TECHNISCHE UNIVERSITÄT WIEN

BACHELORARBEIT

Convergence to equilibrium for the Fokker-Planck equation

Application of the entropy method and ergodic theory of Markov processes

ausgeführt am

Institut für
Analysis und Scientific Computing
TU Wien

unter der Anleitung von

Univ.Prof. Dip.-Ing. Dr.tech. Anton Arnold
und
Univ.Ass. PhD Boris Nectoux

durch

Bernhard Kepka
Matrikelnummer: 01604983
Firmiangasse 37
1130 Wien

Wien, am 3. Juni 2019

Danksagung

Zuerst möchte ich mich bei Prof. Arnold und PhD Nectoux bedanken. Sie haben mich stets unterstützt und sich die Zeit genommen, um hilfreiche (und teilweise lange) Diskussionen zu führen. Dazu kommt, dass mich Prof. Arnold auch weit über diese Arbeit hinaus gefördert hat, wofür ich zutiefst dankbar bin.

Desweiteren möchte ich Prof. Mathias Beiglböck danken, dass er mich seit Beginn meines zweiten Semesters stets gefordert und gefördert hat. Ihm habe ich viel hinsichtlich meiner bisherigen, aber auch meiner noch bevorstehenden Ausbildung zu verdanken.

Zuletzt danke ich meiner Familie für ihren unerschöpflichen Rückhalt. Ohne sie wäre ich nicht so weit gekommen.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Bachelorarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 26. August 2019

Bernhard Kepka

Contents

Abstract

Fokker-Planck equations arise in models in statistical physics which constitute diffusion processes (related to Brownian motion) in the framework of probability theory. In particular, the case of a Brownian motion in a potential raises questions concerning the long-time behavior: Is there a (unique) equilibrium? Does the evolution converge to an equilibrium? If so, can we give a rate of convergence? Such questions emerge for instance when computing averages in statistical physics.

The aim of this Bachelor Thesis is to investigate these questions based on the entropy method (an analytic approach based on techniques in partial differential equations) as well as the theory of Markov processes (a stochastic approach). The connection between the partial differential equation and its corresponding stochastic differential equation allows one to attack this problem from these two angles.

The work is organized as follows. In Chapter 1, we start with a short outline of diffusion models and the context of invariant measures in statistical physics. Then, we will derive the Fokker-Planck equation from stochastic differential equations which are both studied in the following. Furthermore, we discuss the Ornstein-Uhlenbeck process as a toy problem for the study of long-time behavior. Chapter 2 is devoted to the existence and uniqueness of solutions. Here, we will determine the associated semigroup using spectral analysis on one hand. On the other hand, results from stochastic differential equations are used to obtain a solution of the corresponding stochastic differential equation.

In Chapter 3, we will investigate the entropy method in order to study the long-time behavior of solutions. One key result in this investigation is the validity of a convex Sobolev inequality which allows one to extend conclusions to a wider range of situations.

In Chapter 4, we give convergence results from the theory of Markov processes which will then be applied to our particular case. This yields exponential convergence of the the law, subgeometric rates and lower bounds, each under specific conditions. Then, we provide numerical simulations in Chapter 5.

Finally, in the appendix, Chapter A, the reader can find fundamental definitions as well as results concerning self-adjoint operators, Markov processes and stochastic differential equations which are used later on. In addition, some proofs from Chapter 2 and 3 are postponed to Section A.3 and A.4, respectively.

1. Introduction

Diffusion processes arising from stochastic differential equations are used in statistical physics to study models in which both external forces as well as random forces play a significant role. Such models occur for instance in solid-space physics or chemical physics.

In the first section, we will briefly outline two diffusion models described by stochastic differential equations with which we will be concerned in this work. Then, we will briefly recall the context of invariant measures in statistical physics and derive the Fokker-Planck equation, which will be studied in the following chapters too. Finally, we discuss the Ornstein-Uhlenbeck process as a toy problem in the study of long-time behavior.

1.1. A stochastic differential equation in diffusion models

In this work we will consider diffusion processes governed by stochastic differential equations having the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t. \tag{1.1}$$

For instance, the process X_t models positions and/or velocities of particles and $b(X_t)dt$ describes an external force possibly originating from a potential which acts on the particles. The ordinary differential equation $dX_t = b(X_t)dt$ models the major trajectory of the particles. The term $\sigma(X_t)dW_t$, where W_t denotes a Brownian motion, adds random forces due to collisions or thermal fluctuations, for instance, and leads to a perturbation of the trajectory. In the following we will give two examples.

Chemical reactions. It is known that the process X can be used in chemistry to model the positions and/or velocities of atoms, see for instance [SM79, Section 2, p. 605]. Particles rest in a certain configuration with other particles due to molecular bonds which can be viewed as an external force described by the term $b(X_t)dt$. Therefore, the force originate from a potential V, hence $b(X_t) = -\nabla V(X_t)$. When considering forces due to collisions of the particles one adds a white noise via the term $\sigma(X_t)dW_t$.

If the particles form a configuration, the process X_t will oscillate about this equilibrium state. Such a state corresponds to a local minimum of the potential V. However, if the collisions are strong enough such that molecular bonds are broken, then a chemical reaction takes place leading to new configurations. Every possible molecular configuration corresponds to a local minimum of the potential V in $\mathbb{R}^d$. Furthermore, there will be different ways to reach a certain stable equilibrium which lead to different chemical reactions.

Concerning the long-time behavior of the process X_t one might be interested in stable equilibria as well as the speed of reaching those final molecular configurations.

Movements in crystal structures. Yet another model leading to (1.1) is concerned with the migration of impurities or vacant positions in the lattice of a crystal (see [SM79, Section 3, p. 608]). In a crystal atoms are arranged according to a lattice. As an effect of thermal fluctuations they oscillate about their positions. Impurities are stuck between the atoms of the crystal due to atomic forces. However, since the atoms move around their position in the lattice, impurities can squeeze through the crystal structure in order to reach another stable place between atoms. In a similar way, this leads to movements of vacant positions in the lattice. As an effect some of the physical properties of the crystal change.

Like in the previous example atomic forces $b(X_t) = -\nabla V(X_t)$ are modeled via a potential V. Stable positions correspond to local minima of V which can be left due to thermal fluctuations. Furthermore, thermal fluctuations are expressed by a white noise $\sigma(X_t)dW_t$. Moreover, depending on the structure of the crystal certain directions through the lattice are more or less likely, which determines the structure of the matrix $\sigma(X_t)$.

Finally, a study of the long-time behavior might give insights to the physical properties of the crystal in the long run.

1.2. Long-time behavior and invariant measure

In the study of a system with a huge amount of particles it seems impossible to calculate the evolution of each particle both on a conceptional and a numerical level, since most of the initial data is unknown and the number of variables is too large ([Sto15, Subsection 1.3.1, p. 12]). The set of all possible values of the considered degrees of freedom of the particles is called phase space and is denoted by Ω. In statistical physics one is not interested in the actual data (e.g. positions, velocities) of particles, but for instance on averages of the form

$$\mathbb{E}_\mu[A] = \int_\Omega A(q)\, d\mu(q). \tag{1.2}$$

The probability measure μ on Ω enables one to study the system without the whole information. The function A is called an observable which corresponds to some macroscopic effect like temperature or pressure ([Sto15, Subsection 1.1.2, p. 4-5]). The function A might also be the indicator function of some set $B \subset \mathbb{R}^d$ and (1.2) yields the probability of the particles to be found in B.

For the analysis of systems involving in time the probability measure μ is replaced by the law $(\mu_t)_{t\geq 0}$ of some process $(X_t)_{t\geq 0}$. One might ask whether $(\mu_t)_{t\geq 0}$ converges to some equilibrium distribution μ_∞ in the sense that

$$\int_\Omega A(q)\, d\mu_t(q) \to \int_\Omega A(q)\, d\mu_\infty(q) \quad \text{as} \quad t \to \infty.$$

Furthermore, if this is the case, can we also give a rate of speed? For instance, is there an exponential convergence? In such a situation the study of the system (after some ascertained time) can be conducted with μ_∞.

In the study of ergodic properties of Markov processes the measure μ_∞ is an invariant measure of the Markov family $(X_t)_{t\geq 0}$ satisfying $\mathcal{P}_t\mu_\infty = \mu_\infty$ with the Markov kernel $\mathcal{P}_t$

of X. See Section A.2 for the definition and corresponding notions of a Markov process or Markov family.

1.3. Derivation of the Fokker-Planck equation

We consider again the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad \text{or} \quad X_t = X_0 + \int_0^t b(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s$$

for the process $(X_t)_{t\geq 0}$ on $\mathbb{R}^d$, $d \in \mathbb{N}$ the dimension. Here, X_0 is the given initial random variable on $\mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times r}$ for some integer $r \in \mathbb{N}$. The process $(W_t)_{t\geq 0}$ is some r-dimensional Brownian motion. The integral on the right denotes the Itô integral.

Now, our goal is to derive a partial differential equation which is satisfied by the law $(\mu_t)_{t\geq 0}$ of the process $(X_t)_{t\geq 0}$. Therefore, we apply Itô's formula to $\varphi(t, X_t)$ for a test function $\varphi \in C_c^\infty([0, +\infty) \times \mathbb{R}^d)$ and take expectations

$$\begin{aligned}\mathbb{E}[\varphi(t, X_t)] - \mathbb{E}[\varphi(0, X_0)] = \int_0^t \mathbb{E}[\partial_t\varphi(s, X_s)]\,ds + \int_0^t \mathbb{E}[\nabla\varphi(s, X_s) \cdot b(X_s)]\,ds \\ + \sum_{i,j} \int_0^t \mathbb{E}[A_{ij}\partial_{ij}^2\varphi(s, X_s)]\,ds\end{aligned}$$

where we defined the matrix $A := \frac{1}{2}\sigma\sigma^\top \in \mathbb{R}^{d\times d}$. By defining the differential operator

$$L^*\varphi := \sum_{i,j} A_{ij}\partial_{ij}^2\varphi + b \cdot \nabla\varphi$$

we can rewrite this with the law $(\mu_t)_{t\geq 0}$

$$\int_{\mathbb{R}^d} \varphi(t, x)\,d\mu_t(x) - \int_{\mathbb{R}^d} \varphi(0, x)\,d\mu_0(x) = \int_0^t \int_{\mathbb{R}^d} (L^*\varphi(s, x) + \partial_t\varphi(s, x))\,d\mu_t(x)ds.$$

Recall that φ has compact support, so we can send $t \to +\infty$ and obtain

$$-\int_0^{+\infty} \int_{\mathbb{R}^d} \partial_t\varphi(s, x)\,d\mu_t(x)ds - \int_{\mathbb{R}^d} \varphi(0, x)\,d\mu(x) = \int_0^{+\infty} \int_{\mathbb{R}^d} L^*\varphi(s, x)\,d\mu_t(x)ds.$$

Therefore, the law satisfies the *Fokker-Planck equation*

$$\partial_t\mu_t = \sum_{i,j} \partial_{ij}^2(\mu_t\,A_{ij}) + \operatorname{div}(\mu_t b) \tag{1.3}$$

in the distributional sense. This is a partial differential equation of parabolic type and also called *forward Kolmogorov equation* or *Smoluchowski equation* ([Ris96, Subsection 1.2.6]). If the law $(\mu_t)_{t\geq 0}$ has a smooth density $(t, x) \mapsto \rho(t, x)$ with respect to the Lebesgue measure, then ρ satisfied (1.3) in the classical sense.

In our study we will focus on Fokker-Planck equations having the form

$$\partial_t \rho = \operatorname{div}(D(\nabla \rho + \rho \nabla V)) = \operatorname{div}\left(e^{-V} D \nabla \left(\rho e^{V}\right)\right) \tag{1.4}$$

with a diffusion matrix $D : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and a potential $V : \mathbb{R}^d \to \mathbb{R}$. In this case the differential operator L^* from above is

$$L^*\varphi = \sum_{i,j} D_{ij}\partial_{ji}^2\varphi + (\operatorname{div}D - D\nabla V)\cdot\nabla\varphi$$

which is the formal adjoint of the operator on the right in (1.4). As we will see, this operator is the generator of a Markov process satisfying the stochastic differential equation

$$dX_t = (\operatorname{div}D - D\nabla V)(X_t)dt + \sigma(X_t)dW_t \tag{1.5}$$

with $b = \operatorname{div}D - D\nabla V$ and $D = \frac{1}{2}\sigma\sigma^\top$ with the notation from above. Here, $\operatorname{div}D(x) \in \mathbb{R}^d$ is defined by $(\operatorname{div}D)_j := \sum_i \partial_i D_{ij}$.

The stochastic differential equation (1.5) is called *overdamped Langevin equation* ([Sto15, Subsection 4.1.2, p. 57]). The physical interpretation is the same as in Section 1.1. The function V denotes a potential confining particles modeled by the process $(X_t)_{t\geq 0}$, for instance. The matrix σ takes the inhomogeneity of the diffusion into account.

The Fokker-Planck equation (1.4) has e^{-V} as a time-independent solution. In the case $c_V := \int_{\mathbb{R}^d} e^{-V}\,dx < \infty$ the measure $\mu_\infty := \left(e^{-V}/c_V\right)\lambda$ defines an invariant measure for $(X_t)_{t\geq 0}$. Here, λ denotes the Lebesgue measure.

In Chapter 3, we will prove exponential convergence of solutions of (1.4) to e^{-V} in relative entropy (which implies convergence in L^1), based on the entropy method. In Chapter 4, we study the exponential convergence of averages of the form (1.2) (or, as we will refer to it, convergence in weighted L^∞ spaces) as well as subgeometric rates and lower bounds in total variational norm.

1.4. Ornstein-Uhlenbeck Process

In general one cannot calculate solutions of a Fokker-Planck equation explicitly. However, if the diffusion matrix is the identity and the potential is quadratic, one can obtain an explicit solution. Then, the tend to the equilibrium as well as the actual rate of convergence will be explicit from the solution.

We examine the following stochastic differential equation

$$dX_t = -X_t dt + \sqrt{2}dB_t \tag{1.6}$$

and the corresponding partial differential equation

$$\partial_t \rho(t,x) = \operatorname{div}(\nabla\rho(t,x) + x\rho(t,x)) \tag{1.7}$$

satisfied by the density $X_t \sim \rho(t,\cdot)\lambda$ with initial state $X_0 \sim \rho_0\lambda$. The dimension is $d \in \mathbb{N}$, $\rho : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$. The process X_t is called *Ornstein-Uhlenbeck process* ([Ris96, Section 3.2, p. 38], [KS00, Chapter 5, Example 6.8, p. 358]). Let us consider the one-dimensional case $d = 1$. The time-homogeneous solution e^{-V}/c_V from the last section reads $\rho_\infty(x) := e^{-x^2/2}/\sqrt{2\pi}$. In order to solve the stochastic differential equation we make the ansatz

$$X_t = X_0 e^{-t} + e^{-t} \int_0^t a(s) dB_s,$$

which yields together with the product rule

$$dX_t = -X_t dt + e^{-t} a(t) dB_t.$$

By comparing this with (1.6) we obtain $a(t) = \sqrt{2} e^t$ and hence

$$X_t = X_0 e^{-t} + \sqrt{2} \int_0^t e^{s-t} dB_s. \tag{1.8}$$

First of all, observe that $e^{-t} X_0$ has the density $\rho_0(e^t y) e^t$, since

$$\mathbb{P}(e^{-t} X_0 \in A) = \int_{e^t A} f_0(y)\, dy = \int_A f_0(e^t y) e^t\, dy.$$

Furthermore, the stochastic integral on the right in (1.8) is a centered Gaussian random variable, since e^{s-t} is deterministic. In addition, its variance is $(1 - e^{-2t})$ by Itô's isometry.

Now, notice that X_t is the sum of $X_0 e^{-t}$ and a Gaussian random variable for $t > 0$, which are independent. Therefore, the density of X_t is the convolution of the corresponding densities. We obtain

$$\begin{aligned} \rho(t, x) &= \int_{\mathbb{R}} \rho_0(y e^t) e^t \frac{1}{(2\pi(1 - e^{-2t}))^{1/2}} \exp\left(-\frac{(x-y)^2}{2(1 - e^{-2t})}\right) dy \\ &= \int_{\mathbb{R}} \rho_0(y) \frac{1}{(2\pi(1 - e^{-2t}))^{1/2}} \exp\left(-\frac{(x - y e^{-t})^2}{2(1 - e^{-2t})}\right) dy. \end{aligned} \tag{1.9}$$

Note, that we calculated the fundamental solution of (1.7) in (1.9). Formula (1.9) also yields $\rho(t, x) \to \rho_\infty(x)$ as $t \to \infty$ and hence the convergence to the equilibrium. Furthermore, this proves uniqueness of the time-homogeneous solution $\rho_\infty(x) = e^{-x^2/2}/\sqrt{2\pi}$.

In the following we will present another way to determine the solution. Consider the Fokker-Planck equation

$$\partial_t \rho = (\rho' + x\rho)' = (\rho_\infty (\rho \rho_\infty^{-1})')', \tag{1.10}$$

where the prime denotes differentiating with respect to x. Define $g := \rho \rho_\infty^{-1}$ which yields

$$\partial_t g = g'' - x g' = \rho_\infty^{-1} (g' \rho_\infty)' =: Lg. \tag{1.11}$$

The operator L, which is a priori defined on $C_c^\infty(\mathbb{R})$, is symmetric on $L^2(\mathbb{R}, \rho_\infty)$ with scalar product

$$\langle f, g \rangle_{L^2(\mathbb{R}, \rho_\infty)} = \int_{-\infty}^{+\infty} f g \rho_\infty\, dx.$$

Indeed, we have

$$\langle f, Lg\rangle_{L^2(\mathbb{R},\rho_\infty)} = \int_{-\infty}^{+\infty} f(g'\rho_\infty)'\,dx = -\int_{-\infty}^{+\infty} f'g'\rho_\infty\,dx.$$

We will prove in Section 2.1 that L has a self-adjoint extension. More precisely, we will see that the closure of L is self-adjoint, see Proposition 2.2. Now, we will calculate the spectral decomposition of L and therefore consider the following eigenvalue problem

$$\psi'' - x\psi' = \lambda\psi. \tag{1.12}$$

We will prove that the functions, also known as probabilists' Hermite polynomials (see [BGL14, Subsection 2.7.1, p. 105-107] and [Tay11, p. 127-128])

$$H_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}, \quad \lambda_n = -n,\ n \in N_0, \tag{1.13}$$

solve (1.12) with corresponding eigenvalues $(\lambda_n)_n$ and that they satisfy the relation

$$\int_{-\infty}^{+\infty} H_n(x)H_m(x)\rho_\infty(x)\,dx = n!\delta_{nm}. \tag{1.14}$$

First of all, an induction infers that H_n is a polynomial of degree n. Furthermore, we have the following two relations

$$\begin{aligned} H_n'(x) &= xH_n(x) - (-1)^n e^{x^2/2}\frac{d^n}{dx^n}\left[xe^{-x^2/2}\right] \\ &= xH_n(x) - xH_n(x) + nH_{n-1}(x) = nH_{n-1}(x), \end{aligned} \tag{1.15}$$

$$H_{n+1}(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}\left[xe^{-x^2/2}\right] = xH_n(x) - nH_{n-1}(x). \tag{1.16}$$

Therefore, we obtain using (1.15) in the first and (1.16) in the second equality for xH_{n-1}

$$\begin{aligned} H_n'' - xH_n' &= n(n-1)H_{n-2} - nxH_{n-1} \\ &= n(n-1)H_{n-2} - n(H_n + (n-1)H_{n-2}) = -nH_n. \end{aligned}$$

Equation (1.14) now follows from partial integration together with the relation (1.15) (w.l.o.g. $n \geq m$)

$$\begin{aligned} \frac{(-1)^m}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} H_n(x)\frac{d^m}{dx^m}[e^{-x^2/2}]\,dx &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} H_n^{(m)}(x)e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\frac{n!}{(n-m)!}\int_{-\infty}^{+\infty} H_{n-m}(x)e^{-x^2/2}\,dx = n!\delta_{nm}. \end{aligned}$$

In order to prove that the orthogonal functions $(H_n)_n$ are complete, we first note that $\{H_0, \dots H_n\}$ and $\{1, \dots, x^n\}$ span the same linear subspace of $L^2(\rho_\infty)$ for each $n \in \mathbb{N}_0$. Therefore, we have to show that $f \in L^2(\rho_\infty)$ and

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^n f(x)e^{-x^2/2}\,dx = 0, \quad \forall n \in \mathbb{N}_0,$$

implies $f \equiv 0$. Since $f\rho_\infty \in L^1(\mathbb{R}, \lambda)$, we can calculate the Fourier transform

$$\omega \mapsto \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{-i\omega x - x^2/2}\, dx = \sum_{n=0}^{\infty} \frac{(-i\omega)^n}{n!} \frac{1}{2\pi} \int_{-\infty}^{+\infty} x^n f(x) e^{-x^2/2}\, dx = 0.$$

The Fourier transform is injective and we obtain $f\rho_\infty \equiv 0$, hence $f \equiv 0$.

All in all, the functions $\psi_n(x) = H_n(x)/\sqrt{n!}$ are an orthonormal basis of $L^2(\mathbb{R}, \rho_\infty)$ and the solution of the eigenvalue problem (1.12).

The solution of (1.11) for $g_0 \in L^2(\rho_\infty)$ can be written as

$$g(t, x) = \sum_{n=0}^{\infty} e^{-nt} \langle g_0, \psi_n \rangle_{L^2(\rho_\infty)} \psi_n(x)$$

converging in $L^2(\rho_\infty)$. Recalling the transformation $g = \rho\rho_\infty^{-1}$, a solution of the Fokker-Planck equation (1.10) with initial data ρ_0 with $\rho_0\rho_\infty^{-1} \in L^2(\rho_\infty)$ can be written as

$$\rho(t, x) = \sum_{n=0}^{\infty} e^{-nt} \left(\int_{-\infty}^{+\infty} \rho_0(y)\psi_n(y)\, dy \right) \psi_n(x)\rho_\infty(x) \tag{1.17}$$

Note that this series converges in $L^2(\rho_\infty^{-1})$ due to the transformation $g = \rho\rho_\infty^{-1}$ and $\|g\|_{L^2(\rho_\infty)} = \|\rho\|_{L^2(\rho_\infty^{-1})}$. In particular, it converges also in $L^2(\mathbb{R})$.

In the case that ρ_0 is the density of some initial probability distribution, $\int \rho_0\, dx = 1$, the first term, $n = 0$, is $\rho_\infty = e^{-x^2/2}/\sqrt{2\pi}$ due to $\psi_0 = 1$. The rate of convergence now follows from

$$\begin{aligned} \|\rho_t - \rho_\infty\|_{L^1(\mathbb{R}^d)} &= \|\rho_\infty(g_t - 1)\|_{L^1(\mathbb{R}^d)} \le \|\sqrt{\rho_\infty}\|_{L^2(\mathbb{R}^d)} \|g_t - 1\|_{L^2(\rho_\infty)} \\ &\le e^{-t} \left(\sum_{n=1}^{\infty} \left| \langle g_0, \psi_n \rangle_{L^2(\rho_\infty)} \right|^2 \right)^{1/2} = e^{-t} \|g_0 - 1\|_{L^2(\rho_\infty)} . \end{aligned}$$

Note that the second inequality is in fact an equality in the case $g_0 = \psi_1$, $\rho_0(x) = x\rho_\infty(x)$.

In Chapter 3, we will consider the convergence of ρ_t to ρ_∞ via some entropy functional (we will give a definition in Section 3.1). One such entropy functional has the form

$$e(\rho_t \mid \rho_\infty) = \int_{-\infty}^{+\infty} \left(\frac{\rho_t}{\rho_\infty} - 1 \right)^2 \rho_\infty\, dx = \|g_t - 1\|_{L^2(\rho_\infty)}^2 .$$

Hence, we obtain by the above analysis

$$e(\rho_t \mid \rho_\infty) \le e^{-2t} e(\rho_0 \mid \rho_\infty),$$

where the rate is optimal by our previous observation. We conclude this example with the following remarks:

(i) As we saw before the study of the Ornstein-Uhlenbeck process immediately yielded the fundamental solution of the Fokker-Planck equation and hence a representation of the solution. Furthermore, we established convergence towards the equilibrium and its uniqueness.

(ii) The spectral decomposition gives a formula of the solution too, but with the calculation of the eigenfunctions and eigenvalues. If one is merely interested in the rate of convergence to equilibrium only the smallest nonzero eigenvalue in absolute value is needed.

