Couplings and Kantorovich contractions with explicit rates for diffusions

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Raphael Zimmer
aus Wipperfürth

Bonn, Januar 2017

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Andreas Eberle
2. Gutachter: Prof. Dr. Karl-Theodor Sturm

Tag der Promotion: 27.06.2017
Erscheinungsjahr: 2017

Abstract

We consider certain classes of diffusion and McKean-Vlasov processes and provide non-asymptotic quantifications of the longtime behavior using coupling methods. The thesis is divided into three main parts.

In the first part, we consider $\mathbb{R}^d$ valued diffusions of type

$$dX_t \;=\; b(X_t)\,dt \,+\, dB_t.$$

Assuming a geometric drift assumption, we establish Kantorovich contractions with explicit contraction rates for the transition kernels. The results are in the spirit of Mattingly and Hairer's extensions of Harris' theorem, but do not rely on a small set condition. Instead we use reflection coupling and adjust the underlying cost function of the Kantorovich distance in a very specific way to the diffusion model. The resulting rate is given explicitly in terms of a one-sided Lipschitz bound on the drift coefficient and the growth of a chosen Lyapunov function. Consequences include exponential convergence in weighted total variation norms, gradient bounds, bounds for ergodic averages, and Kantorovich contractions for nonlinear McKean-Vlasov processes in the case of sufficiently weak but not necessarily bounded nonlinearities. We also establish quantitative bounds for subgeometric ergodicity assuming a subgeometric drift condition.

In the second part, we show that a related strategy can also be applied for a class of infinite-dimensional and degenerate diffusion processes. Given a separable and real Hilbert space $\mathbb{H}$ and a trace-class, symmetric and non-negative operator $\mathcal{G}:\mathbb{H}\to\mathbb{H}$, we examine the equation

$$dX_t = -X_t\,dt + b(X_t)\,dt + \sqrt{2}\,dW_t, \qquad X_0 = x\in\mathbb{H},$$

where (W_t) is a $\mathcal{G}$-Wiener process on $\mathbb{H}$ and $b:\mathbb{H}\to\mathbb{H}$ is Lipschitz. We assume that there is a splitting of $\mathbb{H}$ into a finite-dimensional space $\mathbb{H}^l$ and its orthogonal complement $\mathbb{H}^h$ such that $\mathcal{G}$ is strictly positive definite on $\mathbb{H}^l$ and the nonlinearity b admits a contraction property on $\mathbb{H}^h$. Assuming a geometric drift condition, we derive a Kantorovich contraction with an *explicit* contraction rate for the corresponding Markov kernels. Our bounds on the rate are based on the eigenvalues of $\mathcal{G}$ on the space $\mathbb{H}^l$, a Lipschitz bound on b and a geometric drift condition.

In the third part, we present a novel approach of coupling two multidimensional and nondegenerate Itô processes (X_t) and (Y_t) which follow dynamics with different drifts. The coupling is *sticky* in the sense that there is a stochastic process (r_t), which solves a one-dimensional stochastic differential equation with a *sticky boundary* behavior at zero, such that almost surely $|X_t - Y_t|\le r_t$ for all $t\ge 0$. The coupling is constructed as a weak limit of Markovian couplings. We provide explicit, non-asymptotic and longtime stable bounds for the probability of the event $\{X_t = Y_t\}$.

Acknowledgement

I want to thank my supervisor Andreas Eberle for his extensive support and guidance during my studies in Bonn and the PhD period. In the same spirit, I want to thank Arnaud Guillin for his support and for his hospitality during several visits in Clermont-Ferrand. Moreover, I want to thank Anton Bovier for his support. I am also grateful to Martin Hairer for his hospitality and guidance while visiting the University of Warwick.

More generally, I want to thank the whole stochastic group for the nice time I had in Bonn. In particular, I want to thank my office mate Becca for a pleasant working environment and for many encouraging words whenever one of my proofs went wrong. Special thanks go to our secretary Mei-Ling Wang who contributes a lot to the well-being of the whole probability group.

My PhD is financially supported by the DFG through a scholarship of the Bonner International Gradauate School (BIGS), the Institute of Applied Mathematics and the Collaborative Research Center 1060. I am grateful for this help!

Finally, I want to thank my parents and my girlfriend for their extensive support.

Contents

List of Figures

0 Introduction

We consider certain classes of diffusion and McKean-Vlasov processes and provide non-asymptotic quantifications of the longtime behavior using coupling methods. This thesis is divided into three main chapters.

In the first chapter, we consider $\mathbb{R}^d$ valued diffusion processes of type

$$dX_t \;=\; b(X_t)\,dt \,+\, dB_t. \tag{0.1}$$

Assuming a geometric drift condition, we establish contractions of the transition kernels in Kantorovich (L^1 Wasserstein) distances with explicit constants. We retrieve constants that are explicit in parameters which can be computed with little effort from one-sided Lipschitz conditions for the drift coefficient and the growth of a chosen Lyapunov function. Consequences include exponential convergence in weighted total variation norms, gradient bounds, bounds for ergodic averages, and Kantorovich contractions for nonlinear McKean-Vlasov diffusions in the case of sufficiently weak but not necessarily bounded nonlinearities. We also establish quantitative bounds for subgeometric ergodicity assuming a subgeometric drift condition.

In the second chapter, we consider a class of infinite-dimensional and possibly degenerate diffusions of type

$$dX_t \;=\; -X_t\,dt \,+\, b(X_t)\,dt \,+\, dW_t$$

on a separable Hilbert space $\mathbb{H}$, where W_t is a $\mathcal{G}$-Wiener process and $\mathcal{G}$ a trace-class, symmetric and non-negative operator. We assume there is a splitting of $\mathbb{H}$ into a finite-dimensional space $\mathbb{H}^l$ and its orthogonal complement $\mathbb{H}^h$ such that $\mathcal{G}$ is strictly positive definite on $\mathbb{H}^l$ and the nonlinearity b admits a contraction property on $\mathbb{H}^h$. Assuming a geometric drift condition, we derive a Kantorovich contraction with an explicit contraction rate for the corresponding Markov transition functions. The bounds on the rate are based on the eigenvalues of $\mathcal{G}$ on the space $\mathbb{H}^l$, a Lipschitz bound on b and a geometric drift condition. In comparison to the diffusions considered in the first chapter, the major difficulty here is that the driving noise is possibly degenerate. The results are obtained by a direct coupling approach.

In chapter three, we present a novel approach of coupling two multidimensional and nondegenerate diffusions (X_t) and (Y_t) which follow dynamics with different drifts. Our coupling is *sticky* in the sense that there is a stochastic process (r_t), which solves a one-dimensional stochastic differential equation with a *sticky boundary* behavior at zero, such that almost surely $|X_t - Y_t| \le r_t$ for all $t \ge 0$. The coupling is constructed as a weak limit of Markovian couplings. We provide explicit, non-asymptotic and longtime stable bounds for the probability of the event $\{X_t = Y_t\}$. Sticky couplings generalize the coupling approach for nonlinear diffusions from the first chapter.

Before we present the contributions of this thesis in detail, we recall some basics. In Section 0.1 we introduce transportation distances. These distances are then used to formulate contraction inequalities for Markov transition functions in Section 0.2. We discuss a few of the many remarkable consequences such inequalities have and make general remarks on how to establish them. In Section 0.3 we recall Harris type theorems which are among the standard tools for studying ergodic properties of Markov processes and can be used to derive contraction inequalities. The statements in Sections 0.1 - 0.3 are formulated for general Markov chains taking values in a Polish space. In Section 0.4 we get more concrete and discuss existing coupling approaches for diffusions in $\mathbb{R}^d$.

All statements in this chapter are essentially known and references are provided in each section. Let us fix a Polish space $(\mathcal{S}, d)$ with Borel σ-algebra $\mathcal{B}(\mathcal{S})$.

0.1 Transportation distances

We introduce transportation distances for measures. The set of probability measures on $\mathcal{B}(\mathcal{S})$ is denoted by $\mathcal{P}(\mathcal{S})$. Let $\mu, \nu \in \mathcal{P}(\mathcal{S})$. A measure $\gamma \in \mathcal{P}(\mathcal{S} \times \mathcal{S})$ is called a *coupling* of the measures (μ, ν) if

$$\gamma(A \times \mathcal{S}) = \mu(A) \quad \text{and} \quad \gamma(\mathcal{S} \times A) = \nu(A) \qquad \text{for any } A \in \mathcal{B}(\mathcal{S}).$$

The set of all such couplings is denoted by $C(\mu, \nu)$. Let $c : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ be a measurable function. The *optimal transportation cost* of two measures $\mu, \nu \in \mathcal{P}(\mathcal{S})$ w.r.t. the *cost function* c is defined by

$$\mathcal{W}_c(\mu, \nu) \;=\; \inf_{\gamma \in C(\mu,\nu)} \int c(x, y)\, \gamma(dx\, dy) \in [0, \infty]. \tag{0.2}$$

If c is lower semicontinuous, then there is always a coupling $\gamma \in C(\mu, \nu)$ which is optimal in the sense that $\mathcal{W}_c(\mu, \nu) = \int c(x, y)\, \gamma(dx\, dy)$, cf. [149, Theorem 4.1]. We give examples of typical cost functions.

Example 1 (Kantorovich distance)**.** *Let $\rho : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ be a metric on $\mathcal{S}$ which is lower semicontinuous. Then $\mathcal{W}_\rho$ is a metric on the set*

$$\mathcal{P}_\rho(\mathcal{S}) \;=\; \left\{ \mu \in \mathcal{P}(\mathcal{S}) : \int \rho(x, y)\, \mu(dy) < \infty \text{ for some } x \in \mathcal{S} \right\},$$

cf. [149, Section 6]. The distance $\mathcal{W}_\rho$ is called Kantorovich distance w.r.t. ρ*. In the sequel, we often work with lower semicontinuous functions ρ which are only* semimetrics*, i.e. which satisfy $\rho(x, y) = \rho(y, x)$ for all $x, y \in \mathcal{S}$ and $\rho(x, y) = 0$ if and only if $x = y$. In this case, the definition of $\mathcal{W}_\rho$ remains meaningful and $\mathcal{W}_\rho$ is a semimetric on $\mathcal{P}_\rho(\mathcal{S})$.*

Example 2 (L^p Wasserstein distances)**.** *Let $p \in [1, \infty)$. More generally, the L^p* Wasserstein distance *of two measures $\mu, \nu \in \mathcal{P}(\mathcal{S})$ is defined by*

$$\mathcal{W}^p(\mu,\nu) \quad = \quad \inf_{\gamma \in C(\mu,\nu)} \left(\int d(x,y)^p \, \gamma(dx\, dy) \right)^{1/p} .$$

One can show that $(\mathcal{W}^p, \mathcal{P}^p(\mathcal{S}))$ is a Polish space, where $\mathcal{P}^p(\mathcal{S})$ is the set of measures $\mu \in \mathcal{P}(\mathcal{S})$ such that $\int d(x,y)^p \, \mu(dy) < \infty$ for some, and hence all, $x \in \mathcal{S}$. Moreover, if a sequence of measures (μ_n) in $\mathcal{P}^p(\mathcal{S})$ converges towards a measure $\mu \in \mathcal{P}^p(\mathcal{S})$, then $\mu_n \to \mu$ weakly, cf. [149, Section 6].

Example 3 (Total variation distance)**.** *The total variation distance is a Kantorovich distance w.r.t. $\rho(x,y) = I_{x \neq y}$, i.e.*

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \in \mathcal{B}(\mathcal{S})} |\mu(A) - \nu(A)| = \mathcal{W}_\rho(\mu,\nu).$$

Transportation distances have a long and comprehensive history. The statements and definitions in this section are based on the book [149, Chapter 1 and Chapter 6]. There, one can also find a historic outline on the development of transportation distances.

0.2 Kantorovich contractions

Fix an index set $I = \mathbb{N}$ or $I = \mathbb{R}_+$. In the following, $(p_t)_{t \in I}$ denotes a *Markov transition function* on $\mathcal{S}$, i.e. a family of probability kernels $p_t : \mathcal{S} \times \mathcal{B}(\mathcal{S}) \to [0,1]$ such that $p_0(x, \cdot) = \delta_x(\cdot)$ and $p_s p_t = p_{s+t}$ for any $s, t \in I$, where $(p_s p_t)(x, A) := \int p_s(x, dy) \, p_t(y, A)$. We use the notation $\mu p_t(dx) = \int p_t(y, dx) \, \mu(dy)$ for measures $\mu \in \mathcal{P}(\mathcal{S})$ and $p_t f(x) = \int f(y) \, p_t(x, dy)$ for functions $f : \mathcal{S} \to \mathbb{R}$, whenever the latter integral is meaningful. Given a Markov transition function $(p_t)_{t \in I}$, one can show that for any $\mu \in \mathcal{P}(\mathcal{S})$ there is a unique probability measure P_μ on the product space $\mathcal{S}^I = \{\omega : I \to \mathcal{S}\}$ such that the canonical process $(X_t)_{t \in I}$, $X_t(\omega) := \omega(t)$, is a Markov process with transition function (p_t) and $P_\mu \circ (X_0)^{-1} = \mu$, cf. e.g. [52]. In this chapter, we always assume that the occurring transitions functions are Feller, i.e. that for any $f \in C_b$ and $t \in I$ we have that $p_t f \in C_b$, where C_b is the set of continuous and bounded functions $f : \mathcal{S} \to \mathbb{R}$.

We are interested in contraction inequalities for Markov transition functions. We say that (p_t) satisfies a *Kantorovich contraction* w.r.t. a semimetric $\rho : \mathcal{S} \times \mathcal{S} \to [0, \infty)$, if there is a constant $c \in (0, \infty)$ such that

$$\mathcal{W}_\rho(\mu p_t, \nu p_t) \quad \leq \quad \mathrm{e}^{-ct} \, \mathcal{W}_\rho(\mu, \nu) \quad \text{holds for any } \mu, \nu \in \mathcal{P}(\mathcal{S}) \text{ and } t \in I. \tag{0.3}$$

The constant c is called *contraction rate*. Inequalities of this type already appear in a work of Dobrushin in the 70s, cf. [40]. If ρ is a metric, then the largest c

such that (0.3) holds is sometimes called Wasserstein curvature of the Markov chain w.r.t. ρ [89, 124] which is motivated by a relation to the concept of Ricci curvature on Riemannian manifolds, cf. [3, 150]. In order to draw from (0.3) conclusions on the longtime behavior of the Markov process, we impose the following additional assumption.

Assumption 1. *The function ρ is lower semicontinuous. Moreover, there are a measurable function $V : \mathcal{S} \to \mathbb{R}_+$ and constants $C_1, C_2, C_3, \lambda \in (0, \infty)$ such that*

$$\begin{aligned} d(x,y) &\leq C_1\, \rho(x,y), &(0.4)\\ \rho(x,y) &\leq C_2\, \left(1 + V(x) + V(y)\right), &(0.5)\\ p_t V(x) &\leq C_3 + \mathrm{e}^{-\lambda t}\, V(x), &(0.6)\end{aligned}$$

for any $x, y \in \mathcal{S}$ and $t \in I$.

