(s,t,y)\big]=0$.

Now the calculation of the elements of the covariance matrix $\Lambda(s,t,y)$ yields

$$\begin{aligned}&\operatorname{cov}\big(\xi^i(s,t,y),\xi^j(s,t,y)\big)\\ &\quad= \mathbb{E}\big[\xi^i(s,t,y)\,\xi^j(s,t,y)\big]\\ &\quad= \mathbb{E}\Big[\Big(\sum_{k=1}^{d_1}\int_s^t \sigma_{ik}(r,y)\,\mathrm{d}W^k(r)\Big)\Big(\sum_{\tilde{k}=1}^{d_1}\int_s^t \sigma_{j\tilde{k}}(\tilde{r},y)\,\mathrm{d}W^{\tilde{k}}(\tilde{r})\Big)\Big]\\ &\quad= \mathbb{E}\Big[\sum_{k=1}^{d_1}\sum_{\tilde{k}=1}^{d_1}\int_s^t\int_s^t \sigma_{ik}(r,y)\sigma_{j\tilde{k}}(\tilde{r},y)\,\mathrm{d}W^k(r)\,\mathrm{d}W^{\tilde{k}}(\tilde{r})\Big]\\ &\quad= \mathbb{E}\Big[\int_s^t\sum_{k=1}^{d_1}\sigma_{ik}(r,y)\sigma_{jk}(r,y)\,\mathrm{d}r\Big]\\ &\quad= \int_s^t (\sigma\sigma^T)_{ij}(r,y)\,\mathrm{d}r.\end{aligned}$$

Furthermore, we have to prove that (4.20) is fulfilled. Therefore, observe that

$$\Lambda(s,t,y) = \int_s^t (\sigma\sigma^T)(r,y)\,\mathrm{d}r$$

holds and that the symmetry of $\Lambda(s,t,y)$ follows because $(\sigma\sigma^T)(r,y)$ is symmetric. In addition, we have

$$\langle\Lambda(s,t,y)\,x,x\rangle_{\mathbb{R}^d} = \Big\langle\int_s^t(\sigma\sigma^T)(r,y)\,\mathrm{d}r\;x,x\Big\rangle_{\mathbb{R}^d} = \int_s^t \underbrace{\langle(\sigma\sigma^T)(r,y)\,x,x\rangle_{\mathbb{R}^d}}_{\substack{\le K\langle x,x\rangle_{\mathbb{R}^d} \\ \ge \varepsilon\,\langle x,x\rangle_{\mathbb{R}^d}}\quad \text{by (4.16)}}\,\mathrm{d}r,$$

and, hence,

$$\begin{aligned}\varepsilon\,(t-s)\,\langle x,x\rangle_{\mathbb{R}^d} &= \int_s^t \varepsilon\,\langle x,x\rangle_{\mathbb{R}^d}\,\mathrm{d}r\\ &\le \langle\Lambda(s,t,y)\,x,x\rangle_{\mathbb{R}^d} \qquad (4.21)\\ &\le \int_s^t K\,\langle x,x\rangle_{\mathbb{R}^d}\,\mathrm{d}r = K\,(t-s)\,\langle x,x\rangle_{\mathbb{R}^d}\end{aligned}$$

holds for every $x \in \mathbb{R}^d$. $\square$

Note that in the case $s = t$ we have $T^*_{t,t} f(y) = f(y) = T_{t,t} f(y)$.

For $s < t$ we can now repeat the calculation of $T_{s,t}$ from **Step 1** in the proof of Lemma 4.7 and obtain

$$\begin{aligned} T_{s,t} f(x) = \int_{\mathbb{R}^d} & f(y)\,(2\pi)^{-\frac{d}{2}} \big(\det \Lambda(s,t,y)\big)^{-\frac{1}{2}} \\ & \cdot \exp\Big(-\frac{1}{2}\langle \Lambda(s,t,y)^{-1}(x-y), x-y\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}y. \end{aligned} \tag{4.22}$$

By using formula (4.22), we can see that the following claim holds.

Claim (2)**.** For any $s < t$ the function $x \longmapsto T_{s,t} f(x)$ is infinitely differentiable. Furthermore,

$$\frac{\partial}{\partial s} T_{s,t} f(x) = -\sum_{i,j=1}^{d} \Big(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t}\, a_{ij}(s,\cdot)\, f(\cdot)\Big)(x) \tag{4.23}$$

holds, where $a_{ij} := \frac{1}{2}(\sigma\sigma^T)_{ij}$.

Proof of Claim (2). We omit the proof and refer to [GK96], where it is claimed that the assertion is fulfilled. $\square$

Step 2: In this step we start to calculate $\mathbb{E}\Big[f\big(Y_n(t)\big)\Big]$.

Consider the map $s \longmapsto \mathbb{E}\Big[\varphi\big(s, Y_n(s)\big)\Big]$, where the function φ is given by $\varphi\big(s, Y_n(s)\big) := T_{s,t} f\big(Y_n(s)\big)$. Then observe that for any $r \in [0,t]$ we can apply the Newton-Leibniz formula and obtain

$$\begin{aligned} \mathbb{E}\Big[f\big(Y_n(t)\big)\Big] = \mathbb{E}\Big[T_{t,t} f\big(Y_n(t)\big)\Big] &= \mathbb{E}\Big[\varphi\big(t, Y_n(t)\big)\Big] \\ &= \mathbb{E}\Big[\varphi\big(r, Y_n(r)\big)\Big] + \int_r^t \frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\Big[\varphi\big(s, Y_n(s)\big)\Big]\,\mathrm{d}s. \end{aligned}$$

From Itô's formula (cf. Theorem A.9 in the Appendix) follows that P-a.s.

$$\begin{aligned} \varphi\big(s, Y_n(s)\big) = \ & \varphi\big(0, Y_n(0)\big) + \int_0^s \frac{\partial}{\partial s}\varphi\big(u, Y_n(u)\big)\,\mathrm{d}u \\ & + \frac{1}{2}\sum_{i,j=1}^{d}\int_0^s a_{ij}\big(u, Y_n\big(\kappa_n(u)\big)\big)\frac{\partial^2}{\partial x^i \partial x^j}\varphi\big(u, Y_n(u)\big)\,\mathrm{d}u \\ & + \underbrace{\int_0^s \big\langle \nabla_x \varphi\big(u, Y_n(u)\big), \sigma\big(u, Y_n\big(\kappa_n(u)\big)\big)\,\mathrm{d}W(u)\big\rangle_{\mathbb{R}^d}}_{=:\, m(s)} \end{aligned}$$

holds, where $m(\varrho)$, $\varrho \in [0,s]$, is a continuous local $(\mathcal{F}_\varrho)$-martingale with $m(0) = 0$.

Claim (3)**.** The continuous local $m(\varrho)$, $\varrho \in [0,s]$, is a martingale for $s < t$.

Proof of Claim (3). We have

$$
\begin{aligned}
m(\varrho) &= \int_0^{\varrho} \big\langle \nabla_x \varphi\big(u, Y_n(u)\big), \sigma\big(u, Y_n\big(\kappa_n(u)\big)\big)\, \mathrm{d}W(u) \big\rangle_{\mathbb{R}^d} \\
&= \sum_{k=1}^{d_1} \underbrace{\int_0^{\varrho} \sum_{j=1}^{d} \sigma_{jk}\big(u, Y_n\big(\kappa_n(u)\big)\big) \frac{\partial}{\partial x^j} \varphi\big(u, Y_n(u)\big)\, \mathrm{d}W^k(u)}_{=:\, m^k(\varrho)} .
\end{aligned}
$$

Now it again suffices to prove that $\mathbb{E}\Big[\big\langle m^k(\cdot)\big\rangle_s\Big] < \infty$ holds for every $k = 1, \ldots, d_1$ (cf. [RY99], Corollary 1.25 on page 130). Namely, we have by the Itô isometry

$$
\begin{aligned}
\mathbb{E}\Big[\big\langle m^k(\cdot)\big\rangle_s\Big] &= \mathbb{E}\bigg[\Big\langle \int_0^{\cdot} \sum_{j=1}^{d} \sigma_{jk}\big(u, Y_n\big(\kappa_n(u)\big)\big) \frac{\partial}{\partial x^j} \varphi\big(u, Y_n(u)\big)\, \mathrm{d}W^k(u) \Big\rangle_s\bigg] \\
&= \mathbb{E}\bigg[\int_0^s \Big| \sum_{j=1}^{d} \sigma_{jk}\big(u, Y_n\big(\kappa_n(u)\big)\big) \frac{\partial}{\partial x^j} \varphi\big(u, Y_n(u)\big)\Big|^2 \mathrm{d}u\bigg] \\
&\leq \mathbb{E}\bigg[\int_0^s \underbrace{\Big(\sum_{j=1}^{d} \Big|\sigma_{jk}\big(u, Y_n\big(\kappa_n(u)\big)\big)\Big|^2\Big)}_{\leq \left\|\sigma\left(u, Y_n(\kappa_n(u))\right)\right\|_{L_2}^2} \Big(\sum_{j=1}^{d} \Big| \frac{\partial}{\partial x^j} \varphi\big(u, Y_n(u)\big)\Big|^2\Big)\, \mathrm{d}u\bigg],
\end{aligned}
$$

where we have used the Cauchy-Schwarz inequality in the last step. Besides,

$$
\big\|\sigma(r, y)\big\|_{L_2}^2 \leq d\, K
$$

holds for every $r > 0$ and $y \in \mathbb{R}^d$ by Lemma A.7 from the Appendix and inequality (4.16). Furthermore, from equation (4.22) it follows that

$$
\begin{aligned}
&\Big| \frac{\partial}{\partial x^j} \varphi\big(u, Y_n(u)\big)\Big|^2 \\
&\quad = \Big| \frac{\partial}{\partial x^j} T_{u,t} f\big(Y_n(u)\big)\Big|^2 \\
&\quad = \bigg| \bigg(\frac{\partial}{\partial x^j} \int_{\mathbb{R}^d} f(y)\, (2\pi)^{-\frac{d}{2}} \big(\det \Lambda(u, t, y)\big)^{-\frac{1}{2}} \\
&\qquad\qquad \cdot \exp\Big(-\frac{1}{2} \langle \Lambda(u, t, y)^{-1}\big(\,\cdot - y\big), \cdot - y\rangle_{\mathbb{R}^d}\Big)\, \mathrm{d}y\bigg) \big(Y_n(u)\big)\bigg|^2 \\
&\quad = \bigg| \int_{\mathbb{R}^d} f(y)\, (2\pi)^{-\frac{d}{2}} \big(\det \Lambda(u, t, y)\big)^{-\frac{1}{2}} \\
&\qquad\qquad \cdot \bigg(\frac{\partial}{\partial x^j} \exp\Big(-\frac{1}{2} \langle \Lambda(u, t, y)^{-1}\big(\,\cdot - y\big), \cdot - y\rangle_{\mathbb{R}^d}\Big)\bigg) \big(Y_n(u)\big)\, \mathrm{d}y\bigg|^2
\end{aligned}
$$

holds since the last step is just the assertion of **Claim** (1) in the proof of Lemma 4.7.