2. Solution theory

In this chapter we undergo the analysis of existence and uniqueness to the Fokker-Planck equation

$$\partial_t \rho = \mathrm{div}(D(\nabla \rho + \rho \nabla V)) = \mathrm{div}(De^{-V}\nabla(\rho e^V)), \quad \rho(0,x) = \rho_0(x) \tag{2.1}$$

as well as the stochastic differential equation

$$dX_t = (\mathrm{div} D - D\nabla V)(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = \xi. \tag{2.2}$$

Similar as in the previous chapter D is a given diffusion matrix and V a given potential. More precisely, $D : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $V : \mathbb{R}^d \to \mathbb{R}$, where $d \in \mathbb{N}$ is the dimension. Furthermore, $\rho : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ and we will sometimes write ρ_t instead of $\rho(t,\cdot)$.

In addition, $D(x) = \frac{1}{2}\sigma\sigma^\top(x) \in \mathbb{R}^{d\times d}$ is symmetric and $W = (W_t, \mathscr{F}_t, t \geq 0)$ denotes some r-dimensional Brownian motion. Here $r \in \mathbb{N}$ is some integer and σ is also matrix-valued, $\sigma(x) \in \mathbb{R}^{d\times r}$. Moreover, ξ is a random variable denoting the initial condition of the process X.

We already outlined the connection between (2.1) and (2.2) in Section 1.3. In the case that the process X has a smooth density ρ, i.e. $X_t \sim \rho(t,\cdot)\lambda$, this solves the Fokker-Planck equation.

We will always impose the following conditions.

Assumption 2.1. For D, V and σ like above it holds:

(A) $D, V, \sigma \in C^\infty(\mathbb{R}^d)$;

(B) D is bounded on $\mathbb{R}^d$ and uniformly elliptic, i.e. there exists $\alpha > 0$ such that $\xi^\top D(x)\xi \geq \alpha|\xi|^2$ for all $x, \xi \in \mathbb{R}^d$;

(C) $\rho_\infty := e^{-V} \in L^1(\mathbb{R}^d)$ and w.l.o.g. $\int \rho_\infty \, dx = 1$.

We already know that ρ_∞ is the density of the equilibrium distribution and prove uniqueness later on.

In the first section we will solve (2.1) via the spectral decomposition of a self-adjoint operator. This yields the corresponding semigroup (like in the case of the Ornstein-Uhlenbeck process in (1.17)). In the second section we will study (2.2) and prove existence of solutions under certain conditions. These solutions are strong Markov processes and one obtains fundamental solutions of the Fokker-Planck equation via their densities (similar to (1.9) for the Ornstein-Uhlenbeck process).

Some (long) proofs are postponed to the appendix, Section A.3. Furthermore, in the appendix (Section A.1, A.2) the reader can find all definitions as well as statements from the theory of self-adjoint operators, Markov processes and stochastic differential equations which are used later on.

2.1. Study of the partial differential equation

At first, we will rewrite (2.1) with the substitution $g := \rho e^V$, which satisfies

$$\partial_t g = e^V \operatorname{div}(De^{-V}\nabla g), \quad g(0,x) = \rho_0(x)e^{V(x)}. \tag{2.3}$$

Let's define the probability measure $\mu := f_\infty \lambda$ as well as the differential operator $L\varphi := e^V \operatorname{div}(De^{-V}\nabla\varphi)$ on $C_c^\infty(\mathbb{R}^d)$. Again, λ is the Lebesgue measure. We observe for $\varphi, \psi \in C_c^\infty(\mathbb{R}^d)$ via partial integration

$$\int_{\mathbb{R}^d} \psi L\varphi \, d\mu = \int_{\mathbb{R}^d} \psi \operatorname{div}(De^{-V}\nabla\varphi)\, dx = -\int_{\mathbb{R}^d} \nabla\psi^\top D\nabla\varphi \, d\mu.$$

Thus L is symmetric in $L^2(\mu)$. The following proposition summarizes important properties of L. See Section A.3.1 for a proof and a definition of the weighted Sobolev space $H^1(\mu)$.

Proposition 2.2. *Let Assumption 2.1 be valid and consider the above operator* $L : C_c^\infty(\mathbb{R}^d) \subset L^2(\mu) \to L^2(\mu)$. *Then,* L *is symmetric and has a self-adjoint extension* L_{ex} *with* $\operatorname{dom} L_{ex} \subset H^1(\mu)$. *Furthermore, this extension is the closure of* L *and hence unique. In addition, we have* $\sigma(L_{ex}) \subset (-\infty, 0]$, $0 \in \sigma_p(L_{ex})$ *and* $\ker L_{ex}$ *contains only constants,* $\dim\ker L_{ex} = 1$.

Remark 2.3. (i) The very last assertion tells us that the equilibrium of (2.3) is (up to normalization) unique in $L^2(\mu)$.

(ii) Under the above assumption, i.e. $-L$ symmetric and monotone ($\langle -L\varphi, \varphi\rangle_{L^2(\mu)} \geq 0$ for all $\varphi \in \operatorname{dom}(-L)$), the Friedrichs' extension theorem applies, inferring the existence of a self-adjoint extension (see [Yos08, XI.7 Theorem 2, p. 317]). However, we will give a proof in the appendix (subsection A.3.1) without using the Friedrichs' extension, since in our concrete case it corresponds to solving an elliptic partial differential equation.

(iii) For our further studies we will write L instead of L_{ex} for the self-adjoint extension.

Now we are able to solve the initial value problem (2.3).

Theorem 2.4. *Let Assumption 2.1 be satisfied and consider (2.3) with initial data* $g_0 \in L^2(\mu)$. *Then, there exists a unique solution* g *in the following sense:* $g \in C([0,\infty); L^2(\mu)) \cap C^1((0,\infty); L^2(\mu))$, $g(t) \in \operatorname{dom} L$ *for all* $t > 0$ *and*

$$\frac{dg}{dt}(t) = Lg(t), \quad g(0) = g_0$$

It can be written in the form $g(t,\cdot) := e^{Lt}g_0(\cdot)$, *where* $e^{Lt} : L^2(\mu) \to L^2(\mu)$ *denotes the strongly continuous contraction semigroup with infinitesimal generator* L. *Furthermore, the solution has the following properties:*

(i) $e^{Lt}g_0 = \int_{\sigma(L)} e^{\lambda t}\, dE(\lambda)g_0 = \int_{(-\infty,0)} e^{\lambda t}\, dE(\lambda)g_0 + \int g_0 \, d\mu$;

(ii) $t \mapsto e^{Lt}g_0$ *is infinitely often differentiable on* $(0,\infty)$;

(iii) $e^{Lt}g_0 \in \operatorname{dom} L^k$ *for all* $t > 0$, $k \in \mathbb{N}$, *hence* $e^{Lt}g_0 \in C^\infty(\mathbb{R}^d)$*;*

(iv) $(t, x) \mapsto e^{Lt}g_0$ *is smooth;*

(v) $\lim_{t\to+\infty} e^{Lt}g_0 = \int g_0 \, d\mu$ *in* $L^2(\mu)$*.*

Remark 2.5. The above assumptions would allow us to use the Hille-Yosida Theorem for dissipative operators in order to prove the existence of the semigroup e^{Lt} (see [Yos08, IX.8, p. 250] or [Bre11, Theorem 7.7, p. 194]). However, the spectral theorem enables us to prove all the other assertions right away. The proof can be found in the appendix, Subsection A.3.1.

Finally, we obtain a solution of (2.1).

Corollary 2.6. *Suppose Assumption 2.1 is satisfied. Then, the solution of (2.1) with initial condition* $\rho_0 \in L^2(\rho_\infty^{-1})$ *is given by the smooth function* $\rho_t = e^{-V} e^{Lt}(\rho_0 e^V)$*. The map* $L^2(\rho_\infty^{-1}) \to L^2(\rho_\infty^{-1}) : \rho_0 \mapsto e^{-V} e^{Lt}(\rho_0 e^V)$ *defines a strongly continuous contraction semigroup. Furthermore, if* $\int \rho_0 \, dx = 1$ *then* $\int \rho_t \, dx = 1$ *for all* $t \geq 0$*. Finally, we have* $\|\rho_t - \rho_\infty\|_{L^2(\rho_\infty^{-1})} \to 0$ *as* $t \to \infty$*.*

Proof. Everything, despite of the last two assertions, follows from Theorem 2.4 and the fact that $\|\rho_t\|_{L^2(\rho_\infty^{-1})} = \left\|\rho_t \rho_\infty^{-1}\right\|_{L^2(\rho_\infty)} = \|g_t\|_{L^2(\mu)}$. For the last but one claim we observe $\rho_t \in L^1(\mathbb{R}^d)$ and for $0 < s \leq t$

$$\int_{\mathbb{R}^d} \rho_t \, dx - \int_{\mathbb{R}^d} \rho_s \, dx = \int_{\mathbb{R}^d} \int_s^t \partial_t \rho_r \, dr dx = \int_s^t \int_{\mathbb{R}^d} \partial_t \rho_r \rho_\infty^{-1} \, d\mu dr = \int_s^t \int_{\mathbb{R}^d} L g_r \, d\mu dr.$$

The last expression vanishes, since $\langle 1, L g_r \rangle_{L^2(\mu)} = 0$. The last equation is also valid for $s = 0$, because

$$\|\rho_s - \rho_0\|_{L^1(\mathbb{R}^d)} = \|g_s - g_0\|_{L^1(\rho_\infty)} \leq \|g_s - g_0\|_{L^2(\rho_\infty)} \to 0$$

as $s \to 0$ we obtain the claim. The last assertion follows from Theorem 2.4 (v). □

Concerning the method given above note that we have to pay a high prize in order to obtain a solution of (2.1). More precisely, we needed $\rho_0 e^V \in L^2(\mu)$ or equivalently $\rho_0^2 e^V \in L^1(\mathbb{R}^d)$.

2.2. Study of the stochastic differential equation

In the following we investigate solutions of (2.2). Sometimes it is convenient to set $b(x) := (\operatorname{div} D - D\nabla V)(x)$. As we will see the corresponding generator $\mathcal{L}$ reads

$$\mathcal{L}\varphi = \sum_{i,j} D_{ij} \partial_{ji}^2 \varphi + (\operatorname{div} D - D\nabla V) \cdot \nabla\varphi. \tag{2.4}$$

Since the coefficients in (2.2) are smooth, hence locally Lipschitz-continuous, we get uniqueness of strong solutions of (2.2) from Theorem A.12. However, locally Lipschitz-continuous is not sufficient for global existence. We employ the following idea, which is familiar from studying ordinary differential equations (see also [Kha12, Chapter 3.4, p. 74-74]): we construct a solution $(X_t^n)_{t\geq 0}$ on the ball $B_n(0)$ ($n \in \mathbb{N}$), on which all conditions in Theorem A.13 are satisfied.

The process may or may not reach the boundary of the ball after some finite time. If it does we stop it. For all n we get a local solution and since those solutions are unique, $(X_t^n)_{t\geq 0}$ will coincide with $(X_t^m)_{t\geq 0}$ for $m < n$ before it reaches the boundary of $B_m(0)$. Thus, we can define a process $(X_t)_{t\geq 0}$ by gluing together all local solutions, which constitutes a well-defined solution. The rigorous proof together with the construction of the filtration $(\mathscr{F}_t)$ is given in the appendix, Subsection A.3.2. We summarize the result in the following proposition.

Proposition 2.7. *Assume that Assumption 2.1 holds and consider the equation (2.2). Given $(\Omega, \mathscr{F}, \mathbb{P})$, a Brownian motion $(W_t)_{t\geq 0}$ and initial data ξ, there exists a $(\mathscr{F}_t)$ stopping time τ and a process $(X_t, \mathscr{F}_t, t < \tau)$ which is a solution to (2.2) in the following sense:*

(i) $(X_t, t < \tau)$ *is adapted and continuous (a.s.),*

(ii) $X_0 = \xi$ *on* $\{\tau > 0\}$*,*

(iii) $\int_0^t (|b_i(X_s)|^2 + |\sigma_{ij}(X_s)|^2)\, ds < \infty$ *a.s. on* $\{t < \tau\}$*,*

(iv) $X_t = \xi + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s$ *a.s. on* $\{t < \tau\}$*.*

Furthermore, $(X_t, \mathscr{F}_t, t < \tau)$ is unique, i.e. if $(Y_t, \mathscr{F}_t, t < \sigma)$ (σ a stopping time) is a solution in the above sense, then $\sigma \leq \tau$ and $X_t = Y_t$ on $\{t < \sigma \wedge \tau\}$ (despite of a null set independent of t).

Remark 2.8. Observe that a priori $\{\tau = \infty\}$ can be rather small (e.g. a null set). For instance, there are ordinary differential equations ($\sigma \equiv 0$) having solutions which "explode" in finite time. We will discuss cases where $\mathbb{P}(\tau = \infty) = 1$ and thus $(X_t)_{t\geq 0}$ is a strong solution (see Definition A.11).

In the following we discuss a condition that allows us to prove $\mathbb{P}(\tau = \infty) = 1$ (see [Kha12, Section 3.4, p.74-77]).

Proposition 2.9. *Under the conditions of Proposition 2.7 assume furthermore that there is a nonnegative function $G \in C^2(\mathbb{R}^d)$ and a constant $c > 0$ such that*

$$\mathcal{L}G \leq cG, \qquad \lim_{R\to+\infty} \inf_{|x|\geq R} G(x) = +\infty.$$

Finally, suppose $\mathbb{E}[G(X_0)] < \infty$. Then, the solution in Proposition 2.7 is defined for all times, i.e. $\mathbb{P}(\tau = \infty) = 1$ and

$$\mathbb{E}[G(X_t)] \leq \mathbb{E}[G(X_0)]e^{ct}.$$

Hence, $(X_t)_{t\geq 0}$ is a strong solution.

Proof. Let $\tau_n := \inf\{t > 0 \ :\ |X_t| \leq n\}$. By construction of $(X_t)_{t\geq 0}$ and Itô's formula we have

$$\mathbb{E}\left[G(X_{t\wedge\tau_n})e^{-c(t\wedge\tau_n)}\right] = \mathbb{E}[G(X_0)] + \mathbb{E}\left[\int_0^{t\wedge\tau_n} (\mathcal{L}G(X_s) - cG(X_s))e^{-cs}\, ds\right] \leq \mathbb{E}[G(X_0)],$$

which implies

$$\mathbb{E}[G(X_{t\wedge\tau_n})] \leq e^{ct}\mathbb{E}[G(X_0)]. \tag{2.5}$$

We further obtain by the nonnegativity of G

$$\mathbb{E}[G(X_{t\wedge\tau_n})] \geq \mathbb{E}[G(X_{\tau_n})\mathbb{1}_{\{\tau_n\leq t\}}] \geq \inf_{|x|\geq n} G(x)\mathbb{P}(\tau_n \leq t)$$

and hence

$$\mathbb{P}(\tau \leq t) \leq \mathbb{P}(\tau_n \leq t) \leq \frac{e^{ct}\mathbb{E}[G(X_0)]}{\inf_{|x|\geq n} G(x)} \to 0 \quad \text{for } n \to \infty.$$

Since t was arbitrary, we obtain $\mathbb{P}(\tau < \infty) = 0$ and by applying Fatou's lemma in (2.5)

$$\mathbb{E}[G(X_t)] \leq e^{ct}\mathbb{E}[G(X_0)].$$

□

Remark 2.10. It is worth to discuss why the solution to (2.2) intuitively should not "explode". First of all, if $D \equiv I$ and $V \equiv 0$ then $(X_t)_{t\geq 0}$ is simply a Brownian motion, which behaves nicely. If we add a confinement potential then the process $(X_t)_{t\geq 0}$ has a drift which, roughly speaking, points to minima of the potential. Thus, $(X_t)_{t\geq 0}$ should intuitively behave even more nicely than a Brownian motion does. In order to describe this property we assume $\langle \nabla V(x), x\rangle \geq 0$.

If the diffusion is not homogeneous, then we will assume $\langle D\nabla V(x), x\rangle \geq 0$ instead. In addition, we will impose D, div$D$ to be bounded, which implies that the diffusion is not arbitrarily strong and D does not oscillate quickly.

The intuitive reasoning actually works as the next lemma points out.

Lemma 2.11. *Let Assumption 2.1 be valid and assume $\mathbb{E}[|X_0|^{2k}] < \infty$ for some $k \in \mathbb{N}$. Furthermore, suppose that D and its first derivatives are uniformly bounded as well as $\langle D\nabla V(x), x\rangle \geq 0$ whenever $|x| \geq M$ for some constant M. Then, the conclusion of Proposition 2.9 holds and for all $t \geq 0$*

$$\mathbb{E}[|X_t|^{2k}] \leq C(t,k)\mathbb{E}[|X_0|^{2k}] + K(t,k)$$

for some constants $C(t,k), K(t,k) \geq 0$.

The aim of the proof is to apply Proposition 2.9 by defining $G(x) = |x|^{2k} + C$ and estimating $\mathcal{L}G$. See A.3.2 for the calculations.

We end this section with some properties of the solution of (2.1) and (2.2). We will use the following maximum principle for parabolic equations. Its proof is provided in [Fri64, Section 2.2, Theorem 5, p. 39].

Theorem 2.12. *Consider a classical solution $u \in C^{1,2}((0,\infty)\times\mathbb{R}^d) \cap C([0,\infty)\times\mathbb{R}^d)$ of the following parabolic partial differential equation with continuous coefficients*

$$\partial_t u(t,x) = \sum_{ij} a_{ij}(x)\partial^2_{ji}u(t,x) + \sum_i b_i(x)\partial_i u(t,x) + c(x)u(t,x).$$

Furthermore, assume that the matrix (a_{ij}) is positive definite at every point and $u \geq 0$. If u admits the value zero at some point (t_0,x_0), then $u(t,x) = 0$ for all $(t,x) \in [0,t_0]\times\mathbb{R}^d$.

Theorem 2.13. *Let Assumption 2.1 together with the condition from the previous Lemma 2.11 be valid. Thus, D and its first derivatives are bounded, $\langle D\nabla V(x), x\rangle \geq 0$ for all $|x| \geq M$, M some constant, and $\mathbb{E}[|X_0|^{2k}] < \infty$ for some $k \in \mathbb{N}$. Let μ_0 be the law of X_0. Then, we have:*

(i) One can define a (time-homogeneous) strong Markov family $(X_t, \mathscr{F}_t, t \geq 0)$, $(\mathbb{P}_x)_x$ on some probability space $(\Omega, \mathscr{F})$ from the solution of (2.2) for every initial condition $\xi \equiv x \in \mathbb{R}^d$. The generator has the form $\mathcal{L}$ as in (2.4), where its domain contains $C_c^2(\mathbb{R}^d)$.

(ii) Suppose the Markov process $(X_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P}_x)$ with initial state $x \in \mathbb{R}^d$ has a density $p(t, x, \cdot)$ with respect to Lebesgue-measure which is smooth in all variables. Then, $(t, y) \mapsto p(t, x, y)$ is a fundamental solution of (2.1) with singularity at $x \in \mathbb{R}^d$, $t = 0$.

(iii) Suppose $(X_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P}_{\mu_0})$ has a density $\rho(t, \cdot)$ with $(t, x) \mapsto \rho(t, x)$ smooth. Then, this is a classical solution of (2.1). If in addition $(X_t)_{t\geq 0}$ has a smooth density p like in (ii), then the solution has the representation

$$\rho(t, x) = \int p(t, y, x)\, d\mu_0(y). \tag{2.6}$$

(iv) Every classical solution $\rho(t, \cdot)$ of (2.1) which is a density of some probability measure satisfies $\rho(t, \cdot) > 0$ for $t > 0$ if $\rho_0 \not\equiv 0$. In particular, we have $p(t, x, \cdot) > 0$, $t > 0$.

Remark 2.14. (i) Under assumptions one can prove that the transition kernel $\mathcal{P}_t$ of the Markov family $(X_t)_{t\geq 0}$ has a smooth density, i.e. there exists for all $x \in \mathbb{R}^d$, $t > 0$ a smooth function $(t, y) \mapsto p(t, x, y)$ with

$$\mathbb{P}_x(X_t \in A) = \int_A p(t, x, y)\, dy.$$

for every $A \in \mathscr{B}(\mathbb{R}^d)$. In essence, one assumes that the smooth coefficients and all their derivatives grow at most polynomially at infinity and that D is uniformly positive definite (see [Hai16, Theorem 6.3, p. 26]).

(ii) Continuous solutions having a representation like (2.6) are also called *mild solutions*. See [SCDM04, Definition 2.1, p. 242, Subsection 3.1, p. 242-245] for the definition and an investigation of the quantum Fokker-Planck equation, where the fundamental solution is calculated for a quadratic potential based on the method of characteristics.

Proof of Theorem 2.13. (i) By virtue of the previous Lemma 2.11 the equation (2.2) has a global solution. Furthermore, by Theorem A.14 we can define a strong Markov family $(X_t, \mathscr{F}_t, t \geq 0)$ with probability measures $(\mathbb{P}_x)_{x\in\mathbb{R}^d}$ on a measure space $(\Omega, \mathscr{F})$, such that $(X_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P}_x)$ is the strong solution with initial state $x \in \mathbb{R}^d$.

In order to prove that the generator $\mathcal{L}$ has the form (2.4), take $f \in C_c^2(\mathbb{R}^d)$. Now, apply Itô's formula to $f(X_t)$ yielding

$$f(X_t) - f(X_0) = \int_0^t \mathcal{L}f(X_s)\, ds + \int_0^t \nabla f(X_s)^\top \sigma(X_s)\, dW_s.$$

We obtain after taking expectations

$$\mathbb{E}_x[f(X_t)] - f(x) = \int_0^t \mathbb{E}_x[\mathcal{L}f(X_s)]\,ds.$$

We divide by t and take the limit $t \to 0$, which exists, since $s \mapsto \mathbb{E}_x[\mathcal{L}f(X_s)]$ is continuous (by dominated convergence) with $\lim_{s\to 0}\mathbb{E}_x[\mathcal{L}f(X_s)] = \mathcal{L}f(x)$. Thus, we establish

$$\lim_{t\to 0}\frac{1}{t}(\mathbb{E}_x[f(X_t)] - f(x)) = \mathcal{L}f(x)$$

and the domain of $\mathcal{L}$ contains every function $f \in C_c^2(\mathbb{R}^d)$.

(ii) We already know that the law solves (2.1) in the sense of distributions. Since p is smooth, it certainly solves the equation in the classical sense (for $t > 0$). Now, choose a bounded function $\varphi \in C(\mathbb{R}^d)$, then we obtain by dominated convergence and the continuity of $(X_t)_{t\geq 0}$

$$\lim_{t\to 0}\int_{\mathbb{R}^d}\varphi(y)p(t,x,y)\,dy = \lim_{t\to 0}\mathbb{E}_x[\varphi(X_t)] = \varphi(x).$$

Therefore, $(t,y) \mapsto p(t,x,y)$ has singularity at x for $t = 0$.

(iii) The Markov process $(X_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P}_{\mu_0})$ has initial distribution μ and its law solves (2.1) in the sense of distribution. Hence, the corresponding smooth density $\rho(t,\cdot)$ is a classical solution of (2.1).

The representation (2.6) follows from

$$\begin{aligned}\int \varphi(x)\rho(t,x)\,dy &= \mathbb{E}_{\mu_0}[\varphi(X_t)] = \int \mathbb{E}_y[\varphi(X_t)]\,d\mu_0(y)\\ &= \int\int \varphi(x)p(t,y,x)\,dx\,d\mu_0(y) = \int \varphi(x)\int p(t,y,x)\,d\mu_0(y)\,dx\end{aligned}$$

for arbitrary φ bounded and continuous.

(iv) Since ρ is a classical solution and a density of some probability measure, $\rho(t,\cdot) \geq 0$, the maximum principle Theorem 2.12 applies.

□

3. Entropy method

In this chapter we study the convergence to equilibrium of solutions of the Fokker-Planck equation (2.1) via the entropy method. The idea is to analyze the long-time behavior of an *entropy* functional e (see Definition 3.1 below) along a solution ρ_t, i.e. $e(\rho_t)$.

In the first section we will briefly introduce entropy functionals and state some properties used later on. Then, we outline the entropy method in Section 3.2 in order to establish one primary convergence result (where the proof is provided in the appendix, see Section A.4). Furthermore, the link with convex Sobolev inequalities will be discussed. This will allow us to extend the results.