The function V is called *Lyapunov function*. We discuss the role of such functions later on in more detail. For the moment one might just think of it as an integrability constraint ensuring that $\sup_{t\in I} E_x[V(X_t)] < \infty$. The main purpose of imposing these conditions is the following statement:

Consequence 1. *If* (0.3) *and Assumption 1 hold true for a semimetric ρ, then there exists a unique measure $\pi \in \mathcal{P}(\mathcal{S})$ such that $\pi p_t = \pi$ for any $t \in I$. Moreover, $\pi \in \mathcal{P}_V(\mathcal{S}) \subset \mathcal{P}_\rho(\mathcal{S})$ and*

$$\mathcal{W}^1(\mu p_t, \pi) \ \leq \ C_1\, \mathcal{W}_\rho(\mu p_t, \pi) \ \leq \ C_1\ \mathrm{e}^{-ct}\, \mathcal{W}_\rho(\mu, \pi) \tag{0.7}$$

for any $\mu \in \mathcal{P}(\mathcal{S})$ and $t \in I$.

Here, $\mathcal{P}_V(\mathcal{S})$ is the set of measures $\mu \in \mathcal{P}(\mathcal{S})$ such that $\int V d\mu < \infty$. The proof is given in the appendix for the readers convenience, cf. page 139. Consequence 1 demonstrates that, in the setting of Assumption 1, one cannot expect inequality (0.3) to hold for an arbitrary Markov transition function. We discuss sufficient conditions in the next section. For the moment, we assume that inequality (0.3) holds for a given Markov transition function together with a semimetric ρ satisfying Assumption 1, and discuss a few consequences.

Given a function $g : \mathcal{S} \to \mathbb{R}$, we define the Lipschitz constant of g w.r.t. ρ by

$$|g|_{\mathrm{Lip}(\rho)} = \sup\left\{ \frac{|g(x) - g(y)|}{\rho(x,y)} : x, y \in \mathcal{S}, x \neq y \right\} \in [0, \infty] \tag{0.8}$$

and write $\mathrm{Lip}(\rho)$ for the set of measurable functions g satisfying $|g|_{\mathrm{Lip}(\rho)} < \infty$. The following result is taken from [70, Proposition 2.8]. Similar statements occur in [25, 152, 124, 31].

Consequence 2 ($L^2(\pi)$ spectral gap)**.** *If π is reversible w.r.t. (p_t), i.e. if*

$$\int I_{A\times B}(x,y)\,\pi(dx)\,p(x,dy) \;=\; \int I_{B\times A}(x,y)\,\pi(dx)\,p(x,dy) \quad \text{for any } A,B\in\mathcal{B}(\mathcal{S}),$$

and if $\mathrm{Lip}(\rho)\cap L^\infty(\pi)$ *is dense in* $L^2(\pi)$*, then*

$$\left|p_t f-\int f\,d\pi\right|^2_{L^2(\pi)} \le \mathrm{e}^{-2ct}\left|f-\int f\,d\pi\right|^2_{L^2(\pi)} \qquad \text{for any } f\in L^2(\pi) \text{ and } t\ge 0.$$

A sufficient condition for the density assumption is that $(\mathcal{S},|\cdot|)$ is a separable Banach space and $d(x,y)=|x-y|$, cf. [70, Theorem 2.15].

Given a probability measure π, a typical question in practice is how to estimate integrals $\int f d\pi$. One strategy to approach such a problem is to construct a Markov chain (X_n) admitting π as the unique invariant measure and then to argue that the ergodic averages converge, i.e. that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_n)\to\int f d\pi \qquad \text{for } n\to\infty.$$

Following the work of Joulin and Ollivier [91, 90, 123], we present a non-asymptotic quantification of this convergence based on Kantorovich contractions.

Consequence 3. *Let $g\in|g|_{\mathrm{Lip}(\rho)}$. Then, for any $n\in\mathbb{N}$,*

$$\left|E_x\left[\frac{1}{n}\sum_{k=1}^{n} g(X_n)\right]-\int g\,d\pi\right| \;\le\; \frac{1}{cn}\,|g|_{\mathrm{Lip}(\rho)}\int\rho(x,y)\,\pi(dy),$$

$$\mathrm{Var}_x\left[\frac{1}{n}\sum_{k=1}^{n} g(X_n)\right] \;\le\; \frac{1}{2(1-\mathrm{e}^{-c})n}\,|g|^2_{\mathrm{Lip}(\rho)}\int\int\rho(y,z)^2\,p_n(x,dy)\,p_n(x,dz).$$

A proof can be found in [91]. Related statements for diffusions are given in [51].

The Kantorovich contraction (0.3) has many other interesting consequences and it is impossible to give a full account here. Hopefully the few examples already show that it is interesting to examine contraction inequalities of type (0.3). Notice that for the above statements, ρ only needs to be a semimetric and the triangle inequality is not needed.

Now we turn to the question how contractions of type (0.3) can be established. Let us for the moment assume that $I=\mathbb{N}$, i.e. that $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with one-step kernel $p=p_1$. We call a probability kernel $\tilde{p}$ on $\mathcal{S}\times\mathcal{S}$ a coupling for p, if $\delta_{(x,y)}\tilde{p}\in C(\delta_x p_n,\delta_y p)$ for any $x,y\in\mathcal{S}$, i.e. if $\delta_{(x,y)}\tilde{p}$ is a coupling of $\delta_x p$ and $\delta_y p$. In classical applications of coupling theory one often takes the ansatz of fixing a Kantorovich distance $\mathcal{W}_\rho$, e.g. the total variation distance, and then one tries to construct a suitable coupling which yields an upper bound on $\mathcal{W}_\rho(\delta_x p_n, \delta_y p_n)$ and allows to conclude that $\mathcal{W}_\rho(\delta_x p_n, \delta_y p_n) \to 0$ for $n \to \infty$. More generally, one can also consider the underlying distance ρ as a variable parameter and look for suitable *combinations* of couplings and distances. Assume that there exist a semimetric ρ, a constant $c \in (0,1)$ and a coupling $\tilde{p}$, such that

$$\tilde{\mathcal{L}}\rho(x,y) \leq -c\,\rho(x,y) \qquad \text{holds for any } x,y \in \mathcal{S}, \tag{0.9}$$

where $\tilde{\mathcal{L}} := \tilde{p} - I$ denotes the generator associated with $\tilde{p}$. Then,

$$\mathcal{W}_\rho(\mu p, \nu p) \;\;\leq\;\; (1-c)\,\mathcal{W}_\rho(\mu,\nu) \qquad \text{holds for any } \mu,\nu \in \mathcal{P}(\mathcal{S}).$$

We see that a contraction inequality can be obtained by aligning couplings and distances such that (0.9) holds. A similar statement can be formulated for continuous-time Markov processes, but there one needs to be more careful making sense of (0.9) and one needs to impose some path regularity, cf. [26, Lemma A.6].

Up to the author's knowledge, the ansatz of aligning couplings and distances in such a way occurs first in the works [29, 28] by Mu-Fa Chen and Feng-Yu Wang in the 90s, see also [26]. There, the approach is used to retrieve bounds on spectral gaps for elliptic operators. In the context of Kantorovich contractions, related approaches can be found in the works of Mattingly, Hairer and Scheutzow [75, 69] and Eberle [50, 51]. Mattingly, Hairer and Scheutzow construct Harris type theorems in a general setup, which allow to establish Kantorovich contractions for a large class of Markov processes under verifiable assumptions. Eberle concentrates on a certain class of diffusions and puts much effort in maximizing the contraction rate. Roughly speaking, the strategy is to pick a reasonable Markovian coupling for the diffusions and then to construct an underlying distance ρ such that (0.9) holds with c being "as large as possible", where $\tilde{\mathcal{L}}$ denotes the generator of the coupling. We now present these results in more detail and start with general Harris type theorems.

0.3 Harris type theorems

In this section, $(X_n)_{n\in\mathbb{N}}$ denotes a Markov chain on $\mathcal{S}$ with one-step kernel p and generator $\mathcal{L} = p - I$.

Harris type theorems are nowadays among the standard tools for studying ergodic properties of Markov processes. The starting point for these theorems is the seminal work [77] from T. E. Harris in the 50s. He investigated existence and uniqueness of invariant measures for Markov chains on general state spaces. The main result is that the Markov transition kernel p admits an, up to multiplication with constants, unique invariant measure π, if there is a set $A \subset \mathcal{S}$ such that $P_x[T_A < \infty] = 1$ for *any* $x \in \mathcal{S}$, where $T_A = \inf\{n \geq 1 : X_n \in A\}$, and if there is a measure $m \in \mathcal{P}(\mathcal{S})$ together with $\alpha \in (0,1)$ such that

$$\inf_{x\in A} p(x,B) \geq \alpha\, m(B) \qquad \text{for any } B \in \mathcal{B}(\mathcal{S}). \tag{0.10}$$

The latter condition is typically called *small set* or *minorization condition* and can be interpreted as a local version of Doeblin's condition, cf. [41, 42]. Harris formulated the result actually under different conditions, but his conditions are equivalent to the ones stated here, cf. [8, Corollary 2.1] and [125].

Nowadays, Harris' type theorems typically use drift conditions to quantify the recurrence behavior of the Markov process and combine them with small set conditions to provide explicit bounds on the speed of convergence to equilibrium. In the case of diffusions such a statement can be found in the work of Khasminskii [79, 96], and in the general case it has been developed systematically by Meyn and Tweedie [119, 120, 118]. For a historical overview, describing the development in more detail and including references to authors which are not named here explicitly, we refer the reader to the commentaries in [119, Chapter 9-13]. We now state a more recent version of Harris' theorem which is due to Mattingly and Hairer [75] and allows to establish contraction inequalities. The first assumption is a geometric drift condition:

Assumption 2. *There exists a measurable function $V : \mathcal{S} \to \mathbb{R}_+$ and constants $C, \lambda \in (0, \infty)$ such that*

$$\mathcal{L}V(x) \;\leq\; C \,-\, \lambda V(x) \qquad \textit{for any } x \in \mathcal{S}. \tag{0.11}$$

For given $R \in (0, \infty)$, we define the level set

$$A_R \;:=\; \{x \in \mathcal{S} : V(x) \leq R\}.$$

If $R > C/\lambda$, i.e. $R = (1 + \delta)C/\lambda$ for some $\delta > 0$, one can show that (0.11) implies that $E_x[\exp(\frac{\lambda\delta}{1+\delta}T_{A_R})] < \infty$ for any $x \in \mathcal{S}$. Notice that (0.11) implies (0.6) with $C_3 = C/\lambda$.

Known result 1 (Harris' theorem)**.** *If Assumption 2 holds true and if condition* (0.10) *is satisfied for a set A_R with $R > 2C/\lambda$ and corresponding $\alpha_R \in (0, 1)$, then there exist $c_R, \epsilon_R \in (0, \infty)$ such that*

$$\mathcal{W}_\rho(\mu p_n, \nu p_n) \;\leq\; e^{-c_R n}\, \mathcal{W}_\rho(\mu, \nu) \quad \textit{for all } \mu, \nu \in \mathcal{P}(\mathcal{S}) \textit{ and } n \in \mathbb{N},$$

where $\rho(x, y) = [1 + \epsilon_R V(x) + \epsilon_R V(y)] \cdot I_{x \neq y}$. The constants are given explicitly by $c = -\log(\max(1 - \alpha_R/2, 1 - \lambda, 1 - \gamma_R))$, $\gamma_R = \epsilon_R(\lambda R - 2C)/(1 + \epsilon_R R)$ and $\epsilon_R = \alpha_R/(4C)$.

The statement is due to Hairer and Mattingly [75]. The formulation and constants have been adapted and differ from the version given in the latter source. The proof given in [75] is quite simple: The minorization condition implies a *local* contraction in total variation distance, i.e.

$$\mathcal{W}_{\rho_1}(\delta_x p, \delta_y p) \;\leq\; (1 - \alpha_R)\, W_{\rho_1}(\delta_x p, \delta_y p) \qquad \text{for any } x, y \in A_R,$$

where $\rho_1(x,y) = I_{x\neq y}$. Moreover, because of (0.11) and $R > 2C/\lambda$, there is $\beta_R \in (0,1)$ such that

$$\mathcal{W}_{\rho_2}(\delta_x p, \delta_y p) \;\leq\; (1-\beta_R) W_{\rho_2}(\delta_x p, \delta_y p) \qquad \text{for any } (x,y) \notin A_R \times A_R,$$

where $\rho_2(x,y) = [V(x)+V(y)] \cdot I_{x\neq y}$. One can then consider a family of distances $\rho_\epsilon := \rho_1 + \epsilon\rho_2$ and, choosing ϵ carefully, it is possible to establish a global contraction.

The result allows to establish *global* contractions based on *local* contractions in total variation norm and a recurrence criteria in form of a geometric drift condition. On locally compact state spaces the local contraction can often be established. We give an example in the case of nondegenerate diffusion processes with values in $\mathbb{R}^d$ further below. However, for Markov chains on infinite-dimensional spaces, the condition might either be hard to verify or even false. Extending the result to a more general setting, Mattingly, Hairer and Scheutzow designed a weak Harris' theorem, where the local contraction in total variation norm is replaced by a local contraction in a possibly weaker distance. The main statement from [69] can be formulated as follows:

Known result 2 (Weak Harris' theorem)**.** *Suppose that Assumption 2 is satisfied for a continuous function $V : \mathcal{S} \to \mathbb{R}_+$ and that there is a lower semicontinuous semimetric $d : \mathcal{S}\times\mathcal{S} \to [0,1]$ which is* locally contracting *for p, i.e. there is $\alpha \in (0,1)$ such that*

$$\mathcal{W}_d(\delta_x p, \delta_y p) \;\leq\; \alpha\, d(x,y) \qquad \textit{for all } x,y \in \mathcal{S} \textit{ with } d(x,y) < 1.$$

Moreover, assume that the levelset $A = \{x \in \mathcal{S} : V(x) \leq 4C/\lambda\}$ is d-small, i.e. there is $\beta \in (0,1)$ such that

$$\mathcal{W}_d(\delta_x p, \delta_y p) \;\leq\; \beta \qquad \textit{for all } x,y \in A.$$

Then there exists $c, \epsilon \in (0,\infty)$ such that

$$\mathcal{W}_\rho(\mu p_n, \nu p_n) \;\leq\; e^{-cn}\, \mathcal{W}_\rho(\mu,\nu) \quad \textit{for any } \mu,\nu \in \mathcal{P}(\mathcal{S}) \textit{ and } n \in \mathbb{N},$$

where $\rho(x,y) = \sqrt{d(x,y)\,(1+\epsilon V(x)+\epsilon V(y))}$.

0.4 Diffusions

The statements in the latter sections have been formulated in an abstract setting. Now we consider diffusion processes in $\mathbb{R}^d$. First, we motivate why it is particularly interesting to derive Kantorovich contractions with explicit contraction rates for such processes. Afterwards, we discuss existing coupling approaches. In this section, $|\cdot|$ and $\langle\cdot,\cdot\rangle$ denote the euclidean norm and inner product on $\mathbb{R}^d$ respectively.