Applying the calculation from inequality (4.7) yields

$$
\begin{aligned}
&\Big|\frac{\partial}{\partial x^j}\varphi\big(u,Y_n(u)\big)\Big|^2\\
&\qquad\le\bigg(\int_{\mathbb{R}^d}|f(y)|\,(2\pi)^{-\frac{d}{2}}\big(\det\Lambda(u,t,y)\big)^{-\frac12}\,\|\Lambda(u,t,y)^{-1}\|_{L_2}\\
&\qquad\qquad\cdot\exp\Big(-\frac{1}{2\,K\,(t-u)}\,\big\|y-Y_n(u)\big\|^2_{\mathbb{R}^d}\Big)\,\big\|y-Y_n(u)\big\|_{\mathbb{R}^d}\,\mathrm{d}y\bigg)^2,
\end{aligned}
$$

where we have used that by (4.21) the inequality $\varepsilon\,(t-u)\ \mathrm{Id}\le\Lambda(u,t,y)\le K\,(t-u)\ \mathrm{Id}$ and, hence, $\frac{1}{\varepsilon\,(t-u)}\ \mathrm{Id}\ge\Lambda(u,t,y)^{-1}\ge\frac{1}{K\,(t-u)}\ \mathrm{Id}$ by Lemma 4.4 i) hold. Moreover, by applying Lemma 4.6 iii) we can see that

$$
\big\|\Lambda(u,t,y)^{-1}\big\|^2_{L_2}=\sum_{i,j=1}^{d}\big|\Lambda_{ij}(u,t,y)^{-1}\big|^2\le\Big(\frac{d}{\varepsilon\,(t-u)}\Big)^2
$$

is fulfilled. Furthermore, since the estimate (4.21) holds, we know that an application of Lemma 4.6 i) yields $\big(\varepsilon\,(t-u)\big)^{-\frac{d}{2}}\ge\big(\det\Lambda(u,t,y)\big)^{-\frac12}\ge\big(K\,(t-u)\big)^{-\frac{d}{2}}$ for every $y\in\mathbb{R}^d$. Hence,

$$
\begin{aligned}
&\Big|\frac{\partial}{\partial x^j}\varphi\big(u,Y_n(u)\big)\Big|^2\\
&\qquad\le\bigg(\int_{\mathbb{R}^d}\|f\|_\infty\,(2\pi)^{-\frac{d}{2}}\big(\varepsilon\,(t-u)\big)^{-\frac{d}{2}}\,\frac{d}{\varepsilon\,(t-u)}\\
&\qquad\qquad\cdot\exp\Big(-\frac{1}{2\,K\,(t-u)}\,\big\|y-Y_n(u)\big\|^2_{\mathbb{R}^d}\Big)\,\big\|y-Y_n(u)\big\|_{\mathbb{R}^d}\,\mathrm{d}y\bigg)^2\\
&\qquad=\bigg((2\pi\varepsilon)^{-\frac{d}{2}}\,\frac{d}{\varepsilon}\,\|f\|_\infty\,\big(t-u\big)^{-\frac{d}{2}}\,(t-u)^{-1}\\
&\qquad\qquad\cdot\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K\,(t-u)}\,\big\|y-Y_n(u)\big\|^2_{\mathbb{R}^d}\Big)\,\big\|y-Y_n(u)\big\|_{\mathbb{R}^d}\,\mathrm{d}y\bigg)^2,
\end{aligned}
$$

and by using the transformation $y\longmapsto\big(t-u\big)^{\frac12}y+Y_n(u)$ with the Jacobian determinant given by $\big(t-u\big)^{\frac{d}{2}}$, we obtain

$$
\begin{aligned}
&\bigg((2\pi\varepsilon)^{-\frac{d}{2}}\,\frac{d}{\varepsilon}\,\|f\|_\infty\,\big(t-u\big)^{-\frac12}\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K}\,\big\|y\big\|^2_{\mathbb{R}^d}\Big)\,\big\|y\big\|_{\mathbb{R}^d}\,\mathrm{d}y\bigg)^2\\
&\qquad=\big(t-u\big)^{-1}\|f\|^2_\infty\cdot\underbrace{\bigg((2\pi\varepsilon)^{-\frac{d}{2}}\,\frac{d}{\varepsilon}\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K}\,\big\|y\big\|^2_{\mathbb{R}^d}\Big)\,\big\|y\big\|_{\mathbb{R}^d}\,\mathrm{d}y\bigg)^2}_{=:\,C_{(4.24)}},
\end{aligned}
\tag{4.24}
$$

where $C_{(4.24)}=C_{(4.24)}(\varepsilon,K,d)$ is a constant. Altogether we have

$$
\begin{aligned}
\mathbb{E}\Big[\big\langle m^k(\cdot)\big\rangle_s\Big]&\le\mathbb{E}\bigg[\int_0^s d^2\,K\,\big(t-u\big)^{-1}\|f\|^2_\infty\,C_{(4.24)}\,\mathrm{d}u\bigg]\\
&=d^2\,K\,\|f\|^2_\infty\,C_{(4.24)}\int_0^s\big(t-u\big)^{-1}\,\mathrm{d}u<\infty
\end{aligned}
$$

since $\int_0^s \big(t-u\big)^{-1}\,\mathrm{d}u = \int_{t-s}^t u^{-1}\,\mathrm{d}u = \ln(t) - \ln(t-s) < \infty$ for $t > s$. $\square$

Therefore, $\mathbb{E}\big[m(s)\big] = 0$ holds such that we conclude

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}\Big[\varphi\big(s,Y_n(s)\big)\Big] \\
&\quad= \frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}\Big[\varphi\big(0,Y_n(0)\big)\Big] + \frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}\bigg[\int_0^s \frac{\partial}{\partial s}\varphi\big(u,Y_n(u)\big)\,\mathrm{d}u\bigg] \\
&\qquad\quad + \frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}\bigg[\frac{1}{2}\sum_{i,j=1}^d \int_0^s a_{ij}\big(u,Y_n\big(\kappa_n(u)\big)\big)\frac{\partial^2}{\partial x^i \partial x^j}\varphi\big(u,Y_n(u)\big)\,\mathrm{d}u\bigg] \\
&\quad= \underbrace{\frac{\mathrm{d}}{\mathrm{d}s}\,\varphi\big(0,Y_0\big)}_{=0} + \mathbb{E}\bigg[\frac{\partial}{\partial s}\varphi\big(s,Y_n(s)\big)\bigg] \\
&\qquad\quad + \frac{1}{2}\sum_{i,j=1}^d \mathbb{E}\bigg[a_{ij}\big(s,Y_n\big(\kappa_n(s)\big)\big)\frac{\partial^2}{\partial x^i \partial x^j}\varphi\big(s,Y_n(s)\big)\bigg],
\end{aligned}
$$

where we have applied Fubini's theorem in the last step to interchange the integral and the expectation by using the same arguments as in the proof of **Claim (1)** in Lemma 4.7 for the necessary finiteness of the integrals. Hence, by applying (4.23) we altogether obtain

$$
\begin{aligned}
\mathbb{E}\Big[f\big(Y_n(t)\big)\Big] &= \mathbb{E}\Big[T_{r,t}f\big(Y_n(r)\big)\Big] \\
&\quad + \int_r^t \mathbb{E}\bigg[\sum_{i,j=1}^d a_{ij}\big(s,Y_n\big(\kappa_n(s)\big)\big)\bigg(\frac{\partial^2}{\partial x^i \partial x^j}T_{s,t}f\bigg)\big(Y_n(s)\big) \\
&\qquad\qquad - \bigg(\frac{\partial^2}{\partial x^i \partial x^j}T_{s,t}\,a_{ij}(s,\cdot)f(\cdot)\bigg)\big(Y_n(s)\big)\bigg]\,\mathrm{d}s.
\end{aligned}
$$

Define

$$
\eta(s,x) := \int_{\kappa_n(s)}^s \sigma(r,x)\,\mathrm{d}W(r).
$$

Then we have the following claim.

Claim (4)**.** If $\kappa_n(s) < s$ and $x \in \mathbb{R}^d$, then the stochastic integral $\eta(s,x)$ is a d-dimensional Gaussian vector with distribution $N\big(0,\Lambda_{\eta(s,x)}\big)$, where $\Lambda_{\eta(s,x)} := \int_{\kappa_n(s)}^s (\sigma\sigma^T)(r,x)\,\mathrm{d}r$ is a covariance matrix.

Proof of Claim (4). Analogous to the proof of **Claim** (1). $\square$

Observe that we can write

$$
\begin{aligned}
Y_n(s) &= Y_n\big(\kappa_n(s)\big) + \int_{\kappa_n(s)}^s \sigma\big(r,\underbrace{Y_n\big(\kappa_n(r)\big)}_{=Y_n(\kappa_n(s))}\big)\,\mathrm{d}W(r) \\
&= Y_n\big(\kappa_n(s)\big) + \eta\Big(s,Y_n\big(\kappa_n(s)\big)\Big).
\end{aligned}
$$

Taking conditional expectation with respect to the sigma-algebra $\sigma\big(Y_n\big(\kappa_n(s)\big)\big)$ yields

$$\begin{aligned}
&\mathbb{E}\Big[f\big(Y_n(t)\big)\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[f\big(Y_n(t)\big)\,\Big|\,\sigma\big(Y_n\big(\kappa_n(s)\big)\big)\Big]\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[T_{r,t}f\big(Y_n(r)\big)\,\Big|\,\sigma\big(Y_n\big(\kappa_n(s)\big)\big)\Big]\Big] \\
&\qquad + \int_r^t \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_{i,j=1}^d a_{ij}\big(s, Y_n\big(\kappa_n(s)\big)\big)\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\right)\big(Y_n(s)\big)\,\Bigg|\,\sigma\big(Y_n\big(\kappa_n(s)\big)\big)\Bigg] \\
&\qquad\qquad - \mathbb{E}\Bigg[\sum_{i,j=1}^d \left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t}\, a_{ij}(s,\cdot) f(\cdot)\right)\big(Y_n(s)\big)\,\Bigg|\,\sigma\big(Y_n\big(\kappa_n(s)\big)\big)\Bigg]\Bigg]\,\mathrm{d}s,
\end{aligned}$$

where we can write the second summand by using the measurability of a_{ij} as

$$\begin{aligned}
\int_r^t \sum_{i,j=1}^d \mathbb{E}\Bigg[a_{ij}\big(s, Y_n\big(\kappa_n(s)\big)\big)\,\mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\right)\big(Y_n(s)\big)\,\Bigg|\,\sigma\big(Y_n\big(\kappa_n(s)\big)\big)\Bigg]\Bigg] \\
- \mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t}\, a_{ij}(s,\cdot) f(\cdot)\right)\big(Y_n(s)\big)\Bigg]\,\mathrm{d}s.
\end{aligned}$$