In the case of a logarithmic entropy functional the corresponding convex Sobolev inequality is the logarithmic Sobolev inequality studied by Gross for Gaussian reference measures (or in our framework equilibrium measures), see [Gro75]. A famous condition, under which a logarithmic Sobolev inequality holds, is due to Bakry and Emery, see for instance [BGL14, Section 5.7]. The following investigation is based on the work [AMTU01] where these results were extended to more general relative entropy functionals, leading to convex Sobolev inequalities mentioned above. In addition, sharpness of these inequalities and applications to non-symmetric linear as well as nonlinear Fokker-Planck equations were studied in [AMTU01].

3.1. Entropy functionals

At first we shall give a definition as well as some properties of an entropy functional. We refer to [AMTU01, Subsection 2.2, p. 12-18] for the proofs of these properties.

Definition 3.1 (relative entropy)**.** Consider a function $\psi \in C([0,\infty]) \cap C^4((0,\infty))$ satisfying

$$\psi(1) = 0, \quad \psi'' \geq 0, \quad \psi'' \not\equiv 0, \quad (\psi''')^2 \leq \frac{1}{2}\psi''\psi^{IV}.$$

We define for two probability density functions ρ_1, ρ_2 with ρ_1/ρ_2 finite $\rho_2\lambda$-a.s. (λ the Lebesgue measure) the relative entropy of ρ_1 with respect to ρ_2 by

$$e_\psi(\rho_1 \mid \rho_2) := \int_{\mathbb{R}^d} \psi\left(\frac{\rho_1}{\rho_2}\right) \rho_2 \, dx \geq 0.$$

The function ψ is called the generating function of e_ψ. Furthermore, if $\psi'(1) = 0$ then ψ is normalized.

Remark 3.2. For every generating function ψ there is its normalization $\widetilde{\psi}(\sigma) = \psi(\sigma) - \psi'(1)(\sigma - 1)$, which generates the same relative entropy as ψ (due to $\int \rho_1 \, dx = \int \rho_2 \, dx$).

For instance, the original relative entropy generated by $\sigma \ln \sigma - \sigma + 1$ and its generalization generated by

$$\chi(\sigma) = \alpha(\sigma + \beta) \ln \left(\frac{\sigma + \beta}{1 + \beta} \right) - \alpha(\sigma - 1) \tag{3.1}$$

for $\alpha > 0$, $\beta \geq 0$ provide relative entropies in the above sense. Another example is given by the generating function $\varphi(\sigma) = \alpha(\sigma - 1)^2$, for $\alpha > 0$. Note that the generated entropy functional e_φ corresponds to the $L^2(\rho_2^{-1})$-norm, i.e.

$$e_\varphi(\rho_1 \mid \rho_2) = \int_{\mathbb{R}^d} \alpha \left(\frac{\rho_1}{\rho_2} - 1 \right)^2 \rho_2 \, dx = \alpha \left(\int_{\mathbb{R}^d} \rho_1^2 \rho_2^{-1} \, dx - 1 \right).$$

In order to cover all admissible entropies in our investigations it will be useful to know certain growth properties of generating functions, which are stated in the following lemma. Furthermore, the lemma contains a continuity property of entropy functionals. We refer to [AMTU01, Remark 2.3, Lemma 2.8, Lemma 2.9, p. 13-18] for the proofs.

Lemma 3.3. *Consider a normalized generator ψ. Then, we have the following properties:*

(i) Suppose $\lim_{j \to +\infty} \rho_j = \rho$ in $L^2(\widetilde{\rho}^{-1})$ for probability density functions $(\rho_j)_j$, ρ, $\widetilde{\rho}$. Then, $\lim_{j \to \infty} e_\psi(\rho_j \mid \widetilde{\rho}) = e_\psi(\rho \mid \widetilde{\rho})$.

(ii) The functions $\sigma \mapsto \sigma\psi'(\sigma)$ and ψ' are increasing. Moreover, $\psi'' > 0$ and ψ'' is decreasing.

(iii) Let $\mu_2 = \psi''(1)$ then for all $\sigma \geq \sigma_0 > 0$ it holds

$$\psi(\sigma) \leq \psi(\sigma_0) \left(\frac{\sigma}{\sigma_0} \right)^2 + \mu_2 \left(\frac{\sigma}{\sigma_0} - 1 \right) (\sigma - 1) \tag{3.2}$$

and we have for all $\sigma_0 \geq \sigma > 0$

$$\psi(\sigma) \leq \psi(\sigma_0) \frac{\sigma}{\sigma_0} + \mu_2 \left(\frac{\sigma}{\sigma_0} - 1 \right) (\sigma - 1). \tag{3.3}$$

The next lemma shows that convergence in entropy to zero, i.e. $e_\psi(\rho_t \mid \rho_\infty) \to 0$, provides convergence in L^1 as well. This is a generalization of the Csiszár-Kullback inequality. The proof can be found in [AMTU01, p. 15].

Lemma 3.4. *Consider a relative entropy e_ψ and two probability density functions ρ_1, ρ_2. Then, we have (with $\mu_2 = \psi''(1)$)*

$$\frac{1}{2} \|\rho_1 - \rho_2\|_{L^1(\mathbb{R}^d)}^2 \leq \frac{1}{\mu_2} e_\psi(\rho_1 \mid \rho_2).$$

In Section 2.1, we studied solutions ρ_t of (2.1) in $L^2(\rho_\infty^{-1})$ via the substitution $g_t := \rho_t / \rho_\infty$, recalling the time homogeneous solution $\rho_\infty := e^{-V}$ of (2.1). The convergence of ρ_t to ρ_∞ in $L^2(\rho_\infty^{-1})$, see Corollary 2.6, implies the following convergence result for the relative entropy (see also [AMTU01, Lemma 2.11, p. 19]).

Lemma 3.5. *Let Assumption 2.1 be satisfied and consider a solution ρ_t of (2.1) with initial distribution $\rho_0 \in L^2(\rho_\infty^{-1})$. For every relative entropy e_ψ it holds*

$$e_\psi(\rho_t \mid \rho_\infty) \to 0 \quad \text{as } t \to \infty.$$

Proof. We know $\psi(\sigma) \le \mu_2(\sigma-1)^2 =: \varphi(\sigma)$ by choosing $\sigma_0 = 1$ in Lemma 3.3 (iii) (recalling $\psi(1) = 0$) and therefore

$$0 \le e_\psi(\rho_t \mid \rho_\infty) \le e_\varphi(\rho_t \mid \rho_\infty) = \mu_2 \left\| \frac{\rho_t}{\rho_\infty} - 1 \right\|^2_{L^2(\rho_\infty)} = \mu_2 \left\| \rho_t - \rho_\infty \right\|^2_{L^2(\rho_\infty^{-1})}.$$

The last expression tends to zero as $t \to \infty$ by Corollary 2.6. □

3.2. Convergence in relative entropy

The entropy method consists of the following steps.

(1) Given an admissible relative entropy e_ψ we calculate the entropy dissipation along the trajectory of a solution ρ_t of (2.1) with initial data $\rho_0 \in L^2(\rho_\infty^{-1})$, i.e. for $t > 0$

$$\frac{d}{dt} e_\psi(\rho_t \mid \rho_\infty) =: I_\psi(\rho_t \mid \rho_\infty).$$

The function $I_\psi(\rho_t \mid \rho_\infty)$ is also called the *relative Fisher information.*

(2) Then, one proves the validity of a *convex Sobolev inequality* under assumptions on D, V for initial data $\rho_0 \in L^2(\rho_\infty^{-1})$, i.e. there holds for $t > 0$

$$e_\psi(\rho_t \mid \rho_\infty) \le -\frac{1}{2\lambda} I_\psi(\rho_t \mid \rho_\infty). \tag{3.4}$$

According to the first step, this implies exponential convergence of the entropy with a given rate 2λ, i.e. for $t \ge 0$

$$e_\psi(\rho_t \mid \rho_\infty) \le e^{-2\lambda t} e_\psi(\rho_0 \mid \rho_\infty). \tag{3.5}$$

(3) Finally, the result can be extended to initial distributions having finite entropy.

Remark 3.6. Set $g_t = \rho_t/\rho_\infty$. A formal calculation shows

$$I_\psi(\rho_t \mid \rho_\infty) = -\int_{\mathbb{R}^d} \psi''(g_t) \nabla g_t^\top D \nabla g_t \rho_\infty \, dx$$

using $\partial_t g_t = e^V \operatorname{div}(D e^{-V} \nabla g_t)$ and partial integration. Therefore, the convex Sobolev inequality (3.4) reads

$$e_\psi(\rho_t \mid \rho_\infty) \le \frac{1}{2\lambda} \int_{\mathbb{R}^d} \psi'' \left(\frac{\rho_t}{\rho_\infty}\right) \nabla \left(\frac{\rho_t}{\rho_\infty}\right)^\top D \nabla \left(\frac{\rho_t}{\rho_\infty}\right) \rho_\infty \, dx \tag{3.6}$$

This becomes Gross' inequality when $\psi(\sigma) = \sigma \ln \sigma - \sigma + 1$, $D \equiv I$ and ρ_∞ is a Gaussian density, i.e. $\rho_\infty(x) = e^{-|x|^2/2}/(2\pi)^{d/2}$, see [Gro75].

We start the above investigation with the following lemma which provides step (1).

Lemma 3.7. *Let Assumption 2.1 be satisfied and consider a solution ρ_t of (2.1) with $\rho_0 \in L^2(\rho_\infty^{-1})$. Furthermore, consider a relative entropy e_ψ. Then, $e_\psi(\rho_t \mid \rho_\infty) < \infty$ and for $0 \le s \le t$ we have*

$$e_\psi(\rho_t \mid \rho_\infty) - e_\psi(\rho_s \mid \rho_\infty) = \int_s^t \left(- \int_{\mathbb{R}^d} \psi''(g_r) \nabla g_r^\top D \nabla g_r \rho_\infty \, dx \right) dr \le 0. \tag{3.7}$$

The proof can be found in the appendix, see Section A.4. The technical point in proving (3.7) is that $\sigma \mapsto \psi'(\sigma)$ and $\sigma \mapsto \psi''(\sigma)$ can have a singularity at $\sigma = 0$. The idea is to use an approximating function $\psi_\varepsilon \to \psi$ for $\varepsilon \to 0$, which does not have a singularity.

In the following we will provide our first convergence result in the case of homogeneous diffusion, i.e. when $D \equiv I$.

Proposition 3.8. *Let us assume Assumption 2.1 together with $D \equiv I$. Furthermore, assume that* Hess $V(x) \ge \lambda I$ *for all $x \in \mathbb{R}^d$. Consider a solution ρ_t of (2.1) with $\rho_0 \in L^2(\rho_\infty^{-1})$ and some relative entropy e_ψ. Then, the convex Sobolev inequality (3.6) holds and for $t \ge 0$*

$$e_\psi(\rho_t \mid \rho_\infty) \le e^{-2\lambda t} e_\psi(\rho_0 \mid \rho_\infty).$$

Remark 3.9. The assumption Hess $V \ge \lambda I$ is a special case of the Bakry-Emery condition, see [BGL14, Theorem 5.7.4, p. 270]. The general condition also considers the curvature of a Riemannian manifold with which one is working (instead of $\mathbb{R}^d$). The primary idea is to study the dissipation of the entropy dissipation, i.e. $\frac{d}{dt} I_\psi(\rho_t \mid \rho_\infty)$. However, there is again the technicality in computing $\frac{d}{dt} I_\psi(\rho_t \mid \rho_\infty)$, since $\sigma \mapsto \psi''(\sigma)$ and $\sigma \mapsto \psi'''(\sigma)$ can have a singularity at $\sigma = 0$. Therefore, we will approximate $\rho_0 \rho_\infty^{-1} = g_0 \in L^2(\rho_\infty)$ by initial data which are bounded away from zero and satisfy a gradient estimate $|\nabla g_t| \le C$ (following [OV00, Section 4, p. 18]). Finally, one can use the uniform convexity of V to obtain the convex Sobolev inequality (3.6), following [AMTU01, Lemma 2.13, p. 21]. The proof is provided in the appendix, see Subsection A.4.1

Finally, step (3) extends the exponential convergence to a wider class of initial distributions. We state it in a separate fashion, in order to point out the independence of the above steps.

Theorem 3.10. *Let Assumption 2.1 be satisfied and consider a relative entropy e_ψ. Suppose that for every solution ρ_t of (2.1) with initial data $\rho_0 \in L^2(\rho_\infty^{-1})$ there is exponential convergence towards the equilibrium with rate 2λ, i.e. it holds for $t \ge 0$*

$$e_\psi(\rho_t \mid \rho_\infty) \le e^{-2\lambda t} e_\psi(\rho_0 \mid \rho_\infty).$$

Then, the same result is true for initial distributions ρ_0 with finite entropy $e_\psi(\rho_0 \mid \rho_\infty) < \infty$.

The proof is provided in [AMTU01, Theorem 2.16, p. 26]. The idea is to approximate a given initial distribution with finite entropy by functions $\rho \in L^2(\rho_\infty^{-1})$.

Remark 3.11. The fact that the convex Sobolev inequality (3.6) is the essential point where the conditions on V and D play a role is of great significance. It shows us that the whole analysis of convergence to equilibrium is in essence a matter of proving a convex Sobolev inequality (for solutions having initial data in $L^2(\rho_\infty^{-1})$).

In order to prove further convergence results it is important that inequality (3.6) can be extended. We state the assertion and refer to [AMTU01, Corollary 2.18, p. 29] for a proof.

Theorem 3.12. *Assume* $\operatorname{Hess} V \geq \lambda I$ *in addition to Assumption 2.1, then for every positive probability density function* $\rho \in C^\infty(\mathbb{R}^d)$*, the convex Sobolev inequality (3.6) is valid for any relative entropy with* $\rho_\infty = e^{-V}$*.*

Remark 3.13. One can weaken the assumption of $\rho \in C^\infty(\mathbb{R}^d)$ and $\rho > 0$. However, one has to give a meaning to the right term in (3.6), see [AMTU01, Remark 2.12, p. 20].

Now, we can cover the case of inhomogeneous diffusion D under the general assumptions (see [AMTU01, Corollary 2.17, p. 29 and Remark 3.1, p. 34]).

Theorem 3.14. *Let Assumption 2.1 be valid (in particular,* $D \geq \alpha I$*) and assume* $\operatorname{Hess} V \geq \lambda I$*. Consider a relative entropy* e_ψ*. Then, any solution* ρ_t *of (2.1) with initial data* ρ_0 *having finite entropy converges exponentially to the equilibrium in relative entropy and hence in* L^1*. More precisely, it holds for* $t \geq 0$

$$e_\psi(\rho_t \mid \rho_\infty) \leq e^{-2\lambda\alpha t} e_\psi(\rho_0 \mid \rho_\infty) \quad \textit{and} \quad \|\rho_t - \rho_\infty\|_{L^1(\mathbb{R}^d)} \leq e^{-\lambda\alpha t}\sqrt{\frac{2}{\mu_2} e_\psi(\rho_0 \mid \rho_\infty)}.$$

Proof. We set $g_t = \rho_t/\rho_\infty$ and obtain for $\rho_0 \in L^2(\rho_\infty^{-1})$ using Lemma 3.7 as well as $D \geq \alpha I$

$$\begin{aligned} e_\psi(\rho_t \mid \rho_\infty) - e_\psi(\rho_0 \mid \rho_\infty) &= -\int_0^t \int_{\mathbb{R}^d} \psi''(g_r) \nabla g_r^\top D \nabla g_r \rho_\infty \, dx dr \\ &\leq -\alpha \int_0^t \int_{\mathbb{R}^d} \psi''(g_r) |\nabla g_r|^2 \rho_\infty \, dx dr. \end{aligned}$$

By Theorem 3.12 it holds

$$\frac{1}{2\lambda} \int_{\mathbb{R}^d} \psi''(g_r) |\nabla g_r|^2 \rho_\infty \, dx \geq e_\psi(\rho_r \mid \rho_\infty)$$

and hence by the previous inequality

$$e_\psi(\rho_t \mid \rho_\infty) - e_\psi(\rho_0 \mid \rho_\infty) \leq -2\lambda\alpha \int_0^t e_\psi(\rho_r \mid \rho_\infty) \, dr.$$

This yields exponential convergence with rate $2\lambda\alpha$ in relative entropy via Gronwall's lemma. Now, Theorem 3.10 applies yielding the result for initial data having finite entropy. Finally, Lemma 3.4 infers for $t \geq 0$

$$\|\rho_t - \rho_\infty\|^2_{L^1(\mathbb{R}^d)} \leq \frac{2}{\mu_2} e^{-2\lambda\alpha t} e_\psi(\rho_0 \mid \rho_\infty).$$

This completes the proof. □

The next extension of these results allows the conclusion that the long-time behavior only depends on the growth of V at infinity. Thus, a solution still tends towards the equilibrium. The following perturbation result is originally due to Holley and Stroock, see [HS87], and was generalized in [AMTU01, Theorem 3.1, p. 38-39] (on which the proof below is based). The statement reads:

Theorem 3.15. *Suppose that for V, D in (2.1), where Assumption 2.1 is satisfied, the convex Sobolev inequality (3.6) holds for some relative entropy e_ψ. Consider $\widetilde{V} = V + v$ for some smooth bounded function v with $\int \widetilde{\rho}_\infty \, dx = 1$ ($\widetilde{\rho}_\infty = e^{-\widetilde{V}}$). Let a, b be such that*

$$0 < a \leq e^{-v} \leq b. \tag{3.8}$$

Then, the convex Sobolev inequality (3.6) is also valid with $\widetilde{\rho}_\infty$ replacing ρ_∞ and with constant $\max\left(b/a^2, b^2/a\right)/2\lambda$ *instead of* $1/2\lambda$.

Proof. In the following we will omit the subscript t in ρ_t and corresponding expressions. It suffices to prove the result only for solutions ρ of (2.1) with initial data $\rho_0 \in L^2(\rho_\infty^{-1}) = L^2(\widetilde{\rho}_\infty^{-1})$, since one can then extend it via Theorem 3.10. Recall that solutions are strictly positive for $t > 0$, see Theorem 2.13 (iv). We want to prove

$$\int_{\mathbb{R}^d} \psi\left(\frac{\rho}{\widetilde{\rho}_\infty}\right) \widetilde{\rho}_\infty \, dx \leq \frac{1}{2\lambda} \max\left(\frac{b}{a^2}, \frac{b^2}{a}\right) \int_{\mathbb{R}^d} \psi''\left(\frac{\rho}{\widetilde{\rho}_\infty}\right) \nabla\left(\frac{\rho}{\widetilde{\rho}_\infty}\right)^\top D \nabla\left(\frac{\rho}{\widetilde{\rho}_\infty}\right) \widetilde{\rho}_\infty \, dx,$$

which contains the term $\rho/\widetilde{\rho}_\infty$. Therefore, we set $\widetilde{f} := \rho/\widetilde{\rho}_\infty$ and observe

$$\int_{\mathbb{R}^d} \widetilde{f} \widetilde{\rho}_\infty \, dx = 1.$$

So it suffices to prove the corresponding inequality for $\widetilde{f} \in L^2(\widetilde{\rho}_\infty)$ with $\|\widetilde{f}\|_{L^1(\widetilde{\rho}_\infty)} = 1$. Note that $\widetilde{f}$ should be smooth and positive too. By the same observation the convex Sobolev inequality corresponding to V, D is true for $f \in L^2(\rho_\infty)$ with $\|f\|_{L^1(\rho_\infty)} = 1$. Furthermore, $L^2(\rho_\infty) = L^2(\widetilde{\rho}_\infty)$ and we can recover unit norm by scaling.

Choose $\widetilde{f}$ like above and set $f := \widetilde{f}/\|\widetilde{f}\|_{L^1(\rho_\infty)}$. We will make us of the following estimate

$$\psi''(\sigma_0) \leq \begin{cases} \psi''(\sigma_1), & \sigma_1 \leq \sigma_0, \\ \psi''(\sigma_1)\dfrac{\sigma_1}{\sigma_0}, & \sigma_1 > \sigma_0, \end{cases} \tag{3.9}$$

which follows either from $\psi''(\sigma)$ being a decreasing function or $\sigma\psi''(\sigma)$ being an increasing function (see Lemma 3.3 (ii)). We set $\sigma_1 = \widetilde{f}$ and $\sigma_0 = f$, which are both positive.

We distinguish two cases. The first one is $\sigma_1 \geq \sigma_0$ or equivalently $\|\widetilde{f}\|_{L^1(\rho_\infty)} \geq 1$. Now, we will use (3.2) in Lemma 3.3 for $\sigma = \sigma_1$ yielding

$$\begin{aligned} \int_{\mathbb{R}^d} \psi(\widetilde{f})\widetilde{\rho}_\infty \, dx &\leq \int_{\mathbb{R}^d} \left\{ \psi(f)\|\widetilde{f}\|^2_{L^1(\rho_\infty)} + \mu_2(\|\widetilde{f}\|_{L^1(\rho_\infty)} - 1)(\widetilde{f} - 1) \right\} \widetilde{\rho}_\infty \, dx \\ &\leq b\|\widetilde{f}\|^2_{L^1(\rho_\infty)} \int_{\mathbb{R}^d} \psi(f)\rho_\infty \, dx \leq \frac{b}{2\lambda}\|\widetilde{f}\|^2_{L^1(\rho_\infty)} \int_{\mathbb{R}^d} \psi''(f)\nabla f^\top D \nabla f \rho_\infty \, dx \end{aligned}$$

where we used $\|f\|_{L^1(\rho_\infty)} = 1$, (3.8) and the convex Sobolev inequality. Using the definition of f, (3.8) and the second inequality in (3.9) we establish

$$\int_{\mathbb{R}^d} \psi(\widetilde{f})\widetilde{\rho}_\infty\, dx \leq \|\widetilde{f}\|_{L^1(\rho_\infty)} \frac{b}{2\lambda a} \int_{\mathbb{R}^d} \psi''(\widetilde{f})\nabla\widetilde{f}^\top D\nabla\widetilde{f}\widetilde{\rho}_\infty\, dx$$

The assertion now follows from $\|\widetilde{f}\|_{L^1(\rho_\infty)} \leq 1/a$.

In the case $\sigma_0 \geq \sigma_1$, one uses (3.3) similarly and obtains

$$\begin{aligned}\int_{\mathbb{R}^d} \psi(\widetilde{f})\widetilde{\rho}_\infty\, dx &\leq \int_{\mathbb{R}^d} \left\{\psi(f)\|\widetilde{f}\|_{L^1(\rho_\infty)} + \mu_2(\|\widetilde{f}\|_{L^1(\rho_\infty)} - 1)(\widetilde{f} - 1)\right\} \widetilde{\rho}_\infty\, dx \\ &\leq b\|\widetilde{f}\|_{L^1(\rho_\infty)} \int_{\mathbb{R}^d} \psi(f)\rho_\infty\, dx \leq \frac{b}{2\lambda}\|\widetilde{f}\|_{L^1(\rho_\infty)} \int_{\mathbb{R}^d} \psi''(f)\nabla f^\top D\nabla f\rho_\infty\, dx\end{aligned}$$

Using the first inequality in (3.9) and $1/b \leq \|\widetilde{f}\|_{L^1(\rho_\infty)}$, we conclude

$$\leq \frac{b}{2\lambda a}\|\widetilde{f}\|^{-1}_{L^1(\rho_\infty)} \int_{\mathbb{R}^d} \psi''(\widetilde{f})\nabla f^\top D\nabla f\widetilde{\rho}_\infty\, dx \leq \frac{b^2}{2\lambda a} \int_{\mathbb{R}^d} \psi''(\widetilde{f})\nabla\widetilde{f}^\top D\nabla\widetilde{f}\widetilde{\rho}_\infty\, dx.$$

□

We conclude this chapter with the following remarks.

Remark 3.16. Theorem 3.15 seems to be very satisfying, since it extends the convergence results based on Theorem 3.14 to potentials which are uniformly convex with constant λ out of some bounded set, for instance.

There are conditions which imply the existence of a spectral gap λ_0 (i.e. $\sigma(L)\backslash\{0\}$ has distance λ_0 from 0) and hence infer exponential convergence with rate λ_0. For instance, following [AMTU01, Section 2, p. 9-11] one uses the substitution $h_t = \rho_t/\sqrt{\rho_\infty}$ in (2.1) yielding

$$\partial_t h_t = -\widetilde{L}h_t = \operatorname{div}(D\nabla h_t) - \widetilde{V}h_t, \quad \widetilde{V} := -\frac{1}{2}e^{V/2}\operatorname{div}(e^{-V/2}D\nabla V).$$

The operator $\widetilde{L}$ is densely defined on $L^2(\mathbb{R}^d)$ and can be extended to a self-adjoint operator (see [AMTU01, Section 2, p. 10]). Now, if $\widetilde{V}(x) \to \infty$ as $|x| \to \infty$, then nearly the same proof in [RS78, Theorem XIII.67, p. 249] implies that $\widetilde{L}$ has compact resolvent, i.e. $(\widetilde{L} - \lambda)^{-1}$ is a compact operator for all $\lambda \notin \sigma(\widetilde{L})$. This implies that the spectrum of $\widetilde{L}$ is discrete and the eigenvalues diverge to infinity. Hence, $\widetilde{L}$ has a spectral gap. The same assertions are then also true for L.