0.4.1 Motivation: Langevin equations

Assume that we are interested in a probabilty measure $\pi \in \mathcal{P}(\mathbb{R}^d)$ determined by

$$\pi(dx) \quad \propto \quad \exp\left(-U(x)\right) dx,$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is a smooth function satisfying $\int \exp\left(-U(x)\right) dx < \infty$. Typical questions in practice are how to generate samples from such a distribution or how to approximate integrals $\int f d\pi$ for which a direct computation might either not be possible or feasible. Classical *Markov chain Monte Carlo* (MCMC) methods tackle these question by constructing a Markov process (Z_t) with transition function (p_t) admitting π as an invariant probability measure and such that $\delta_x p_t \to \pi$ for $t \to \infty$. For sufficiently large t, one can then use Z_t as an approximate sample of π. Similarly, one can use ergodic averages $\frac{1}{n}\sum_{k=1}^n f(Z_k)$ to approximate integrals $\int f d\pi$. A comprehensive introduction into MCMC methods, including a historical overview, can be found in [2].

Among the important dynamics used for these purposes are Langevin equations. The Langevin equation describes in statistical physics the evolution of a particle in $\mathbb{R}^d$ subject to damping, environmental influences and random collisions, cf. [53, 102]. In terms of stochastic differential equations (SDEs) the position (X_t) and velocity (V_t) of the particle satisfy the equations

$$\begin{aligned} dX_t &= V_t \, dt, \\ dV_t &= -\gamma \, V_t \, dt \, - \, m^{-1} \, \nabla U(X_t) \, dt \, + \, \sqrt{2\gamma m^{-1}} \, dB_t. \end{aligned} \tag{0.12}$$

Here, $U : \mathbb{R}^d \to \mathbb{R}$ is a given potential, $m \in (0, \infty)$ denotes the mass of the particle, $\gamma \in (0, \infty)$ determines the friction and (B_t) is a d-dimensional Brownian motion. Closely related is the *overdamped Langevin* equation which is given by

$$dX_t \quad = \quad -\nabla U(X_t) \, dt \, + \, \sqrt{2} \, dB_t. \tag{0.13}$$

The latter equation is formally obtained from (0.12) by setting $\gamma = m^{-1}$ and passing to the limit $m \to 0$, cf. e.g. [126]. One thing which makes the Langevin equations (0.12) and (0.13) particularly interesting is that, under reasonable assumption on the potential U, the unique invariant probability measures on $\mathbb{R}^{2d}$ and $\mathbb{R}^d$ are given by $\pi \otimes \mathcal{N}(0, m^{-1})$ and π respectively. It is therefore not surprising that the dynamics are the foundation for several MCMC techniques, see e.g. [136, 61, 36, 45, 15].

From this point of view it is particularly interesting to obtain explicit and sharp bounds on the speed of convergence towards the invariant distribution. To this end, there exist several approaches. One approach to study the speed of convergence is based on functional inequalities, cf. e.g. [6] for results in this direction regarding reversible diffusion processes and [148] for applications in non-reversible settings. We focus in the following on direct coupling approaches.

0.4.2 Couplings and Kantorovich contractions

We now consider diffusions in $\mathbb{R}^d$ of type

$$dX_t \;=\; b(X_t)\,dt \,+\, dB_t, \qquad X_0 = x_0, \tag{0.14}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitz continuous and (B_t) is a d-dimensional Brownian motion. We assume non-explosiveness. Thus, for any given Brownian motion (B_t) and any initial value $x_0 \in \mathbb{R}^d$, there is a unique and strong solution (X_t) of (0.14), i.e. a stochastic process with continuous trajectories such that, for almost every trajectory, $X_0 = x_0$ and

$$X_t - X_0 \;=\; \int_0^t b(X_s)\,ds \,+\, B_t, \qquad t \geq 0.$$

If (X_t) and $(\tilde{X}_t)$ are solutions of (0.14) with the same initial value and w.r.t. the same Brownian motion, then $P[X_t = \tilde{X}_t \ \forall t \geq 0] = 1$. More generally, if (X_t) and $(\tilde{X}_t)$ are solutions of (0.14) with the same initial value (but possibly defined on different probability spaces), then the laws of the processes on $C([0,\infty),\mathbb{R}^d)$ coincide. The solution is a Markov process and we write (p_t) and $\mathcal{L} = \frac{1}{2}\Delta + \langle b, \nabla\rangle$ for the corresponding Markov transition function and generator respectively, cf. e.g. [87, Chapter IV]. Given solutions (X_t) and (Y_t) of (0.14), a coupling of the processes is a random variable $(\tilde{X}_t, \tilde{Y}_t)$ with values in $C([0,\infty),\mathbb{R}^{2d})$ such that the marginal laws $(\tilde{X}_t)$ and $(\tilde{Y}_t)$ on $C([0,\infty),\mathbb{R}^d)$ coincide with the laws of (X_t) and (Y_t) respectively.

We discuss conditions and approaches to establish Kantorovich contractions for (p_t). One possibility is to interpret (X_t) as a discrete-time Markov process $(X_n)_{n\in\mathbb{N}}$ and to apply general Harris type theorems for the one-step kernels $p = p_1$. As we have seen, two assumptions are needed for this: A drift condition and a minorization condition. The drift condition can often be verified by elementary computations and explicit bounds for the resulting constants can be obtained using the representation of $\mathcal{L}$ as a second order differential operator. For processes of type (0.14) with locally Lipschitz drift b satisfying a non-explosion criteria, the minorization condition (0.10) holds true for any compact set $A \subset \mathbb{R}^d$: One can argue that there is a continuous and strictly positive density $(x,y) \mapsto f(x,y)$ such that $p_1(x,B) = \int_B f(x,y)\,dy$ for any $B \in \mathcal{B}$, cf. [11, 10], and thus condition (0.10) is satisfied with m being the uniform distribution on A and $\alpha = \lambda(A)\min_{x,y\in A} f(x,y) > 0$, where $\lambda(A)$ denotes the Lebesgue measure of the set A, cf. e.g. [101, Discussion after Remark 1.29]. However, the diameter of the set A_R occurring in Harris' theorem, see Known result 1 further above, typically depends on the dimension and trying to quantify α, one is likely to end up with bounds which are exponentially small in the dimension, even for seemingly well behaved drifts b. Moreover, the minorization condition is not transparent in the sense that it is unclear how a perturbation of the drift b effects the corresponding α in Harris' theorem. In this sense, Harris' theorem is often applied in a qualitative, rather then a quantitative way. A noteworthy exception is the work [135] by Roberts and Rosenthal, who provide a way of quantifying the α in the

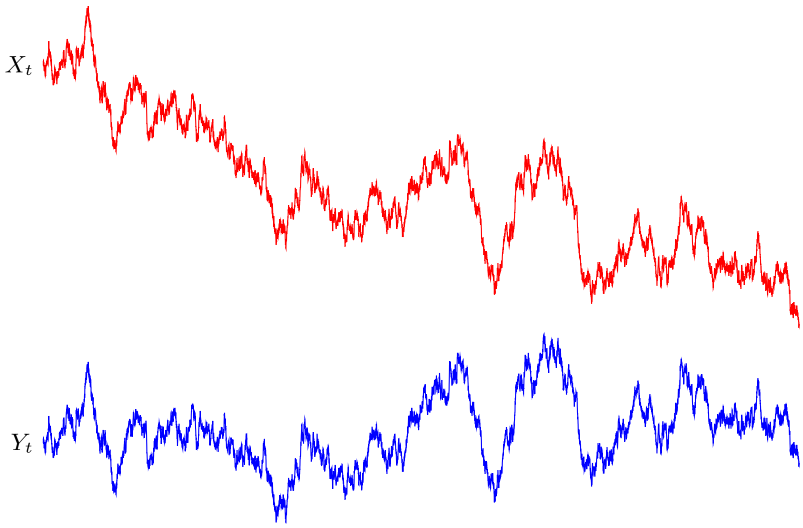

Figure 0.1: Synchronous coupling of one-dimensional diffusions

minorization condition for diffusions of type (0.14) using coupling arguments. There, the resulting α is expressed in terms of "how much the drift b varies on the set A_R". More explicitly, in one-dimension, the authors call a set $C \subset \mathbb{R}$ a "$[a, c]$-medium set", if $a \le b(x) \le c$ for any $x \in C$. If A_R is $[a, c]$-medium, the resulting α given in [135] depends on the difference $D = c - a$ and vanishes exponentially fast for $D \to \infty$. In a multidimensional setting one can formulate similar statements, cf. [135, Theorem 9]. In the first part of this thesis, we go one step further and develop a Harris' theorem for diffusions where we replace the minorization condition by a *one-sided* Lipschitz bound which yields more precise estimates in general. To this end, we now discuss direct coupling approaches for diffusions.

A comprehensive discussion of couplings for diffusions has started in the 80s, cf. [106, 107, 39, 27]. We recall two important couplings and applications regarding contractions. Given initial values $(x_0, y_0) \in \mathbb{R}^{2d}$ and a d-dimensional Brownian motion (B_t), we define a *synchronous coupling* of two solutions of (0.14) as a diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$ solving

$$\begin{aligned} dX_t &= b(X_t)\,dt + dB_t, \qquad & X_0 &= x_0, \\ dY_t &= b(Y_t)\,dt + dB_t, \qquad & Y_0 &= y_0, \end{aligned}$$

i.e. both processes (X_t) and (Y_t) are driven by the same Brownian motion. Figure 0.1 shows a trajectory of a synchronous coupling for two one-dimensional diffusions.

Let us for the moment assume that there is $c > 0$ such that

$$\langle b(x) - b(y), x - y\rangle \;\le\; -c\,|x-y|^2 \qquad \text{holds for any } x, y \in \mathbb{R}^d. \tag{0.15}$$

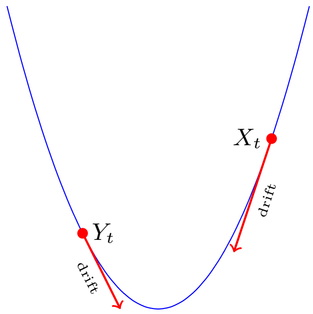

Figure 0.2: Diffusions inside a strictly convex potential

The condition is satisfied if $b = -\nabla U$ for a strictly convex function $U \in C^2$. Let (X_t, Y_t) be a synchronous coupling. The definition of the coupling and (0.15) imply that, for almost every trajectory, the difference process $t \mapsto X_t - Y_t$ is continuously differentiable and that

$$|X_t - Y_t|^2 \quad \leq \quad \mathrm{e}^{-2ct}\, |x_0 - y_0|^2 \qquad \text{for any } t \geq 0. \tag{0.16}$$

This *pathwise contraction* implies in particular a L^p Wasserstein contraction for any $p \in [1, \infty)$, i.e.

$$\mathcal{W}^p(\delta_{x_0} p_t, \delta_{y_0} p_t) \quad \leq \quad e^{-ct}\, \mathcal{W}^p(\delta_{x_0}, \delta_{y_0}) \qquad \text{for any } t \geq 0. \tag{0.17}$$

A similar statement can already be found in the work of Mu Fa Chen and Shao Fu Li [27] from 1989. If $b = -\nabla U$ for a C^2 function U, then strict convexity is also a necessary condition for (0.17) to hold, cf. the more recent result of von Renesse and Sturm [150]. Let us stress two aspects: First of all, (0.16) also holds for solutions (X_t, Y_t) of the deterministic dynamics

$$dX_t = b(X_t)\, dt, \qquad dY_t = b(Y_t)\, dt, \qquad (X_0, Y_0) = (x_0, y_0),$$

and therefore the driving Brownian motion plays no important role for the contraction result. Secondly, synchronous couplings do not necessarily meet in finite time, cf. [27]. Consider for example the case where $b(x) = -c\, x$ with $c > 0$. A solution with intial value x_0 is given by $X_t = e^{-c\,t} x_0 + \mathrm{e}^{-ct} \int_0^t \mathrm{e}^{ct}\, dB_s$. If (X_t, Y_t) is a synchronous coupling, then, due to pathwise uniqueness, $X_t - Y_t = e^{-c\,t}(x_0 - y_0)$. We see that (X_t, Y_t) is an *asymptotic coupling* in the sense that $X_t - Y_t \to 0$ for $t \to \infty$, but, unless $x_0 = y_0$, we have $X_t \neq Y_t$ for any $t > 0$.

We have seen that synchronous couplings are particularly useful if the underlying deterministic dynamics admits a contraction property. Nevertheless, if the underlying deterministic system is not globally contractive or if one wants to obtain bounds on the total variation distance $\|\delta_{x_0} p_t - \delta_{y_0} p_t\|_{\mathrm{TV}}$, then purely synchronous couplings are in general not a good choice.

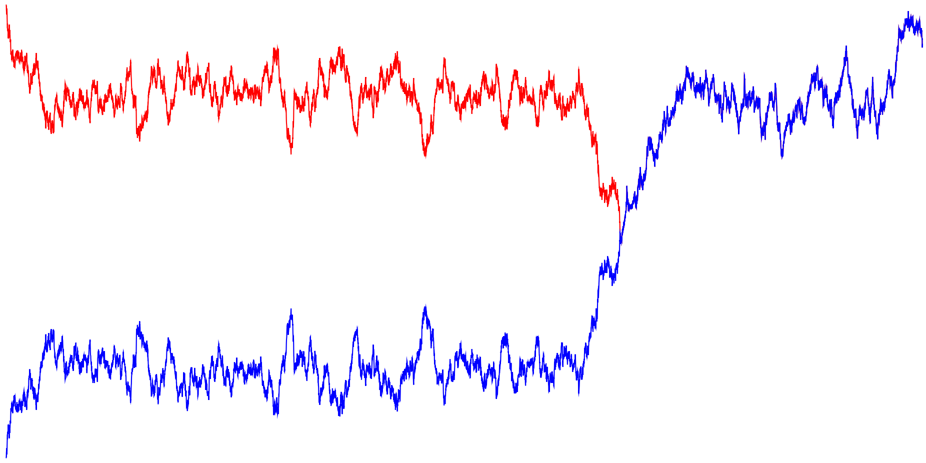

Figure 0.3: Reflection coupling of diffusions inside a double-well

Another important coupling for diffusions has been introduced by Lindvall and Roger [107] in the 80s: Given initial values $(x_0, y_0) \in \mathbb{R}^{2d}$ and a d-dimensional Brownian motion (B_t), a *reflection coupling* of two solutions of (0.14) is a diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$ satisfying

$$\begin{aligned} dX_t &= b(X_t)\,dt + dB_t, && \\ dY_t &= b(Y_t)\,dt + (I - 2\,e_t\,\langle e_t, \cdot\rangle)\,dB_t && \text{for } t < T, \\ Y_t &= X_t && \text{for } t \geq T, \\ (X_0, Y_0) &= (x_0, y_0), && \end{aligned}$$

where $T = \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time and, for $t < T$, e_t is the unit vector given by $e_t = (X_t - Y_t)/|X_t - Y_t|$. Generalizations of this coupling for diffusions on manifolds have been constructed in [32, 95]. Figure 0.3 shows a trajectory of a reflection coupling for two one-dimensional diffusions inside a double-well potential and Figure 0.4 shows such a coupling for two-dimensional Brownian motions without drift in the plane.