Now observe that

$$\begin{aligned}
&\mathbb{E}\Bigg[a_{ij}\big(s, Y_n\big(\kappa_n(s)\big)\big)\,\mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\right)\big(Y_n(s)\big)\,\Bigg|\,\sigma\big(Y_n\big(\kappa_n(s)\big)\big)\Bigg]\Bigg] \\
&\quad = \int_{\mathbb{R}^d} a_{ij}(s,x)\,\mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\right)\big(x + \eta(s,x)\big)\,\Bigg|\,\sigma(x)\Bigg]\, p_n\big(\kappa_n(s), x\big)\,\mathrm{d}x \\
&\quad = \int_{\mathbb{R}^d} a_{ij}(s,x)\,\mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\right)\big(x + \eta(s,x)\big)\Bigg]\, p_n\big(\kappa_n(s), x\big)\,\mathrm{d}x
\end{aligned}$$

holds if we write the expectation as an integral with respect to the density $p_n\big(\kappa_n(s), \cdot\big)$ of the transition semigroup. Hence, altogether we have

$$\begin{aligned}
&\mathbb{E}\Big[f\big(Y_n(t)\big)\Big] \\
&\quad = \mathbb{E}\Big[T_{r,t}f\big(Y_n(r)\big)\Big] \\
&\qquad + \int_r^t \sum_{i,j=1}^d \mathbb{E}\Bigg[a_{ij}\big(s, Y_n\big(\kappa_n(s)\big)\big)\,\mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\right)\big(Y_n(s)\big)\Bigg] \\
&\qquad\qquad - \mathbb{E}\Bigg[\left(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t}\, a_{ij}(s,\cdot) f(\cdot)\right)\big(Y_n(s)\big)\Bigg]\Bigg]\,\mathrm{d}s.
\end{aligned} \tag{4.25}$$

Therefore, based on the last summand of (4.25), we consider the expression

$$\int_r^t \mathbb{E}\Bigg[\sum_{i,j=1}^d H_{ij}\big(s,t,Y_n\big(\kappa_n(s)\big)\big)\Bigg]\,\mathrm{d}s = \int_r^t \int_{\mathbb{R}^d} \sum_{i,j=1}^d H_{ij}(s,t,x)\, p_n\big(\kappa_n(s), x\big)\,\mathrm{d}x\,\mathrm{d}s,$$

where

$$H_{ij}(s,t,x) := a_{ij}(s,x)\,\mathbb{E}\bigg[\bigg(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t} f\bigg)\big(x+\eta(s,x)\big)\bigg] \\ - \mathbb{E}\bigg[\bigg(\frac{\partial^2}{\partial x^i \partial x^j} T_{s,t}\, a_{ij}(s,\cdot) f(\cdot)\bigg)\big(x+\eta(s,x)\big)\bigg].$$

Step 3: In this step we determine an estimate for the term $\big|H_{ij}(s,t,x)\big|$ by applying Lemma 4.7.

Claim (5)**.** For $H_{ij}(s,t,x)$ the necessary assumptions from Lemma 4.7 hold.

Proof of Claim (5).
1) $f\colon \mathbb{R}^d \longrightarrow \mathbb{R}$ is a bounded Borel-measurable function.
2) The role of g is taken by the function $a_{ij}(s,\cdot)\colon \mathbb{R}^d \longrightarrow \mathbb{R}$. Therefore, we have to prove its Hölder continuity, which is induced by (4.17). Namely, by using the submultiplicativity of $\|\cdot\|_{L_2}$, we have

$$\begin{aligned} & \|\sigma(s,x)-\sigma(s,y)\|_{L_2} \le K\,\|x-y\|^{\alpha}_{\mathbb{R}^d} \\ \Leftrightarrow \quad & \|\sigma(s,x)-\sigma(s,y)\|_{L_2}\,\|\sigma^T(s,x)\|_{L_2} \le K\,\|\sigma^T(s,x)\|_{L_2}\,\|x-y\|^{\alpha}_{\mathbb{R}^d} \\ \Rightarrow \quad & \Big\|\frac{1}{2}\sigma\sigma^T(s,x) - \frac{1}{2}\sigma(s,y)\,\sigma^T(s,x)\Big\|_{L_2} \le \frac{1}{2}\,K\,\|\sigma^T(s,x)\|_{L_2}\,\|x-y\|^{\alpha}_{\mathbb{R}^d} \end{aligned}$$

and since $\|\sigma(s,x)-\sigma(s,y)\|_{L_2} = \|\big(\sigma(s,x)-\sigma(s,y)\big)^T\|_{L_2} = \|\sigma^T(s,x)-\sigma^T(s,y)\|_{L_2}$ also

$$\begin{aligned} & \|\sigma^T(s,x)-\sigma^T(s,y)\|_{L_2} \le K\,\|x-y\|^{\alpha}_{\mathbb{R}^d} \\ \Leftrightarrow \quad & \|\sigma(s,y)\|_{L_2}\,\|\sigma^T(s,x)-\sigma^T(s,y)\|_{L_2} \le K\,\|\sigma(s,y)\|_{L_2}\,\|x-y\|^{\alpha}_{\mathbb{R}^d} \\ \Rightarrow \quad & \Big\|\frac{1}{2}\sigma(s,y)\,\sigma^T(s,x) - \frac{1}{2}\sigma\sigma^T(s,y)\Big\|_{L_2} \le \frac{1}{2}\,K\,\|\sigma(s,y)\|_{L_2}\,\|x-y\|^{\alpha}_{\mathbb{R}^d}. \end{aligned}$$

From this we can conclude that

$$\begin{aligned} \|a(s,x)-a(s,y)\|_{L_2} &= \Big\|\frac{1}{2}\sigma\sigma^T(s,x) - \frac{1}{2}\sigma\sigma^T(s,y)\Big\|_{L_2} \\ &\le \Big\|\frac{1}{2}\sigma\sigma^T(s,x) - \frac{1}{2}\sigma(s,y)\,\sigma^T(s,x)\Big\|_{L_2} + \Big\|\frac{1}{2}\sigma(s,y)\,\sigma^T(s,x) - \frac{1}{2}\sigma\sigma^T(s,y)\Big\|_{L_2} \\ &\le \underbrace{\frac{1}{2}\,K\,\big(\|\sigma^T(s,x)\|_{L_2} + \|\sigma(s,y)\|_{L_2}\big)}_{=:\tilde{K}}\,\|x-y\|^{\alpha}_{\mathbb{R}^d} \end{aligned}$$

holds for $x,y \in \mathbb{R}^d$, $s>0$ with a constant $\tilde{K}>0$.
3) The properties of ξ and η have already been proved in **Claim** (1) and **Claim** (4). But we still have to prove their independence. Therefore, notice that ξ and η are stochastic integrals with non-random integrands which only differ in their bounds of integration. Since the intervals from $\kappa_n(s)$ up to s and from s up to t do not intersect, the independence follows from the independence of the increments of the Wiener process W. □

Therefore, we obtain the estimate

$$\big|H_{ij}(s,t,x)\big| \leq C_{(4.26)}(t-s)^{-\frac{d}{2p}-1+\frac{\alpha}{2}}\|f\|_{L^p} \tag{4.26}$$

from Lemma 4.7, where $C_{(4.26)}$ is a constant.

At this point we have to insert an important remark about the dependencies of $C_{(4.26)}$.

Remark. The constant $C_{(4.26)}$ depends among others on the variables λ_η^{min} and λ_η^{max} that are introduced in Lemma 4.7. In this lemma we have proved the assertion with a constant $C_{(4.26)} = C_{(4.26)}\big(K, \varepsilon, d, p, \lambda_\eta^{min}, \lambda_\eta^{max}\big)$ and a covariance matrix Λ_η corresponding to a Gaussian vector η, which was introduced as an abstract object being independent of s and t. In our case the covariance matrix is given by

$$\Lambda_{\eta(s,x)} = \int_{\kappa_n(s)}^{s} (\sigma\sigma^T)(r,x)\,\mathrm{d}r.$$

We can now use the estimate

$$\begin{aligned}
\varepsilon\,\big(s-\kappa_n(s)\big)\,\langle x,x\rangle_{\mathbb{R}^d} &= \int_{\kappa_n(s)}^{s} \varepsilon\,\langle x,x\rangle_{\mathbb{R}^d}\,\mathrm{d}r\\
&\leq \langle \Lambda_{\eta(s,x)}\,x, x\rangle_{\mathbb{R}^d}\\
&\leq \int_{\kappa_n(s)}^{s} K\,\langle x,x\rangle_{\mathbb{R}^d}\,\mathrm{d}r = K\,\big(s-\kappa_n(s)\big)\,\langle x,x\rangle_{\mathbb{R}^d}
\end{aligned}$$

which is induced by the inequality (4.16). Consequently, we consider

$$\varepsilon\,\big(s-\kappa_n(s)\big)\ \mathrm{Id} \leq \Lambda_{\eta(s,x)} \leq K\,\big(s-\kappa_n(s)\big)\ \mathrm{Id}$$

for the estimate of Λ_η in this case. Therefore, the constants λ_η^{min} and λ_η^{max} from Lemma 4.7 are now replaced by $\varepsilon\,\big(s-\kappa_n(s)\big)$ and $K\,\big(s-\kappa_n(s)\big)$, respectively. Hence, we conclude that we cannot exclude that the constant $C_{(4.26)}$ may be depending on s, i.e. $C_{(4.26)} = C_{(4.26)}\big(K, \varepsilon, d, p, s\big)$.

This remark about the possible dependence of $C_{(4.26)}$ on s is a problem for the completion of the proof. In **Step 4** we will show how to finish the proof, but under the crucial assumption that $C_{(4.26)}$ does not depend on s.

Step 4: In this step we prove the estimate (4.18) with a constant depending on an upper time bound $T \in [0,\infty[$ (assuming that $C_{(4.26)}$ is independent of s).