In the case $D \equiv I$ the above condition reads

$$-\frac{1}{2}\Delta V(x) + \frac{1}{4}|\nabla V(x)|^2 \to +\infty, \quad \text{for } |x| \to +\infty.$$

We note that such a growth condition will enable us to prove exponential convergence in Section 4.3.

In the general case of inhomogeneous diffusion ($D \not\equiv I$) the above condition lack of physical meaning. In contrast to that, we assumed V to be uniformly convex and D to be uniformly elliptic while using the entropy method.

On the other hand, one can show that a convex Sobolev inequality (3.6) implies a spectral gap λ_0, which is in general larger than λ. More precisely, the convex Sobolev inequality implies a Poincaré inequality, see for instance [Rot81, Section 3]. However, the entropy method enables one to give a rate λ right away from the equation rather than calculating the spectral gap.

4. Ergodic methods

In this chapter, we will deal with the long-time behavior of solutions to the stochastic differential equation (2.2). As we already know (under conditions on D, V, see Theorem 2.13) they constitute a strong, time-homogeneous Markov family $(X_t, \mathscr{F}_t, t \geq 0, \mathbb{P}_x)$ with infinitesimal generator $\mathcal{L}$, see (2.4). We denote the transition kernel by $\mathcal{P}_t$.

In the first and second section, we outline results from the theory of Markov processes. We give general conditions which infer exponential convergence of the law to the invariant measure in weighted L^∞ spaces. In addition, we state a theorem yielding convergence in total variational norm with a subgeometric rate and prove lower bounds.

In the final section, we apply the above statements to the Markov process which solves (2.2) by imposing conditions on the potential V and the diffusion matrix D.

4.1. Exponential convergence

The following result of Markov processes are formulated in discrete time and can then be applied to continuous time without any losses. Therefore, suppose we were given a time-homogeneous Markov chain $X = (X_n, \mathscr{F}_n, \mathbb{P}_x)$ with transition kernel $\mathcal{P}$. We will need the following two assumptions for the major theorem.

(i) Lyapunov condition: There is a function $G : \mathbb{R}^d \to [0, \infty)$ and constants $K \geq 0$, $\gamma \in (0, 1)$ such that for all $x \in \mathbb{R}^d$

$$(\mathcal{P}G)(x) \leq \gamma G(x) + K.$$

(ii) Doeblin's condition: There exists a constant $\alpha \in (0, 1)$ and a probability measure π with

$$\inf_{x \in \mathcal{C}} \mathcal{P}(x, \cdot) \geq \alpha \pi(\cdot),$$

where $\mathcal{C} = \left\{ x \in \mathbb{R}^d \; : \; G(x) \leq R \right\}$ for some $R > 2K/(1-\gamma)$.

Remark 4.1. We give a brief intuitive explanation of the above conditions and assume that $G(x) \to +\infty$ for $|x| \to +\infty$ (as it will be the case in Section 4.3). Roughly speaking, the first condition prohibits arbitrary big jumps of X during one time step when starting in the "center" $\{G \leq R\}$ of G ($R > 0$ some number). The second assumption allows one to bound the probability of what is happening in the next time step uniformly with respect to starting points in the center of G.

Thus, the Markov chain cannot leave the center immediately and the behavior in the following time step is similar when starting in this region.

With the function G one defines the weighted L^∞_G-norm by

$$\|\varphi\|_{L^\infty_G} = \sup_x \frac{|\varphi(x)|}{1+G(x)}$$

for measurable functions φ. Furthermore, define L^∞_G to be the space of measurable functions for which the above norm is finite. Certainly, L^∞_G is complete and contains L^∞.

Now, the major theorem, which provides convergence with exponential rate in the weighted space L^∞_G, reads as follows:

Theorem 4.2. *Consider a time-homogeneous Markov chain $(X_n, \mathscr{F}_n, n \in \mathbb{N}, \mathbb{P}_x)$ which satisfies the above conditions. Then, there exists a unique invariant measure μ_∞. It satisfies*

$$\int G \, d\mu_\infty < \infty.$$

In addition, there exists $C > 0$, $\rho \in (0,1)$ with

$$\left\| \mathcal{P}_n \varphi - \int \varphi \, d\mu_\infty \right\|_{L^\infty_G} \leq C\rho^n \left\| \varphi - \int \varphi \, d\mu_\infty \right\|_{L^\infty_G}$$

for all $\varphi \in L^\infty_G$.

The proof is provided in [Hai16, Theorem 3.6, p.13] or [Sto15, Theorem 3.4, p. 47].

Now we want to apply this theorem to the solution of (2.2), which is a time-continuous Markov processes having continuous paths. We will consider for some fix time $T > 0$ the embedded Markov chain $(X^T_n, \mathscr{F}^T_n, n \in \mathbb{N}, \mathbb{P}_x)$ defined by $X^T_n := X_{nT}$, $\mathscr{F}^T_n := \mathscr{F}_{nT}$. The transition kernel is therefore $\mathcal{P} = \mathcal{P}_T$.

Furthermore, it is convenient to adapt the Lyapunov condition in terms of the generator of the time-continuous Markov process. One sufficient condition is (see [Hai16, Exercise 3.3, p. 12]):

Lemma 4.3. *Consider the time-homogeneous Markov family $(X_t, \mathscr{F}_t, t \geq 0)$ solving (2.2). Suppose there exists a function $G : C^2(\mathbb{R}^d; [0,\infty))$ and constants $c, K > 0$ with $\mathcal{L}G \leq K - cG$. Then, for every fixed time $T > 0$ the above Lyapunov condition holds for the embedded Markov chain $(X^T_n)_{n\in\mathbb{N}}$ with transition kernel $\mathcal{P}_T$.*

Proof. Applying Itô's rule to $e^{ct}G(X_t)$ yields

$$e^{-cT}G(X_T) - G(X_0) \leq \int_0^T e^{-cs}(K - cG(X_s)) \, ds + \int_0^T \nabla G(X_s)^\top \sigma(X_s) \, dW_s.$$

The stochastic integral is a local martingale. Therefore, after picking a localizing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ and taking expectations we obtain

$$\mathbb{E}_x\left[e^{c(T\wedge\tau_n)}G(X_{T\wedge\tau_n})\right] - G(x) \leq \mathbb{E}_x\left[\int_0^{T\wedge\tau_n} e^{-cs}(K - cG(X_s)) \, ds\right] \leq \frac{K}{c}\left(e^{cT} - 1\right).$$

Since the left integrand is nonnegative and continuous, we have by Fatou's Lemma

$$\mathbb{E}_x[e^{cT}G(X_T)] \le G(x) + \frac{K}{c}\left(e^{cT}-1\right).$$

This is equivalent to

$$P_T G(x) \le e^{-cT}G(x) + \frac{K}{c}\left(1-e^{-cT}\right) \le e^{-cT}G(x) + \frac{K}{c}. \tag{4.1}$$

Since $c, T > 0$, we have $\gamma := e^{-cT} \in (0,1)$, which concludes the proof. □

Now, we can provide a time-continuous version of the above theorem.

Theorem 4.4. *Consider the time-homogeneous Markov family $(X_t, \mathscr{F}_t, t \ge 0)$ which solves (2.2) and assume that it satisfies the assumption of the previous Lemma 4.3. Furthermore, assume that for some $T > 0$ the transition kernel $\mathcal{P}_T$ of the corresponding embedded Markov chain satisfies Doeblin's condition. Then, the conclusion of Theorem 4.2 is also true for $(X_t)_{t\ge 0}$. In particular, we have the estimate*

$$\left\| \mathcal{P}_t\varphi - \int \varphi\, d\mu_\infty \right\|_{L^\infty_G} \le C\rho^{\lfloor t/T \rfloor} \left\| \varphi - \int \varphi\, d\mu_\infty \right\|_{L^\infty_G}$$

for constants $\rho \in (0,1)$, $C > 0$ and for all $\varphi \in L^\infty_G$.

Remark 4.5. Certainly, by changing the constant C and defining $\lambda := -\ln\rho > 0$ we obtain $Ce^{-\lambda t}$ on the right side.

Proof. We fix $T > 0$ and consider again the corresponding embedded Markov chain $(X_n^T)_{n\in\mathbb{N}}$ satisfying the assumptions of Theorem 4.2. Therefore, there exists an invariant measure μ_∞^T for $(X_n^T)_{n\in\mathbb{N}}$. Define

$$\mu_\infty(A) := \frac{1}{T}\int_0^T \mathcal{P}_t\mu_\infty^T(A)\, dt$$

and observe that μ_∞ is invariant for X ([Hai16, Proposition 2.8, p. 11]). Indeed,

$$\mathcal{P}_s\mu_\infty(A) = \frac{1}{T}\int_0^T \mathcal{P}_{t+s}\mu_\infty^T(A)\, dt = \frac{1}{T}\int_s^{T+s} \mathcal{P}_t\mu_\infty^T(A)\, dt = \mu_\infty(A)$$

since

$$\int_T^{T+s} \mathcal{P}_t\mu_\infty^T(A)\, dt = \int_0^s \mathcal{P}_t\mathcal{P}_T\mu_\infty^T(A)\, dt = \int_0^s \mathcal{P}_t\mu_\infty^T(A)\, dt$$

using the invariance of μ_∞^T. Note that

$$t \mapsto \mathcal{P}_t\mu_\infty^T(A) = \int_A \mathbb{P}_x(X_t \in A)\, d\mu_\infty^T(x)$$

is measurable, since for open sets A this follows from the continuity and hence for all measurable sets $A \in \mathscr{B}(\mathbb{R}^d)$.

Furthermore, we have by (4.1)

$$\int G\,d\mu_\infty = \frac{1}{T}\int_0^T \int \mathcal{P}_t G(x)\,d\mu_\infty^T(x)\,dt \le \frac{1}{T}\int_0^T \int \left(\frac{K}{c} + e^{-ct}G(x)\right)\,d\mu_\infty^T(x)\,dt < \infty.$$

Uniqueness will follow from the convergence towards the invariant measure.

Fix $t > 0$ and set $n \in \mathbb{N}$ such that $t = nT + h$ with $h < T$. We have for any $\varphi \in L_G^\infty$

$$\mathcal{P}_{h+Tn}\varphi(x) - \int \varphi\,d\mu_\infty = \int \mathcal{P}_h\varphi(y)\;(d\mathcal{P}_n(x,y) - d\mu_\infty(y))\,.$$

Now, set $\phi := \mathcal{P}_h\varphi$ and observe by (4.1)

$$\sup_x \frac{|\phi(x)|}{1+G(x)} \le \sup_x \frac{\|\varphi\|_{L_G^\infty}}{1+G(x)} \int (1+G(y))\,d\mathcal{P}_h(x,y) \le \|\varphi\|_{L_G^\infty} \sup_x \frac{1+K/c+e^{-ch}G(x)}{1+G(x)}$$

which is bounded by $(1+K/c)\,\|\varphi\|_{L_G^\infty}$, hence $\phi \in L_G^\infty$. This also shows $\|\mathcal{P}_h\|_{L_G^\infty} \le 1 + K/c$. Thus, we obtain from the convergence result of the embedded Markov chain

$$\left\|\mathcal{P}_n\phi - \int \phi\,d\mu_\infty\right\|_{L_G^\infty} \le C\rho^n \left\|\mathcal{P}_h\varphi - \int \mathcal{P}_h\varphi\,d\mu_\infty\right\|_{L_G^\infty} = C\rho^n \left\|\mathcal{P}_h\left(\varphi - \int \varphi\,d\mu_\infty\right)\right\|_{L_G^\infty}$$

which is

$$\left\|\mathcal{P}_t\phi - \int \phi\,d\mu_\infty\right\|_{L_G^\infty} \le (1+K/c)C\rho^n \left\|\varphi - \int \varphi\,d\mu_\infty\right\|_{L_G^\infty}.$$

Redefining C concludes the proof. □

4.2. Subgeometric convergence and lower bounds

The Lyapunov condition in the previous section was strong in the sense that it implies exponential convergence. However, sometimes this rate of convergence is not true and it makes sense to consider other rates of convergence as well as lower bounds.

The first statement relies on a Lyapunov condition together with a minorization condition (see [Hai16, Theorem 4.1, p. 16]).

(i) Lyapunov condition: There exists a continuous function $G : \mathbb{R}^d \to [1, \infty)$ such that $\{G \le c\}$ is compact for all $c \ge 1$. Furthermore, it holds $\mathcal{L}G \le K - \varphi(G)$ for some constant K and a strictly concave function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$. In addition, $\varphi(0) = 0$ and φ is increasing to infinity.

(ii) Minorization condition: For every $C > 0$ there is $\alpha > 0$, $T > 0$ such that

$$\|\mathcal{P}_T(x,\cdot) - \mathcal{P}_T(y,\cdot)\|_{TV} \le 2(1-\alpha)$$

for every x, y with $G(x) + G(y) \le C$.

Theorem 4.6. *Consider a strong Markov family $(X_t, \mathscr{F}_t, t \geq 0)$ with continuous paths. Under the above assumptions there exists a unique invariant measure μ_∞, which furthermore satisfies $\int \varphi(G)\, d\mu_\infty < \infty$.*

Define

$$H_\varphi(u) := \int_1^u \frac{ds}{\varphi(s)},$$

then we have

$$\|\mathcal{P}_t(x, \cdot) - \mu_\infty\|_{TV} \leq \frac{CG(x)}{H_\varphi^{-1}(t)} + \frac{C}{(\varphi \circ H_\varphi^{-1})(t)}$$

for some constant C and all $x \in \mathbb{R}^d$.

Remark 4.7. The proof is provided in [Hai16, Theorem 4.1, p. 16-22] and is based on the so called coupling method. The idea is the following. One takes two realizations $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ of the Markov process at hand with initial condition x_0, y_0, respectively, and makes use of the coupling inequality

$$\|\mathcal{P}_t(x_0, \cdot) - \mathcal{P}_t(y_0, \cdot)\|_{TV} \leq 2\mathbb{P}(X_t \neq Y_t).$$

(This inequality follows directly from the definition of the total variational norm). Now it depends on our realizations $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ whether $\mathbb{P}(X_t \neq Y_t)$ converges to zero or not. For instance, if we build our probability space together with $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ in such a way, that they are independent, then it does not tend to zero. So, one has to find two realizations which are "independent enough" such that the above probability converges and that one obtains estimates. Such constructions are called couplings of the law of $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$.

Now, we turn to lower bounds of convergence ([Hai16, Section 5, p. 23-24]).

Theorem 4.8. *Suppose μ_∞ is an invariant measure of some Markov family $(X_t, \mathscr{F}_t, t \geq 0)$ with continuous paths. Furthermore, assume that there is a function $G : \mathbb{R} \to [1, \infty)$ satisfying:*

- *(i) There is some function $f : [1, \infty) \to [0, 1]$ with $\mathrm{Id} \cdot f : y \mapsto yf(y)$ increasing to infinity and $\mu_\infty(G \geq y) \geq f(y)$ for every $y \geq 1$.*
- *(ii) There exists a function $g : \mathbb{R} \times \mathbb{R}_+ \to [1, \infty)$ increasing in its second argument with $\mathbb{E}_x G(X_t) \leq g(x, t)$ for every initial condition $x \in \mathbb{R}^d$.*

Then, we have

$$\|\mathcal{P}_t(x, \cdot) - \mu_\infty\|_{TV} \geq \frac{1}{2} f\left((\mathrm{Id} \cdot f)^{-1}\left(2g(x, t)\right)\right)$$

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

Proof. First we observe for fixed $y \geq 1$ and $t \geq 0$

$$\|\mathcal{P}_t(x,\cdot) - \mu_\infty\|_{TV} \geq |\mathcal{P}_t(x,\cdot) - \mu_\infty|(\{G(x) \geq y\}) \geq \mu_\infty(G(x) \geq y) - \mathcal{P}_t(x, \{G(x) \geq y\}).$$

By Markov's inequality we have

$$\mathcal{P}_t(x, \{G(x) \geq y\}) = \mathbb{P}_x(G(X_t) \geq y) \leq \frac{g(x,t)}{y},$$

where the last inequality uses the second assumption. Together with the first condition we obtain

$$\|\mathcal{P}_t(x,\cdot) - \mu_\infty\|_{TV} \geq f(y) - \frac{g(x,t)}{y}.$$

By assumption there exists y_0 with $(\mathrm{Id} \cdot f)(y_0) = 2g(x,t)$, hence $f(y_0) = 2g(x,t)/y_0$. After setting $y = y_0$ above we establish

$$\|\mathcal{P}_t(x,\cdot) - \mu_\infty\|_{TV} \geq \frac{g(x,t)}{y_0} = \frac{1}{2} f(y_0) = \frac{1}{2} f\left((\mathrm{Id} \cdot f)^{-1} (2g(x,t)) \right).$$

□

4.3. Application to the Fokker-Planck equation

As it is always the case we will impose some conditions on the coefficients V, D in order to use the previous results. Those assumptions also provide the global existence of the Markov family (i.e. Proposition 2.9 applies). First of all, we check the validity of Doeblin's condition as well as the minorization condition. Suppose we already have a Lyapunov function G such that $G(x) \to +\infty$ for $|x| \to +\infty$ (as this will be the case later on).

Both conditions then follow from the existence of a smooth density $p(t,x,y)$ of $\mathcal{P}_t$ under $\mathbb{P}_x$ (see Remark 2.14). We know that $p(t,x,y) > 0$ for $t > 0$ by Theorem 2.13 (iv). By continuity there is for every compact set $\mathcal{C} \subset \mathbb{R}^d$ a constant $\alpha > 0$ with $p(T,x,y) \geq \alpha$ for all $x, y \in \mathcal{C}$ (and any fixed time $T > 0$). Hence,

$$p(T,x,\cdot) \geq \alpha\lambda(\mathcal{C})\pi(\cdot), \quad \pi(\cdot) := \frac{1}{\lambda(\mathcal{C})}\lambda(\cdot \cap \mathcal{C})$$

whenever $x \in \mathcal{C}$. At last we have to ensure that $\mathcal{C} := \left\{x \in \mathbb{R}^d \ : \ G(x) \leq R\right\}$ is compact for every $R \geq 0$ in order to provide Doeblin's condition. But this will always be the case.

Furthermore, the minorization condition is also satisfied. Again consider a Lyapunov function like in Subsection 4.2 and fix any time $T > 0$. Again we know $p(T,x,z) > 0$. Fix some $z^* \in \mathbb{R}^d$. Choose $C > 0$ arbitrary and define the set

$$\mathcal{C} := \{(x,y) \ : \ G(x) + G(y) \leq C\} \subset \{G \leq C\}^2,$$

where $\{G \leq C\}$ is compact by the assumptions on G. Therefore, $\{G \leq C\} \times \overline{B_1(z^*)}$ is also compact and there is some $c > 0$ with

$$p(T,x,z) \geq c$$

for all $(x,z) \in \{G \le C\} \times \overline{B_1(z^*)}$, where $B_1(z^*)$ denotes the unit ball with center z^*. Fix $(x,y) \in \mathcal{C}$ and define $\mathcal{X}^+ := \{p(T,x,\cdot) \ge p(T,y,\cdot)\}$. Then,

$$\begin{aligned}\int |p(T,x,z) - p(T,y,z)|\,dz &\le 1 - \int_{\mathcal{X}^+} p(T,y,z)\,dz + 1 - \int_{(\mathcal{X}^+)^c} p(T,x,z)\,dz \\ &\le 2 - c\lambda(\mathcal{X}^+ \cap B_1(z^*)) - c\lambda((\mathcal{X}^+)^c \cap B_1(z^*)) = 2 - c\lambda(B_1(z^*)).\end{aligned}$$

Define $\alpha := c\lambda(B_1(z^*))/2 > 0$ which does not depend on x, y (whereas $\mathcal{X}^+$ does). So we finally proved

$$\|\mathcal{P}_T(x,\cdot) - \mathcal{P}_T(y,\cdot)\|_{TV} \le 2(1-\alpha)$$

for all $(x,y) \in \mathcal{C}$.

The only thing which is left, is the existence of a Lyapunov function G such that $\{G \le C\}$ is compact for every $C > 0$. At first we deal with a homogeneous diffusion, i.e. $D \equiv I$. The general situation works then similar when imposing further assumptions.

Homogeneous diffusion

Here we follow [Hai16, Sec 7, p. 34-35] and [Sto15, Subsection 4.3.3, p. 66-67].

We will make the following assumptions (in addition Assumption 2.1): there are constants c_1, $c_2 > 0$ and $k \in (0,\infty)$ with

$$\langle \nabla V(x), x\rangle \ge c_1 |x|^{2k}, \quad |\text{Hess } V(x)| \le c_2 |x|^{2k-2} \tag{4.2}$$

for $|x| \ge M$, for some constant $M \ge 0$. (One might even assume that V and all its derivatives grow at most polynomially in order to ensure the existence of a smooth density, see Remark 2.14 or [Hai16, Sec 7, p. 34]). Thus, V should grow like $|x|^{2k}$ at infinity. Note that under those assumptions the Markov process exists for all times (Lemma 2.11).

The generator of the Markov process reads

$$\mathcal{L}\varphi = \Delta\varphi - \nabla V \cdot \nabla\varphi.$$

There are a few Lyapunov functions that work out. However, one has to consider the case $k \ge 1/2$ and $k \in (0, 1/2)$.

Case I ($k \ge 1/2$): Define $G(x) = \exp(\alpha V(x))$, yielding

$$\mathcal{L}G(x) = \alpha(\Delta V(x) + (\alpha - 1)|\nabla V(x)|^2)\exp(\alpha V(x)). \tag{4.3}$$

With Cauchy's inequality and the assumptions one observes $|\nabla V(x)|^2 \ge C|x|^{4k-2}$ as well as $|\Delta V(x)| \le C|x|^{2k-2}$ for some constant $C > 0$ and $|x| \ge M$. If $k > 1/2$ the gradient diverges faster than Hess V. If $k = 1/2$ then $|\Delta V(x)|$ tends to zero for $|x| \to \infty$ whereas $|\nabla V(x)|^2 \ge C > 0$. Hence, when fixing $\alpha \in (0,1)$ we have for some $c > 0$ and $\widetilde{M} \ge 0$ large enough

$$\mathcal{L}G(x) \le -cG(x)$$

whenever $|x| \geq \widetilde{M}$. By continuity we can find $K \geq 0$ with

$$\mathcal{L}G(x) \leq K - cG(x)$$

for all $x \in \mathbb{R}^d$ and Theorem 4.4 applies.

Another type of Lyapunov function would be $G(x) = |x|^n$ for $n \in \mathbb{N}$, $n \geq 2$. We have for some constants $c, K > 0$ like above

$$\begin{aligned}\mathcal{L}G(x) &= n(d+n-2)|x|^{n-2} - n\langle \nabla V(x), x\rangle |x|^{n-2} \\ &\leq n(d+n-2)|x|^{n-2} - nc_1|x|^{n-2+2k} \leq K - cG(x)\end{aligned}$$

if $k \geq 1$. (But we only needed the growth condition on ∇V).

One could also use $G(x) = \exp(\alpha|x|^n)$ with $\alpha > 0$ and $n \in \mathbb{N}$, $n \geq 2$ in order to obtain

$$\begin{aligned}\mathcal{L}G(x) &= (\alpha n(d+n-2)|x|^{n-2} + n^2\alpha^2|x|^2|x|^{2n-4} - \alpha n\langle \nabla V(x), x\rangle |x|^{n-2})\exp(\alpha|x|^n) \\ &\leq \alpha n((d+n-2)|x|^{n-2} + n\alpha|x|^{2n-2} - c_1|x|^{n-2+2k})\exp(\alpha|x|^n)\end{aligned}$$

for large $|x|$. This works out if $2k > n$.