A reflection coupling has many remarkable properties: One crucial property is that the process $r_t := |X_t - Y_t|$ satisfies almost surley the SDE

$$dr_t \;\; = \;\; r_t^{-1}\,\langle X_t - Y_t, b(X_t) - b(Y_t)\rangle\,dt + 2\;dW_t, \quad t < T, \tag{0.18}$$

where (W_t) is a one-dimensional Brownian motion, cf. e.g. [51]. In particular, the driving noise (W_t) has a direct impact on $|X_t - Y_t|$. Moreover, the question of whether the two d-dimensional processes (X_t) and (Y_t) meet in finite time can be reduced to a *one-dimensional* problem by considering (0.18) and, under relatively mild assumptions, they actually do so, cf. [107, Lemma 1] for a precise statement. In the case $b = 0$, i.e. if one just considers Brownian motions without drift, reflection coupling is

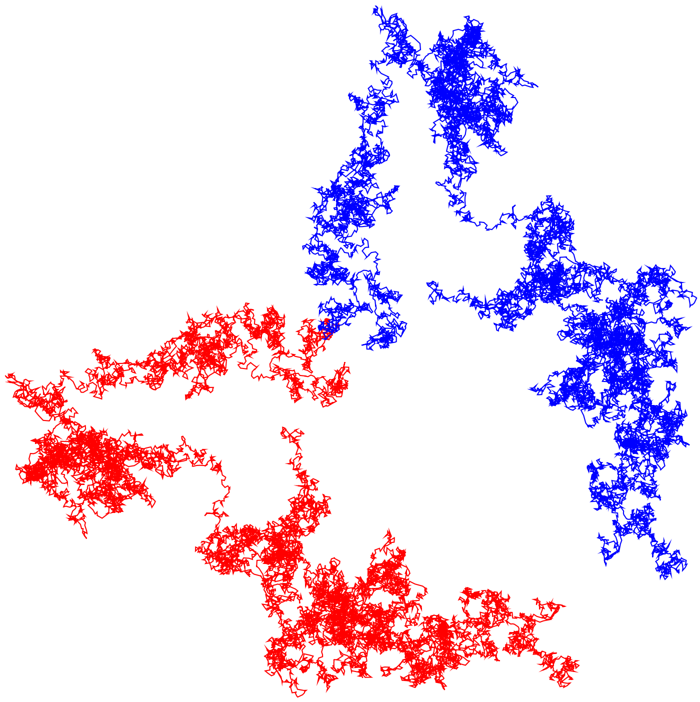

Figure 0.4: Reflection coupling of two-dimensional Brownian motions

optimal for the total variation distance, i.e.

$$\|\delta_{y_0}p_t - \delta_{x_0}p_t\|_{\mathrm{TV}} \quad = \quad E_{(x_0,y_0)}[I(X_t \neq Y_t)] \text{ for any } t \geq 0, \text{ cf. [27].}$$

We have seen that condition (0.15) implies a L^p Wasserstein contraction for the corresponding Markov transition functions and that synchronous couplings provide an elegant way of proving this. At least in the case $p = 1$, one can also use a reflection coupling to show that (0.17) holds by exploiting (0.18). However, in this case we do not have a pathwise contraction of $|X_t - Y_t|$, but only a contraction *on average*, i.e. we have that $E[|X_t - Y_t|] \leq e^{-ct}\,|x_0 - y_0|$ for any $t \geq 0$.

Unfortunately, condition (0.15) is too restrictive for many applications. In the recent works [50, 51], Eberle studies diffusion of type (0.14) assuming the condition (0.15) only outside of a bounded set and derives Kantorovich contractions for the corresponding transition functions. The main result, cf. [51, Theorem 2.2 and Corollary 2.3], can be stated as follows:

Known result 3. *Set*

$$\kappa(r) \quad := \quad \inf\left\{2\frac{\langle x-y, b(x)-b(y)\rangle}{|x-y|^2} : x, y \in \mathbb{R}^d \text{ with } |x-y| = r\right\}.$$

Assume that $\kappa(r) : (0,\infty) \to \mathbb{R}$ is continuous, that $\int_0^1 r\kappa(r)^+dr < \infty$ and that

$$\limsup_{r\to\infty} \kappa(r) < 0. \tag{0.19}$$

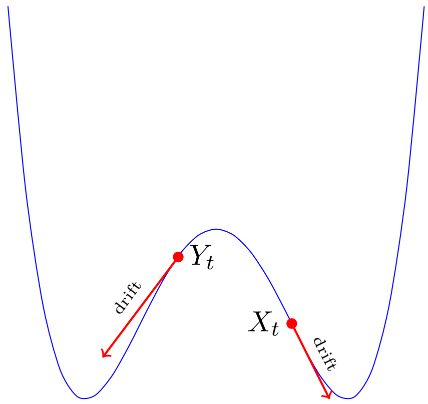

Figure 0.5: Diffusions inside a double-well potential

Then, there is $c \in (0,\infty)$ *such that for any* $t \geq 0$ *and any* $\mu, \nu \in \mathcal{P}(\mathcal{S})$*,*

$$\begin{aligned} \mathcal{W}_\rho(\mu p_t, \nu p_t) &\leq \mathrm{e}^{-ct}\,\mathcal{W}_\rho(\mu,\nu) \qquad \textit{and} \\ \mathcal{W}^1(\mu p_t, \nu p_t) &\leq 2\phi(R_0)^{-1}\,\mathrm{e}^{-ct}\; W^1(\mu,\nu). \end{aligned}$$

Here, $\rho(x,y) = f(|x-y|)$ *where* f *is a strictly increasing, continuous and concave function with* $f(0) = 0$*. The functions* f*,* ϕ *and the rate* c *are given by*

$$\begin{aligned} f(r) &= \int_0^r \phi(s)\, g(s)\, ds, \quad c^{-1} = \int_0^{R_1} \Phi(s)\,\phi(s)\, ds \\ \phi(r) &= \exp\left(-\frac{1}{4}\int_0^r s\,\kappa(s)^+\, ds\right), \quad \Phi(r) = \int_0^r \phi(s)\, ds \\ g(r) &= 1 - \frac{1}{2}\int_0^{r\wedge R_1} \frac{\Phi(s)}{\phi(s)}\, ds \Big/ \int_0^{R_1} \frac{\Phi(s)}{\phi(s)}\, ds\,, \end{aligned}$$

and the constant $R_0, R_1 \in (0,\infty)$ *are given by*

$$\begin{aligned} R_0 &= \inf\left\{R \geq 0 : \kappa(r) \leq 0 \quad \forall r \geq R\right\}, \\ R_1 &= \inf\left\{R \geq R_0 : \kappa(r) R (R - R_0) \leq -8 \quad \forall r \geq R\right\}. \end{aligned}$$

Notice that the definition of κ above differs from the definition given in [51] by a factor -1.

The result yields *explicit bounds* on the contraction rate c and these bounds turn out to be remarkable sharp in several situations, cf. [51, Lemma 2.9 and Remark 2.10] for precise statements. Notice that the condition (0.19) is satisfied if $b = -\nabla U$ for a C^2 function U which is strictly convex outside of a bounded set. In particular, double-well potentials (see Figure 0.5) are covered. For such potentials, the underlying deterministic system is locally noncontractive and thus a Kantorovich contraction can only be established by exploiting the noise. A reflection coupling is quite useful for this purpose, since by (0.18), the driving noise has a direct impact on $r_t = |X_t - Y_t|$. The idea leading to the above result is to use reflection coupling and then to carefully construct a concave function f, such that $f(|X_t - Y_t|)$ is contracting *on average*, i.e. such that

$$E_{(x_0,y_0)}[f(|X_t - Y_t|)] \quad \leq \quad \mathrm{e}^{-ct} f(|x_0 - y_0|) \qquad \text{for all } t \geq 0. \tag{0.20}$$

To this end, one may assume that f is an increasing and concave function with $f(0) = 0$ so that $(x, y) \mapsto f(|x - y|)$ is itself a distance. Assuming that $f \in C^2$, one can apply Itô's formula to $f(r_t)$ and conclude that, almost surely,

$$\begin{aligned} df(r_t) \quad &= \quad f'(r_t)\, r_t^{-1} \langle X_t - Y_t, b(X_t) - b(Y_t)\rangle \; dt \; + \; 2\, f''(r_t)\, dt \; + \; 2f'(r_t)\, dW_t \\ &\leq \quad (f'(r_t)\, \kappa(r_t)\, r_t + 2f''(r_t)\ )\ dt \; + \; 2f'(r_t)\, dW_t \qquad \text{for } t < T \end{aligned}$$

Thus, if f satisfies the inequality

$$f'(r)\, \kappa(r)\, r + 2f''(r) \quad \leq \quad -c\, f(r), \qquad \text{for } r \in (0, \infty),$$

one can conclude (0.20). This ordinary differential (in)equality is solved explicitly in [51] with a focus on maximizing the rate c. The resulting contraction is based on two arguments. For small distances, $\kappa(r)$ might be positive and thus one has to choose f sufficiently concave, so that the noise provides a contraction on average. For large distances, (0.19) implies that $\kappa(r)$ is strictly negative and hence the underlying deterministic system provides a contraction. This is reflected in choosing f to be linear for large distances.

Notice that the combination of concave distance functions and reflection coupling has been exploited by other authors before to obtain bounds on total variation distances, cf. e.g. [107, 33].

Finally, we remark that it is possible to construct "*hybrid couplings*" (X_t, Y_t) who behave in some regions of the state space as reflection couplings and in other regions like synchronous couplings. A rigoros way to define such couplings is given in [51, Section 6]: Fix $\delta > 0$ and let (B_t^1) and (B_t^2) be independent d-dimensional Brownian motions. Moreover, fix Lipschitz functions $\mathrm{rc}, \mathrm{sc} : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ satisfying $\mathrm{rc}^2 + \mathrm{sc}^2 = 1$ and $\mathrm{sc}(x, y) = 1$ for $|x - y| \leq \delta$. Let $u \in \mathbb{R}^d$ be an arbitrary unit vector. Then, the diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$ solving the SDE

$$\begin{aligned} dX_t \;&=\; b(X_t)\; dt \,+\, \mathrm{rc}\,(U_t)\;\; dB_t^1 &&+\; \mathrm{sc}\,(U_t)\;\; dB_t^2, \\ dY_t \;&=\; b(Y_t)\; dt \;+\; \mathrm{rc}\,(U_t)\;\; (\mathrm{Id}_{\mathbb{R}^d} - 2\, e_t\, \langle e_t, \cdot\rangle)\;\; dB_t^1 &&+\; \mathrm{sc}\,(U_t)\;\; dB_t^2, \end{aligned}$$

is a coupling of (0.21), where $U_t := (X_t, Y_t)$, $e_t := (X_t - Y_t)/\,|X_t - Y_t|$ for $|X_t - Y_t| > 0$ and $e_t = u$ for $|X_t - Y_t| = 0$. Notice that the concrete choice of u is not relevant for the dynamics, since $\mathrm{rc}(x, x) = 0$. In [51] such couplings are used to establish Kantorovich contractions for interacting particle systems. We demonstrate in this thesis that such mixtures of synchronous and reflection couplings can be applied in various situations to obtain explicit bounds on Kantorovich distances.

0.5 Outline: Quantitative Harris type theorems for diffusions

In the first main chapter, we discuss quantitative Harris type theorems for non-degenerate diffusions and McKean-Vlasov processes on $\mathbb{R}^d$. The results have been distributed prior as a research paper on the online-portal ArXiv:

A. Eberle, A. Guillin, and R. Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. *ArXiv e-print 1606.06012*, June 2016

The article is a joint work with Andreas Eberle (University of Bonn) and Arnaud Guillin (Université Blaise Pascal). Chapter 1 contains the article mostly as it has been distributed on ArXiv subject to minor modifications in formulations and formatting. One exception is, that the version presented here has an additional Section 1.6 which is not part of the original article and gives slight extensions of the main results. In this section, we give an outline of the main results with a focus on presenting ideas. Mathematical precise statements and comparisons with the literature are given in Chapter 1.

0.5.1 Diffusions

Let (B_t) be a d-dimensional Brownian motion. We consider diffusions of type

$$dX_t \;=\; b(X_t)\,dt \,+\, dB_t \tag{0.21}$$

with values in $\mathbb{R}^d$ and assume that the drift $b : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitz. We assume that non-explosiveness holds and denote the corresponding Markov transition function and generator by (p_t) and $\mathcal{L} = \langle b, \nabla\rangle + \frac{1}{2}\Delta$ respectively. The euclidean norm and inner product on $\mathbb{R}^d$ are called $|\cdot|$ and $\langle\cdot,\cdot\rangle$ respectively.

In Section 0.3, we have introduced classical Harris type theorems and, in Section 0.4.2, we have explained how one can use them to establish Kantorovich contractions for diffusions. However, as pointed out, those theorems are typically applied in a non-quantitative way since a quantification of the minorization condition is not trivial. Our aim here is to establish a more quantitative version of Harris' theorem for diffusion of type (0.21) which is based on two main assumptions: a geometric drift condition and a *one-sided* Lipschitz bound on b. We do *not* impose a minorization condition.

In Section 0.4.2, we have seen a recent result by Eberle who establishes Kantorovich contractions for diffusions of type (0.21) with explicit and in several cases sharp contraction rates, cf. [50, 51]. The approach is based on using reflection coupling for the diffusions and adapting the underlying cost function for the Kantorovich distance carefully to the chosen coupling and the diffusion. One of the main assumptions imposed in [50, 51] is the "contractivity at infinity condition" (0.19), which is satisfied if the drift has the form $b = -\nabla U$ for a C^2 function U which is strictly convex outside a compact set. One might ask if it is possible to replace this assumption by a more general recurrence condition. Besides the wish to establish a quantitative Harris theorem for diffusions, there are several reasons why this question is interesting. First of all, perturbations or approximations of a drift satisfying (0.19) do not necessarily inherit this property. One can therefore ask how stable the techniques from [50, 51] are. Secondly, in more complicated diffusion models the condition (0.19) is typically not satisfied, while a geometric drift condition often holds. We will see an example in the second part of this thesis, where related techniques are used to derive Kantorovich contractions for a class of infinite-dimensional and degenerate diffusions.

We now give an outline of the main results. Our main assumptions are:

Assumption 3 (Geometric drift condition). *There is a C^2 function $V : \mathbb{R}^d \to \mathbb{R}_+$ as well as constants $C, \lambda \in (0, \infty)$ such that $V(x) \to \infty$ as $|x| \to \infty$, and*

$$\mathcal{L}V(x) \;\leq\; C - \lambda V(x) \qquad \text{for any } x \in \mathbb{R}^d. \tag{0.22}$$

Assumption 4 (Generalized one-sided Lipschitz condition). *There is a continuous function $\kappa : (0, \infty) \to [0, \infty)$ such that $\int_0^1 r\,\kappa(r)\,dr \;<\; \infty$, and*

$$\langle x - y, b(x) - b(y)\rangle \;\leq\; \kappa(|x - y|) \cdot |x - y|^2 \qquad \text{for any } x, y \in \mathbb{R}^d.$$

Notice that for constant κ, this is just a one-sided Lipschitz bound. Given these conditions, we aim to establish a Kantorovich contraction for (p_t) with a contraction rate which can be computed given C, λ, V and κ. At the end of Section 0.2, we have seen that contraction inequalities can be established by finding reasonable combinations of couplings and distances.