Let $T \in [0,\infty[$. Consider the equation

$$\mathbb{E}\Big[f\big(Y_n(t)\big)\Big] = \mathbb{E}\Big[T_{r,t}f\big(Y_n(r)\big)\Big] + \int_r^t \mathbb{E}\Big[\sum_{i,j=1}^{d} H_{ij}\big(s,t,Y_n\big(\kappa_n(s)\big)\big)\Big]\,\mathrm{d}s \tag{4.27}$$

for $r = 0$ and $t \in]0, T]$. Then we have

$$\begin{aligned} \Big|\mathbb{E}\Big[f\big(Y_n(t)\big)\Big]\Big| &= \Big|\mathbb{E}\Big[T_{0,t}f\big(Y_n(0)\big)\Big] + \int_0^t \mathbb{E}\Big[\sum_{i,j=1}^d H_{ij}\big(s,t,Y_n\big(\kappa_n(s)\big)\big)\Big]\,\mathrm{d}s\Big| \\ &\le \mathbb{E}\Big[\big|T_{0,t}f\big(Y_0\big)\big|\Big] + \int_0^t \mathbb{E}\Big[\sum_{i,j=1}^d \Big|H_{ij}\big(s,t,Y_n\big(\kappa_n(s)\big)\big)\Big|\Big]\,\mathrm{d}s. \end{aligned}$$

At first we prove an estimate for $\big|T_{0,t}f\big(Y_0\big)\big|$ for $t > 0$. Note that by equation (4.20) and Lemma 4.6 i) we have the inequalities $\varepsilon\, t\ \mathrm{Id} \le \Lambda(0,t,y) \le K\,t\ \mathrm{Id}$ and $\big(\varepsilon\, t\big)^d \le \det\big(\Lambda(0,t,y)\big) \le \big(K\,t\big)^d$. Therefore, we can calculate as before in the proof of Lemma 4.7

$$\begin{aligned} \big|T_{0,t}f\big(Y_0\big)\big| &= \Big|\int_{\mathbb{R}^d} f(y)\,(2\pi)^{-\frac{d}{2}}\big(\det\Lambda(0,t,y)\big)^{-\frac{1}{2}} \\ &\qquad\qquad \cdot \exp\Big(-\frac{1}{2}\langle\Lambda(0,t,y)^{-1}(y-Y_0), y-Y_0\rangle_{\mathbb{R}^d}\Big)\,\mathrm{d}y\Big| \\ &\le (2\pi)^{-\frac{d}{2}}\int_{\mathbb{R}^d}|f(y)|\,(\varepsilon\,t)^{-\frac{d}{2}}\ \exp\Big(-\frac{1}{2\,K\,t}\|y-Y_0\|^2_{\mathbb{R}^d}\Big)\,\mathrm{d}y \\ &\le (2\pi\,\varepsilon)^{-\frac{d}{2}}\,t^{-\frac{d}{2}}\,\|f\|_{L^p}\Big(\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K\,t}\|y-Y_0\|^2_{\mathbb{R}^d}\Big)^q\,\mathrm{d}y\Big)^{\frac{1}{q}}, \end{aligned}$$

where the last step follows by applying Hölder's inequality for $q \in]1, \frac{d}{d-\alpha}[$ and its conjugate $p \in]\frac{d}{\alpha}, \infty[$. Since

$$\Big(\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K\,t}\|y-Y_0\|^2_{\mathbb{R}^d}\Big)^q\,\mathrm{d}y\Big)^{\frac{1}{q}} = t^{\frac{d}{2q}}\Big(\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K}\|y\|^2_{\mathbb{R}^d}\Big)^q\,\mathrm{d}y\Big)^{\frac{1}{q}}$$

holds by using a shift $y \longmapsto y + Y_0$ and the transformation $y \longmapsto t^{\frac{1}{2}}y$ with the Jacobian determinant given by $t^{\frac{d}{2}}$ as seen before in the proof of Lemma 4.7, we obtain

$$\begin{aligned} \big|T_{0,t}f\big(Y_0\big)\big| &\le \underbrace{t^{-\frac{d}{2}}\,t^{\frac{d}{2q}}}_{=t^{-\frac{d}{2p}}}\,\|f\|_{L^p}\,\underbrace{(2\pi\,\varepsilon)^{-\frac{d}{2}}\Big(\int_{\mathbb{R}^d}\exp\Big(-\frac{1}{2\,K}\|y\|^2_{\mathbb{R}^d}\Big)^q\,\mathrm{d}y\Big)^{\frac{1}{q}}}_{=:C_{(4.28)}} \\ &= C_{(4.28)}\,t^{-\frac{d}{2p}}\,\|f\|_{L^p}, \end{aligned} \tag{4.28}$$

where $C_{(4.28)} = C_{(4.28)}(K,\varepsilon,d,p)$ is a constant. Next we have by estimate (4.26)

$$\begin{aligned} &\int_0^t \mathbb{E}\Big[\sum_{i,j=1}^d\Big|H_{ij}\big(s,t,Y_n\big(\kappa_n(s)\big)\big)\Big|\Big]\,\mathrm{d}s \\ &\qquad\qquad \le \int_0^t\sum_{i,j=1}^d C_{(4.26)}(t-s)^{-\frac{d}{2p}-1+\frac{\alpha}{2}}\,\|f\|_{L^p}\,\mathrm{d}s \\ &\qquad\qquad = C_{(4.29)}\,\|f\|_{L^p}\int_0^t(t-s)^{-\frac{d}{2p}-1+\frac{\alpha}{2}}\,\mathrm{d}s, \end{aligned} \tag{4.29}$$

where $C_{(4.29)} := d^2\, C_{(4.26)}$.

Now we have to take a closer look at the power of $(t-s)$ in the integral. By assumption we have $1 < q < \frac{d}{d-\alpha}$ and, hence,

$$q < \frac{d}{d-\alpha} \Leftrightarrow (d-\alpha)\, q < d \Leftrightarrow (q-1)\, d < \alpha q \Leftrightarrow \frac{d}{\alpha} < \underbrace{\frac{q}{q-1}}_{=p}$$

since $d \in \mathbb{N}$ and $\alpha \in]0,1[$. Therefore,

$$\frac{d}{\alpha} < p \Leftrightarrow d < \alpha\, p \Leftrightarrow 0 < \frac{\alpha\, p}{2p} - \frac{d}{2p} \Leftrightarrow -1 < -\frac{d}{2p} - 1 + \frac{\alpha}{2}.$$

Hence, we can compute the integral by substitution, i.e. we obtain

$$\begin{aligned}\int_0^t (t-s)^{-\frac{d}{2p}-1+\frac{\alpha}{2}}\, \mathrm{d}s &= -\int_t^0 z^{-\frac{d}{2p}-1+\frac{\alpha}{2}}\, \mathrm{d}z = \frac{1}{-\frac{d}{2p}+\frac{\alpha}{2}}\, z^{-\frac{d}{2p}+\frac{\alpha}{2}}\Big|_0^t \\ &= C_{(4.30)}\, t^{-\frac{d}{2p}+\frac{\alpha}{2}}\end{aligned} \tag{4.30}$$

with a constant $C_{(4.30)} = C_{(4.30)}(d,\alpha,p)$. Altogether we have

$$\begin{aligned}\Big|\mathbb{E}\Big[f\big(Y_n(t)\big)\Big]\Big| &= \mathbb{E}\Big[\big|T_{0,t}f\big(Y_0\big)\big|\Big] + \int_0^t \mathbb{E}\Big[\sum_{i,j=1}^d \Big|H_{ij}\big(s,t,Y_n\big(\kappa_n(s)\big)\big)\Big|\Big]\, \mathrm{d}s \\ &\le C_{(4.28)}\, t^{-\frac{d}{2p}}\, \|f\|_{L^p} + C_{(4.29)}\, C_{(4.30)}\, t^{-\frac{d}{2p}+\frac{\alpha}{2}}\, \|f\|_{L^p} \\ &\le \underbrace{\big(C_{(4.28)} + C_{(4.29)}\, C_{(4.30)}\, T^{\frac{\alpha}{2}}\,\big)}_{=:\, C_{(4.31)}}\, t^{-\frac{d}{2p}}\, \|f\|_{L^p} \\ &\le C_{(4.31)}\, \|f\|_{L^p}\, (t^{-\frac{d}{2p}} + 1),\end{aligned} \tag{4.31}$$

where $C_{(4.31)} = C_{(4.31)}(K,\varepsilon,d,\alpha,p,\lambda_\eta^{min},\lambda_\eta^{max},T)$.

Now we will show how this implies the estimate from the assertion. Consider the linear functional

$$\Phi\big(p_n(t,\cdot)\big) : \begin{array}{rl} L^p(\mathbb{R}^d;\mathbb{R}) \longrightarrow & \mathbb{R} \\ f \longmapsto & \displaystyle\int_{\mathbb{R}^d} f(y)\, p_n(t,y)\, \mathrm{d}y \end{array}$$

and observe that we have proved that

$$\Big|\Phi\big(p_n(t,\cdot)\big)(f)\Big| = \Big|\int_{\mathbb{R}^d} f(y)\, p_n(t,y)\, \mathrm{d}y\Big| = \Big|\mathbb{E}\Big[f\big(Y_n(t)\big)\Big]\Big| \le C_{(4.31)}\, \|f\|_{L^p}\, (t^{-\frac{d}{2p}} + 1)$$

holds for bounded Borel-measurable functions f. By an approximation argument we also get this inequality for $f \in L^p$ because we can consider the sequence $(f_m)_{m\in\mathbb{N}}$ of bounded Borel-measurable functions given by $f_m := f\, \mathbb{1}_{\{-m \le f \le m\}}$, which converges by

Lebesgue's dominated convergence theorem to f in L^p. Hence, we can conclude that $\Phi\big(p_n(t,\cdot)\big) \in \big(L^p(\mathbb{R}^d;\mathbb{R})\big)'$.

Therefore, by using the duality of the L^p-spaces for $1 \leq p < \infty$, we can consider the isometric isomorphism $T\colon L^q(\mathbb{R}^d;\mathbb{R}) \longrightarrow \big(L^p(\mathbb{R}^d;\mathbb{R})\big)'$ that provides the general form of the linear functional $\Phi\big(p_n(t,\cdot)\big)$ (see e.g. [Bog07], Theorem 4.4.1 on page 262 or [Alt12], Theorem 4.12 on page 183). Hence, there exists a $g \in L^q(\mathbb{R}^d;\mathbb{R})$ such that

$$\int_{\mathbb{R}^d} f(y)\,g(y)\,\mathrm{d}y =: (T\,g)(f) = \Phi\big(p_n(t,\cdot)\big)(f),$$

the isometry property $\big\|T\,g\big\| = \|g\|_{L^q}$ and

$$\left|\int_{\mathbb{R}^d} f(y)\,g(y)\,\mathrm{d}y\right| \leq \|g\|_{L^q}\|f\|_{L^p}$$

from Hölder's inequality are fulfilled.

Furthermore, we know that $\big\|T\,g\big\| = \inf\big\{c \geq 0 \,\big|\, \big|(T\,g)f\big| \leq c\,\|f\|_{L^p}\big\}$ is given by the definition of the operator norm. Thus we realise that in fact

$$\|g\|_{L^q} \leq C_{(4.31)}\,(t^{-\frac{d}{2p}} + 1)$$

has to hold. Hence, by using that $g = p_n(t,\cdot)$, we can obtain the estimate

$$\|p_n(t,\cdot)\|_{L^q} \leq C_{(4.31)}\,(t^{-\frac{d}{2p}} + 1) \tag{4.32}$$

from the assertion with a constant depending on T.

□

5. Existence and Uniqueness (Non-degeneracy Version)

This last chapter will focus on the second main theorem (see Theorem 5.2) in which we will change the assumptions from the first main theorem (see Theorem 3.7) slightly. We base the following considerations on [GK96], where it is claimed that such a theorem holds.

Unfortunately, in order to keep this thesis within reasonable length, we cannot go into details concerning the proof such that we just have to refer to the one given by Gyöngy and Krylov in [GK96]. There the authors also mention the similarity to the proof of Theorem 3.7 such that we have already worked out the necessary essential ideas in the previous chapters.

5.1. Main theorem (non-degeneracy version)

In this section we will extend the assumptions from the framework in Section 2.1 by a so-called *non-degeneracy* condition for the diffusion coefficient σ. With this new condition we can change the continuity assumptions on b and σ later in Theorem 5.2.