Theorem 4.4 implies that there is a unique invariant measure μ_∞, which we already know to be $e^{-V}\lambda$, with $\int G\,d\mu_\infty < \infty$. Furthermore, for all $x \in \mathbb{R}^d$

$$\left|\mathcal{P}_t\varphi(x) - \int \varphi\,d\mu_\infty\right| \leq (1+G(x))C\rho^{\lfloor t/T\rfloor}\left\|\varphi - \int \varphi\,d\mu_\infty\right\|_{L^\infty_G}$$

for some $C > 0$, $\rho \in (0,1)$ and all $\varphi \in L^\infty_G$.

If we want to consider an initial condition ξ with distribution μ we just consider the Markov process X on $(\Omega, \mathscr{F}, \mathbb{P}_\mu)$ with

$$\mathbb{P}_\mu(F) := \int \mathbb{P}_x(F)\,d\mu(x)$$

for $F \in \mathscr{F}$. Thus, by integrating the above inequality

$$\left|\mathbb{E}_\mu[\varphi(X_t)] - \int \varphi\,d\mu_\infty\right| \leq \left(1 + \int G\,d\mu\right)C\rho^{\lfloor t/T\rfloor}\left\|\varphi - \int \varphi\,d\mu_\infty\right\|_{L^\infty_G} \tag{4.4}$$

Certainly, one has to ensure $\int G\,d\mu < \infty$. For instance, in the case $k \geq 1$ one can choose $G(x) = |x|^2$ $(n = 2)$ and obtains exponential convergence for initial data having finite second order moments.

Case II ($k \in (0, 1/2)$): In the following, we consider $k \in (0, 1/2)$ which implies that $|\text{Hess}\,V|$ converges to zero for $|x| \to \infty$. In addition to the above conditions, assume that there are constants $A, B > 0$ with

$$A|x|^{2k} \leq V(x) \leq B|x|^{2k} \tag{4.5}$$

for $|x| \geq M$ and some $M > 0$. For notational simplicity we write $|x|^{2k} \lesssim V(x) \lesssim |x|^{2k}$ instead of (4.5) and similar inequalities since the constants do not matter. We choose again $G(x) = \exp(\alpha V(x))$ with $\alpha \in (0,1)$ and obtain

$$\mathcal{L}G(x) \leq K - c|x|^{4k-2} \exp(\alpha V(x)).$$

Since $|x|^{2k} \lesssim \ln G(x) \lesssim |x|^{2k}$, we have

$$|x|^{4k-2} \lesssim (\ln G(x))^{2-1/k} \lesssim |x|^{4k-2} \tag{4.6}$$

and

$$\mathcal{L}G(x) \leq K - c\frac{G(x)}{(\ln G(x))^{1/k-2}}$$

redefining c. The function $\varphi(x) := cx/(\ln x)^{1/k-2}$ is strictly concave (with continuous extension $\varphi(0) = 0$) and increases to infinity. Certainly, $G \geq 1$ also increases to infinity. We can apply Theorem 4.6 and therefore set

$$H_\varphi(u) = \int_1^u \frac{(\ln x)^{1/k-2}}{cx}\,dx = \int_0^{\ln u} \frac{y^{1/k-2}}{c}\,dy = \frac{(\ln u)^{1/k-1}}{c(1/k-1)}$$

with $H_\varphi^{-1}(v) = \exp(\beta v^{k/(1-k)})$, $\beta > 0$. Furthermore,

$$(\varphi \circ H_\varphi^{-1})(t) = c\frac{\exp(\beta t^{k/(1-k)})}{\beta t^{(1-2k)/(1-k)}}$$

and hence $1/(\varphi \circ H_\varphi^{-1})(t)$ is bounded. We obtain the following convergence result from Theorem 4.6 (absorbing the second summand in the first one by increasing the constant independently of x)

$$\left\|p(t,x,\cdot) - e^{-V}\right\|_{L^1} = \|\mathcal{P}_t(x,\cdot) - \mu_\infty\|_{TV} \leq C\exp(\alpha V(x) - \beta t^{\frac{k}{1-k}}). \tag{4.7}$$

Here, we can even give a lower bound using Theorem 4.8. Again with $G(x) = \exp(\alpha V(x))$ and (4.3), (4.6) we obtain for any $\alpha > 0$ the bounded

$$\mathcal{L}G \leq c|x|^{4k-2}G(x) \leq c\frac{G(x)}{(\ln G(x))^{1/k-2}}$$

with $c > 0$. Since $\mathcal{L}$ is the generator of X, we obtain (by using Fatou's lemma on the left and monotone convergence on the right side for the localizing sequence of stopping times)

$$\mathbb{E}_x[G(X_t)] - G(x) \leq \mathbb{E}_x\left[\int_0^t c\frac{G(X_s)}{(\ln G(X_s))^{1/k-2}}\,ds\right].$$

Recalling the definition of φ and the fact that it is concave, we use Jensen's inequality yielding

$$\mathbb{E}_x[G(X_{t+h})] - \mathbb{E}_x[G(X_t)] \leq \int_t^{t+h} \varphi(\mathbb{E}_x[G(X_s)])\,ds$$

which implies

$$\limsup_{h\searrow 0}\frac{1}{h}(\mathbb{E}_x[G(X_{t+h})]-\mathbb{E}_x[G(X_t)])\leq \varphi(\mathbb{E}_x[G(X_t)]).$$

The limit on the left is the right, upper derivative of $\mathbb{E}_x[G(X_t)]$. The function $g(x,t) = G(x)\exp(ct^{k/(1-k)})$ satisfies $\partial_t g = \varphi(g)$ and is therefore an upper solution with initial data $G(x)$. From the differential inequality we obtain

$$\mathbb{E}_x[G(X_t)]\leq g(x,t).$$

Note that g satisfies condition (ii) in Theorem 4.8 and only (i) is left. We observe for large $y\geq 1$ (in order to use (4.5))

$$\mu_\infty(G\geq y)=\mu_\infty(\alpha V\geq \ln y)\geq \int_{\{|x|^{2k}\geq \ln y/\alpha A\}} e^{-B|x|^{2k}}\,dx = C\int_{\{r^{2k}\geq \ln y/\alpha A\}} e^{-Br^{2k}}\,dr.$$

Recalling $2k-1<0$, we further estimate (where the constant C might change from line to line)

$$\begin{aligned} C\int_{\{r^{2k}\geq \ln y/\alpha A\}}\frac{r^{2k-1}}{r^{2k-1}}e^{-Br^{2k}}\,dr &\geq C\left(\frac{\ln y}{\alpha A}\right)^{(1-2k)/2k}\int_{\{r^{2k}\geq \ln y/\alpha A\}} r^{2k-1}e^{-Br^{2k}}\,dr\\ &= C\left(\frac{\ln y}{\alpha A}\right)^{1/2k-1} y^{-B/\alpha A}\geq Cy^{-B/\alpha A}\alpha^{1-1/2k}, \end{aligned}$$

where the last inequality uses the fact that any polynomial grows faster than the logarithm (and one can adapt the constant $C>0$). If we decrease C further we establish

$$\mu_\infty(G\geq y)\geq Cy^{-B/\alpha A}\alpha^{1-1/2k}$$

for all $y\geq 1$. Now, define $f(y) := Cy^{-B/\alpha A}\alpha^{1-1/2k}$ and increase α such that $\gamma := B/\alpha A < 1$ holds. Then, $yf(y)$ increases to infinity. Thus, Theorem 4.8 applies and in the following we calculate the lower bound. We have

$$(\mathrm{Id}\cdot f)^{-1}(z)=\left(\frac{z}{C\alpha^{1-1/2k}}\right)^{1/(1-\gamma)},$$

hence

$$(\mathrm{Id}\cdot f)^{-1}(2g(x,t))=\left(\frac{2}{C\alpha^{1-1/2k}}\right)^{1/(1-\gamma)}\exp\left(\frac{\alpha}{1-\gamma}V(x)+\frac{c}{1-\gamma}t^{k/(1-k)}\right).$$

Finally, we obtain

$$\left\|p(t,x,\cdot)-e^{-V}\right\|_{L^1}=\|\mathcal{P}_t(x,\cdot)-\mu_\infty\|_{TV}\geq C\exp\left(-\frac{\alpha\gamma}{1-\gamma}V(x)-\frac{c\gamma}{1-\gamma}t^{k/(1-k)}\right). \quad (4.8)$$

For fixed x the bound is of the type $C\exp(ct^{k/(1-k)})$ which is the same as in (4.7). Since $k/(1-k)<1$, there is no exponential convergence in the total variational norm (or in L^1 concerning the densities).

Inhomogeneous diffusion

The generator of the Markov process is

$$\mathcal{L}\varphi = \sum_{i,j} D_{ij}\partial^2_{ji}\varphi + (\mathrm{div}D - D\nabla V)\cdot\nabla\varphi.$$

All calculations from Case I in the previous analysis with the Lyapunov function $G(x) = \exp(\alpha V(x))$ are similar if we make the following assumptions: D and $\mathrm{div}D$ are uniformly bounded. Furthermore, we have the growth conditions

$$\langle D\nabla V(x), x\rangle \geq c_1|x|^{2k}, \quad |\mathrm{Hess}\ V(x)| \leq c_2|x|^{2k-2}$$

and $A|x|^{2k} \leq V(x) \leq B|x|^{2k}$.

We conclude this chapter with the following remarks.

Remark 4.9. (i) The results above based on the growth condition (4.2) cover a variety of situations. (Recall that we only needed the first inequality in the case $k \geq 1$, $D \equiv I$ when choosing the Lyapunov function $G(x) = |x|^n$). In particular, we established exponential convergence for potentials which have roughly speaking at least linear growth. In addition to that, not only did we prove subgeometric rates, but also showed the necessity of studying them via lower bounds for potentials having weaker growth properties.

Concerning the physical meaning of those conditions, the first inequality in (4.2) (in Case I with $k \geq 1$) is reminiscent of the convexity condition required in the last chapter. (In fact, the uniform convexity implies it with $k = 1$).

(ii) Although we gave properties enabling us to apply the theorems from the previous section, it does not seem to be easy to apply those results in a specific situation. One always has to find a Lyapunov function, which might not be a simple task.

(iii) One drawback of the *presented* results are the undetermined constants occurring in our final estimates (4.4), (4.7), (4.8). In particular, the rate ρ in (4.4) remains unknown here. In contrast to that, the entropy method allows one to give at least a certain order of magnitude for the rate in the case that V is not uniformly convex, but satisfies the conditions in Theorem 3.15.

5. Numerical simulations

5.1. Numerical procedure and examples

In this chapter, we discuss two numerical examples and apply the preceding convergence results. We consider the one-dimensional Fokker-Planck equation with homogeneous diffusion matrix

$$\partial_t u(t,x) = \partial_x^2 u(t,x) + V'(x)\partial_x u(t,x) + V''(x)u(t,x). \tag{5.1}$$

In the first example, the potential is a double-well potential, i.e. $V_1(x) = x^4 - x^2$. In the second example, we consider the potential

$$V_2(x) = \begin{cases} |x|^{9/10} & |x| \geq 1/10, \\ p(x) & |x| \leq 1/10. \end{cases}$$

Here p is some polynomial such that V is C^2. See Figure 5.1 for a plot of these two potentials. We also plot the effect on the first potential after adding a perturbation which allows us to use the perturbation result in Theorem 3.15. Below we also provide plots concerning the convergence in L^1, relative entropy as well as in weighted L^∞ spaces.

Before we focus on the numerical results we briefly explain the numerical methods used here. More precisely, we use both a finite difference scheme and the Euler-Maruyama method.

(i) *Finite difference scheme:* Take an interval $[-L, L]$ with $L > 0$ large enough such that values of the initial condition as well as the density of the invariant measure ρ_∞ outside of $[-L, L]$ can be neglected. In order to solve (5.1) on $[-L, L]$ we consider homogeneous Dirichlet boundary conditions and discretize derivatives according to a equidistant grid $-L = x_0 < x_1 < \cdots < x_N = L$, $h = x_{i+1} - x_i$ by

$$\partial_x^2 u(t,x_i) = \frac{1}{h^2}\left(u(t,x_{i+1}) - 2u(t,x_i) + u(t,x_{i-1})\right),$$
$$\partial_x u(t,x_i) = \frac{1}{2h}\left(u(t,x_{i+1}) - u(t,x_{i-1})\right).$$

This constitutes a discrete differential operator (i.e. a matrix) L_h yielding the linear ordinary differential equation in time

$$\frac{d}{dt}\bar{u}(t) = L_h \bar{u}(t) \tag{5.2}$$

for the time-dependent vector $\bar{u}(t) \in \mathbb{R}^{N+1}$ with initial condition $\bar{u}(0)$. The matrix L_h acts on $v \in \mathbb{R}^{N+1}$ as follows: $(L_h v)_0 = v_0$, $(L_h v)_N = v_N$ and for $i = 1, \ldots, N-1$

$$(L_h v)_i = \left(\frac{1}{h^2} - \frac{V'(x_i)}{2h}\right) v_{i-1} + \left(-\frac{2}{h^2} + V''(x_i)\right) v_i + \left(\frac{1}{h^2} + \frac{V'(x_i)}{2h}\right) v_{i+1}.$$

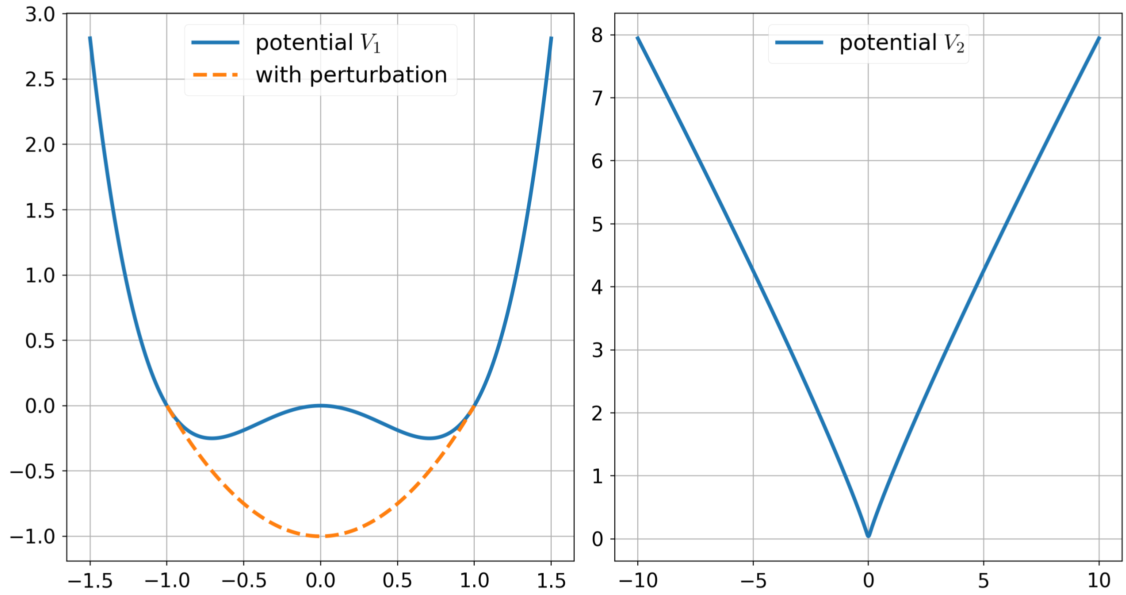

Figure 5.1.: Left: Plot of the potential $V_1(x) = x^4 - x^2$ and the change of V_1 after adding a perturbation. Right: Plot of the potential $V_2(x) = |x|^{9/10}$ for $|x| \geq 1/10$.

Note that the boundary conditions imply $\bar{u}_0(t) = \bar{u}_N(t) = 0$ and hence we only need to calculate the other $N-1$ vector entries. A numerical calculation shows that the eigenvalues of L_h are negative and growing in absolute value. Therefore, we use an L-stable Runge-Kutta method, which prevails the stability of the linear ordinary differential equation (5.2). (Here we used a three staged, fourth order DIRK method). Finally, we calculate the L^1 (or total variational) difference and the relative entropy with respect to $\rho_\infty = e^{-V}$ with a composite Simpson's rule based on the grid $(x_i)_i$.

(ii) *Euler-Maruyama scheme:* We discretized the stochastic differential equation

$$dX_t = -V'(X_t)dt + \sqrt{2}dW_t$$

according to a fixed time step τ. Given the value X_n one calculates

$$X_{n+1} = X_n - \tau V'(X_n) + \sqrt{2}(W_{(n+1)\tau} - W_{n\tau}),$$

which is in essence one step in the explicit Euler method. For the last term, we know that $S_{n+1} := W_{(n+1)\tau} - W_{n\tau}$ is centered Gaussian with standard deviation $\sqrt{\tau}$ independent of the past up to time $n\tau$. Now the scheme works as follows: one samples the initial condition yielding the value X_0. In order to obtain X_{n+1} from X_n one samples $S_{n+1} \sim \mathcal{N}(0, \tau)$ and sets

$$X_{n+1} = X_n - \tau V'(X_n) + \sqrt{2}S_{n+1}.$$

This simulation is performed N times, which gives N discrete sample paths $(X_0^i, \ldots, X_T^i)$, $i = 1, \ldots, N$, up to time $T\tau$ $(T \in \mathbb{N})$.

Then we calculate the empirical or statistical expectation of some test function φ by

$$\mathbb{E}_{\text{emp}}[\varphi(X_n)] := \frac{1}{N}\sum_{i=1}^{N}\varphi(X_n^i)$$

at the time point $n\tau$.

Example I: We consider the double-well potential $V_1(x) = x^4 - x^2$ and the initial condition $\rho_0(x) = e^{-(x-30)^2/2}/\sqrt{2\pi}$, a Gaussian density with mean 30 and variance one.

From our investigation in Section 4.3, see (4.4), we obtain exponential convergence of $\mathbb{E}_\mu[\varphi(X_t)]$ for all bounded functions φ. In Figure 5.2, the difference

$$\left|\mathbb{E}_{\text{emp}}[\varphi(X_n)] - \int \varphi \, d\mu_\infty\right| \tag{5.3}$$

is plotted based on the Euler-Maruyama scheme. The test functions are $\varphi_1(x) = x$ in the first (hence we calculate the mean) and $\varphi_2 = \mathbb{1}_{[-1,0]}$ in the second case. One can see that the exponential rate is approximately 1.7.

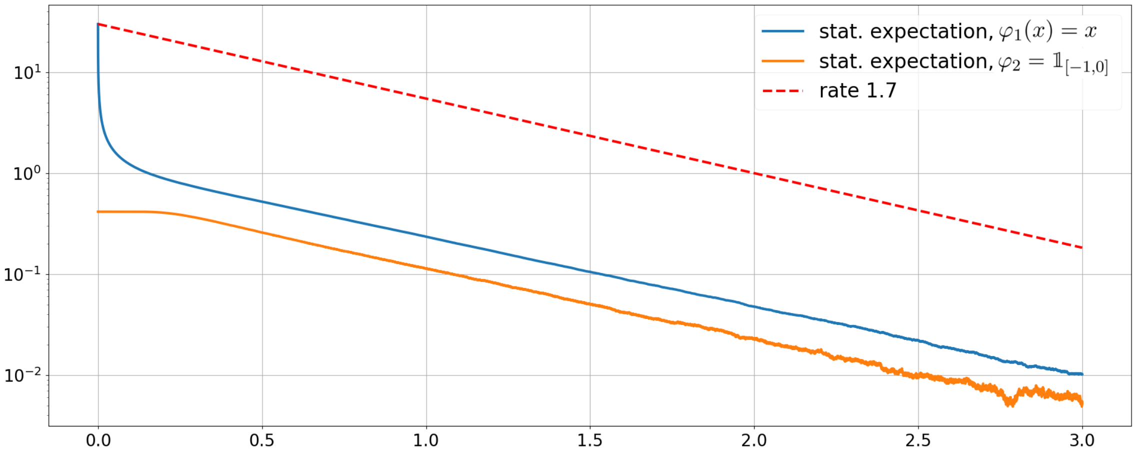

Figure 5.2.: Semilogarithmic[1] plot of absolute difference between statistical expectation and expectation with respect to the invariant measure against time (Example I). The test functions are $\varphi_1(x) = x$ and the indicator function $\varphi_2 = \mathbb{1}_{[-1,0]}$.

Moreover, one can add a bounded perturbation v in order to obtain a uniformly convex potential, see again Figure 5.1. Then Theorem 3.15 yields exponential convergence in L^1 (total variation) as well as in relative entropy.

In Figure 5.3, we plotted the L^1-difference as well as the relative entropy based on the finite difference scheme. The entropy here is generated by (like in (3.1) with $\alpha = \beta = 1$)

$$\chi(\sigma) = (\sigma + 1)\ln\left(\frac{\sigma+1}{2}\right) - (\sigma - 1).$$

[1] In a semilogarithmic plot only the second coordinate axis is logarithmic.

One can observe the exponential rate to be approximately $\lambda = 1.5$. We expect that the relative entropy decreases like $\exp(-2\lambda t)$ and due to Lemma 3.4 the L^1-norm like $\exp(-\lambda t)$, which is indeed the case in Figure 5.3.

Furthermore, observe that the rate is larger between the times 0 and approximately 0.3 in the right plot in Figure 5.3. This is because the initial density was "far" away from the center and hence the "attraction" is much bigger due to the term x^4.

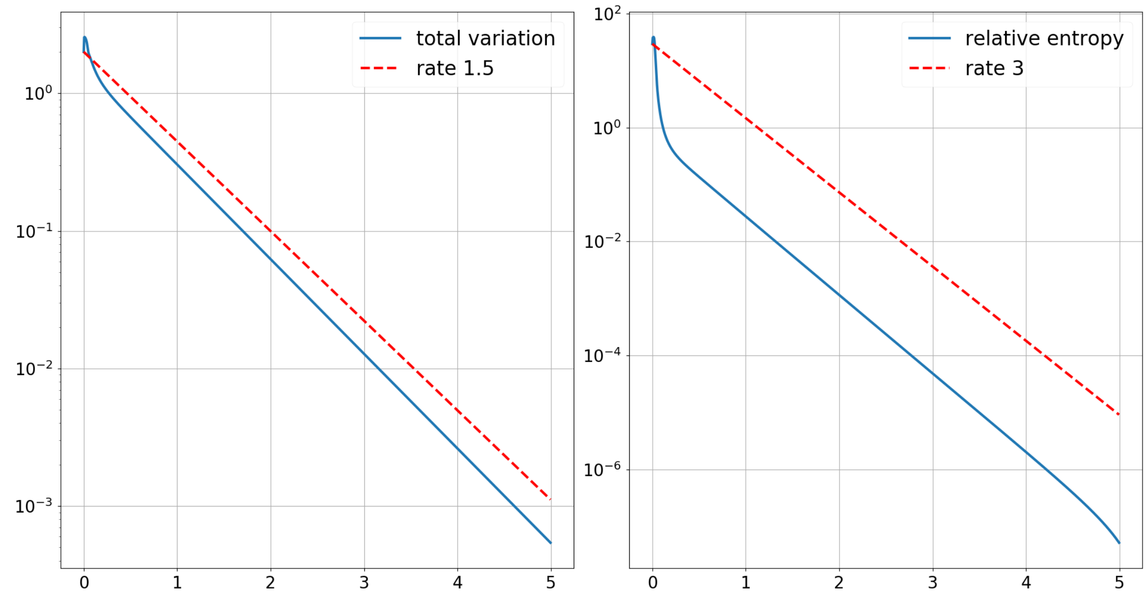

Figure 5.3.: Semilogarithmic plot of total variational distance and relative entropy of a solution with respect to the equilibrium distribution against time (Example I).

Example II: In the following we consider the potential

$$V_2(x) = \begin{cases} |x|^{9/10} & |x| \geq 1/10, \\ p(x) & |x| \leq 1/10. \end{cases}$$

Note again that p is some polynomial such that V is C^2. We have the initial condition $\rho_0(x) = e^{-(x-50)^2/2}/\sqrt{2\pi}$. In Section 4.3, Case II, we gave a subgeometric rate (4.7) as well as a lower bound (4.8) of the same form for the total variation when considering the fundamental solution. In this example, the growth rate of V is $k = 9/20$ (with the notation from Section 4.3) and the bounds have the form

$$C \exp\left(\alpha V_2(x) - \beta t^{9/11}\right), \quad C' \exp\left(-\alpha' V_2(x) - \beta' t^{9/11}\right).$$

The constants C, C', β, β' are positive. Note that, although it is not an exponential bound, the decay is quite fast (depending on β and β'). Furthermore, since we are interested in the process with the above initial condition the constants C and C' are, respectively, proportional to

$$\int \exp(\alpha V_2)\rho_0 \, dx, \qquad \int \exp(-\alpha' V_2)\rho_0 \, dx.$$

This follows from (4.7) and (4.8) by an integration with respect to $\rho_0 \lambda$. Therefore, by the unboundedness of the potential V we expect that the first one will be quite large and the second one small if one starts far away from the origin (like we do). However, even in this example it is difficult to apply the bounds since the constants (in particular β, β') are unknown (as we already discussed in Remark 4.9).