Let us first think about the coupling strategy. We aim at using a "hybrid coupling" (X_t, Y_t), i.e. a coupling which behaves in some regions of the state space as a synchronous coupling and in some regions as a reflection coupling, cf. page 16. We describe how we want this coupling to behave in different regions of the state space. Roughly speaking, the geometric drift condition allows us to find a compact set S which is recurrent for the marginal processes (X_t) and (Y_t) of any such hybrid coupling. While $X_t \in S$ and $Y_t \in S$, we use a reflection coupling of the processes with the aim of driving X_t and Y_t together. If either $X_t \notin S$ or $Y_t \notin S$, we use a synchronous or reflection coupling depending on the application and the chosen distance. However, in many cases, the concrete coupling in this situation is not particularly relevant, since the recurrence property is a result of the drift and not the noise. The coupling approach is visualized in Figure 0.6.

Let us now think about distances. We first consider a distance of type

$$\rho_1(x, y) \;=\; [\, f(|x - y|) + \epsilon\, V(x) + \epsilon\, V(y)\,] \cdot I_{x \neq y}, \tag{0.23}$$

where f is a concave function, V the Lyapunov function and $\epsilon \in (0, \infty)$ a constant. The distance is partially motivated by Mattingly and Hairer's extension of Harris'

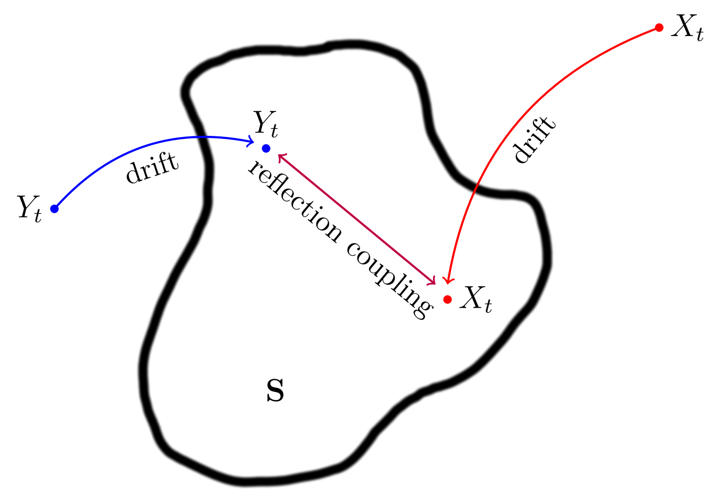

Figure 0.6: Coupling approach for a quantitative Harris' theorem

theorem [75]. The usage of the concave function is inspired by [50, 51]. Imposing a growth condition on the chosen Lyapunov function, our first result states that it is possible to choose f and ϵ in an explicit way such that

$$\mathcal{W}_{\rho_1}(\mu p_t, \nu p_t) \quad \leq \quad e^{-ct}\, \mathcal{W}_{\rho_1}(\mu, \nu) \tag{0.24}$$

holds for all $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $t \geq 0$, cf. Theorem 1 in Chapter 1. Moreover, we provide an explicit expression for the contraction rate c which can be quantified with little effort given λ, C and V from Assumption 3 and κ from Assumption 4. Consequences include exponential convergence towards the unique stationary distribution in weighted total variation distances (Corollary 1), exponential convergence in L^p Wasserstein distances (Remark 3) and quantifications of ergodic averages (Corollary 2). In the case of convex potentials, i.e. where we can choose $\kappa \equiv 0$ in Assumption 4, we obtain contraction rates with a polynomial dimension dependance, cf. Section 1.3.3 for precise statements. We also consider replacing the geometric drift condition by a subgeometric one. Using the distance ρ_1, we obtain in this case explicit bounds on the decay of $\|\delta_x p_t - \delta_y p_t\|_{\mathrm{TV}}$ for $t \to \infty$ (Theorem 5).

The "additive distance" ρ_1 is very simple and contractions w.r.t. $\mathcal{W}_{\rho_1}$ have many interesting consequences. However, the distance has the disadvantage that in general $\rho_1(x, y) \not\to 0$ as $x \to y$. Therefore, a contraction w.r.t. $\mathcal{W}_{\rho_1}$ can only be expected to hold if there is a coupling (X_t, Y_t) such that $P(X_t = Y_t) \to 1$ as $t \to \infty$. In the case of nondegenerate diffusions as in (0.21), it is not difficult to construct such a coupling. However, for degenerate, infinite-dimensional or nonlinear diffusions such couplings might either be difficult or even impossible to construct. Partially motivated by the weak Harris' theorem [69] by Mattingly, Hairer and Scheutzow, we also consider contractions w.r.t. a semimetric of type

$$\rho_2(x,y) \quad = \quad f(|x-y|)\,(1+\epsilon V(x)+\epsilon V(y))\,. \tag{0.25}$$

This distance allows to derive quantitative bounds for *asymptotic couplings* in the sense of [72, 115, 69], i.e., for couplings (X_t, Y_t) where X_t and Y_t get arbitrarily close to each other but do not necessarily meet in finite time. It is therefore also suited for applications in more complicated diffusion models and it is used within this thesis to derive contractions for McKean-Vlasov processes, as well as a class of infinite-dimensional and degenerate diffusions. The corresponding contraction result for diffusions of type (0.21) w.r.t. this semimetric is given in Theorem 3. Corollary 4 demonstrates that this contraction can be used to derive gradient bounds for the transition semigroup.

0.5.2 McKean-Vlasov processes

We also consider contractions for nonlinear diffusions satisfying an SDE of type

$$\begin{aligned} dX_t \quad &= \quad b(X_t)\,dt + \tau \int \vartheta(X_t,y)\,\mu_t^x(dy)\,dt \,+\, dB_t, \qquad X_0 \,=\, x, \\ \mu_t^x \quad &= \quad \mathrm{Law}(X_t). \end{aligned} \tag{0.26}$$

Here $\tau \in \mathbb{R}$ is a given (small) constant, $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\vartheta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ are Lipschitz functions, and (B_t) is a d-dimensional Brownian motion. Under the imposed assumptions, Equation (0.26) has a unique solution (X_t) which is a nonlinear Markov process in the sense of McKean, i.e., the future development after time t depends both on the current state X_t and on the law of X_t, cf. [144, 117]. Such SDEs arise naturally as marginal limits of mean field interacting particle systems

$$dX_t^i \;=\; b(X_t^i)\,dt \,+\, \frac{\tau}{n}\sum_{j=1}^n \vartheta(X_t^i, X_t^j)\,dt \,+\, dB_t^i, \qquad i=1,\dots n, \tag{0.27}$$

as $n \to \infty$. Here, the (B_t^i) are independent Brownian motions. For such interacting systems, assuming that $|\tau|$ is sufficiently small, that b satisfies condition (0.19) and that ϑ is Lipschitz, Eberle derived in [51] Kantorovich contractions w.r.t. an underlying distance of type $\rho(x,y) = \sum_{i=1}^n f(|x^i - y^i|)$, where f is a suitable concave function. The coupling approach leading to this contraction can be described as follows: One considers componentwise couplings (X_t^i, Y_t^i). Given a parameter $\delta > 0$, one uses a reflection coupling if $|X_t^i - Y_t^i| > \delta$ and a synchronous coupling otherwise. The contraction result is obtained when considering the limit of the resulting bounds as $\delta \downarrow 0$.

We wanted to understand this coupling approach for the interacting particle system in more detail and see whether a similar strategy can be applied directly to the nonlinear SDE to establish contractions for the corresponding laws. It turns out that

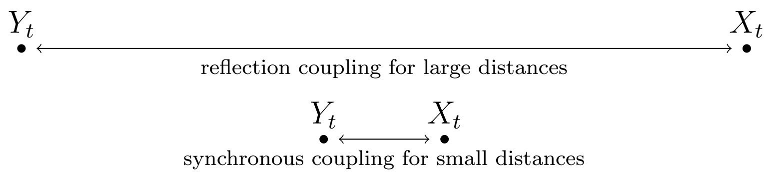

Figure 0.7: Coupling approach for nonlinear diffusions

this is indeed true (Theorem 3). Moreover, using the multiplicative semimetric ρ_2, we are able to relax assumption (0.19) on the drift b to a geometric drift condition and establish a (local) contraction result for (0.26) (Theorem 4). We are of course not the first to study convergence to equilibrium for such equations. References to existing results and comparisons are given in Section 1.2.3 and Section 1.3.5. Notice that our contraction results are obtained under the assumption that $|\tau|$ is sufficiently small. This assumption is natural, since for large τ, Equation (0.26) can have several distinct stationary solutions. Nevertheless, we do not claim that our bound on τ is sharp. The coupling approach for nonlinear diffusions considered here is extended and generalized in Chapter 3, where we construct *sticky couplings* for diffusions with different drifts.

0.6 Outline: Explicit contraction rates for a class of degenerate diffusions

In the second chapter, we establish Kantorovich contractions for a class of degenerate and infinite-dimensional diffusion processes. The results have been distributed prior as a research paper on the online-portal ArXiv:

R. Zimmer. Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. *ArXiv e-print 1605.07863*, May 2016

Chapter 2 contains the article mostly as it has been distributed on ArXiv subject to small changes. In this section, we give an outline of the main results with a focus on presenting ideas. Mathematical precise statements and comparisons with the literature are given in Chapter 2.

In the first chapter of this thesis, we derive Kantorovich contractions for diffusions of type (0.21) on a finite-dimensional state space. In the second chapter, we consider diffusions on a separable and real Hilbert space $(\mathbb{H}, |\cdot|)$. Let $\mathcal{G} : \mathbb{H} \to \mathbb{H}$ be a trace-class, symmetric and nonnegative operator. Such a bounded and linear operator can be "diagonalized", i.e. there is an orthonormal basis $(\mathbf{e}_k)_{k \in \mathbb{N}_+}$ of $\mathbb{H}$ and nonnegative real numbers (λ_k), such that $\mathcal{G}\mathbf{e}_k = \lambda_k \mathbf{e}_k$, see e.g. [132]. Moreover, the trace-class property implies that $\sum_{k=1}^{\infty} \lambda_k < \infty$. Denote by (W_t) a $\mathcal{G}$-Wiener process on $\mathbb{H}$, i.e.

let $W_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, B_t^k\, \mathbf{e}_k$ for independent Brownian motions (B_t^k). We are interested in the equation

$$dX_t = -X_t\, dt + b(X_t)\, dt + \sqrt{2}\, dW_t, \qquad X_0 = x \in \mathbb{H}, \tag{0.28}$$

where $b : \mathbb{H} \to \mathbb{H}$ is Lipschitz. In this setting, one can show (very similar as for finite-dimensional diffusions) that (0.28) admits a unique and strong solution (X_t), cf. e.g. [103].

The motivation for studying equation (0.28) is, that it has a natural appearance in the domain of sampling problems and acts as a diffusion limit for Markov chain Monte Carlo (MCMC) methods, see [71, 114, 38, 143, 70] and the references therein. In particular, if $U : \mathbb{H} \to \mathbb{R}_+$ is a smooth function, if $\mathcal{G}$ is positive definite and if we choose the nonlinearity $b(x) = -\mathcal{G}\nabla_{\mathbb{H}} U(x)$ in (0.28), then the results from [71] imply that the unique invariant probability measure π for the equation is determined by $\pi(dx) \propto \exp(-U(x))\, \mathcal{N}(0, \mathcal{G})(dx)$, where $\mathcal{N}(0, \mathcal{G})$ denotes a centered normal distribution on $\mathbb{H}$ with covariance operator $\mathcal{G}$. Such measures appear for example in the area of diffusion bridges, cf. [71].

From this point of view it is particularly interesting to obtain explicit bounds on the speed of convergence to equilibrium for solutions of (0.28) in the above setting. It is most likely possible to use the weak Harris' theorem [69] to obtain a Kantorovich contraction for the corresponding Markov transition functions, see in this context also [70]. However, it is not clear how to apply such a general framework to obtain explicit bounds. Instead, we use a more direct coupling approach and construct a simple and very explicit asymptotic coupling using a mixture of reflection and synchronous couplings. In order to establish a Kantorovich contraction, we adapt the underlying (semi)distance carefully to the chosen coupling and model following the strategy from [50, 51] and the first chapter of this thesis.

We discuss this approach in more detail. Inspired by the sampling setup, we work in the following setting: We consider a splitting of the Hilbert space $\mathbb{H}$ into a space $\mathbb{H}^l = \langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$, spanned by the first n basis vectors, and its orthogonal complement $\mathbb{H}^h$, i.e. $\mathbb{H} = \mathbb{H}^l \oplus \mathbb{H}^h$. Here, $n \in \mathbb{N}_+$ is some fixed number. We call $\mathbb{H}^l$ *low*-dimensional and $\mathbb{H}^h$ *high*-dimensional space. Given $x \in \mathbb{H}$, we denote by x^l and x^h the orthogonal projections onto $\mathbb{H}^l$ and $\mathbb{H}^h$ respectively.

Assumption 5. *There are constants* $0 \le H_h < 1$ *and* $L_l, L_h, H_l \ge 0$ *such that*

$$\begin{aligned} \left|b^h(x) - b^h(y)\right| &\le H_l \left|x^l - y^l\right| + H_h \left|x^h - y^h\right| \quad \text{and} && (0.29) \\ \left|b^l(x) - b^l(y)\right| &\le L_l \left|x^l - y^l\right| + L_h \left|x^h - y^h\right| \quad \text{for any } x, y \in \mathbb{H}. && (0.30) \end{aligned}$$

Assumption 6. $\mathcal{G}$ *is strictly positive definite on* $\mathbb{H}^l$*, i.e. for any* $k \in \mathbb{N}$ *with* $1 \le k \le n$*, we have* $\lambda_k > 0$.

In the sampling setup described above, assuming that the map $x \mapsto \nabla U(x)$ is Lipschitz on $\mathbb{H}$, it is always possible to find a splitting $\mathbb{H} = \mathbb{H}^l \oplus \mathbb{H}^h$ such that

Assumptions 5 and 6 are satisfied. In addition to the above conditions, we assume a geometric drift condition. Based on these assumptions, we derive quantitative Kantorovich contractions for the associated Markov transition functions (Theorem 9). The resulting contraction rates are given explicitly in terms of the eigenvalues of $\mathcal{G}$ on the space $\mathbb{H}^l$, the constants from Assumption 5 and a geometric drift condition. In comparison to the first chapter, the main difficulty here is that the driving noise is possibly degenerate on the space $\mathbb{H}^h$.