Therefore, we introduce **A4)** as the fourth main assumption of this thesis.

A4) For every $k \in \mathbb{N}$ the domain D_k is bounded and convex and

$$\sum_{i,j=1}^{d} (\sigma\sigma^T)_{ij}(t,x)\lambda_i\lambda_j \geq \varepsilon_k\, M_k(t) \sum_{i=1}^{d} |\lambda_i|^2$$

holds for every $t \in [0,k]$, $x \in D_k$ and $\lambda_i \in \mathbb{R}$ for $i = 1,\dots,d$, where $\varepsilon_k > 0$ are some constants.

But before we can state the theorem, we have to define the local Hölder continuity, which we will assume for the diffusion coefficient σ in the following.

Definition 5.1 (Local Hölder continuity)**.** *Let $n \in \mathbb{N}$. A function $f\colon \mathbb{R}_+ \times D \longrightarrow \mathbb{R}^n$ is called locally Hölder continuous in $x \in D$ (with exponent $\alpha \in]0,1]$) if for every $k \in \mathbb{N}$, $t \geq 0$ and $x, y \in D_k$ we have*

$$\|f(t,x) - f(t,y)\|_{\mathbb{R}^n}^2 \leq M_k(t)\|x-y\|_{\mathbb{R}^d}^{2\alpha}.$$

If $\alpha = 1$, we say that f is locally Lipschitz continuous in $x \in D$.

Now we can finally state the second main and simultaneously last theorem of this thesis, that can be found in [GK96].

Theorem 5.2 (cf. [GK96], Theorem 2.8 on page 149)**.** *Let the assumptions from the framework in Section 2.1 and in addition* ***A4)*** *be fulfilled. Suppose moreover that* σ *is locally Hölder continuous in* $x \in D$ *with some exponent* $\alpha \in]0,1]$. *If* $\alpha \neq 1$, *assume that pathwise uniqueness holds for the equation* (2.1)*. Then we have:*

1) There exists a process $(X(t))_{t\geq 0}$ *such that* $X_n(t) \xrightarrow[n\to\infty]{p} X(t)$ *uniformly in* t *on bounded intervals.*

2) $(X(t))_{t\geq 0}$ *is the unique solution of equation* (2.1) *(up to* P*-indistinguishability).*

Proof. At this point we refer to [GK96] on page 157. First of all, we note that Corollary 4.3 on page 156 still has to be proved before. Then the reader can comprehend that the main idea is to prove Lemma 5.1, which is stated on page 157. □

A. Appendix

The Appendix contains several fundamental lemmas and theorems, which we use within this thesis. On the one hand we will recall well-known basic facts that are mentioned here to ensure the completeness and comprehensibility of the proofs and also to state their intended version exactly (e.g. Theorem A.20, A.21, A.3 and A.6). But on the other hand we will also prove some helpful assertions (e.g. Lemma A.14 and A.22) that have been extracted from the previous chapters, for example due to their length or simplicity.

A.1. Basic theorems

At the beginning we will recall some basic theorems like the continuous mapping theorems, a generalised Young inequality and a generalised Minkowski inequality for integrals, where we usually give references for their proofs.

Theorem A.1 (Continuous mapping theorem, convergence in distribution)**.** *Let X, Y be topological spaces and $f\colon X \longrightarrow Y$ be a continuous function. Let μ_n, for $n \in \mathbb{N}$, and μ be distributions on X such that $\mu_n \xrightarrow[n\to\infty]{w} \mu$. Then on Y we have $\mu_n \circ f^{-1} \xrightarrow[n\to\infty]{w} \mu \circ f^{-1}$ for the image distributions.*

Proof. We refer to [Dud02], Theorem 9.3.7 on page 296. $\square$

Remark. Note that for a sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, P)$ with $Z_n \xrightarrow[n\to\infty]{d} Z$, i.e. $P_{Z_n} \xrightarrow[n\to\infty]{w} P_Z$, we have $f(Z_n) \xrightarrow[n\to\infty]{d} f(Z)$ for continuous functions f because

$$P_{f(Z_n)}[A] = P\big[f(Z_n) \in A\big] = P\big[Z_n \in f^{-1}(A)\big] = P_{Z_n}\big[f^{-1}(A)\big] = \big(P_{Z_n} \circ f^{-1}\big)[A]$$

and, therefore, $P_{f(Z_n)} = P_{Z_n} \circ f^{-1}$ holds.

Theorem A.2 (Continuous mapping theorem, P-a.s. convergence and convergence in probability)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(S_1, \rho_1), (S_2, \rho_2)$ be separable metric spaces and let $f\colon S_1 \longrightarrow S_2$ be a continuous function. Assume that $(Z_n)_{n\in\mathbb{N}}$ is a sequence of S_1-valued random variables. Then the assertions*

i) $Z_n \xrightarrow[n\to\infty]{p} Z$ *implies* $f\big(Z_n\big) \xrightarrow[n\to\infty]{p} f\big(Z\big)$,

ii) $Z_n \xrightarrow[n\to\infty]{P-a.s.} Z$ *implies* $f\big(Z_n\big) \xrightarrow[n\to\infty]{P-a.s.} f\big(Z\big)$

hold.

The proof is an adapted version of the one stated in [vdV98] (see Theorem 2.3 on page 8).

Proof. (cf. [vdV98] on page 8)
"*i*)" Let $\varepsilon > 0$. For every $\delta > 0$ define the set

$$B_\delta := \Big\{ x \in S_1 \,\Big|\, \exists\, y \in S_1 :\ \rho_1(x,y) < \delta \text{ and } \rho_2\big(f(x), f(y)\big) \geq \varepsilon \Big\}$$

for which $\delta \downarrow 0$ implies $B_\delta \downarrow \emptyset$. Note that $Z \notin B_\delta$ and $\rho_2\big(f(Z_n), f(Z)\big) \geq \varepsilon$ imply $\rho_1\big(Z_n, Z\big) \geq \delta$. Hence,

$$P\Big[\rho_2\big(f(Z_n), f(Z)\big) \geq \varepsilon\Big] \leq P\big[Z \in B_\delta\big] + P\Big[\rho_1\big(Z_n, Z\big) \geq \delta\Big].$$

The second summand converges to zero for $n \to \infty$ by assumption. By letting $\delta \downarrow 0$ the first summand also tends to zero since $B_\delta \downarrow \emptyset$.
"*ii*)" Note that $\lim\limits_{n\to\infty} Z_n(\omega) = Z(\omega)$ for $\omega \in \Omega$ implies $\lim\limits_{n\to\infty} f\big(Z_n(\omega)\big) = f\big(Z(\omega)\big)$ by the continuity of f. Hence,

$$P\Big[\lim_{n\to\infty} \rho_2\big(f(Z_n), f(Z)\big)\Big] \geq P\Big[\lim_{n\to\infty} \rho_1\big(Z_n, Z\big)\Big] = 1.$$

□

Lemma A.3 (Generalised Young inequality)**.** *Let X be a vector space, $p \geq 1$ and let $f\colon X \longrightarrow \mathbb{R}$ be a convex function which is homogeneous of degree p (i.e. $f(\alpha x) = \alpha^p f(x)$ for every $\alpha > 0$, $x \in X$). Then*

$$f(a+b) \leq 2^{p-1}\left(f(a) + f(b)\right)$$

for every $a, b \in X$.

In particular, for a normed space $(X, \|\cdot\|)$ we have $\|a+b\|^p \leq 2^{p-1}\left(\|a\|^p + \|b\|^p\right)$ for every $a, b \in X$.

Proof. We compute

$$f(a+b) = 2^p\, f\Big(\frac{1}{2}\,a + \frac{1}{2}\,b\Big) \leq 2^p\left(\frac{1}{2}\,f(a) + \frac{1}{2}\,f(b)\right) = 2^{p-1}\left(f(a) + f(b)\right)$$

by using the homogeneity in the first and the convexity in the second step. □

Lemma A.4 (Reverse Fatou lemma)**.** *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of non-negative, $\bar{\mathbb{R}}$-valued, measurable functions on a measure space $(S, \mathcal{S}, \mu)$. Suppose there exists a function $g \in L^1(S; \bar{\mathbb{R}})$ such that $f_n \leq g$ for every $n \in \mathbb{N}$. Then*

$$\limsup_{n\to\infty} \int_S f_n \,\mathrm{d}\mu \leq \int_S \limsup_{n\to\infty} f_n \,\mathrm{d}\mu$$

holds.

Proof. This version follows immediately from the original Fatou lemma (see e.g. [Dud02], Lemma 4.3.3 on page 131) by considering $g - f_n$. □

Now we have to mention the space $\mathcal{N}_W$ from the general stochastic integration theory in [PR07] (see Section 2.3 starting on page 21) and fit it in our framework such that we can state a Burkholder-Davis-Gundy type inequality afterwards. Therefore, we define

$$\mathcal{N}_W\big(0,T;\mathbb{R}^d\big) := \left\{\Phi\colon [0,T]\times\Omega \longrightarrow \mathbb{R}^{d\times d_1} \,\middle|\, \Phi \text{ is predictable and } P\bigg[\int_0^T \big\|\Phi(s)\big\|_{L_2}^2 \,\mathrm{d}s < \infty\bigg] = 1\right\}$$

as in [PR07] on page 30. Recall that

$$\mathcal{P}_T = \sigma\Big(Y\colon [0,T]\times\Omega \longrightarrow \mathbb{R} \,\Big|\, Y \text{ is left-continuous and adapted to } \mathcal{F}_t, t\in[0,T]\Big)$$

is the so-called *predictable σ-algebra*, and for any separable Hilbert space H a process $Y\colon [0,T]\times\Omega \longrightarrow H$ is said to be (H-)predictable if it is $\mathcal{P}_T/\mathcal{B}(H)$-measurable.

Lemma A.5 (Burkholder-Davis-Gundy type inequality)**.** *Assume that $p \geq 2$ and $\Phi \in \mathcal{N}_W\big(0,T;\mathbb{R}^d\big)$. Then we have*

$$\mathbb{E}\bigg[\sup_{t\in[0,T]}\Big\|\int_0^t \Phi(s)\,\mathrm{d}W(s)\Big\|_{\mathbb{R}^d}^p\bigg]^{\frac{1}{p}} \leq p\left(\frac{p}{2(p-1)}\right)^{\frac{1}{2}}\bigg(\int_0^T \mathbb{E}\Big[\big\|\Phi(s)\big\|_{L_2}^p\Big]^{\frac{2}{p}}\,\mathrm{d}s\bigg)^{\frac{1}{2}}.$$

At this point we recall that the predictability assumption on $\Phi \in \mathcal{N}_W\big(0,T;\mathbb{R}^d\big)$ can be replaced by assuming progessive measurability, i.e. $\Phi|_{[0,t]\times\Omega}$ is $\mathcal{B}([0,t])\otimes\mathcal{F}_t/\mathcal{B}\big(\mathbb{R}^{d\times d_1}\big)$-measurable for every $t\in[0,T]$, since we consider the Wiener process W as an integrator. For further details we refer to [PR07] on page 42.