We turn to the numerical results. In Figure 5.4, the total variation is plotted based on the finite difference scheme, where the left one is semilogarithmic. One can see that an exponential convergence does not take place at the beginning. In Figure 5.4, to the right we added two upper bounds of the above form by choosing the constants according to the plot of the total variation. Hence, the correct error estimate (if one would know the values of the constants appearing in the upper bound) might or might not look like that. Nevertheless, observe that for these "guessed" bounds the constant C is indeed large.

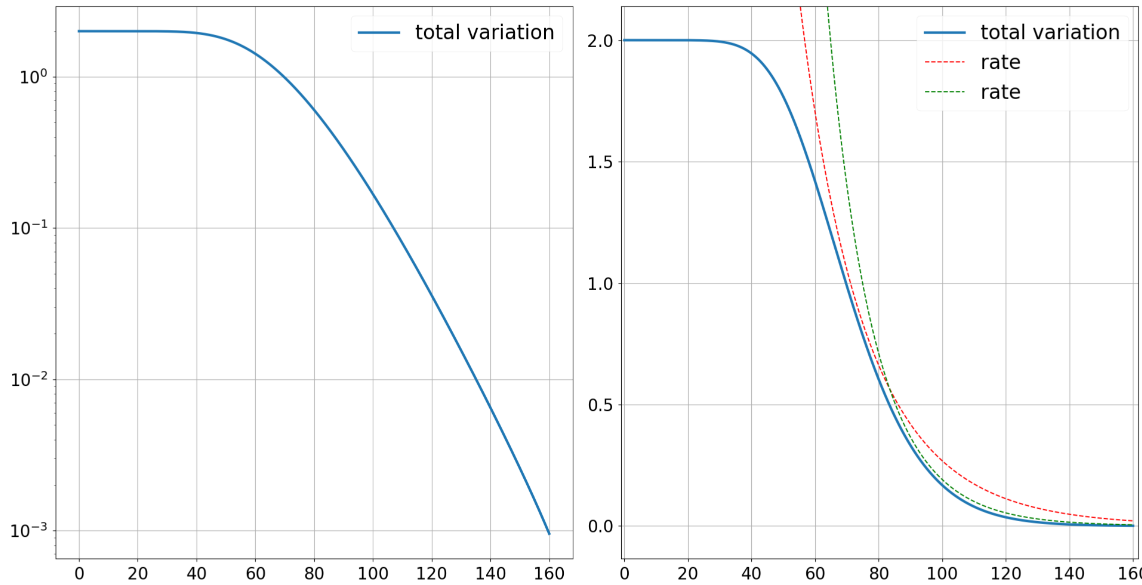

Figure 5.4.: Left: Semilogarithmic plot of total variational distance of the solutions with respect to the equilibrium against time (Example II). Right: Plot of the total variational distance together with two "guessed" rates against time.

In addition, we plotted the difference (5.3) in Figure 5.5 with the test functions $\varphi_3(x) = x$ (hence the mean) and $\varphi_4 = \mathbb{1}_{[10,60]}$. For this simulation we used the Euler-Maruyama scheme. Observe that there is (at least up to the time point 100 in Figure 5.5) no exponential convergence.

Note that both Figure 5.4 and Figure 5.5 suggest that the convergence is better than the lower bound, since the logarithmic plots show a concave shaped curve. But the lower bound of the form $C' \exp(-\beta' t^{9/11})$ would have a convex logarithmic plot. One reason for this might be that the constant C' is small, which we would expect as mentioned above. Furthermore, also the case that $\beta' > 0$ is large would allow our observation.

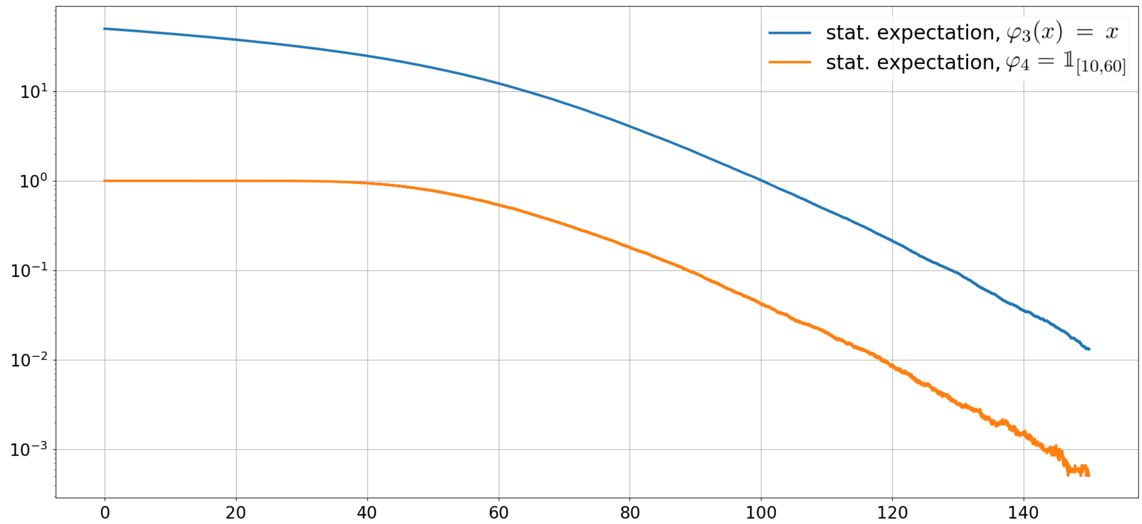

Figure 5.5.: Semilogarithmic plot of absolute difference between statistical average and the corresponding integral with respect to the invariant measure against time (Example II). The test functions are $\varphi_3(x) = x$ and the indicator function $\varphi_4 = \mathbb{1}_{[10,60]}$.

5.2. Conclusion

As we saw in Chapter 3, the entropy method enables us to give simple conditions implying the exponential tend to equilibrium. In particular, the perturbation result in Theorem 3.15 allowed us to widely extend the applicability.

The analysis via Markov processes allowed us to infer the same in weighted L^∞-spaces. Furthermore, we could also study the case where an exponential convergence does not take place via subgeometric rates and lower bounds. More precisely, the L^1-difference $\|\rho_t - \rho_\infty\|_{L^1}$ decays like $\exp(-\beta t^\alpha)$ for $0 < \alpha < 1$, $\beta > 0$. However, in our second example we could observe that conceptually the decay was faster than the lower bound of the form $C' \exp(-\beta' t^\alpha)$ would allow. At the very end of the last section we suggested that this is due to the constants C', β'. Unfortunately, this remains an unsatisfactory point in the preceding analysis.

In comparison, one advantage of the entropy method is the fact that it allowed us to calculate rates whereas such constants remain unknown in the stochastic approach. In particular, the constants appearing in the subgeometric and lower estimates remain unknown in Example II in the previous section. However, the growth conditions in Section 4.3 cover a lot of situations.

A. Appendix

A.1. Digression: Spectral theory

We give the definition of a symmetric/self-adjoint (unbounded) operator on a Hilbert space H. Furthermore, we recall a few properties of spectral measures and state the spectral theorem for unbounded self-adjoint operators. We make references to [Kal18], [Yos08] for full statements and proofs. Furthermore, we refer to [Kal18, Chapter 5, p. 100-118], [Yos08, IX.1-8, p. 231-251] for the theory of semigroups.

Definition A.1. A linear operator $A : \operatorname{dom} A \subset H \to H$, densely defined, is called symmetric if and only if $A \subset A^*$ (in the sense of linear relations), i.e. $\operatorname{dom} A \subset \operatorname{dom} A^*$ and for all $x, y \in \operatorname{dom} A$ it holds $(Ax, y) = (x, Ay)$. The operator A will be called self-adjoint if $A = A^*$, i.e. $\operatorname{dom} A = \operatorname{dom} A^*$ and A is symmetric.

For the definition of a spectral measure and the integral of an (un-)bounded measurable function with respect to a spectral measure see [Kal18, Definition 2.1.1, p. 31], [Kal18, Lemma 2.1.5, p. 36], [Kal18, Definition 4.5.2, p. 90].

Proposition A.2. *Let E denote a spectral measure on $(\Omega, \mathcal{A})$ and H. The mapping $\phi \mapsto \int_\Omega \phi \, dE$, $\phi : \Omega \to \mathbb{C}$ measurable, has the following properties:*

(i) If ϕ is bounded, then so is $\int \phi \, dE$.

(ii) $\phi \mapsto \int_\Omega \phi \, dE$ is a C^*-homomorphism, in particular for all ϕ, ψ bounded, measurable

$$\int_\Omega \phi \, dE \int_\Omega \psi \, dE = \int_\Omega \phi\psi \, dE$$

(iii) Whenever $\phi_n \to \phi$ pointwise and $\sup_n \|\phi_n\|_\infty < \infty$ for bounded, measurable functions ϕ_n, ϕ, then $\int \phi_n \, dE \to \int \phi \, dE$ in the strong operator topology.

(iv) If $g \in \operatorname{dom} \left(\int_\Omega \phi \, dE \right)$ then for all $h \in H$

$$\left\langle \int_\Omega \phi \, dE g, h \right\rangle = \int_\Omega \phi \, dE_{g,h},$$

where the Borel measure on the right side is defined by $A \mapsto E_{g,h}(A) := \langle E(A)g, h \rangle$.

(v) $g \in \operatorname{dom} \left(\int_\Omega \phi \, dE \right)$ if and only if $\int |\phi|^2 \, dE_{g,g} < \infty$.

(vi) If ϕ, ψ are bounded and $g, h \in H$ then

$$\int_\Omega \phi \, dE_{\int_\Omega \psi \, dE g, h} = \int_\Omega \phi\psi \, dE_{g,h}.$$

(vii) We have $\sigma\left(\int_\Omega \phi \, dE\right) \subset \overline{\phi(\Omega)}$ *with the closure in* $\mathbb{C} \cup \{+\infty\}$.

For the proof of the following theorem see [Kal18, Theorem 4.6.1, p. 96] or [Yos08, XI.6, p. 313].

Theorem A.3 (spectral theorem)**.** *Consider a self-adjoint operator* $A : \operatorname{dom} A \subset H \to H$. *Then there exists a unique spectral measure* E *on* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *and* H*, such that*

$$A = \int \lambda \, dE(\lambda).$$

Furthermore, $\sigma(A)$ *contains the support of* E*,* $E(\mathbb{R} \backslash \sigma(A)) = 0$.

A.2. Digression: Markov processes and stochastic differential equations

In this section we recall the definition of a (strong) Markov process and as an example Brownian motion (see [KS00, Section 2.5-2.6, p. 47-48, 71-89]). Furthermore, we give statements from the theory of stochastic differential equations.

Definition A.4 (Markov process)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and μ a Borel measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. A progressively measurable process $(X_t, \mathscr{F}_t, t \geq 0)$ with values in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ on this probability space is called a Markov process with initial distribution μ if

(i) $\mathbb{P}(X_0 \in A) = \mu(A)$ for all $A \in \mathscr{B}(\mathbb{R}^d)$,

(ii) $\mathbb{P}(X_{t+s} \in A \mid \mathscr{F}_s) = \mathbb{P}(X_{t+s} \in A \mid X_s)$ a.s. for all $s, t \geq 0$ and $A \in \mathscr{B}(\mathbb{R}^d)$.

Furthermore, X is called a strong Markov process if in addition

$$\mathbb{P}(X_{S+t} \in A \mid \mathscr{F}_S) = \mathbb{P}(X_{S+t} \in A \mid X_S)$$

is satisfied for all $t \geq 0$, $A \in \mathscr{B}(\mathbb{R}^d)$ and $(\mathscr{F}_t)$ stopping times S.

Definition A.5 (strong Markov family)**.** A d-dimensional, time-homogeneous strong Markov family is a progressively measurable process $(X_t, \mathscr{F}_t; t \geq 0)$ with values in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ on some measurable space $(\Omega, \mathscr{F})$ together with a family of probability measures $(\mathbb{P}_x)_{x \in \mathbb{R}^d}$ on $(\Omega, \mathscr{F})$ with the properties:

(i) $x \mapsto \mathbb{P}_x(F)$ is $\mathscr{B}(\mathbb{R}^d)$-$\mathscr{B}[0,1]$-measurable for every fixed $F \in \mathscr{F}$;

(ii) $\mathbb{P}_x(X_0 = x) = 1$ for all $x \in \mathbb{R}^d$;

(iii) for every $x \in \mathbb{R}^d$, $A \in \mathscr{B}(\mathbb{R}^d)$, $t \geq 0$ and every $(\mathscr{F}_t)$ stopping time S

$$\mathbb{P}_x(X_{S+t} \in A \mid \mathscr{F}_S) = \mathbb{P}_{X_S}(X_t \in A).$$

The advantage of a (strong) Markov family is that one obtains a (strong) Markov process with initial distribution μ by defining the probability measure

$$\mathbb{P}_\mu(F) := \int \mathbb{P}_x(F)\, d\mu(x),\ F \in \mathscr{F},$$

and considering the process X on $(\Omega, \mathscr{F}, \mathbb{P}_\mu)$.

Definition A.6. Consider a (strong) Markov family $(X_t, \mathscr{F}_t; t \geq 0)$ on $(\Omega, \mathscr{F})$. Then we define the following notions. Here, $B(\mathbb{R}^d; \mathbb{R})$ denotes the space of functions $\mathbb{R}^d \to \mathbb{R}$ which are essentially bounded and measurable.

(i) The transition or Markov kernel is defined by $\mathcal{P}_t(x, A) := \mathbb{P}_x(X_t \in A)$ for every $t \geq 0$, $A \in \mathscr{B}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$. For $t \geq 0$ fixed this extends to an operator $\mathcal{P}_t : B(\mathbb{R}^d; \mathbb{R}) \to B(\mathbb{R}^d; \mathbb{R})$ with operator norm one by defining $\mathcal{P}_t f(x) := \mathbb{E}_x[f(X_t)]$.

(ii) The Markov kernel acts also on probability measures μ on $\mathbb{R}^d$ according to

$$\mathcal{P}_t\mu(A) = \int_{\mathbb{R}^d} \mathcal{P}_t(x, A)\, d\mu(x)$$

for every $A \in \mathscr{B}(\mathbb{R}^d)$ and $t \geq 0$. Hence, $\mathcal{P}_t\mu$ defines a probability measure on $\mathbb{R}^d$.

(iii) The generator of the Markov process is an operator $\mathcal{A}$ acting on a suitable subspace of $B(\mathbb{R}^d; \mathbb{R})$ such that

$$\lim_{t\to 0} \frac{1}{t}(\mathbb{E}_x[f(X_t)] - f(x)) = \mathcal{A}f(x), \quad \forall x \in \mathbb{R}^d.$$

(iv) The process is said to have continuous paths if $t \mapsto X_t(\omega)$ is continuous for all $\omega \in \Omega$.

Remark A.7. Property (iii) in Definition A.5 for the deterministic stopping time $S \equiv s$ infers $\mathcal{P}_s(\mathcal{P}_t f) = \mathcal{P}_{s+t} f$ for every $f \in B(\mathbb{R}^d; \mathbb{R})$ and $t, s \geq 0$. In the same way one has $\mathcal{P}_s(\mathcal{P}_t\mu) = \mathcal{P}_{s+t}\mu$ for every probability measure μ on $\mathbb{R}^d$. Observe also that the probability measure $\mathcal{P}_t\mu$ is just the law of X_t on $(\Omega, \mathscr{F}, \mathbb{P}_\mu)$.

Definition A.8 (Brownian motion)**.** An adapted process $(W_t, \mathscr{F}_t, t \geq 0)$ with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a d-dimensional standard Brownian motion if

(i) $\mathbb{P}(W_0 = 0) = 1$,

(ii) $W_t - W_s$ is independent from $\mathscr{F}_s$ for all $t \geq s \geq 0$,

(iii) $W_t - W_s \sim \mathcal{N}(0, (t-s)I)$, i.e. centered Gaussian with covariance $(t-s)I$,

(iv) $(W_t)_{t\geq 0}$ has (a.s.) continuous sample paths.

If we define $W^x = (W_t + x, \mathscr{F}_t, t \geq 0)$ then this process is a Brownian motion starting at $x \in \mathbb{R}^d$. Brownian motion is an example of a strong Markov process and together with the previous family of processes $(W^x)_{x\in\mathbb{R}^d}$ it constitutes a strong Markov family ([KS00, Section 2.6, p. 79-89]).

Remark A.9. Most of the time when working with a stochastic process we will ensure that the filtration satisfies the *usual conditions*, i.e. F_0 contains all $\mathbb{P}$-null sets and $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$ (right-continuity). In the case of Brownian motion this can be done with the augmented filtration (see [KS00, Section 2.7, p. 89-94]). The augmentation of some filtration $(\mathscr{G}_t)$ is defined by $\mathscr{F}_t := \sigma(\mathscr{G}_t \cup \mathscr{N})$, where $\mathscr{N}$ is the set of all $\mathbb{P}$-null sets.

The following definition is fundamental for our study in Chapter 4.

Definition A.10. A probability measure μ on $\mathbb{R}^d$ is called an *invariant measure* for the Markov family $(X_t, \mathscr{F}_t; t \geq 0)$ if $\mathcal{P}_t\mu = \mu$ for all $t \geq 0$.

The remainder of this section is devoted to stochastic differential equations of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{A.1}$$

with Borel measurable functions $b : \mathbb{R} \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times r}$ for fixed integers $d, r \in \mathbb{N}$. The process $(X_t)_{t\geq 0}$ takes values in $\mathbb{R}^d$ and $(W_t)_{t\geq 0}$ is some r-dimensional Brownian motion. For convenience we define $|\sigma(x)|^2 := \sum_{i,j} |\sigma_{ij}(x)|^2$. We refer to [KS00, Section 5.1, p. 284-291] for the proofs.

Definition A.11 (strong solution)**.** Consider a given probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a Brownian motion W as well as an initial random vector ξ. Define $\mathscr{F}_t$ to be the augmentation of the filtration $\mathscr{G}_t := \sigma(\xi, W_s : 0 \leq s \leq t)$. A continuous process $(X_t, t \geq 0)$ on this space is called a strong solution if

(i) $(X_t)_{t\geq 0}$ is adapted to $(\mathscr{F}_t)$,

(ii) $\mathbb{P}(X_0 = \xi) = 1$,

(iii) $\mathbb{P}\left[\int_0^t (|b_i(X_s)|^2 + |\sigma_{ij}(X_s)|^2)\, ds < \infty\right] = 1$ for all $1 \leq i \leq d$, $1 \leq j \leq r$,

(iv) (A.1) is satisfied for all $0 \leq t < \infty$, i.e.

$$X_t = X_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s \qquad \mathbb{P}\text{-a.s.}$$

Note that since a strong solution $(X_t)_{t\geq 0}$ is adapted and continuous, it is also progressively measurable ([KS00, Prop. 1.13, p. 5]).

Theorem A.12 (strong uniqueness)**.** *If b, σ are locally Lipschitz-continuous, then there is at most one (up to indistinguishability) strong solution to* (A.1)*.*

Theorem A.13 (strong existence)**.** *Assume that b, σ are globally Lipschitz-continuous, i.e.*

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq K(|x - y|)$$

for some constant K. Then for every probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a Brownian motion $(W_t)_{t\geq 0}$ and an initial condition ξ there exists a strong solution $(X_t)_{t\geq 0}$ to (A.1)*.*

Let $(Y_t^x)_{t\geq 0}$ be the strong solution with initial state $x \in \mathbb{R}^d$. Since $(Y_t^x)_{t\geq 0}$ has (a.s.) continuous paths we can consider the canonical version of this process. We define $\Omega := C[0,\infty)^d$ with its Borel σ-algebra $\mathscr{F} := \mathscr{B}(C[0,\infty)^d)$ (the space is endowed with the topology of uniform convergence on compact sets) together with the family of probability measures $\mathbb{P}_x := \mathbb{P}(Y^x)^{-1}$ on Ω. Furthermore, we define $X_t(\omega) := \omega(t)$ for every $\omega \in \Omega$, $t \geq 0$ and the natural filtration $\mathscr{F}_t := \sigma(X_s : 0 \leq s \leq t)$. One can prove the following ([KS00, Theorem 5.4.20 and Remark 4.21, p. 322]).

Theorem A.14. *Suppose b, σ are bounded on compact sets (and independent of t) and there is a unique strong solution $(Y_t^x)_{t\geq 0}$ for every initial state $x \in \mathbb{R}^d$. Consider the canonical version $(X_t)_{t\geq 0}$ on $(C[0,\infty)^d, \mathscr{B}(C[0,\infty)^d))$ with the family $\mathbb{P}_x$. Then, $(X_t)_{t\geq 0}$ defines a (time-homogeneous) strong Markov family.*

A.3. Postponed proofs from Chapter 2

A.3.1. Proof of Proposition 2.2 and Theorem 2.4

We consider

$$\partial_t g = e^V \operatorname{div}(De^{-V}\nabla g), \quad g(0,x) = f_0(x)e^{V(x)},$$

where $g = f/f_\infty$ and f satisfies the Fokker-Planck equation (2.1). We recall $\mu := f_\infty \lambda$ and $L\varphi := e^V \operatorname{div}(De^{-V}\nabla\varphi)$ on $C_c^\infty(\mathbb{R}^d)$. The operator L is symmetric on $L^2(\mu)$. The following proof, in which we define a self-adjoint extension, is based on solving elliptic partial differential equations in the weak sense.

Therefore, we define the weighted Sobolev space (see [Tur00, Section 2.1, p. 16])

$$H^1(\mathbb{R}^d,\mu) := H^1(\mu) := \left\{ f \in L^2(\mu) \ : \ \partial_{x_i} f \in L^2(\mu),\, i = 1,\dots,d \right\},$$

where derivatives are taken in the sense of distribution. A norm is defined by

$$\|u\|_{H^1(\mu)}^2 := \|u\|_{L^2(\mu)}^2 + \sum_{i=0}^{d} \|\partial_{x_i} u\|_{L^2(\mu)}^2$$

such that $H^1(\mu)$ is a Hilbert space (like the familiar Sobolev space $H^1(\mathbb{R}^d,\lambda)$). The scalar product in $L^2(\mu)$ will be denoted by $\langle\cdot,\cdot\rangle$. We also note the following

Lemma A.15. *Under the considered Assumption 2.1 $C_c^\infty(\mathbb{R}^d)$ is dense in $L^2(\mu)$, $H^1(\mu)$.*

Proof. First observe that $H^1(\mu) \subset L^1_{loc}(\mathbb{R}^d)$, since e^{-V} is continuous and positive. Therefore, the product rule $\partial_{x_i}(u\varphi) = \partial_{x_i} u\varphi + u\partial_{x_i}\varphi$ holds for $u \in H^1(\mu)$, $\varphi \in C_c^\infty(\mathbb{R}^d)$. Consider $\eta * \mathbb{1}_{B_k(0)}$ with a ball $B_k(0)$ centred at the origin with radius k and a mollifier η, such that $\eta * \mathbb{1}_{B_k(0)} = 1$ on $B_{k-1}(0)$ and $= 0$ on $B_{k+1}(0)^c$. Then

$$\left\|\partial_{x_i} u - \partial_{x_i}[u(\eta * \mathbb{1}_{B_k(0)})]\right\|_{L^2(\mu)} \leq \left\|\partial_{x_i} u - \partial_{x_i} u(\eta * \mathbb{1}_{B_k(0)})\right\|_{L^2(\mu)} + \left\|u\partial_{x_i}(\eta * \mathbb{1}_{B_k(0)})\right\|_{L^2(\mu)}$$

and note that $\partial_{x_i}(\eta * \mathbb{1}_{B_k(0)})$ is supported in $B_{k+1}(0)\setminus B_{k-1}(0)$. Now, both terms converge to zero by dominated convergence. Since $u(\eta * \mathbb{1}_{B_k(0)})$ has compact support and approximates

u in $H^1(\mu)$, it suffices to consider $u \in H^1(\mu)$ with compact support $K \subset \mathbb{R}^d$ in order to prove that $C_c^\infty(\mathbb{R}^d)$ is dense in $H^1(\mu)$. Observe that $u \in H^1(\mathbb{R}^d)$. By choosing $u_k = \eta_k * u \in C_c^\infty(\mathbb{R}^d)$ with a sequence of mollifiers (η_k) such that $\|\partial_{x_i} u - \eta_k * \partial_{x_i} u\|_{L^2} \to 0$ we obtain

$$\|\partial_{x_i} u - \partial_{x_i} u_k\|_{L^2(\mu)} = \|\partial_{x_i} u - \eta_k * \partial_{x_i} u\|_{L^2(\mu)} \le \sup_{x\in K} e^{-V(x)} \|\partial_{x_i} u - \eta_k * \partial_{x_i} u\|_{L^2} \to 0$$

as $k \to \infty$. This implies $\|u - u_k\|_{H^1(\mu)} \to 0$ for $k \to \infty$, since also $\|u - \eta_k * u\|_{L^2(\mu)} \to 0$. □

Proof of Proposition 2.2. Self-adjoint extension: We already know that $\langle L\varphi, \psi\rangle_{L^2(\mu)} = \langle \varphi, L\psi\rangle_{L^2(\mu)}$ as well as $\langle L\varphi, \varphi\rangle_{L^2(\mu)} \le 0$ for $\varphi, \psi \in C_c^\infty(\mathbb{R}^d)$. It suffices to prove the existence of a self-adjoint extension of $A := I - L$. Consider the (stationary) partial differential equation $Au = f$ and its weak formulation: find $u \in H^1(\mu)$ such that for all $v \in H^1(\mu)$

$$a(u,v) := \int \nabla v^\top D \nabla u \, d\mu + \int uv \, d\mu = \int fv \, d\mu =: F(v)$$

with $f \in L^2(\mu)$. The bilinear form $a(\cdot,\cdot)$ on $H^1(\mu)$ is continuous and coercive, i.e. $|a(u,v)| \le \|u\|_{H^1(\mu)} \|v\|_{H^1(\mu)}$, $a(u,u) \ge \min(\alpha,1)\|u\|^2_{H^1(\mu)}$. Since F is continuous on $H^1(\mu)$, the Lax-Milgram lemma implies that there exists a unique solution $u \in H^1(\mu)$. Accordingly, we can define an operator $S : L^2(\mu) \to L^2(\mu)$, where $S(f)$ is the weak solution to $Au = f$. Furthermore, we have $\|S\| \le \min(\alpha,1)^{-1}$. We conclude from uniqueness $\ker S = \{0\}$.