Let us briefly explain the coupling strategy leading to the contraction result: Assume for a moment that (X_t, Y_t) is a *synchronous coupling* of solutions to (0.28), i.e. let the processes (X_t) and (Y_t) be driven by the same noise. We argue pathwise. Assume that $X_t - Y_t$ satisfies for some $t \geq 0$ the inequality

$$H_l \left|X_t^l - Y_t^l\right| \;\leq\; (1 - H_h) \left|X_t^h - Y_t^h\right| /2, \tag{0.31}$$

then Assumption 5 implies that

$$\left|b^h(X_t) - b^h(Y_t)\right| \;\leq\; H_l \left|X_t^l - Y_t^l\right| \;+\; H_h \left|X_t^h - Y_t^h\right| \;\leq\; (1 + H_h) \left|X_t^h - Y_t^h\right| /2,$$

where $(1 + H_h)/2 < 1$ by assumption. In particular, as long as $X_t - Y_t$ satisfies (0.31), $\left|X_t^h - Y_t^h\right|$ decreases exponentially fast. At some point, as time increases, $X_t - Y_t$ might not satisfy (0.31) any more. Then, we use a reflection coupling of X_t^l and Y_t^l in the space $\mathbb{H}^l$ (on which the noise is nondegenerate by Assumption 6) with the aim of decreasing $\left|X_t^l - Y_t^l\right|$ relative to $\left|X_t^h - Y_t^h\right|$. Iterating these two arguments, it is possible to construct an asymptotic coupling (X_t, Y_t) such that, for almost every trajectory, $X_t - Y_t \to 0$ for $t \to \infty$. The coupling approach is visualized in Figure 2.1 on page 90. On an abstract level, the strategy can be summarized as follows: We identify regions of the state space, where the underlying deterministic system of (0.28) admits a contraction property and then, we use the noise to position the coupling inside of those regions. Up to the author's knowledge, asymptotic couplings have been used first by Mattingly and Hairer [72, 115, 65, 67, 68] in settings related to the stochastic 2D Navier-Stokes equation to prove exponential mixing for degenerate systems. The strategy of splitting the underlying Hilbert space into a finite-dimensional space $\mathbb{H}^u$ of "unstable modes", where the dynamics is forced directly with noise, and an infinite-dimensional complement $\mathbb{H}^s$ of "stable modes", where the driving noise can be degenerate, also occurs in the literature regarding the stochastic 2D Navier-Stokes equation, cf. [157, 115, 113, 100, 99, 158, 17, 16] and also [72]. In comparison to those results, we use a more direct and very explicit coupling approach leading to explicit bounds on the speed of convergence. A more comprehensive discussion of the existing literature is given in Chapter 2.

The above coupling is combined with an underlying semimetric of type

$$\rho(x, y) = f\left(\left|x^l - y^l\right| + \alpha \left|x^h - y^h\right|\right) (1 + \epsilon V(x) + \epsilon V(y)),$$

to establish a Kantorovich contraction. As before, f is a concave function, $\epsilon \in (0, \infty)$ and V is a chosen Lyapunov function. The constant $\alpha \in [1, \infty)$ is used to put additional weight on the components in the space $\mathbb{H}^h$. This enables us to exploit the contraction property provided by Assumption 5.

In Section 2.3.2, we argue that a similar strategy can also be applied for finite-dimensional diffusion approximations of (0.28) yielding contractions with a *dimension-independent* and explicit rate.

0.7 Outline: Sticky couplings of diffusions

In the third chapter, we introduce a novel approach of coupling multidimensional diffusions with different drifts. The results have been distributed prior as a research article on the online-portal ArXiv:

> A. Eberle and R. Zimmer. Sticky couplings of multidimensional diffusions with different drifts. *ArXiv e-print 1612.06125*, December 2016

The article is a joint work with Andreas Eberle (University of Bonn). Chapter 3 contains the article mostly as it has been distributed on ArXiv subject to minor modifications in formulations and formatting. In this section, we give a brief outlook of the main results with a focus on motivation and presenting ideas in form of a collage. Mathematical precise statements and more comprehensive comparisons with the literature are given in Chapter 3.

Let (B_t) and $(\tilde{B}_t)$ be d-dimensional Brownian motions. We consider two diffusion processes with values in $\mathbb{R}^d$ which follow dynamics with different drifts, i.e.

$$dX_t = b(t, X_t)\,dt + dB_t, \qquad X_0 = x, \tag{0.32}$$

$$dY_t = \tilde{b}(t, Y_t)\,dt + d\tilde{B}_t, \qquad Y_0 = y. \tag{0.33}$$

We assume that the drift coefficients $b, \tilde{b} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ are locally Lipschitz. Moreover, we impose assumptions which imply that a geometric Lyapunov drift condition holds for (0.32) and that there is a constant $M > 0$ such that uniformly $|b - \tilde{b}| \le M$.

Diffusions with different drifts occur in many application areas. For example, one can consider a Langevin diffusion (X_t) and a perturbation or approximation (Y_t) of the latter. Other natural examples are McKean-Vlasov processes, as introduced in Section 0.5.2 further above. A natural question arising is how to obtain explicit bounds for the distance of X_t and Y_t in Kantorovich distances, e.g. in total variation norm. There are a few articles which try to answer this question in a general setting: Using Girsanov's theorem and coupling on the path space, the works [92, 104, 105] establish bounds on the total variation norm of such diffusions. In [12] bounds for the distance between transition probabilities of diffusions with different drifts are derived using analytic arguments, see also the related work [112]. The drawback of these approaches is that the derived bounds are typically only useful for small time horizons and are not longtime stable. The article [9] provides bounds for the distance between stationary measures of diffusions with different drifts. Coupling methods are used in [46] to provide longtime stable bounds on the distance between a Langevin diffusion and its Euler approximation. Howitt constructs in [84] a *sticky coupling* of two one-dimensional Brownian motions with different drifts using time-change arguments which are restricted to the one-dimensional setting.

In Chapter 3, we discuss a novel approach of constructing couplings (X_t, Y_t) of solutions to (0.32) and (0.33) in a multidimensional setting. Let us point out the difficulty of this problem: Consider the case where $\tilde{b}$ differs from b by a non-zero constant m, i.e., $\tilde{b}(t,x) = b(t,x) + m$ for some $m \in \mathbb{R}^d$, and let (X_t) and (Y_t) be solutions of (0.32) and (0.33) respectively. In this case, whenever X_t and Y_t meet, the drift forces the processes to immediately move apart from each other. It is clear that, regardless of how the processes are coupled, one cannot hope for the existence of an almost surely finite stopping time T such that $P[X_t = Y_t \ \forall t \geq T] = 1$. Nevertheless, we construct a coupling such that for any given $t > 0$, we have $P[X_t = Y_t] > 0$ and the coupling is *sticky* in the sense that there is a continuous semimartingale (r_t) which solves a one-dimensional stochastic differential equation with a *sticky boundary* behavior at zero such that almost surely $|X_t - Y_t| \leq r_t$ for all $t \geq 0$. This allows us to establish explicit, non-asymptotic and longtime stable bounds for the probability of the event $\{X_t = Y_t\}$, cf. Theorem 10. A visualization of a sticky coupling for one-dimensional diffusions is presented in Figure 3.1 on page 109.

Let us describe the setting in more detail. We assume that the drift coefficients $b, \tilde{b} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ are locally Lipschitz and impose the following conditions:

Assumption 7. *There is a constant $M \in [0,\infty)$ such that*

$$\left| b(t,x) - \tilde{b}(t,x) \right| \leq M \qquad \text{for any } x \in \mathbb{R}^d \text{ and } t \geq 0.$$

Assumption 8. *There is a Lipschitz function $\kappa : [0,\infty) \to \mathbb{R}$ such that*

$$\langle x - y, b(t,x) - b(t,y) \rangle \leq \kappa(|x-y|) \cdot |x-y|^2 \quad \text{for any } x, y \in \mathbb{R}^d \text{ and } t \geq 0.$$

Outside of a bounded interval, the function κ is constant and strictly negative.

Our sticky coupling generalizes the coupling approach for interacting particle systems in [51] and for McKean-Vlasov processes in Chapter 1 respectively. This approach is outlined in Section 0.5.2 further above. The idea is to construct a family of couplings (X_t^δ, Y_t^δ), $\delta > 0$, where reflection coupling is applied for $\left|X_t^\delta - Y_t^\delta\right| > \delta$ and a synchronous coupling for $X_t^\delta = Y_t^\delta$. One can define such a "hybrid coupling" similar to the one described at the end of Section 0.4.2 on page 16. Moreover, we construct for any $\delta > 0$ a one-dimensional diffusion (r_t^δ) on $\mathbb{R}_+$ such that $\left|X_t^\delta - Y_t^\delta\right| \leq r_t^\delta$ holds almost surely for all $t \geq 0$. We argue, that the sequence $(X_t^\delta, Y_t^\delta, r_t^\delta)$ converges weakly (along a subsequence) towards a limiting process (X_t, Y_t, r_t) as $\delta \downarrow 0$ and that $|X_t - Y_t| \leq r_t$ holds almost surely for all $t \geq 0$. It turns out that the process (r_t) is a solution of the equation

$$dr_t \;=\; (M \;+\; \kappa(r_t)\, r_t)\; dt \;+\; 2\, I(r_t > 0)\; dW_t, \qquad r_0 \;=\; |x-y|\,, \quad (0.34)$$

on $\mathbb{R}_+$, where (W_t) is a one-dimensional Brownian motion, and M and κ are given by Assumption 7 and Assumption 8 respectively (Theorem 13). If $M = 0$, then zero is an absorbing boundary for the diffusion process (r_t). In the case $M > 0$, it is a sticky reflecting boundary. Let us explain this in more detail: Suppose that (r_t) is a solution of (0.34) with $M > 0$. An application of the Itô-Tanaka formula to $f(r_t)$ with the function $f(x) = \max(0, x)$ and a comparison with (0.34) shows that almost surely,

$$\int_0^t M\ I(r_s = 0)\ ds \ = \ \frac{1}{2}\ell_t^0(r), \qquad 0 \ \le \ t \ < \ \infty, \tag{0.35}$$

where $\ell_t^0(r) = \lim_{\epsilon \downarrow 0} \epsilon^{-1} \int_0^t I(0 \le r_s \le \epsilon)\, d\, [r]_s$ is the right local time of (r_t). Equation (0.35) shows that there is *reflection* at zero. Moreover, for almost all trajectories, the Lebesgue measure of the set $\{0 \le s \le t : r_s = 0\}$ increases whenever $\ell_t^0(r)$ increases. In this sense (r_t) is *sticky* at zero, cf. [54] for a similar argument. The discovery of a sticky boundary behavior for one-dimensional diffusions seems to go back to Feller [56, 57]. An overview over the literature regarding sticky processes is given in Section 3.3. In order to obtain bounds on the probability of the event $\{X_t = Y_t\}$ for the limiting coupling, we argue that equation (0.34) admits an invariant measure π which puts positive mass on the point zero, i.e. $\pi(\{0\}) > 0$ (Lemma 23). We prove that there exists a synchronous coupling for equation (0.34) (Theorem 13) and use this coupling to provide explicit and non-asymptotic bounds on the speed of convergence to equilibrium for the process (r_t) (Theorem 14). The above outline gives only a very rough sketch of the arguments. The precise arguments are given in Chapter 3, where we also provide references to the works of other authors. Finally, we remark that it is possible to use sticky couplings to extend the result for McKean-Vlasov processes from Chapter 1 (Theorem 3) and prove an exponential decay of the corresponding laws in total variation norm (Theorem 12).

1 Quantitative Harris type theorems for diffusions and McKean-Vlasov processes

We consider $\mathbb{R}^d$ valued diffusion processes of type

$$dX_t \;=\; b(X_t)\,dt \,+\, dB_t.$$

Assuming a geometric drift condition, we establish contractions of the transition kernels in Kantorovich (L^1 Wasserstein) distances with explicit constants. Our results are in the spirit of Hairer and Mattingly's extension of Harris' Theorem. In particular, they do not rely on a small set condition. Instead we combine Lyapunov functions with reflection coupling and concave distance functions. We retrieve constants that are explicit in parameters which can be computed with little effort from one-sided Lipschitz conditions for the drift coefficient and the growth of a chosen Lyapunov function. Consequences include exponential convergence in weighted total variation norms, gradient bounds, bounds for ergodic averages, and Kantorovich contractions for nonlinear McKean-Vlasov diffusions in the case of sufficiently weak but not necessarily bounded nonlinearities. We also establish quantitative bounds for subgeometric ergodicity assuming a subgeometric drift condition.

A. Eberle, A. Guillin, and R. Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. *ArXiv e-print 1606.06012*, June 2016

Financial support from DAAD and French government through the PROCOPE program, and from the German Science foundation through the *Hausdorff Center for Mathematics* is gratefully acknowledged.

1.1 Introduction

We consider $\mathbb{R}^d$ valued diffusion processes of type

$$dX_t \;=\; b(X_t)\,dt \,+\, dB_t, \tag{1.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitz, and (B_t) is a d-dimensional Brownian motion. We assume non-explosiveness, and we denote the transition function of the corresponding Markov process by (p_t).

The classical Harris' Theorem [77, 118] gives simple conditions for geometric ergodicity of Markov processes. In the case of diffusion processes on $\mathbb{R}^d$ it goes back to Khasminskii [79, 96], in the general case it has been developed systematically by Meyn and Tweedie [119, 120, 118]. For solutions of (1.1), it is often not difficult to verify the assumptions in Harris' Theorem, a minorization condition for the transition probabilities on a bounded set, and a global Lyapunov type drift condition. However, it is not at all easy to quantify the constants in Harris' Theorem, and, even worse, the resulting bounds are far from sharp, and they usually have a very bad dimensional dependence. Therefore, although Harris' Theorem has become a standard tool in many application areas, it is mostly used in a purely qualitative way, a noteworthy exception being Roberts and Rosenthal [135].

Besides the Harris' approach, there is a standard approach for studying mixing properties of Markov processes based on spectral gaps, logarithmic Sobolev inequalities, and more general functional inequalities, see for example the monograph [6]. This approach has the advantage of providing sharp bounds in simple model cases but it sometimes yields slightly weaker, and less probabilistically intuitive results. Recent attempts [5, 4] to connect these functional inequalities to Lyapunov conditions have been proven successful but they are restricted to the reversible setting (or the explicit knowledge of the invariant measure). The concept of the intrinsic curvature of a diffusion process in the sense of Bakry-Emery leads to sharp bounds and many powerful results in the case where there is a strictly positive lower curvature bound κ [150]. In our context, this means that $\partial b(x) \leq -\kappa \, I_d$ for all x in the sense of quadratic forms.

Several of the bounds in the positive curvature case can be derived in an elegant probabilistic way by considering synchronous couplings and contraction properties in L^2 Wasserstein distances. In general, Wasserstein distances have been proven crucial in the study of linear and nonlinear diffusions both via coupling techniques [27, 109, 22], or via analytic techniques based on profound results on optimal transportation, see [20, 21, 149, 13, 14] and references therein. In the case where the curvature is only strictly positive outside of a compact set, reflection coupling has been applied successfully to obtain total variation bounds for the distance to equilibrium [107] as well as explicit contraction rates of the transition semigroup in Kantorovich distances [50, 51].

An important question is how to apply a Harris' type approach in order to obtain explicit bounds that are close to sharp in certain contexts. A breakthrough towards the applicability to high- and infinite dimensional models has been made by Hairer and Mattingly in [67], and in the subsequent publications [69, 75]. The key idea was to replace the underlying couplings with finite coupling time by asymptotic couplings where the coupled processes only approach each other as $t \to \infty$ [72, 115], and the minorization condition by a contraction in an appropriately chosen Kantorovich distance. In recent years, the resulting weak Harris' theorem has been applied successfully to prove (sub)geometric ergodicity in infinite dimensional models, see e.g. [19], and to quantify the dimension dependence in high dimensional problems [70]. Nevertheless, in contrast to the approach based on functional inequalities, the constants in applications of the weak Harris theorem are usually far from optimal.

This is in particular due to the fact that the corresponding Kantorovich distance is still chosen in a somehow ad hoc way.