Proof. We refer to [DZ92], Lemma 7.7 on page 195. □

We call $(X,\mathcal{A},\mu)$ a σ-finite measure space if it is the countable union of $\mathcal{A}$-measurable sets with finite measure. As usual, we define the essential supremum of a function $f\colon X\longrightarrow \bar{\mathbb{R}}$ by

$$\operatorname*{ess\,sup}_{x\in X} f(x) := \inf\Big\{a\in\mathbb{R}\,\Big|\,\mu\big[x\in X \mid f(x) > a\big] = 0\Big\}.$$

Theorem A.6 (Generalised Minkowski integral inequality, cf. [Sch07], Theorem 13.14 on page 130)**.** *Let $(X,\mathcal{A}_1,\mu)$ and $(Y,\mathcal{A}_2,\nu)$ be σ-finite measure spaces and $f\colon X\times Y\longrightarrow \bar{\mathbb{R}}$ be a $\mathcal{A}_1\otimes\mathcal{A}_2$-measurable function. Then*

$$\bigg(\int_X\bigg(\int_Y |f(x,y)|\,\mathrm{d}\nu(y)\bigg)^p\mathrm{d}\mu(x)\bigg)^{\frac{1}{p}} \leq \int_Y\bigg(\int_X |f(x,y)|^p\,\mathrm{d}\mu(x)\bigg)^{\frac{1}{p}}\mathrm{d}\nu(y)$$

holds for every $p\in[1,\infty[$, with equality for $p=1$.
For $p=\infty$ we have the modified inequality

$$\operatorname*{ess\,sup}_{x\in X}\int_Y |f(x,y)|\,\mathrm{d}\nu(y) \leq \int_Y \operatorname*{ess\,sup}_{x\in X}|f(x,y)|\,\mathrm{d}\nu(y).$$

Proof. For $p \in [1, \infty[$ we refer to [Sch07], Theorem 13.14 on page 130 or [HLP67], Theorem 202 on page 148. Observe that the inequality is obvious in the case $p = \infty$. $\square$

Lemma A.7. *Let $A \in \mathbb{R}^{n \times m}$ be a matrix such that*

$$\varepsilon \text{ Id} \leq AA^T \leq K \text{ Id}$$

for some constants $\varepsilon, K > 0$. Then the inequality

$$\varepsilon \leq \|A\|_{L_2}^2 \leq \min\{n, m\} K$$

holds.

Proof. In this proof we could choose any induced matrix norm (operator norm) as a help for the estimate, but we consider a special one for convenience. Therefore, let $\|\cdot\|_2$ be the spectral norm for matrices, i.e. $\|A\|_2 := \sup\limits_{\|x\|_{\mathbb{R}^m}=1} \|Ax\|_{\mathbb{R}^n}$. Then we have $\|A\|_2 \leq \|A\|_{L_2} \leq \sqrt{\min\{n, m\}}\, \|A\|_2$ (cf. [GL13], Inequality (2.3.7) on page 72) and

$$\varepsilon = \varepsilon \sup_{\|x\|_{\mathbb{R}^n}=1} \|x\|_{\mathbb{R}^n}^2 \leq \sup_{\|x\|_{\mathbb{R}^n}=1} \langle AA^T x, x\rangle_{\mathbb{R}^n} \leq K \sup_{\|x\|_{\mathbb{R}^n}=1} \|x\|_{\mathbb{R}^n}^2 = K.$$

Since

$$\sup_{\|x\|_{\mathbb{R}^n}=1} \langle AA^T x, x\rangle_{\mathbb{R}^n} = \sup_{\|x\|_{\mathbb{R}^n}=1} \langle A^T x, A^T x\rangle_{\mathbb{R}^m} = \sup_{\|x\|_{\mathbb{R}^n}=1} \|A^T x\|_{\mathbb{R}^m}^2 = \|A^T\|_2^2 = \|A\|_2^2,$$

it follows that

$$\varepsilon \leq \|A\|_2^2 \leq \|A\|_{L_2}^2 \leq \min\{n, m\}\|A\|_2^2 \leq \min\{n, m\} K$$

holds. $\square$

Finally, we state a well-known lemma about some fundamental inequalities for lim sup, lim inf and probability measures, that the reader should keep in mind.

Lemma A.8. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A_i \in \mathcal{F}$ for every $i \in \mathbb{N}$. Then we have*

i) $\limsup\limits_{n\to\infty} P[A_n] \leq P\big[\limsup\limits_{n\to\infty} A_n\big]$,

ii) $P\big[\liminf\limits_{n\to\infty} A_n\big] \leq \liminf\limits_{n\to\infty} P[A_n]$.

Proof.
"*i)*" Since $\bigcup\limits_{m\geq n} A_m$ is a decreasing sequence in n, we conclude by the continuity from above of P that

$$P\Big[\limsup_{n\to\infty} A_n\Big] = P\Big[\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m\Big] = \lim_{n\to\infty} P\Big[\bigcup_{m\geq n} A_m\Big] \geq \underbrace{\lim_{n\to\infty} \sup_{m\geq n} P[A_m]}_{= \limsup\limits_{n\to\infty} P[A_n]}$$

holds. Assertion "*ii)*" can be proved analogously. $\square$

A.2. Itô's formula and Itô's product rule

In preparation of the application of the well-known Itô formula and Itô product rule for semimartingales, we recall the corresponding theorems from Karatzas/Shreve in [KS05] and Revuz/Yor in [RY99]. Besides, we deduce in Corollary A.10 the explicit representation of the Itô formula for processes which satisfy the SDE (2.1).

Theorem A.9 (cf. [KS05], Theorem 3.6 on page 153). *Let $(Z(t))_{t\geq 0}$ be an $(\mathcal{F}_t)$-adapted, $\mathbb{R}^d$-valued continuous local martingale with $Z(0) = 0$, $(B(t))_{t\geq 0}$ be an $(\mathcal{F}_t)$-adapted, $\mathbb{R}^d$-valued process of bounded variation with $B(0) = 0$ and Y_0 be an $\mathcal{F}_0$-measurable random vector with values in $\mathbb{R}^d$. Set $Y(t) := Y_0 + Z(t) + B(t)$ for $t \in [0,\infty[$ and let $F \in C^{1,2}\big(\mathbb{R}_+ \times \mathbb{R}^d;\mathbb{R}\big)$. Then we have P-a.s.*

$$\begin{aligned}
F(t,Y(t)) = F(0,Y_0) &+ \int_0^t \frac{\partial}{\partial t} F(s,Y(s))\,\mathrm{d}s \\
&+ \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x^i} F(s,Y(s))\,\mathrm{d}B^i(s) \\
&+ \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x^i} F(s,Y(s))\,\mathrm{d}Z^i(s) \\
&+ \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \int_0^t \frac{\partial}{\partial x^i \partial x^j} F(s,Y(s))\,\mathrm{d}\langle Z^i, Z^j\rangle_s
\end{aligned}$$

for all $t \in [0,\infty[$.

Corollary A.10. *Let $k \in \mathbb{N}$ and let $(X(t))_{t\geq 0}$ be a process satisfying SDE (2.1) for every $t \leq \tau^k := \inf\big\{t \geq 0 \,\big|\, X(t) \notin D_k\big\} \wedge k$. Suppose that the assumptions from Section 2.1 are fulfilled and that $F \in C^{1,2}\big(\mathbb{R}_+ \times D;\mathbb{R}\big)$. Then we have P-a.s.*

$$\begin{aligned}
F(t,X(t)) = F(0,X(0)) &+ \int_0^t \frac{\partial}{\partial t} F(s,X(s))\,\mathrm{d}s \\
&+ \sum_{i=1}^d \int_0^t b_i(s,X(s)) \frac{\partial}{\partial x^i} F(s,X(s))\,\mathrm{d}s \\
&+ \int_0^t \big\langle \nabla_x F(s,X(s)), \sigma(s,X(s))\,\mathrm{d}W(s)\big\rangle_{\mathbb{R}^d} \\
&+ \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \int_0^t (\sigma\sigma^T)_{ij}(s,X(s)) \frac{\partial}{\partial x^i \partial x^j} F(s,X(s))\,\mathrm{d}s
\end{aligned}$$

for all $t \in [0,\tau^k]$.

Proof. By satisfying SDE (2.1), we know that $(X(t))_{t\geq 0}$ is a semimartingale with the representation $Y_0 := X(0)$, $B(t) := \int_0^t b(s,X(s))\,\mathrm{d}s$ and $Z(t) := \int_0^t \sigma(s,X(s))\,\mathrm{d}W(s)$ for $t \in [0,\tau^k]$ from Theorem A.9. Note that $(Z(t))_{t\geq 0}$ is a continuous local martingale by construction of the Itô integral and $(B(t))_{t\geq 0}$ is of bounded variation by **A1)**.

Indeed, we obtain

$$\sup_{\Pi} \sum_{i\colon s_{i+1}\in\Pi} \big\|B(s_{i+1}) - B(s_i)\big\|_{\mathbb{R}^d} = \sup_{\Pi} \sum_{i\colon s_{i+1}\in\Pi} \Big\| \int_{s_i}^{s_{i+1}} b(s, X(s))\,\mathrm{d}s\Big\|_{\mathbb{R}^d}$$
$$\leq \int_0^t \big\|b(s, X(s))\big\|_{\mathbb{R}^d}\,\mathrm{d}s \leq \int_0^t M_k(s)\,\mathrm{d}s < \infty,$$

where Π is a partition of $[0,t]$. Furthermore, we have

$$B^i(t) = \int_0^t b_i(s, X(s))\,\mathrm{d}s \quad \text{and} \quad Z^i(t) = \sum_{k=1}^{d_1} \int_0^t \sigma_{ik}(s, X(s))\,\mathrm{d}W^k(s)$$

for $i = 1, \ldots, d$. Therefore, we use

$$\mathrm{d}B^i(s) = b_i(s, X(s))\,\mathrm{d}s \quad \text{as well as} \quad \mathrm{d}Z^i(s) = \sum_{k=1}^{d_1} \sigma_{ik}(s, X(s))\,\mathrm{d}W^k(s)$$

and compute the covariation

$$\begin{aligned}
\langle Z^i, Z^j\rangle_s &= \Big\langle \sum_{k=1}^{d_1} \int_0^{\cdot} \sigma_{ik}(r, X(r))\,\mathrm{d}W^k(r), \sum_{\tilde{k}=1}^{d_1} \int_0^{\cdot} \sigma_{j\tilde{k}}(r, X(r))\,\mathrm{d}W^{\tilde{k}}(r)\Big\rangle_s \\
&= \sum_{k=1}^{d_1}\sum_{\tilde{k}=1}^{d_1} \int_0^s \sigma_{ik}(r, X(r))\sigma_{j\tilde{k}}(r, X(r))\, \underbrace{\mathrm{d}\langle W^k, W^{\tilde{k}}\rangle_r}_{=\delta_{k\tilde{k}}\,\mathrm{d}r} \\
&= \int_0^s \sum_{k=1}^{d_1} \sigma_{ik}(r, X(r))\sigma_{jk}(r, X(r))\,\mathrm{d}r \\
&= \int_0^s (\sigma\sigma^T)_{ij}(r, X(r))\,\mathrm{d}r,
\end{aligned}$$

where $\delta_{k\tilde{k}}$ means the Kronecker delta of k and $\tilde{k}$. Hence,

$$\mathrm{d}\langle Z^i, Z^j\rangle_s = (\sigma\sigma^T)_{ij}(s, X(s))\,\mathrm{d}s.$$