Now, we define $A_{ex} : \operatorname{ran} S \to L^2(\mu) : u \mapsto S^{-1}u$, which is clearly an extension of A by uniqueness. Certainly, $C_c^\infty(\mathbb{R}^d) \subset \operatorname{dom} A_{ex} := \operatorname{ran} S \subset H^1(\mu)$ dense in $L^2(\mu)$. One can see that S is symmetric, hence self-adjoint. Thus A_{ex} is self-adjoint too. Indeed, we have $\langle A_{ex}x, y\rangle = \langle A_{ex}x, SA_{ex}y\rangle = \langle SA_{ex}x, A_{ex}y\rangle = \langle x, A_{ex}y\rangle$ for $x, y \in \operatorname{dom} A_{ex}$, so A_{ex} is symmetric. Take $x \in \operatorname{dom} A_{ex}^*$ arbitrary. By definition there is $x^* \in L^2(\mu)$ with $\langle A_{ex}y, x\rangle = \langle y, x^*\rangle$ for all $y \in \operatorname{dom} A_{ex}$. Consequently, we have $\langle A_{ex}Sz, x\rangle = \langle Sz, x^*\rangle$ for all $z \in L^2(\mu)$, which implies $\langle z, x\rangle = \langle z, Sx^*\rangle$. Thus $Sx^* = x \in \operatorname{ran} S = \operatorname{dom} A_{ex}$ and A_{ex} is self-adjoint. We can define a self-adjoint extension of L by $L_{ex} := I - A_{ex}$.

Closure of L and uniqueness: In order to prove that L_{ex} is the closure of L it suffices to prove the same relation for A_{ex} and A from above. At first we will show that $\operatorname{ran} A$ is dense in $L^2(\mu)$, which follows if we can prove $H^1(\mu) \subset \overline{\operatorname{ran} A}$, because $H^1(\mu)$ is dense in $L^2(\mu)$. Suppose that there is some $0 \ne x \in H^1(\mu)$ with $x \notin \overline{\operatorname{ran} A}$, then $\overline{\operatorname{ran} A} \subset (\operatorname{span}\{x\})^\perp$. But this infers $a(u,x) = \langle Au, x\rangle = 0$ for all $u \in \operatorname{dom} A$. Since $\operatorname{dom} A = C_c^\infty(\mathbb{R}^d)$ is dense in $H^1(\mu)$ and a is a continuous bilinear form on $H^1(\mu)$ we obtain $a(u,x) = 0$ for all $u \in H^1(\mu)$. By uniqueness of the weak formulation it follows $x = 0$, a contradiction. Therefore, we establish $\overline{\operatorname{ran} A} = L^2(\mu)$.

Now, observe that A_{ex} contains the closure of A, because it is closed as a consequence of the self-adjointness. For the other inclusion take $u \in \operatorname{dom} A_{ex}$ and $f := A_{ex}u$. Since $\operatorname{ran} A$ is dense we can find a sequence $Au_n \to f$ in $L^2(\mu)$. Hence, we obtain $u_n = SAu_n \to Sf = u$ and therefore, A_{ex} is contained in the closure of A.

For uniqueness let $\widetilde{L}$ be another self-adjoint extension of L. The operator $\widetilde{L}$ is closed and hence $\widetilde{L}$ extends also L_{ex}. But then it follows that L_{ex}^* extends $\widetilde{L}^*$ which yields $L_{ex} = \widetilde{L}$.

Spectrum of L_{ex}: Take $\lambda > 0$ arbitrary and observe $\lambda I - L_{ex} = A_{ex} + (\lambda - 1)I$. We can conclude like above $\langle A_{ex}u + (\lambda-1)u, u\rangle \geq \min(\alpha,\lambda)\,\|u\|^2_{H^1(\mu)}$. Again with the Lax-Milgram lemma there is a bounded operator S_λ, $\|S_\lambda\| \leq \min(\alpha,\lambda)^{-1}$, such that $S_\lambda(f)$ is the unique weak solution of $A_{ex}u + (\lambda-1)u = f$. We want to prove that S_λ is the inverse operator of $A_{ex} + (\lambda-1)I$.

Therefore, set $f := A_{ex}u + (\lambda-1)u$ for an arbitrary $u \in \operatorname{dom} A_{ex}$. Since $\langle A_{ex}u, v\rangle = a(u,v)$, u is also a weak solution and thus $u = S_\lambda(A_{ex} + (\lambda-1)I)u$, i.e. $S_\lambda(A_{ex} + (\lambda-1)I) = I$.

In order to prove $(A_{ex} + (\lambda-1)I)S_\lambda = I$, we first observe $\operatorname{dom} A_{ex} \subset \operatorname{ran} S_\lambda$ and by uniqueness $\ker S_\lambda = \{0\}$. Now, if $u \in \operatorname{ran} S_\lambda$, then there is $f \in L^2(\mu)$ with $S_\lambda f = u$ such that

$$\langle f, v\rangle = a(u,v) + (\lambda-1)\langle u, v\rangle$$

for all $v \in H^1(\mu)$. Uniqueness of the weak solution implies $S(f - (\lambda-1)u) = u \in \operatorname{dom} A_{ex}$. Since $S_\lambda f = u$ this implies, $f = (A_{ex} + (\lambda-1)I)S_\lambda f$ and thus $A_{ex}u + (\lambda-1)I$ is continuously invertible.

This proves $\lambda \in \rho(L_{ex})$. In addition, we know that $\mathbb{C}\backslash\mathbb{R} \subset \rho(L_{ex})$, because L_{ex} is self-adjoint. Certainly, $0 \in \sigma(L_{ex})$ and if $u \in \ker L_{ex}$, then $0 = \langle -L_{ex}u, u\rangle = \langle A_{ex}u - u, u\rangle \geq \alpha\,\|\nabla u\|^2_{L^2(\mu)}$. Consequently, u is constant.

□

For the proof of Theorem 2.4 we will need the following regularity result.

Lemma A.16. *Suppose Assumption 2.1 is satisfied and consider the weak solution $u \in H^1(\mu)$ of $Lu = f$ with $f \in H^1(\mu)$. If $f \in H^k_{loc}(\mathbb{R}^d)$ for some fixed $k \in \mathbb{N}$ then $u \in H^{k+2}_{loc}(\mathbb{R}^d)$.*

Proof. Fix some $\Omega \subset \mathbb{R}^d$ open, bounded. We choose $\varphi \in C_c^\infty(\Omega)$ as a test function in the weak formulation and obtain

$$\int_\Omega \nabla u^\top D \nabla\varphi\, d\mu = \int_\Omega f\varphi\, d\mu,$$

which is then also true for every $\varphi \in H^1_0(\Omega)$. Thus, the restriction of u to Ω, $u \in H^1(\Omega)$, is a weak solution of the equation $-\operatorname{div}(e^{-V} D\nabla u) = fe^{-V}$ on Ω. Note that the matrix-valued function $e^{-V}D$ is uniformly elliptic on $\overline{\Omega}$ and $fe^{-V} \in H^k(\Omega)$ by smoothness of e^{-V}. Now, the familiar interior regularity results (see [Eva10, Section 6.3, Theorem 2, p. 314]) imply $u \in H^{k+2}_{loc}(\Omega)$. □

In order to prove Theorem 2.4 we follow [Bau14, p. 103-104, 106-107]. We will use the spectral theorem as well as properties of a spectral measure stated in Section A.1.

Proof of Theorem 2.4. *Uniqueness:* We obtain for two solutions u, v by differentiating the norm

$$\frac{d}{dt}\|u - v\|^2_{L^2(\mu)} = 2\langle L(u-v), u - v\rangle \leq 0.$$

Together with the initial condition $(u-v)(0) = 0$ we conclude $\|(u-v)(t)\|_{L^2(\mu)} = 0$ for all $t \geq 0$.

Semigroup: By setting $e^{Lt} := \int_{\sigma(L)} e^{\lambda t}\, dE(\lambda)$ we define a family of bounded operators with

$$e^{Lt}e^{Ls} = \int_{\sigma(L)} e^{\lambda t}e^{\lambda s}\, dE(\lambda) = e^{L(t+s)}, \quad \|e^{Lt}\| = \sup_{\lambda\in\sigma(L)} |e^{\lambda t}| \leq 1$$

for all $t, s \geq 0$. Since the continuous functions $e^{\lambda t}$ are uniformly bounded on the spectrum of L, we obtain strongly continuity of e^{Lt}.

Regularity in time: After fixing $t > 0$ we obtain from Taylor's formula

$$\lambda^k \left(\frac{e^{\lambda(t+h)} - e^{\lambda t}}{h} \right) = \lambda^{k+1}e^{\lambda t} + \lambda^{k+2}he^{\lambda\vartheta(h)}$$

with $\vartheta(h) \in (t - |h|, t + |h|)$ and $k \in \mathbb{N}_0$. Now, the right side is uniformly bounded in $\lambda \in \sigma(L)$ and h, whenever $|h| < t$, respectively $\vartheta(h) > 0$. Furthermore, for $h \to 0$ it converges pointwise to $\lambda^{k+1}e^{\lambda t}$. We conclude

$$\lim_{h\to 0} \left\| \int_{\sigma(L)} \lambda^k \frac{e^{\lambda(t+h)} - e^{\lambda t}}{h}\, dE(\lambda)g_0 - \int_{\sigma(L)} \lambda^{k+1}e^{\lambda t}\, dE(\lambda)g_0 \right\|_{L^2(\mu)} = 0$$

for all $k \in \mathbb{N}_0$ and $g_0 \in L^2(\mu)$. This is

$$\frac{d^k}{dt^k}e^{Lt}g_0 = \int_{\sigma(L)} \lambda^k e^{\lambda t}\, dE(\lambda)g_0.$$

Regularity in space: At first we have to prove $e^{Lt}g_0 \in \operatorname{dom} L^k$ for $g_0 \in L^2(\mu)$, $t > 0$ and recall $L^k = \int_{\sigma(L)} \lambda^k\, dE(\lambda)$. This statement is equivalent to

$$\int_{\sigma(L)} |\lambda|^{2k}\, dE_{e^{Lt}g,e^{Lt}g} = \int_{\sigma(L)} |\lambda|^{2k}e^{2\lambda t}\, dE_{g,g} < \infty,$$

which is satisfied for $t > 0$. In addition,

$$Le^{Lt}g_0 = \int_{\sigma(L)} \lambda e^{\lambda t}\, dE(\lambda)g_0 = \frac{d}{dt}e^{Lt}g_0,$$

thus $e^{Lt}g_0$ is a solution of (2.3).

Furthermore, for every $k \in \mathbb{N}$, $k \geq 1$ we have $L^k e^{Lt}g_0 \in \operatorname{dom} L \subset H^1(\mu)$, hence $L^k e^{Lt}g_0 \in\in H^1(\Omega)$ for every $\Omega \subset \mathbb{R}^d$ open and bounded. Lemma A.16 yields $L^{k-1}e^{Lt}g_0 \in H^{1+2}(\Omega)$ and inductively $e^{Lt}g_0 \in H^{2k+1}(\Omega)$ for every $\Omega \subset \mathbb{R}^d$ open, bounded and every $k \in \mathbb{N}$. We conclude $e^{Lt}g_0 \in C^\infty(\mathbb{R}^d)$ by Sobolev's embedding theorem ([Eva10, Theorem 6, Subsection 5.6.3, p. 284]).

Joint regularity: We sketch this part and consider only joint continuity (see also [Bau14, Prop. 4.20, p. 106]). Fix $t_0 > 0$. By Sobolev's inequality in Sobolev's embedding theorem we can choose an integer $n \in \mathbb{N}$ such that for every compact set $K \subset \mathbb{R}^d$ there is a constant $C \geq 0$ with

$$\sup_{x\in K}\left|e^{-V(x)/2}\left(e^{Lt}g_0(x)-e^{Lt_0}g_0(x)\right)\right| \leq C\left\|e^{-V(x)/2}\left(e^{Lt}g_0(x)-e^{Lt_0}g_0(x)\right)\right\|_{H^n(K)}.$$

Since $e^{-V/2}$ and all its derivatives up to order n are bounded on K and $e^{-V/2}$ is also bounded away from zero on K, we obtain by redefining the constant C

$$\sup_{x\in K}\left|e^{Lt}g_0(x)-e^{Lt_0}g_0(x)\right| \leq C\sum_{k=0}^{n}\left\|L^k e^{Lt}g_0 - L^k e^{Lt_0}g_0\right\|_{L^2(\mu)}.$$

Now, $\lambda^k e^{\lambda t} \to \lambda^k e^{\lambda t_0}$ as $t \to t_0$ and $\lambda^k e^{\lambda t}$ is uniformly bounded in λ, t, where $\lambda \in \sigma(L)$. Therefore, we obtain for $t \to t_0$

$$\left\|L^k e^{Lt}g_0 - L^k e^{Lt_0}g_0\right\|_{L^2(\mu)} \to 0,$$

hence

$$\sup_{x\in K}\left|e^{Lt}g_0(x)-e^{Lt_0}g_0(x)\right| \to 0.$$

This implies the joint continuity. The joint continuity of the t, x-derivatives works similar.

Long-time behavior: Whenever $\lambda < 0$, $e^{\lambda t} \to 0$ for $t \to +\infty$ and $e^{\lambda t}$ is bounded uniformly in λ, t. Thus, we obtain strong convergence of e^{Lt}, i.e.

$$\lim_{t\to+\infty}\left\|e^{Lt}g_0 - \int g_0\,d\mu\right\|_{L^2(\mu)} = \lim_{t\to+\infty}\left\|\int_{(-\infty,0)} e^{\lambda t}\,dE(\lambda)g_0\right\|_{L^2(\mu)} = 0.$$

This concludes the proof.

□

A.3.2. Proof of Proposition 2.7 and Lemma 2.11

The following proof gives the rigorous justifications in order to obtain Proposition 2.7. Recall the idea of gluing together local solution of the stochastic differential equation (2.2) discussed at the beginning of Section 2.2. The filtration $(\mathscr{F}_t)$ in the statement is constructed as follows: If $(W_t, \mathscr{G}_t, t \geq 0)$ is a Brownian motion on $(\Omega, \mathscr{F}, \mathbb{P})$ and ξ is a random variable on it, then $\widetilde{\mathscr{F}_t} := \sigma(G_s \cup \xi \;\; s \leq t)$. Finally, $\mathscr{F}_t$ is defined to be the augmentation of $\widetilde{\mathscr{F}_t}$ like in Remark A.9 and therefore satisfies the usual conditions.

Proof of Proposition 2.7. Existence: For $n \in \mathbb{N}$ define b_n and σ_n such that

$$b_n(x) = b(x), \quad \sigma_n(x) = \sigma(x), \quad \text{for } x \in B_n(0)$$

and b_n, σ_n fulfill the conditions in Theorem A.13, i.e. are globally Lipschitz-continuous and grow at most linearly. By virtue of Theorem A.13 there are continuous processes Y^n satisfying

$$Y_t^n = \xi + \int_0^t b_n(Y_s^n)\, ds + \int_0^t \sigma_n(Y_s^n)\, dW_s \tag{A.2}$$

as well as the other assertions in Definition A.11. For the sake of notation, we will assume that equation (A.2) is satisfied for all $\omega \in \Omega$. Set $\tau_n := \inf\{t > 0 \ : \ Y^n \in B_n(0)^c\}$, which are $(\mathscr{F}_t)$ stopping times. We show that $Y_t^n = Y_t^m$ on $\{t \leq \tau_n\}$ for $n < m$.

First of all we have $b_n(x) = b_m(x), \sigma_n(x) = \sigma_m(x)$ for $x \in B_n(0)$ and therefore on $\{t \leq \tau_n\}$

$$Y_t^m = \xi + \int_0^t b_n(Y_s^m)\, ds + \int_0^t \sigma_n(Y_s^m)\, dW_s.$$

Strong uniqueness (Theorem A.12) implies $Y_t^m = Y_t^n$ on $\{t \leq \tau_n\}$ (despite of a null set independent of t). Furthermore, we observe $\tau_n \leq \tau_m$ and define the stopping time $\tau := \sup_n \tau_n$ as well as $\widetilde{\Omega} := \bigcap_{1 \leq n < m} \{Y_t^m = Y_t^n, \forall t \leq \tau_n\}$. For every $\omega \in \widetilde{\Omega}$ and $t < \tau(\omega)$ we define a continuous process by $X_t(\omega) := Y_t^n(\omega)$ with n such that $t \leq \tau_n(\omega)$. This definition does not depend on n we pick to ensure $t \leq \tau_n(\omega)$.

We obtain for any $\omega \in \widetilde{\Omega}$, $t < \tau(\omega)$ and n with $t \leq \tau_n(\omega)$

$$\begin{aligned} X_t(\omega) &= \left(\xi + \int_0^t b(Y_s^n)\, ds + \int_0^t \sigma(Y_s^n)\, dW_s \right)(\omega) \\ &= \left(\xi + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s \right)(\omega). \end{aligned}$$

That is on $\{t < \tau\}$

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

Furthermore, on $\{\tau > 0\}$ we have $X_0 = \xi$ and on $\{t < \tau\}$

$$\int_0^t (|b_i(X_s)|^2 + |\sigma_{ij}(X_s)|^2)\, ds < \infty,$$

since this is satisfied for all Y^n. Because $X_t = Y_t^n$ on $\{t \leq \tau_n\} \cap \widetilde{\Omega}$ and $\widetilde{\Omega}^c \in \mathscr{F}_0$ by the usual conditions, we conclude that $X_{t \wedge \tau_n}$ is adapted for all n. But then so is $(X_t, \mathscr{F}_t, t < \tau)$.

Uniqueness: Let $(Z_t, \mathscr{F}_t, t < \sigma)$ together with a stopping time σ be another solution. By strong uniqueness we observe $Y_t^n = Z_t$ on $\{t < \sigma \wedge \tau_n\}$ for all $n \in \mathbb{N}$. But this is $X_t = Z_t$ on $\{t < \sigma \wedge \tau\}$. Assume for a moment that $\sigma > \tau$. Take ω such that $(X_t)_{t \geq 0}, (Z_t)_{t \geq 0}$ are equal for all $t < \sigma(\omega) \wedge \tau(\omega) = \tau(\omega)$ and $\tau(\omega) < \infty$. Moreover, define the sequence $t_n := \tau_n(\omega)$ and observe $t_n < \tau(\omega)$, $t_n \to \tau(\omega)$. By definition of τ_n we have $|Y_{t_n}(\omega)| = |X_{t_n}(\omega)| \geq n$. Since Y is continuous, we obtain $|Y_{\tau(\omega)}(\omega)| = \infty$. Therefore, $\sigma \leq \tau$ must hold.

This concludes the proof.

□

At the end of this section we give all calculations needed to prove Lemma 2.11 based on Proposition 2.9 which yields the existence of a global solution to the stochastic differential equation (2.2).

Proof of Lemma 2.11. First we will consider $k = 1$. We define the function $G(x) := |x|^2 + C$ for some constant $C \geq 0$ to be determined later. We have for $|x| \geq M$

$$\begin{aligned}\mathcal{L}G &= \sum_{i,j} 2D_{ij} + 2\langle \text{div}D - D\nabla V(x), x\rangle \leq 2d^2 \|D\|_\infty + 2|\text{div}D||x| - \langle D\nabla V(x), x\rangle \\ &\leq 2d^2 \|D\|_\infty + d\|\text{div}D\|_\infty + |x|^2\end{aligned}$$

and for $|x| \leq M$

$$\mathcal{L}G \leq 2d^2 \|D\|_\infty + d\|\text{div}D\|_\infty + 2\|D\|_\infty \sup_{|x|\leq M} |\nabla V(x)|M + |x|^2.$$

If we define C accordingly, we obtain $\mathcal{L}G \leq G$ and Proposition 2.9 applies.

For $k > 1$ it works similarly. We define $G(x) := |x|^{2k} + C$, then we have

$$\nabla G(x) = 2kx|x|^{2k-2}, \quad \text{Hess } G(x) = 2kI|x|^{2k-2} + 2k(2k-2)xx^\top |x|^{2k-4}$$

and thus

$$\sum_{i,j} D_{ij} \left(2k|x|^{2k-2} + 2k(2k-2)x_i x_j |x|^{2k-4}\right) \leq c_k \|D\|_\infty |x|^{2k-2} \leq c_k \|D\|_\infty (1 + |x|^{2k})$$

for some constant c_k by using $x_i x_j \leq (x_i^2 + x_j^2)/2$. In the same manner

$$2k\langle \text{div}D, x\rangle |x|^{2k-2} \leq 2k\|\text{div}D\|_\infty |x|^{2k-1} \leq 2k\|\text{div}D\|_\infty (1 + |x|^{2k}).$$

Together with

$$2k\langle D\nabla V(x), x\rangle |x|^{2k-2} \leq 2k\|D\|_\infty \sup_{|x|\leq M} |V(x)|M^{2k-1} =: c_m$$

for $|x| \leq M$ we obtain

$$\mathcal{L}G(x) \leq (c_k\|D\|_\infty + 2k\|\text{div}D\|_\infty)(1 + |x|^{2k}) + c_m \leq c_k G(x)$$

by redefining c_k and defining C properly. Again Proposition 2.9 applies. □

A.4. Postponed proofs from Chapter 3

In the following we will give a proof of Lemma 3.7 and Proposition 3.8.

For the first one, recall that ρ is solution of (2.1) with initial data $\rho_0 \in L^2(\rho_\infty^{-1})$, where Assumption 2.1 holds. Recall $\mu = \rho_\infty \lambda$, $g = \rho/\rho_\infty$ from Section 2.1 and Theorem 2.4 for properties of the semigroup.

Proof of Lemma 3.7. In order to prove $e_\psi(\rho_t \mid \rho_\infty) < \infty$, we first observe by Lemma 3.3 (iii) for $\sigma_0 = 1$ that $\psi(\sigma) \le \psi(1)(\sigma-1)^2$ for $\sigma \ge 0$. Therefore, we infer

$$e_\psi(\rho_t \mid \rho_\infty) = \int_{\mathbb{R}^d} \psi(g_t) d\mu \le \int_{\mathbb{R}^d} \psi(1)(g_t - 1)^2 \, d\mu = \psi(1) \, \|g_t - 1\|^2_{L^2(\mu)} < \infty.$$

Since we deal with normalized generating functions, we have $\psi \ge 0$ and hence the claim.