It turns out that a key for making the bounds more quantitative is to adapt the underlying metric on $\mathbb{R}^d$ and the corresponding Kantorovich metric on the space of probability measures in a very specific way to the problem under consideration. For diffusion processes solving (1.1), this approach has been discussed in [51] assuming strict contractivity for the corresponding deterministic dynamics outside a ball. Our goal here is to replace this "contractivity at infinity" condition by a Lyapunov condition, thus providing a more specific quantitative version of the (weak) Harris' theorem. Indeed, we will define explicit metrics on $\mathbb{R}^d$ depending both on the drift coefficient b and the Lyapunov function V such that the transition semigroup is contractive with an explicit contraction rate c w.r.t. the corresponding Kantorovich distances. Such a contraction property has remarkable consequences including not only a non-asymptotic quantification of the distance to equilibrium, but also non-asymptotic bounds for ergodic averages, gradient bounds for the transition semigroup, stability under perturbations etc.

Outline: In Section 1.2, we present our main results. Section 1.3 contains a discussion of the results including more detailed comparisons to the existing literature. The couplings considered here are introduced in Section 1.4 and the proofs of our results are given in Section 1.5. Extensions of the results from Section 1.2 to a more general setup are discussed in Section 1.6.

1.2 Main results

Let $\langle\cdot,\cdot\rangle$ and $|\cdot|$, respectively, denote the euclidean inner product and the corresponding norm on $\mathbb{R}^d$. We assume a generalization of a global one-sided Lipschitz condition for b combined with a Lyapunov condition. These assumptions can be weakened, cf. Section 1.6 further below for an extension of the results to a more general setup.

Assumption 9 (Generalized one-sided Lipschitz condition)**.** *There is a continuous function $\kappa : (0,\infty) \to [0,\infty)$ such that $\int_0^1 r\,\kappa(r)\,dr < \infty$, and*

$$\langle x-y, b(x)-b(y)\rangle \;\le\; \kappa(|x-y|)\cdot|x-y|^2 \qquad \text{for any } x,y\in\mathbb{R}^d. \qquad (1.2)$$

Notice that for constant κ, (1.2) is a one-sided Lipschitz condition. In particular, if $b = -\nabla U$ for some function $U \in C^2(\mathbb{R}^d)$, then the assumption with constant κ is equivalent to a global lower bound on the Hessian of U. If U is strictly convex outside a ball in $\mathbb{R}^d$, then we can choose $\kappa(r) = 0$ in (1.2) for sufficiently large r. Let $\mathcal{L} = \frac{1}{2}\Delta + \langle b(x),\nabla\rangle$ denote the generator of the diffusion process.

Assumption 10 (Geometric drift condition)**.** *There is a C^2 function $V : \mathbb{R}^d \to \mathbb{R}_+$ as well as constants $C,\lambda\in(0,\infty)$ such that $V(x)\to\infty$ as $|x|\to\infty$, and*

$$\mathcal{L}V(x) \;\le\; C-\lambda V(x) \qquad \text{for any } x\in\mathbb{R}^d. \qquad (1.3)$$

It is well-known that Assumption 10 implies the non-explosiveness of solutions for (1.1), see e.g. [96, 120]. The function V in (1.3) is called a *Lyapunov function.*

Remark 1 (Choice of Lyapunov functions)**.** *Assume there are $\mathcal{R} > 0$, $\gamma > 0$ and $q \geq 1$ such that*

$$\langle b(x), x\rangle \quad \leq \quad -\gamma\, |x|^q \qquad \text{for any } |x| \geq \mathcal{R}.$$

Then Lyapunov functions of the following form can be chosen depending on the values of q and γ:

- *Let $\alpha > 0$. If V is a C^2 function with $V(x) = \exp(\alpha\, |x|^q)$ outside of a compact set, then (1.3) holds for arbitrary $\lambda > 0$ with a finite constant $C(\alpha, \lambda)$ provided $q > 1$ and $\alpha \in (0, 2\gamma/q)$, or $q = 1$ and $\gamma > \frac{\alpha}{2} + \frac{\lambda}{\alpha}$.*

- *Let $\alpha > 0$ and $p \in [1, q)$. If V is C^2 with $V(x) = \exp(\alpha\, |x|^p)$ outside of a compact set, then (1.3) holds for arbitrary $\lambda > 0$ with a finite constant $C(\alpha, \lambda)$.*

- *Let $q \geq 2$ and $p > 0$. If V is C^2 with $V(x) = |x|^p$ outside of a compact set, then (1.3) holds with a finite constant $C(\lambda, p)$ if $q > 2$ and $\lambda > 0$, or if $q = 2$ and $\lambda \in (0, p\gamma)$.*

Besides the two key assumptions made above, we will need a growth assumption on the Lyapunov function, cf. Assumption 11 below for our first main result, or Assumption 12 below for our second main result.

The Kantorovich distance of two probability measures μ and ν on a metric space $(\mathcal{S}, \rho)$ is defined by

$$\mathcal{W}_\rho(\nu, \mu) \quad = \quad \inf_{\gamma \in C(\nu,\mu)} \int \rho(x, y)\, \gamma(dx\, dy),$$

where the infimum is taken over all couplings with marginals ν and μ respectively. $\mathcal{W}_\rho(\nu, \mu)$ can be interpreted as the L^1 transportation cost between the probability measures ν and μ w.r.t. the underlying cost function $\rho(x, y)$. As such, it is also well-defined if ρ is only a semimetric, i.e., a function on $\mathcal{S} \times \mathcal{S}$ that is symmetric and nonnegative with $\rho(x, y) > 0$ for $x \neq y$ but that does not necessarily satisfy the triangle inequality. In Subsections 1.2.1 and 1.2.2, we derive contractions of the transition semigroup with respect to $\mathcal{W}_\rho$ based on two different types of underlying cost functions ρ. The first one, called the “additive distance”, is a metric, whereas the second one, called the “multiplicative distance”, in general is only a semimetric. We then consider a variation of our approach that applies to McKean-Vlasov diffusions, cf. Subsection 1.2.3. Subsection 1.2.4 discusses replacing the geometric by a subgeometric Lyapunov condition.

1.2.1 Geometric ergodicity with explicit constants: First main result

We first consider an underlying distance of the form

$$\rho_1(x,y) \;=\; [\, f(|x-y|) + \epsilon\, V(x) + \epsilon\, V(y)\,] \cdot I_{x\neq y} \tag{1.4}$$

where f is a suitable bounded, non-decreasing and concave continuous function satisfying $f(0) = 0$, and $\epsilon \in (0,\infty)$ is a positive constant. The choice of a distance is partially motivated by [75], where an underlying metric of the form $(x,y) \mapsto [2 + \epsilon V(x) + \epsilon V(y)]\, I_{x\neq y}$ is considered in order to retrieve a Kantorovich contraction based on a small set condition. We define functions $\phi, \Phi : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\phi(r) \;=\; \exp\left(-\frac{1}{2}\int_0^r t\,\kappa(t)\,dt\right) \qquad \text{and} \qquad \Phi(r) \;=\; \int_0^r \phi(t)\,dt \tag{1.5}$$

with κ as in Assumption 9. For constant κ, we have $\phi(r) = \exp(-\kappa r^2/4)$. We will choose the function f to be constant outside a finite interval $[0, R_2]$ where R_2 is defined in (1.10) below. Inside the interval, the function f satisfies

$$\frac{1}{2}\Phi(r) \;\le\; f(r) \;\le\; \Phi(r).$$

We consider a set S_1 which is recurrent for any Markovian coupling (X_t, Y_t) of solutions of (1.1):

$$S_1 \;=\left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : V(x) + V(y) \le 4C/\lambda\right\}. \tag{1.6}$$

The set S_1 is chosen such that for $(x,y) \in \mathbb{R}^{2d} \setminus S_1$,

$$\mathcal{L}V(x) + \mathcal{L}V(y) \;\le\; -\frac{\lambda}{2}\,(\,V(x) + V(y)\,)\,.$$

Here the factor $1/2$ is, to some extent, an arbitrary choice. The "diameter"

$$R_1 \;=\; \sup\{|x-y| : (x,y) \in S_1\} \tag{1.7}$$

of the set S_1 determines our choice of ϵ in (1.4):

$$\epsilon^{-1} \;=\; 4C\int_0^{R_1} \phi(r)^{-1}\,dr \;=\; 4C\int_0^{R_1} \exp\left(\frac{1}{2}\int_0^r t\,\kappa(t)\,dt\right) dr. \tag{1.8}$$

Notice that R_1 is always finite since $V(x) \to \infty$ as $|x| \to \infty$. An upper bound is given by

$$R_1 \;\le\; 2\,\sup\{\,|x| : V(x) \le 4C/\lambda\}.$$

We now state our third key assumption that links κ and V:

Assumption 11 (Growth condition)**.** *There exist a constant $\alpha > 0$ and a bounded set $S_2 \supseteq S_1$ such that for any $(x, y) \in \mathbb{R}^{2d} \setminus S_2$, we have*

$$V(x) + V(y) \;\geq\; \frac{4C}{\lambda}\left(1 + \alpha \int_0^{R_1} \phi(r)^{-1}\, dr\; \Phi(|x-y|)\right). \tag{1.9}$$

Assumptions linking curvature and Lyapunov functions already appeared in [24] to prove a logarithmic Sobolev inequality in the reversible setting for the case where the curvature may explode (polynomially). Similarly to R_1, we define

$$R_2 \;=\; \sup\{|x-y| : (x,y) \in S_2\}\,. \tag{1.10}$$

Notice that Φ grows at most linearly. If one chooses $\alpha^{-1} = \int_0^{R_1} \phi(r)^{-1}\, dr$, then Condition (1.9) takes the simple form

$$V(x) + V(y) \;\;\geq\;\; \frac{4C}{\lambda}\left(1 + \int_0^{|x-y|} \exp\left(-\frac{1}{2}\int_0^r t\,\kappa(t)\, dt\right) dr\right).$$

Lemma 1. *If there exists a finite constant $\mathcal{R} \geq R_1$ such that*

$$V(x) \;\;\geq\;\; 4C\lambda^{-1}\,(1 + 2\,|x|) \qquad \textit{for any } x \in \mathbb{R}^d \textit{ with } |x| \geq \mathcal{R},$$

then Assumption 11 is satisfied with $\alpha^{-1} = \int_0^{R_1} \phi(r)^{-1}\, dr$ and

$$S_2 \;\;=\;\; \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : \max(|x|\,, |y|\} < \mathcal{R}).$$

The proofs of Lemma 1 and the subsequent results in Section 1.2.1 are given in Section 1.5.1 below. Let $\mathcal{P}_V(\mathbb{R}^d)$ denote the set of all probability measures μ on $\mathbb{R}^d$ such that $\int V\, d\mu < \infty$. We can now state our first main result:

Theorem 1 (Contraction rates for additive metric)**.** *Suppose that Assumptions 9, 10 and 11 hold true. Then there exist a concave, bounded and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and constants $c, \epsilon \in (0, \infty)$ s.t.*

$$\mathcal{W}_{\rho_1}(\mu p_t, \nu p_t) \;\;\leq\;\; e^{-ct}\, \mathcal{W}_{\rho_1}(\mu, \nu) \quad \textit{for any } \mu, \nu \in \mathcal{P}(\mathbb{R}^d) \textit{ and } t \geq 0. \tag{1.11}$$

Here the underlying distance ρ_1 is defined by (1.4) with ϵ determined by (1.8), and $c = \min(\beta, \alpha, \lambda)/2$ where

$$\beta^{-1} \;\;=\;\; \int_0^{R_2} \Phi(r)\,\phi(r)^{-1} ds = \int_0^{R_2}\int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa(u)\, du\right) dr\, ds. \tag{1.12}$$

The function f is constant for $r \geq R_2$, and

$$\frac{1}{2} \;\;\leq\;\; f'(r)\,\exp\left(\frac{1}{2}\int_0^r t\,\kappa(t)\, dt\right) \;\;\leq\;\; 1 \qquad \textit{for any } r \in (0, R_2).$$

The precise definition of the function f is given in the proof in Section 1.5.1.

Example 4 (Bounds under global one-sided Lipschitz condition)**.** *Suppose that there is a* constant $\kappa \geq 0$ *such that for any* $x, y \in \mathbb{R}^d$*, we have*

$$\langle x-y, b(x)-b(y)\rangle \;\leq\; \kappa\,|x-y|^2 .$$

Then we can state our result in a simplified form. Suppose that Assumption 10 holds, and there is a bounded set $S_2 \supseteq S_1$ *such that for any* $(x,y) \notin S_2$*,*

$$V(x)+V(y) \;\geq\; \frac{4C}{\lambda}\left(1+\int_0^{|x-y|}\exp\left(-\kappa r^2/4\right)dr\right).$$

Then (1.11) *holds with* $c = \min\left(2R_2^{-2}, R_1^{-1}, \lambda\right)/2$ *for* $\kappa = 0$*, and*

$$c = \frac{1}{2}\min\left(\sqrt{\frac{\kappa}{\pi}}\left(\int_0^{R_2}\exp\left(\kappa r^2/4\right)\,dr\right)^{-1}, \left(\int_0^{R_1}\exp\left(\kappa r^2/4\right)dr\right)^{-1}, \lambda\right) \tag{1.13}$$

for $\kappa > 0$*. Here* R_1 *and* R_2 *are defined as above, and the underlying distance* ρ_1 *is given by* (1.4) *with* $\epsilon^{-1} = 4C\int_0^{R_1}\exp\left(\kappa s^2/4\right)ds$ *and a concave, bounded and increasing continuous function* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *satisfying* $f(0) = 0$ *and*

$$1/2 \;\leq\; f'(r)\exp\left(\kappa r^2/4\right) \;\leq\; 1 \qquad \text{for } 0 < r < R_2.$$

Remark 2. *For* $\kappa = 0$ *and, more generally, for* $\kappa R_2^2 = O(1)$*, the lower bound* c *for the contraction rate in the example is of the optimal order* $\Omega(\min(R_2^{-2}, \lambda))$*. In general, under the assumptions made above, the bound on the contraction rate given by* (1.13) *is of optimal order in* λ*, and of optimal order in* R_2 *up to polynomial factors, see the discussion below Lemma 1 in [51].*

It is well-known, see e.g. [73], that the local Lipschitz assumption on b and Assumption 10 imply that (p_t) has a unique invariant measure $\pi \in \mathcal{P}_V(\mathbb{R}^d)$ satisfying $\int V\,d\pi \leq C/\lambda$. A result from [75, Lemma 2.1] then shows that the Kantorovich contraction in Theorem 1 implies bounds for the distance of μp_t and π in a weighted total variation norm.