Finally, we write ∇_x for the gradient in the second component and $\langle\cdot,\cdot\rangle_{\mathbb{R}^d}$ for the Euclidean inner product. Hence,

$$\begin{aligned}
&\sum_{i=1}^{d}\sum_{k=1}^{d_1} \int_0^t \frac{\partial}{\partial x^i} F(s, X(s))\sigma_{ik}(s, X(s))\,\mathrm{d}W^k(s) \\
&\quad = \int_0^t \big\langle \nabla_x F(s, X(s)), \sigma(s, X(s))\,\mathrm{d}W(s)\big\rangle_{\mathbb{R}^d}.
\end{aligned}$$

□

Theorem A.11 (cf. [RY99], Proposition 3.1 on page 146)**.** *Let $(Y_1(t))_{t\geq 0}$ and $(Y_2(t))_{t\geq 0}$ be continuous semimartingales on a probability space $(\Omega, \mathcal{F}, P)$. Then we have P-a.s.*

$$Y_1(t)\,Y_2(t) = Y_1(0)\,Y_2(0) + \int_0^t Y_1(s)\,\mathrm{d}Y_2(s) + \int_0^t Y_2(s)\,\mathrm{d}Y_1(s) + \langle Y_1, Y_2\rangle_t.$$

Proof. We refer to [RY99], Proposition 3.1 on page 146. □

A.3. Convergence

In this section we will state Lemma A.12 and A.13 about implications between the different modes of convergence of sequences of random variables, that are used in this thesis. Furthermore, we will also prove Lemma A.14 and A.15, which we apply in the proof of Theorem 3.7.

Lemma A.12 (Implications between modes of convergence, cf. [vdV98], Theorem 2.7 on page 10)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, (S, ρ) be a separable metric space and let $c \in S$. Assume that Z and Z_n, for $n \in \mathbb{N}$, are S-valued random variables on $(\Omega, \mathcal{F}, P)$. Then the following assertions hold.*

i) $Z_n \xrightarrow[n\to\infty]{P-a.s.} Z$ *implies* $Z_n \xrightarrow[n\to\infty]{p} Z$,

ii) $Z_n \xrightarrow[n\to\infty]{p} Z$ *implies* $Z_n \xrightarrow[n\to\infty]{d} Z$,

iii) $Z_n \xrightarrow[n\to\infty]{d} c$ *if and only if* $Z_n \xrightarrow[n\to\infty]{p} c$.

Proof. We refer to [vdV98], Theorem 2.7 on page 10. □

Lemma A.13 (cf. [vdV98], Theorem 2.7 vi) on page 10)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and (S_1, ρ_1), (S_2, ρ_2) be separable metric spaces. Assume that Z^1, Z_n^1, for $n \in \mathbb{N}$, are S_1-valued and Z^2, Z_n^2, for $n \in \mathbb{N}$, are S_2-valued random variables on $(\Omega, \mathcal{F}, P)$ such that $Z_n^i \xrightarrow[n\to\infty]{p} Z^i$ for $i = 1, 2$. Then we also have the convergence of the joint random variable (Z_n^1, Z_n^2), i.e. $(Z_n^1, Z_n^2) \xrightarrow[n\to\infty]{p} (Z^1, Z^2)$.*

Proof. (cf. [vdV98], Theorem 2.7 vi) on page 10)
Let ρ^* be the metric on the product space $S_1 \times S_2$ given by

$$\rho^*\big((X_n, Y_n), (X, Y)\big) := \rho_1\big(X_n, X\big) + \rho_2\big(Y_n, Y\big).$$

Then we have

$$P\Big[\rho^*\big((Z_n^1, Z_n^2), (Z^1, Z^2)\big) \geq \varepsilon\Big] \leq P\Big[\rho_1\big(Z_n^1, Z^1\big) \geq \frac{\varepsilon}{2}\Big] + P\Big[\rho_2\big(Z_n^2, Z^2\big) \geq \frac{\varepsilon}{2}\Big]$$

such that the assertion follows by the assumed convergence of the individual sequences. □

Lemma A.14. *Let $(a_n)_{n\in\mathbb{N}}$ be a $[0,1]$-valued sequence and $c \in [0,1]$. If for every subsequence $(a_{n_m})_{m\in\mathbb{N}}$ of $(a_n)_{n\in\mathbb{N}}$ there exists a subsequence $(a_{n_{m_j}})_{j\in\mathbb{N}}$ such that $\limsup_{j\to\infty} a_{n_{m_j}} \leq c$, then $\limsup_{n\to\infty} a_n \leq c$.*

Proof. Assume that $d := \limsup_{n\to\infty} a_n > c$. Then we can choose a subsequence $(a_{n_m})_{m\in\mathbb{N}}$ such that $|a_{n_m} - d| \leq \frac{d-c}{2}$ for every $m \in \mathbb{N}$, since d is an accumulation point. But then there must exist a subsequence $(a_{n_{m_j}})_{j\in\mathbb{N}}$ such that $\limsup_{j\to\infty} a_{n_{m_j}} \leq c$. That is impossible by our choice of $(a_{n_m})_{m\in\mathbb{N}}$. Hence, $\limsup_{n\to\infty} a_n \leq c$. □

Lemma A.15. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $r \in \mathbb{N}$ and $d_i \in \mathbb{N}$ for $1 \leq i \leq r$. For $i = 1, \cdots, r$ let $\big(X_n^i\big)_{n \in \mathbb{N}}$ be sequences of stochastic processes such that $X_n^i \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^{d_i}\big)$ for every $n \in \mathbb{N}$. Furthermore, assume that $X^i \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^{d_i}\big)$ are stochastic processes such that one of the following convergences*

i) $\big(X_n^1, \ldots, X_n^r\big) \xrightarrow[n\to\infty]{d} \big(X^1, \ldots, X^r\big)$,

ii) $\big(X_n^1, \ldots, X_n^r\big) \xrightarrow[n\to\infty]{p} \big(X^1, \ldots, X^r\big)$,

iii) $\big(X_n^1, \ldots, X_n^r\big) \xrightarrow[n\to\infty]{P-a.s.} \big(X^1, \ldots, X^r\big)$

holds. In each of these cases we have for every $q \in \mathbb{N}$ with $q \leq r$ and every $\big\{j_1, \ldots, j_q\big\} \subseteq \big\{1, \ldots, r\big\}$, where $j_l \neq j_\ell$ for $l \neq \ell$, that

i) $\big(X_n^{j_1}, \ldots, X_n^{j_q}\big) \xrightarrow[n\to\infty]{d} \big(X^{j_1}, \ldots, X^{j_q}\big)$,

ii) $\big(X_n^{j_1}, \ldots, X_n^{j_q}\big) \xrightarrow[n\to\infty]{p} \big(X^{j_1}, \ldots, X^{j_q}\big)$,

iii) $\big(X_n^{j_1}, \ldots, X_n^{j_q}\big) \xrightarrow[n\to\infty]{P-a.s.} \big(X^{j_1}, \ldots, X^{j_q}\big)$.

Proof. We consider the map

$$\phi \colon \begin{array}{c} C\big([0,T];\mathbb{R}^{d_1}\big) \times \cdots \times C\big([0,T];\mathbb{R}^{d_r}\big) \longrightarrow C\big([0,T];\mathbb{R}^{d_{j_1}}\big) \times \cdots \times C\big([0,T];\mathbb{R}^{d_{j_q}}\big), \\ (f^1, \ldots, f^r) \longmapsto (f^{j_1}, \ldots, f^{j_q}), \end{array}$$

which is a continuous function. Hence, by the continuous mapping theorems (cf. Theorem A.1 and Theorem A.2) we obtain the assertion. $\square$

A.4. Tightness

In this section we will consider some necessary facts concerning tightness. This includes on the one hand the tightness criteria A.20 and A.21 as well as on the other hand the important Lemma A.22 about tightness of joint distributions.

First of all, we recall the definition of a tight family of probability measures and Prokhorov's theorem about the implications between tightness and relative compactness.

Definition A.16 (tight). *A family $\mathcal{M}$ of probability measures on a metric space (S, ρ) is called tight if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq S$ such that*

$$\mu(K_\varepsilon) \geq 1 - \varepsilon$$

for all $\mu \in \mathcal{M}$.

Theorem A.17 (Prokhorov). *Let (S, ρ) be a metric space and $\mathcal{M} \subseteq \mathcal{M}_1(S)$ be a family of probability measures. Then the tightness of the family $\mathcal{M}$ implies the relative compactness of $\mathcal{M}$. If S is a Polish space, these properties are equivalent.*

Proof. We refer to [Dur96], Chapter 8.2 starting on page 276 and [Bil99], Theorem 5.1 on page 59 and Theorem 5.2 on page 60. □

Lemma A.18. *Let (S, ρ) be a metric space and $\mathcal{S}$ be the Borel-σ-algebra on S. If S is separable and complete, then every single probability measure on $(S, \mathcal{S})$ is tight.*

Proof. We refer to [Bil99], Theorem 1.3 on page 7. □

Let $T \in [0, \infty[$ and $d \in \mathbb{N}$. Then the space $C\big([0,T]; \mathbb{R}^d\big)$ of continuous functions equipped with the supremum norm $\|\cdot\|_\infty$ is a separable and complete normed space.

Now we define the modulus of continuity that is used in the following tightness criterion.

Definition A.19 (Modulus of continuity). *For every $T \geq 0$, $f \in C\big([0,T]; \mathbb{R}^d\big)$ and $\delta > 0$ define a so-called modulus of continuity by*

$$\mathcal{W}_\delta(f) := \sup_{\substack{s,t \in [0,T] \\ |s-t| \leq \delta}} \|f(s) - f(t)\|_{\mathbb{R}^d}.$$

Theorem A.20 (Tightness criterion). *A sequence of probability measures $(\mu_n)_{n \in \mathbb{N}}$ on $C\big([0,T]; \mathbb{R}^d\big)$ is tight if and only if for every $\varepsilon > 0$ there exist constants $n_0 \in \mathbb{N}$, $R > 0$ and $\delta > 0$ such that*

i) $\mu_n\Big[\big\{f \in C\big([0,T]; \mathbb{R}^d\big) \mid \|f(0)\|_{\mathbb{R}^d} > R\big\}\Big] \leq \varepsilon,$

ii) $\mu_n\Big[\big\{f \in C\big([0,T]; \mathbb{R}^d\big) \mid \mathcal{W}_\delta(f) > \varepsilon\big\}\Big] \leq \varepsilon$

hold for every $n \geq n_0$.