Recall that ψ' and ψ'' can have a singularity at $\sigma = 0$. Therefore, we consider the approximation for $\varepsilon < 1$

$$\psi''_\varepsilon(\sigma) = \begin{cases} \psi''(\sigma), & \sigma \ge \varepsilon, \\ \psi''(\varepsilon) & \sigma < \varepsilon. \end{cases}$$

Note that ψ'' is decreasing (see Lemma 3.3 (ii)) and hence $\psi''_\varepsilon \le \psi''$. Furthermore, $\psi''_\varepsilon \nearrow \psi''$ as $\varepsilon \to 0$. We define for $\varepsilon < 1$

$$\psi_\varepsilon(\sigma) = \int_1^\sigma \left(\int_1^\tau \psi''_\varepsilon(\eta) \, d\eta \right) \, d\tau.$$

Note that $\psi_\varepsilon(\sigma) = \psi(\sigma)$ for $\sigma \ge \varepsilon$, since $\varepsilon < 1$, $\psi(1) = \psi'(1) = 0$ (see Definition 3.1). We infer $\psi_\varepsilon \le \psi$ and $\psi_\varepsilon \nearrow \psi$ as $\varepsilon \to 0$. In addition, ψ'_ε, ψ_ε do not have a singularity at $\sigma = 0$.

Furthermore, we also need the following estimate $\psi'_\varepsilon(\sigma) = \psi'(\sigma) \le \psi''(1)\sigma$ for $\sigma \ge 1$. This follows, since ψ'' is decreasing and hence with $\psi'(1) = 0$

$$\psi'(\sigma) = \int_1^\sigma \psi''(\tau) \, d\tau \le \sigma \psi''(1).$$

Recall that ψ' is an increasing function. This yields the estimate $c(\varepsilon) \le \psi'_\varepsilon(\sigma) \le C(\varepsilon)\sigma$ for all $\sigma \ge 0$ for constants $c(\varepsilon), C(\varepsilon)$ depending on ε. Note that we also have the bounds $0 \le \psi''_\varepsilon \le C'(\varepsilon)$ for all $\sigma \ge 0$ for some constant $C'(\varepsilon)$.

Now, since $\psi(g_t)$ is smooth for $t > 0$ we obtain for $0 < s \le t$

$$\psi_\varepsilon(g_t) - \psi_\varepsilon(g_s) = \int_s^t \psi'_\varepsilon(g_r) \partial_t g_r \, dr$$

and integrating with respect to μ yields

$$e_{\psi_\varepsilon}(\rho_t \mid \rho_\infty) - e_{\psi_\varepsilon}(\rho_s \mid \rho_\infty) = \int_{\mathbb{R}^d} \int_s^t \psi'_\varepsilon(g_r) \partial_t g_r \, dr d\mu.$$

The right integral exists, since $\psi'_\varepsilon(g_r) \le C(\varepsilon) g_r$ and $g_r \partial_t g_r \in L^1(\mu)$. We obtain by a partial integration

$$\int_s^t \left\langle \psi'_\varepsilon(g_r), \partial_t g_r \right\rangle_{L^2(\mu)} \, dr = \int_s^t \left\langle \psi'_\varepsilon(g_r), L g_r \right\rangle_{L^2(\mu)} \, dr = -\int_s^t \int_{\mathbb{R}^d} \psi''_\varepsilon(g_r) \nabla g_r^\top D \nabla g_r \, d\mu dr,$$

which is justified, since $\psi'_\varepsilon(g_r) \in H^1(\mu)$, $0 \le \psi''_\varepsilon \le C'(\varepsilon)$. Finally, we let $\varepsilon \to 0$ and use monotone convergence on the right side of the previous equation and in $e_{\psi_\varepsilon}(\rho_t \mid \rho_\infty)$ yielding

$$e_\psi(\rho_t \mid \rho_\infty) - e_\psi(\rho_s \mid \rho_\infty) = -\int_s^t \int_{\mathbb{R}^d} \psi''(g_r) \nabla g_r^\top D \nabla g_r \, d\mu dr.$$

□

A.4.1. Proof of Proposition 3.8

The given proof is based on [OV00, Section 4, p. 18] (for Step 1-3 below) and [AMTU01, Lemma 2.13, p. 21] (for Step 4). Recall that we assume $D \equiv I$, Hess $V \geq \lambda I$ together with Assumption 2.1.

At first we reduce our investigation to “nice” functions and derive bounds on the solution $g = \rho/\rho_\infty$ of (2.3). Then we calculate the entropy dissipation and its dissipation. In the final step we use the condition Hess $V \geq \lambda I$ to derive the convex Sobolev inequality, which implies convergence in relative entropy.

Step 1. We will consider $g = \rho/\rho_\infty$ for a solution ρ of (2.1) with $\rho_0 \in L^2(\rho_\infty^{-1})$, $\rho \geq 0$ first. Recall that it satisfies (2.3), i.e.

$$\partial_t g = \Delta g - \nabla V \cdot \nabla g = \rho_\infty^{-1} \mathrm{div}(\rho_\infty \nabla g) = Lg \tag{A.3}$$

Now, we want reduce the problem to solutions g which are bounded away form zero and infinity and satisfy a gradient bound. Therefore, define

$$g_0^n(x) := \begin{cases} 1/n & g_0(x) < 1/n, \\ g_0(x) & 1/n \leq g_0(x) \leq n, \\ n & g_0(x) > n. \end{cases}$$

Obviously, $g_0^n \to g_0$ in $L^2(\rho_\infty)$. Define $g_0^{n,l}$ to be a smooth approximation of g_0^n in $L^2(\rho_\infty)$ such that $|\nabla g_0^{n,l}| \leq K(n,l)$ for some constant $K(n,l) > 0$ depending on n, l and $1/n \leq g_0^{n,l} \leq n$. This can be done with mollifies. Hence, $g_0^{n,n}$ is a smooth approximation of g_0 in $L^2(\rho_\infty)$ with

$$0 < c_n \leq g_0^{n,n} \leq C_n, \quad |\nabla g_0^{n,n}| \leq K_n.$$

Therefore, $e^{Lt} g_0^{n,n} \to e^{Lt} g_0$ in $L^2(\rho_\infty)$ for each $t \geq 0$ by the continuity of e^{Lt}. This implies $\rho_t^n \to \rho_t$ in $L^2(\rho_\infty^{-1})$ for the corresponding solutions of (2.1).

Suppose that we already proved

$$e_\psi(\rho_t^n \mid \rho_\infty) \leq e^{-2\lambda t} e_\psi(\rho_0^n \mid \rho_\infty).$$

Then Lemma 3.3 infers the same inequality for ρ_t. Furthermore, the convex Sobolev inequality (3.6) follows from the entropy decay for solutions ρ_t with $\rho_0 \in L^2(\rho_\infty^{-1})$. Indeed, we have the decay estimate for $0 \leq s \leq t$

$$e_\psi(\rho_t \mid \rho_\infty) \leq e^{-2\lambda(t-s)} e_\psi(\rho_s \mid \rho_\infty),$$

since we can consider ρ_t to be the solution after time $(t-s)$ with initial data ρ_s by the uniqueness of solutions (see Theorem 2.4). Hence, by Lemma 3.7 we have

$$\int_s^t \left(-\int_{\mathbb{R}^d} \psi''(g_r) \nabla g_r^\top D \nabla g_r \rho_\infty \, dx \right) dr \leq (e^{-2\lambda(t-s)} - 1) e_\psi(\rho_s \mid \rho_\infty)$$

Dividing by $(t-s)$ and sending $t \searrow s$ yields the inequality.

Step 2. Consider $g_0 \in L^2(\rho_\infty)$ smooth with

$$0 < c \leq g_0 \leq C, \quad |\nabla g_0| \leq M.$$

The condition Hess $V \geq \lambda I$ allows us to use Theorem 2.13. Take $x \in \mathbb{R}^d$, $s \in \mathbb{R}$ and observe (as a function of s using Taylor's theorem)

$$\begin{aligned} V(x - sx) &= V(x) - \langle \nabla V(x), x\rangle + \frac{1}{2}\langle x, \text{Hess } V(\theta)x\rangle \\ &\geq V(x) - \langle \nabla V(x), x\rangle + \frac{\lambda}{2}|x|^2. \end{aligned}$$

At last set $s = 0$ yielding $\langle \nabla V(x), x\rangle \geq \lambda |x|^2/2 \geq 0$.

By Theorem 2.13 (iii) the solution can be written as

$$g_t(x) = e^{V(x)} \int p(t, y, x)\rho_0(y)\, dy = e^{V(x)} \int p(t, y, x)e^{-V(y)} g_0(y)\, dy$$

Using the bounds on g_0 and the fact that e^{-V} is the density of the invariant measure with respect to the corresponding Markov process, i.e. $\int p(t, y, x)e^{-V(y)}\, dy = e^{-V(x)}$, we infer

$$0 < c \leq g_t \leq C.$$

In addition to the above bounds, the condition Hess $V \geq \lambda I$ was used in [OV00, Section 4, p. 18] to derive the gradient estimate $\sup |\nabla g_t| \leq e^{-2\lambda t} \sup |\nabla g_0| \leq M$ for all $t \geq 0$. (Such a gradient estimate was also proved in [BGL14, Subsection 3.2.3, p. 143-149] under the same condition).

Step 3. Now, we calculate the entropy dissipation and its dissipation for g_t from the previous step. At first, we have for fixed $t > 0$ and a function $\phi \in C_c^\infty(\mathbb{R}^d)$

$$\frac{d}{dt}\int_{\mathbb{R}^d} \psi(g_t)\phi\, d\mu = \int \psi'(g_t)Lg_t\phi\, d\mu = -\int \psi''(g_t)|\nabla g_t|^2\phi\, d\mu - \int \psi'(g_t)\nabla g_t \cdot \nabla\phi\, d\mu.$$

The interchange as well as the partial integration can be performed since ϕ has compact support. Choose ϕ^n such that $\phi^n \to 1$, $\nabla\phi^n \to 0$ and $|\nabla\phi^n|$ uniformly bounded. For instance, define $\phi^n = \mathbb{1}_{B_n(0)} * \eta$ with the ball $B_n(0)$ of radius n and a fixed mollifier η.

Since the integrands are bounded we obtain

$$\lim_n \frac{d}{dt}\int_{\mathbb{R}^d} \psi(g_t)\phi^n\, d\mu = -\int \psi''(g_t)|\nabla g_t|^2\, d\mu.$$

Observe that the above limit with respect to n is uniform with respect to t.

Thus, $t \mapsto \int \psi(g_t)\phi^n\, d\mu$ is differentiable for every t, n and its derivative converges uniformly in t. This implies

$$\frac{d}{dt}\int_{\mathbb{R}^d} \psi(g_t)\, d\mu = I_\psi(\rho_t \mid \rho_\infty) = -\int \psi''(g_t)|\nabla g_t|^2\, d\mu.$$

Now we turn to $\frac{d}{dt}I_\psi(\rho_t \mid \rho_\infty)$. We have

$$\frac{d}{dt} - \int \psi''(g_t)|\nabla g_t|^2\phi_n \, d\mu = -\int \psi'''(g_t)\partial_t g_t|\nabla g_t|^2\phi_n \, d\mu - \int \psi''(g_t)\nabla g_t \cdot \nabla\partial_t g_t\phi_n \, d\mu$$

In order to obtain the uniform convergence in t we use the bounds for the first respectively the second integral (K is some constant)

$$\int \psi'''(g_t)|\partial_t g_t||\nabla g_t|^2|\phi_n - 1| \, d\mu \leq K \left\|\partial_t g_t\right\|_{L^2(\mu)} \left\|\phi_n - 1\right\|_{L^2(\mu)},$$
$$\int \psi''(g_t)|\nabla g_t \cdot \nabla\partial_t g_t||\phi_n - 1| \, d\mu \leq K \left\|\nabla\partial_t g_t\right\|_{L^2(\mu)} \left\|\phi_n - 1\right\|_{L^2(\mu)}.$$

It follows from the definition of the semigroup e^{Lt} via the spectral decomposition that $\|\partial_t g_t\|_{L^2(\mu)}$ and $\|\nabla\partial_t g_t\|_{L^2(\mu)}$ are locally uniformly bounded in t. Again we have

$$\frac{d}{dt}I_\psi(\rho_t \mid \rho_\infty) = R_\psi(\rho_t \mid \rho_\infty) = -\int \psi'''(g_t)\partial_t g_t|\nabla g_t|^2 \, d\mu - \int \psi''(g_t)\nabla g_t \cdot \nabla\partial_t g_t \, d\mu. \quad \text{(A.4)}$$

Step 4. In this step, we turn to the exponential convergence of the relative entropy with rate 2λ. First, we prove such a decay for the entropy dissipation and therefore analyze $R_\psi(\rho_t \mid \rho_\infty)$. By virtue of (A.3) we start with partial integration in the first integral in (A.4),

$$\begin{aligned} R_1 &:= \int_{\mathbb{R}^d} \nabla\left(\psi'''(g_t)|\nabla g_t|^2\right) \cdot \nabla g_t \, d\mu \\ &= \int_{\mathbb{R}^d} \psi^{IV}(g_t)|\nabla g_t|^4 \, d\mu + 2\int_{\mathbb{R}^d} \psi'''(g_t)\nabla g_t^\top \left(\text{Hess } g_t\right) \nabla g_t \, d\mu. \end{aligned}$$

The surface integral vanishes since the integrand is bounded and ρ_∞ vanishes at infinity.

For the second integral in (A.4), we observe at first

$$\begin{aligned} \nabla g_t \cdot \nabla\partial_t g_t &= \nabla g_t \cdot \nabla(\Delta g_t - \nabla V \cdot \nabla g_t) \\ &= \nabla g_t \cdot \nabla(\Delta g_t) - \nabla g_t^\top (\text{Hess } g_t)\nabla V - \nabla g_t^\top(\text{Hess } V)\nabla g_t \end{aligned} \quad \text{(A.5)}$$

The first and third term will be very important. The latter allows us to make use of the condition, whereas the first one will lead to a wise application of partial integration. First we calculate

$$\begin{aligned} \frac{1}{2}\Delta(|\nabla g_t|^2) &= \text{div}((\text{Hess } g_t)\nabla g_t) = \text{tr}\left((\text{Hess } g_t)(\text{Hess } g_t)\right) + \nabla g_t \cdot \nabla\Delta g_t \\ &= \|\text{Hess } g_t\|_F^2 + \nabla g_t \cdot \nabla\Delta g_t. \end{aligned}$$

We denoted by $\|\cdot\|_F$ the Frobenius norm. Using this expression for $\nabla g_t \cdot \nabla\Delta g_t$ in (A.5) yields

$$\nabla g_t \cdot \nabla\partial_t g_t = \frac{1}{2}\Delta(|\nabla g_t|^2) - \nabla g_t^\top(\text{Hess } g_t)\nabla V - \|\text{Hess } g_t\|_F^2 - \nabla g_t^\top(\text{Hess } V)\nabla g_t$$

Because of

$$\frac{1}{2}\Delta(|\nabla g_t|^2)\rho_\infty - \nabla g_t^\top (\text{Hess } g_t)\nabla V \rho_\infty = \frac{1}{2}\text{div}\left(\nabla(|\nabla g_t|^2)\rho_\infty\right)$$

we obtain

$$\nabla g_t \cdot \nabla \partial_t g_t \rho_\infty = \frac{1}{2}\text{div}\left(\nabla(|\nabla g_t|^2)\rho_\infty\right) - \|\text{Hess } g_t\|_F^2 \rho_\infty - \nabla g_t^\top (\text{Hess } V)\nabla g_t \rho_\infty.$$

Plugging this into the second integral in (A.4) gives

$$\begin{aligned} R_2 := &- \int_{\mathbb{R}^d} \psi''(g_t)\text{div}(\nabla(|\nabla g_t|^2)\rho_\infty)\,dx \\ &+ 2\int_{\mathbb{R}^d} \psi''(g_t)\left\{\|\text{Hess } g_t\|_F^2 + \nabla g_t^\top (\text{Hess } V)\nabla g_t\right\}\,d\mu. \end{aligned}$$

Now, we obtain after partial integration in the first integral and using the positive definiteness of Hess V in the second one

$$R_2 \geq \int_{\mathbb{R}^d} \psi'''(g_t)\nabla g_t \cdot \nabla(|\nabla g_t|^2)\,d\mu + 2\int_{\mathbb{R}^d} \psi''(g_t)\left\{\|\text{Hess } g_t\|_F^2 + \lambda|\nabla g_t|^2\right\}\,d\mu.$$

The surface integral vanishes since $|\nabla(|\nabla g_t|^2)| = |2(\text{Hess } g_t)\nabla g_t| \in L^1(\rho_\infty)$ and $\psi'''(g_t)$ is bounded.

The very last integral is

$$2\int_{\mathbb{R}^d} \psi''(g_t)\lambda|\nabla g_t|^2\,dx = -2\lambda I_\psi(\rho_t \mid \rho_\infty).$$

Combining all integrals $R_\psi(\rho_t \mid \rho_\infty) = R_1 + R_2$ we establish

$$\begin{aligned} &R_\psi(\rho_t \mid \rho_\infty) \geq -2\lambda I_\psi(\rho_t \mid \rho_\infty) + \int_{\mathbb{R}^d} \psi^{IV}(g_t)|\nabla g_t|^4\,d\mu \\ &+ \int_{\mathbb{R}^d} \psi'''(g_t)\left[2\nabla g_t^\top (\text{Hess } g_t)\nabla g_t + \nabla g_t \cdot \nabla(|\nabla g_t|^2)\right]\,d\mu + 2\int_{\mathbb{R}^d} \psi''(g_t)\,\|\text{Hess } g_t\|_F^2\,d\mu \end{aligned}$$

Observe

$$2\nabla g_t^\top (\text{Hess } g_t)\nabla g_t + \nabla g_t \cdot \nabla(|\nabla g_t|^2) = 4\nabla g_t^\top (\text{Hess } g_t)\nabla g_t.$$

The integrand in the last three integrals is

$$\psi^{IV}(g_t)|\nabla g_t|^4 + 4\psi'''(g_t)\nabla g_t^\top (\text{Hess } g_t)\nabla g_t + 2\psi''(g_t)\,\|\text{Hess } g_t\|_F^2$$

This can expressed by

$$\text{tr}\left(\begin{pmatrix}\psi^{IV}(g_t) & 2\psi'''(g_t)\\ 2\psi'''(g_t) & 2\psi''(g_t)\end{pmatrix}\begin{pmatrix}|\nabla g_t|^4 & \nabla g_t^\top (\text{Hess } g_t)\nabla g_t\\ \nabla g_t^\top (\text{Hess } g_t)\nabla g_t & \|\text{Hess } g_t\|_F^2\end{pmatrix}\right) = \text{tr}\,(XY)$$

and we obtain

$$R_\psi(\rho_t \mid \rho_\infty) \geq -2\lambda I_\psi(\rho_t \mid \rho_\infty) + \int_{\mathbb{R}^d} \operatorname{tr}(XY)\, d\mu.$$

We have $\psi^{IV} \geq 0$ and $2\psi^{IV}\psi'' - 4(\psi''')^2 \geq 0$, since ψ generates a relative entropy, and hence the matrix X is positive semi-definite. Because of

$$\nabla g_t^\top (\text{Hess } g_t)\nabla g_t \leq |\nabla g_t|\, |(\text{Hess } g_t)\nabla g_t| \leq \|\text{Hess } g_t\|_F\, |\nabla g_t|^2$$

the determinant of the second matrix Y is nonnegative, hence this matrix is positive semi-definite. Therefore, $\operatorname{tr}(XY) \geq 0$ and we establish

$$R_\psi(\rho_t \mid \rho_\infty) \geq -2\lambda I_\psi(\rho_t \mid \rho_\infty).$$

By virtue of $I_\psi(\rho_t \mid \rho_\infty) \leq 0$ this is

$$\frac{d}{dt}|I_\psi(\rho_t \mid \rho_\infty)| \leq -2\lambda |I_\psi(\rho_t \mid \rho_\infty)| \tag{A.6}$$

and we obtain with Gronwall's Lemma

$$|I_\psi(\rho_t \mid \rho_\infty)| \leq e^{-2\lambda} |I_\psi(\rho_0 \mid \rho_\infty)|.$$

By Lemma 3.5 we know $e_\psi(\rho_t \mid \rho_\infty) \to 0$ and by the above inequality $|I_\psi(\rho_t \mid \rho_\infty)| \to 0$ as $t \to +\infty$. Now, we integrate (A.6) from t to $+\infty$ yielding

$$I_\psi(\rho_t \mid \rho_\infty) = \frac{d}{dt} e_\psi(\rho_t \mid \rho_\infty) \leq -2\lambda e_\psi(\rho_t \mid \rho_\infty).$$

This implies again by Gronwall's Lemma

$$e_\psi(\rho_t \mid \rho_\infty) \leq e^{-2\lambda} e_\psi(\rho_0 \mid \rho_\infty).$$

which concludes the proof. □

Bibliography

[AMTU01] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex sobolev inequalities and the rate of convergence to equilibrium for fokker-planck type equations. *Communications in Partial Differential Equations*, 26(1-2):43–100, 2001.

[Bau14] F. Baudoin. *Diffusion processes and stochastic calculus.* EMS textbooks in mathematics. European Math. Soc., Zürich, 2014.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften. Springer, Cham [u.a.], 2014.

[Bre11] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. Springer, New York, NY [u.a.], 2011.

[Eva10] L. C. Evans. *Partial differential equations.* Graduate studies in mathematics. American Math. Soc., Providence, RI, 2. edition, 2010.

[Fri64] A. Friedman. *Partial differential equations of parabolic type.* Prentice-Hall, Englewood Cliffs, NJ, 1964.

[Gro75] L. Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics, 1 December 1975, Vol.97(4), pp.1061-1083*, 1975.

[Hai16] M. Hairer. *Convergence of Markov Processes. Lecture notes.* University of Warwick, 2016.

[HS87] R. Holley and D. Stroock. Logarithmic sobolev inequalities and stochastic ising models. *Journal of Statistical Physics, 1987, Vol.46(5), pp.1159-1194*, 1987. We use logarithmic Sobolev inequalities to study the ergodic properties of stochastic Ising models both in terms of large deviations and in terms of convergence in distribution.

[Kal18] M. Kaltenbäck. *Funktionalanalysis 2. Vorlesungsskript.* TU Wien, 2018.

[Kha12] R. Khasminskii. *Stochastic stability of differential equations.* Stochastic modelling and applied probability. Springer, Berlin [u.a.], 2., completely rev. and enl. ed. edition, 2012.

[KS00] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.* Springer study edition. Springer, New York, 2. edition, 2000.

[OV00] F. Otto and C. Villani. Generalization Of An Inequality By Talagrand, And Links With The Logarithmic Sobolev Inequality. *Journal of Functional Analysis*, 173:361–400, 06 2000.

[Ris96] H. Risken. *The Fokker-Planck equation : methods of solution and applications.* Springer series in synergetics. Springer, Berlin [u.a.], 2. ed., 3. print. edition, 1996.

[Rot81] O. Rothaus. Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators. *Journal of Functional Analysis*, 42:110–120, 06 1981.

[RS78] M. Reed and B. Simon. *Methods of modern mathematical physics.* Academic Pr., San Diego, Calif. [u.a.], 1978.

[SCDM04] C. Sparber, J. A. Carrillo, J. Dolbeault, and P. A. Markowich. On the long-time behavior of the quantum Fokker-Planck equation. *Monatshefte für Mathematik*, 141(3):237–257, Mar. 2004.

[SM79] Z. Schuss and B. J. Matkowsky. The exit problem: A new approach to diffusion across potential barriers. *SIAM Journal on Applied Mathematics*, 36(3):604–623, 1979.

[Sto15] G. Stoltz. *Introduction à la physique statistique numérique.* CERMICS, 2015.

[Tay11] M. E. Taylor. *Partial differential equations II : qualitative studies of linear equations.* Applied mathematical sciences Volume 116. Springer, New York, NY Dordrecht Heidelberg London, second edition. edition, 2011.

[Tur00] Bengt Ove Turesson. *Nonlinear potential theory and weighted Sobolev spaces.* Lecture notes in mathematics ; 1736. Springer, Berlin [u.a.], 2000.

[Yos08] K. Yoshida. *Functional analysis.* Classics in mathematics. Springer, Berlin Heidelberg New York, reprint of the 1980 edition, 6. edition, 2008.