Remark. Observe that assertion ii) of Theorem A.20 can be stated equivalently with possibly different $\eta, \xi > 0$ instead of taking the same $\varepsilon > 0$, since ii) certainly holds for $\varepsilon = \eta \wedge \xi$ in that case.

Remark (cf. [Bil99], Theorem 7.3 and Equation (7.8) on page 82)**.** We can restate the assertions i) and ii) in a more compact form, i.e.

i') $\lim\limits_{R\uparrow\infty} \limsup\limits_{n\to\infty} \mu_n\Big[\big\{f \in C\big([0,T];\mathbb{R}^d\big) \mid \|f(0)\|_{\mathbb{R}^d} > R\big\}\Big] = 0,$

ii') $\lim\limits_{\delta\downarrow 0} \limsup\limits_{n\to\infty} \mu_n\big[\mathcal{W}_\delta > \varepsilon\big] = 0$ for every $\varepsilon > 0$.

Proof of Theorem A.20. We refer to [Dur96], Theorem 3.4 on page 284 for a generalisable version of a proof for the space $C\big([0,1];\mathbb{R}^d\big)$. □

If we consider the distributions of stochastic processes with continuous sample paths, we can restate condition ii) from Theorem A.20 in another different way. The new condition is closely related to the theorem of Kolmogorov-Chentsov.

Theorem A.21 (Tightness criterion for distributions of stochastic processes with continuous sample paths)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(X_n)_{n\in\mathbb{N}}$ be a sequence of stochastic processes with $X_n \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^d\big)$. If there exist some constants $\alpha, \beta, K > 0$ such that the inequality*

$$\mathbb{E}\Big[\big\|X_n(t) - X_n(s)\big\|_{\mathbb{R}^d}^\beta\Big] \leq K|t-s|^{1+\alpha}$$

is fulfilled for every $n \in \mathbb{N}$ and $s,t \in [0,T]$, then assertion ii) from the tightness criterion A.20 holds for the sequence of distributions $(P_{X_n})_{n\in\mathbb{N}}$.

Proof. We refer to [KS05], Theorem 4.10 and Problem 4.11 on page 63-64 or [Dur96], Theorem 3.6 on page 284 for proofs in similar settings. □

Finally, we state and prove the previously mentioned Lemma A.22 about the tightness of joint distributions.

Lemma A.22. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, $T \geq 0$, $d_1, d_2 \in \mathbb{N}$ and assume that $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ are sequences of stochastic processes such that $X_n \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^{d_1}\big)$ and $Y_n \colon \Omega \longrightarrow C\big([0,T];\mathbb{R}^{d_2}\big)$ for every $n \in \mathbb{N}$. If both sequences of distributions $(P_{X_n})_{n\in\mathbb{N}}$ and $(P_{Y_n})_{n\in\mathbb{N}}$ are tight, then the sequence of joint distributions $(P_{(X_n,Y_n)})_{n\in\mathbb{N}}$ is a tight family of measures on $C\big([0,T];\mathbb{R}^{d_1+d_2}\big)$.*

Proof. We will just verify the two conditions from Theorem A.20. Let $\varepsilon > 0$.
"*i)*": Choose $n_1 \in \mathbb{N}$ and $R > 0$ such that

$$\begin{aligned} P\left[\|X_n(0)\|_{\mathbb{R}^{d_1}}^2 > \frac{R^2}{2}\right] &= P\left[\|X_n(0)\|_{\mathbb{R}^{d_1}} > \frac{R}{\sqrt{2}}\right] \leq \frac{\varepsilon}{2} \quad \text{and} \\ P\left[\|Y_n(0)\|_{\mathbb{R}^{d_2}}^2 > \frac{R^2}{2}\right] &= P\left[\|Y_n(0)\|_{\mathbb{R}^{d_2}} > \frac{R}{\sqrt{2}}\right] \leq \frac{\varepsilon}{2} \end{aligned} \tag{A.1}$$

for every $n \geq n_1$. Then

$$\begin{aligned}
P\left[\left\| \begin{pmatrix} X_n(0) \\ Y_n(0) \end{pmatrix} \right\|_{\mathbb{R}^{d_1+d_2}} > R\right] &= P\left[\left\| \begin{pmatrix} X_n(0) \\ Y_n(0) \end{pmatrix} \right\|^2_{\mathbb{R}^{d_1+d_2}} > R^2\right] \\
&= P\Big[\|X_n(0)\|^2_{\mathbb{R}^{d_1}} + \|Y_n(0)\|^2_{\mathbb{R}^{d_2}} > R^2\Big] \\
&\leq P\left[\|X_n(0)\|^2_{\mathbb{R}^{d_1}} > \frac{R^2}{2}\right] + P\left[\|Y_n(0)\|^2_{\mathbb{R}^{d_2}} > \frac{R^2}{2}\right] \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

follows from (A.1) for all $n \geq n_1$.
"$ii)$": Choose $n_2 \in \mathbb{N}$ and $\delta > 0$ such that

$$P\Big[\mathcal{W}_\delta(X_n) > \frac{\varepsilon}{\sqrt{2}}\Big] \leq \frac{\varepsilon}{2} \quad \text{and} \quad P\Big[\mathcal{W}_\delta(X_n) > \frac{\varepsilon}{\sqrt{2}}\Big] \leq \frac{\varepsilon}{2} \tag{A.2}$$

for every $n \geq n_2$. Then by using the monotonicity of the square function, it follows

$$\begin{aligned}
&P\Big[\mathcal{W}_\delta\big((X_n, Y_n)\big) > \varepsilon\Big] \\
&\quad = P\Big[\mathcal{W}_\delta\big((X_n, Y_n)\big)^2 > \varepsilon^2\Big] \\
&\quad = P\left[\left(\sup_{\substack{s,t\in[0,T] \\ |s-t|\leq\delta}} \left\| \begin{pmatrix} X_n(t) - X_n(s) \\ Y_n(t) - Y_n(s) \end{pmatrix} \right\|_{\mathbb{R}^{d_1+d_2}}\right)^2 > \varepsilon^2\right] \\
&\quad = P\left[\sup_{\substack{s,t\in[0,T] \\ |s-t|\leq\delta}} \left\| \begin{pmatrix} X_n(t) - X_n(s) \\ Y_n(t) - Y_n(s) \end{pmatrix} \right\|^2_{\mathbb{R}^{d_1+d_2}} > \varepsilon^2\right] \\
&\quad = P\left[\sup_{\substack{s,t\in[0,T] \\ |s-t|\leq\delta}} \Big(\|X_n(t) - X_n(s)\|^2_{\mathbb{R}^{d_1}} + \|Y_n(t) - Y_n(s)\|^2_{\mathbb{R}^{d_2}}\Big) > \varepsilon^2\right] \\
&\quad \leq P\Big[\mathcal{W}_\delta(X_n)^2 + \mathcal{W}_\delta(Y_n)^2 > \varepsilon^2\Big] \\
&\quad \leq P\Big[\mathcal{W}_\delta(X_n)^2 > \frac{\varepsilon^2}{2}\Big] + P\Big[\mathcal{W}_\delta(Y_n)^2 > \frac{\varepsilon^2}{2}\Big].
\end{aligned}$$

Hence by (A.2) we have

$$P\Big[\mathcal{W}_\delta\big((X_n, Y_n)\big) > \varepsilon\Big] \leq P\Big[\mathcal{W}_\delta(X_n) > \frac{\varepsilon}{\sqrt{2}}\Big] + P\Big[\mathcal{W}_\delta(Y_n) > \frac{\varepsilon}{\sqrt{2}}\Big] \leq \varepsilon$$

for every $n \geq n_2$. By taking $n_0 := \max\{n_1, n_2\}$, we finish the proof. □

Bibliography

[AE08] Herbert Amann and Joachim Escher. *Analysis III.* Birkhäuser, Basel, et al., Second edition, 2008.

[Alt12] Hans Wilhelm Alt. *Lineare Funktionalanalysis.* Springer, Berlin, et al., Sixth edition, 2012.

[Bau02] Heinz Bauer. *Wahrscheinlichkeitstheorie.* de Gruyter, Berlin, et al., Fifth edition, 2002.

[Bil95] Patrick Billingsley. *Probability and Measure.* Wiley, New York, et al., Third edition, 1995.

[Bil99] Patrick Billingsley. *Convergence of Probability Measures.* Wiley, New York, et al., Second edition, 1999.

[Bog07] Vladimir I. Bogachev. *Measure Theory.* Springer, Berlin, et al., 2007.

[Dud02] Richard M. Dudley. *Real Analysis and Probability.* Cambridge University Press, Boca Raton, FL, et al., 2002.

[Dur96] Richard Durrett. *Stochastic Calculus.* CRC Press, Boca Raton, FL, et al., 1996.

[DZ92] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, Cambridge, et al., 1992.

[GK96] István Gyöngy and Nicolai V. Krylov. Existence of strong solution for Itô's stochastic equations via approximations. *Probability Theory and Related Fields*, 105:143–158, 1996.

[GL13] Gene H. Golub and Charles F. Van Loan. *Matrix computations.* Johns Hopkins University Press, Baltimore, MD, Fourth edition, 2013.

[Gri09] Alexander Grigor'yan. *Heat Kernel and Analysis on Manifolds.* AMS/IP studies in advanced mathematics; 47. American Math. Soc., et al., Providence, RI, 2009.

[Gyö98] István Gyöngy. A Note on Euler's Approximations. *Potential Analysis*, 8:205–216, 1998.

[HJ85] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, et al., 1985.

[HLP67] Godfrey H. Hardy, John E. Littlewood, and George Pólya. *Inequalities.* Cambridge University Press, Cambridge, et al., Second edition, 1967.

[Kha80] Rafail Z. Khasminskii. *Stochastic Stability of Differential Equations.* Sijthoff & Noordhoff, Alphen aan den Rijn, et al., 1980.

[Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer, Berlin, et al., 2006.

[Kry99] Nicolai V. Krylov. On Kolmogorovs Equations for Finite Dimensional Diffusions. *Stochastic PDEs and Kolmogorov Equations in Infinite Dimensions. Lecture Notes in Mathematics (Cetraro, Italy 1998)*, 1715:1–63, 1999.

[KS05] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, New York, et al., Second edition, 2005.

[PR07] Claudia Prévôt and Michael Röckner. *A Concise Course on Stochastic Partial Differential Equations.* Springer, Berlin, et al., 2007.

[Pro05] Philip E. Protter. *Stochastic Integration and Differential Equations.* Springer, Berlin, et al., Second edition, 2005.

[RY99] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion.* Springer, Berlin, et al., Third edition, 1999.

[Sch07] René L. Schilling. *Measures, Integrals and Martingales.* Cambridge University Press, Cambridge, et al., 2007.

[Shr04] Steven E. Shreve. *Stochastic Calculus for Finance II.* Springer, New York, et al., 2004.

[Sko65] Anatoliy V. Skorokhod. *Studies in the Theory Random Processes.* Addison-Wesley, Reading, Mass., 1965.

[SV79] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional Diffusion Processes.* Springer, Berlin, et al., 1979.

[vdV98] Aad W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, Cambridge, et al., 1998.