Corollary 1 (Exponential Convergence in Weighted Total Variation Norm)**.** *Under the assumptions of Theorem 1, there exists a unique stationary distribution* $\pi \in \mathcal{P}_V(\mathbb{R}^d)$*, and*

$$\int_{\mathbb{R}^d} V\,d\,|\mu p_t - \pi| \;\leq\; \epsilon^{-1}\exp(-c\,t)\,\mathcal{W}_{\rho_1}(\mu,\pi) \quad \text{for any } \mu \in \mathcal{P}(\mathbb{R}^d).$$

In particular, for any $\delta > 0$ *and* $x \in \mathbb{R}^d$*, the mixing time*

$$\tau(\delta, x) \;=\; \inf\{t \geq 0 \,:\, \int_{\mathbb{R}^d} V\,d\,|p_t(x,\bullet) - \pi| < \delta\,\}$$

is bounded from above by

$$\tau(\delta,x) \;\leq\; c^{-1}\log^+\left[\frac{R_2\,\epsilon^{-1} + V(x) + C/\lambda}{\delta}\right].$$

Remark 3 (Exponential Convergence in L^p Wasserstein distances). *For $p \in [1,\infty)$, the standard L^p Wasserstein distance $\mathcal{W}^p$ can be controlled by a weighted total variation norm:*

$$\mathcal{W}^p(\mu,\nu) \;\leq\; 2^{1/q} \left(\int |x|^p \; |\mu-\nu| \, (dx) \right)^{1/p},$$

where $1/q + 1/p = 1$, see e.g. [149, Theorem 6.15]. Thus if there is a constant $K > 0$ such that $|x|^p \leq K\, V(x)$ holds for all $x \in \mathbb{R}^d$, then Corollary 1 also implies exponential convergence in L^p Wasserstein distance.

Following ideas from [90, 91, 51], we show that Theorem 1 can be used to control the bias and the variance of ergodic averages. Let

$$|g|_{\mathrm{Lip}(\rho)} \;=\; \sup\left\{ |g(x)-g(y)| \,/ \rho(x,y) \,:\, x,y \in \mathbb{R}^d,\, x \neq y \right\} \tag{1.14}$$

denote the Lipschitz norm of a measurable function $g : \mathbb{R}^d \to \mathbb{R}$ w.r.t. a metric ρ.

Corollary 2 (Ergodic averages). *Suppose that the assumptions of Theorem 1 hold true. Then for any $x \in \mathbb{R}^d$ and $t > 0$,*

$$\left| E_x \left[\frac{1}{t} \int_0^t g(X_s)\, ds - \int g \, d\pi \right] \right| \;\leq\; \frac{1-e^{-ct}}{c\,t} \, |g|_{\mathrm{Lip}(\rho_1)} \left(R_2 + \epsilon\, V(x) + \epsilon \, \frac{C}{\lambda} \right).$$

If, moreover, the function $x \mapsto V(x)^2$ satisfies the geometric drift condition

$$(\mathcal{L}V^2)(x) \;\leq\; C^* - \lambda^* V^2(x) \qquad \text{for any } x \in \mathbb{R}^d \tag{1.15}$$

with constants $C^*, \lambda^* \in (0,\infty)$, then

$$\mathrm{Var}_x \left[\frac{1}{t} \int_0^t g(X_s) ds \right] \;\leq\; \frac{3}{c\,t} \, |g|^2_{\mathrm{Lip}(\rho_1)} \left(R_2^2 + 2\epsilon^2 \left[C^*/\lambda^* + e^{-\lambda^* t} V(x)^2 \right] \right).$$

1.2.2 Geometric ergodicity with explicit constants: Second main result

The additive distance $\mathcal{W}_{\rho_1}$ defined in (1.4) is very simple, and contractivity w.r.t. $\mathcal{W}_{\rho_1}$ even implies bounds in weighted total variation norms. However, this distance has the disadvantage that in general $\rho_1(x,y) \not\to 0$ as $x \to y$. Therefore, a contraction w.r.t. $\mathcal{W}_{\rho_1}$ as stated in (1.11) can only be expected to hold if there is a coupling (X_t, Y_t) such that $P(X_t = Y_t) \to 1$ as $t \to \infty$. In the case of nondegenerate diffusions as in (1.1), it is not difficult to construct such a coupling, but for generalizations to infinite dimensional or nonlinear diffusions, such couplings might either not be natural or difficult to construct, see e.g. the results in Chapter 2 and Chapter 3, and Section 1.2.3 further below.

Partially motivated by the weak Harris Theorem in [69], we will now replace the additive metric by a multiplicative semimetric. This will allow us to derive quantitative bounds for asymptotic couplings in the sense of [72, 69], i.e., couplings for which X_t and Y_t get arbitrarily close to each other but do not necessarily meet in finite time. To this end we consider an underlying distance-like function

$$\rho_2(x,y) \quad = \quad f(|x-y|)\,(1+\epsilon V(x)+\epsilon V(y)) \tag{1.16}$$

where f is a suitable, non-decreasing, bounded and concave continuous function satisfying $f(0)=0$. Note that in general, the function ρ_2 is a semimetric but not necessarily a metric, since the triangle inequality may be violated. Nevertheless, the Lipschitz norm w.r.t. ρ_2 is still well-defined by (1.14). In [69, Lemma 4.14], conditions are given under which ρ_2 satisfies a weak triangle inequality, i.e., under which there is a constant $C>0$ such that $\rho_2(x,z)\le C(\rho_2(x,y)+\rho_2(y,z))$ holds for all $x,y,z\in\mathbb{R}^d$. This is for example the case if V is strictly positive and grows at most polynomially, or if $V(x)=\exp(\alpha\,|x|)$ for large $|x|$. In any case, ρ_2 has the desirable property that $\rho_2(x,y)\to 0$ as $x\to y$.

As before, we assume that Assumptions 9 and 10 hold true. The growth condition on the Lyapunov function in Assumption 11 is now replaced by the following condition:

Assumption 12. *The logarithm of V is Lipschitz continuous, i.e.,*

$$\sup\frac{|\nabla V|}{V} \quad < \quad \infty.$$

Notice that in contrast to Assumption 11, Assumption 12 does not depend on κ. The global bound on ∇V can be replaced by a local bound, see Section 1.6 for an extension. Assumption 12 is satisfied if, for example, V is strictly positive, and, outside of a compact set, $V(x)=|x|^\alpha$ or $V(x)=\exp(\alpha\,|x|)$ for some $\alpha>0$.

We define a bounded non-decreasing function $Q:(0,\infty)\to[0,\infty)$ by

$$Q(\epsilon) \quad = \quad \sup\frac{|\nabla V|}{\max(V,1/\epsilon)}. \tag{1.17}$$

In contrast to Section 1.2.1, we now allow the constant ϵ in (1.16) to be chosen freely inside a given range. We require that

$$(4C\epsilon)^{-1} \quad \ge \quad \int_0^{R_1}\int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa(u)\,du\;+2\,Q(\epsilon)\,(s-r)\right)dr\;ds \tag{1.18}$$

with C and κ given by Assumptions 10 and 9, respectively. Notice that since Q is non-decreasing, (1.18) is always satisfied for ϵ sufficiently small. Further below, we demonstrate how the freedom to choose ϵ can be exploited to optimize the resulting contraction rate in certain cases. Similarly to Section 1.2.1, we define functions $\phi, \Phi : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\phi(r) \;=\; \exp\left(-\frac{1}{2}\int_0^r t\,\kappa(t)\,dt \,-2\,Q(\epsilon)\,r\right), \quad \Phi(r) \;=\; \int_0^r \phi(t)\,dt. \tag{1.19}$$

The function f in (1.16) will be chosen such that

$$\frac{1}{2}\Phi(r) \le f(r) \le \Phi(r) \quad \text{for } r \le R_2, \qquad \text{and} \qquad f(r) = f(R_2) \quad \text{for } r \ge R_2,$$

where we define

$$S_1 \;=\; \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : V(x) + V(y) \le 2C/\lambda\right\}, \tag{1.20}$$
$$S_2 \;=\; \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : V(x) + V(y) \le 4C(1+\lambda)/\lambda\right\}, \tag{1.21}$$
$$R_i \;=\; \sup\left\{|x-y| : (x,y) \in S_i\right\}, \qquad i = 1,2. \tag{1.22}$$

Here the sets S_1 and S_2 have been chosen such that

$$\begin{aligned} \mathcal{L}V(x) + \mathcal{L}V(y) \;&<\; 0 \qquad \text{for } (x,y) \notin S_1, \qquad \text{and} \\ \epsilon\mathcal{L}V(x) + \epsilon\mathcal{L}V(y) \;&<\; -\frac{\lambda}{2}\min\left(1, 4C\epsilon\right)\left(1 + \epsilon V(x) + \epsilon V(y)\right) \quad \text{for } (x,y) \notin S_2. \end{aligned}$$

We now state our second main result.

Theorem 2 (Contraction rates for multiplicative semimetric)**.** *Suppose that Assumptions 9, 10, and 12 hold true. Fix $\epsilon \in (0,\infty)$ such that (1.18) is satisfied. Then there exist a concave, bounded and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and a constant $c \in (0,\infty)$ such that*

$$\mathcal{W}_{\rho_2}(\mu p_t, \nu p_t) \;\le\; e^{-ct}\,\mathcal{W}_{\rho_2}(\mu,\nu) \qquad \text{for any } \mu,\nu \in \mathcal{P}(\mathbb{R}^d) \text{ and } t \ge 0. \tag{1.23}$$

Here $c = \min\left(\beta, \lambda, 4C\epsilon\lambda\right)/2$ where

$$\begin{aligned} \beta^{-1} \;&=\; \int_0^{R_2} \Phi(r)\phi(r)^{-1}\,dr \\ &=\; \int_0^{R_2}\int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa(u)\,du \,+2\,Q(\epsilon)\,(s-r)\right) dr\; ds, \end{aligned}$$

the distance ρ_2 is defined by (1.16), and f is constant for $r \ge R_2$ and satisfies

$$\frac{1}{2} \;\le\; f'(r)\,\exp\left(\frac{1}{2}\int_0^r u\,\kappa(u)\,du \,+2\,Q(\epsilon)\,r\right) \;\le\; 1 \qquad \textit{for } r \in (0, R_2).$$

The precise definition of the function f is given in the proof in Section 1.5.2.

In order to optimize our bounds by choosing ϵ appropriately, we replace Assumption 12 by a slightly stronger condition:

Assumption 13.

$$\frac{\nabla V(x)}{V(x)} \to 0 \qquad as \qquad |x| \to \infty.$$

If Assumption 13 holds then $Q(\epsilon) \to 0$ as $\epsilon \to 0$. Therefore, by choosing ϵ sufficiently small, we can ensure that the term $Q(\epsilon)(s-r)$ occurring in the exponents in (1.18) and in the definition of β is bounded by 1. Explicitly, we choose

$$\epsilon \;=\; \min\left(Q^{-1}(R_2^{-1}),\; \left(4Ce^2\, I(R_1)\right)^{-1}\right), \tag{1.24}$$

where $Q^{-1}(t) = \sup\{\epsilon > 0 \,:\, Q(\epsilon) \le t\} \in (0,\infty]$ for $t > 0$ by Assumption 13, and

$$I(r) \;=\; \int_0^r \int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa(u)\,du\right) dr\; ds.$$

Corollary 3 (Contraction rates for multiplicative semimetric II)**.**
Suppose that Assumptions 9, 10, and 13 hold true. Then the assertion of Theorem 2 is satisfied with ϵ given by (1.24) *and*

$$c \;\ge\; \frac{1}{2}\min\left(e^{-2}/I(R_2),\, \lambda,\, \lambda e^{-2}/I(R_1),\, 4C\lambda\, Q^{-1}\left(1/R_2\right)\right).$$

The corollary is particularly useful if $b = -\nabla U$ for a convex (but not strictly convex) function U. In this case we can choose $\kappa = 0$, and hence $I(r) = r^2/2$:

Example 5 (Convex case)**.** *Let $b(x) = -\nabla U(x)$ for a convex function $U \in C^2(\mathbb{R}^d)$, and suppose that Assumption 10 holds with V satisfying $V(x) = |x|^p$ outside of a compact set for some $p \in [1,\infty)$. Then there is a constant $A \in (0,\infty)$ such that $Q^{-1}(t) \ge A\,t^p$ for any $t > 0$, and hence*

$$c \;\ge\; \min\left(e^{-2}R_2^{-2},\, \lambda/2,\, \lambda e^{-2}R_1^{-2},\, 2C\lambda\, AR_2^{-p}\right).$$

In particular, $c^{-1} = O(R_2^2)$ if $V(x) = |x|^2$ outside a compact set.

Similarly as in Corollary 2 above, the bounds in Theorem 2 can be used, among other things, to control the bias and variance of ergodic averages. Furthermore, a statement as in (1.23) implies gradient bounds for the transition kernel:

Corollary 4 (Gradient bounds for the transition semigroup)**.** *Suppose that the assumptions in Theorem 2 are satisfied. Then*

$$|p_t g|_{\mathrm{Lip}(\rho_2)} \;\le\; e^{-ct}\; |g|_{\mathrm{Lip}(\rho_2)}$$

holds for any $t \ge 0$ and for any function $g : \mathbb{R}^d \to \mathbb{R}$ that is Lipschitz continuous w.r.t. ρ_2. In particular, if $p_t g$ is differentiable at x, then

$$|\nabla p_t g(x)| \le |g|_{\mathrm{Lip}(\rho_2)}\,\left(1 + 2\,\epsilon\, V(x)\right) e^{-ct}. \tag{1.25}$$

The proof is included in Section 1.5.2 further below.

1.2.3 McKean-Vlasov diffusions

We now apply our approach to nonlinear diffusions on $\mathbb{R}^d$ satisfying an SDE of type

$$\begin{aligned} dX_t &= b(X_t)\,dt + \tau \int \vartheta(X_t, y)\,\mu_t^x(dy)\,dt + dB_t, \qquad X_0 = x, \\ \mu_t^x &= \mathrm{Law}(X_t). \end{aligned} \tag{1.26}$$

Here $\tau \in \mathbb{R}$ is a given constant and (B_t) is a d-dimensional Brownian motion. Under appropriate conditions on the coefficients b and ϑ, Equation (1.26) has a unique solution (X_t) which is a nonlinear Markov process in the sense of McKean, i.e., the future development after time t depends both on the current state X_t and on the law of X_t [144, 117]. To ensure that standard existence and uniqueness results apply, we assume that b is Lipschitz (this can probably be relaxed), and in Assumption 14 below we will also assume that ϑ is Lipschitz (this is an essential assumption for our results below). Corresponding nonlinear SDEs arise naturally as marginal limits of mean field interacting particle systems

$$dX_t^i = b(X_t^i)\,dt + \frac{\tau}{n}\sum_{j=1}^{n} \vartheta(X_t^i, X_t^j)\,dt + dB_t^i, \qquad i = 1, \dots n, \tag{1.27}$$

as $n \to \infty$. Here, the (B_t^i) are independent Brownian motions.

Convergence to equilibrium, or contractivity, for the nonlinear equation and the particle system are longstanding problems. Assuming $b = -\nabla V$ and $\vartheta(x, y) = \nabla W(y) - \nabla W(x)$ with smooth potentials V and W, the convex case for the nonlinear equation was tackled by Carrillo, McCann and Villani [20, 21] using PDE techniques, and by Malrieu [109] and Cattiaux, Guillin and Malrieu [22] using coupling arguments. More recently, using direct control of the derivative of the Wasserstein distance, Bolley, Gentil and Guillin [14] have proven an exponential trend to equilibrium for small bounded and Lipschitz perturbations of the strictly convex case. In the spirit of Meyn-Tweedie's approach, and via nonlinear Markov chains, Butkovsky [18] established exponential convergence to equilibrium in the bounded perturbation case. In [51, Corollary 3.4], a contraction property for the particle system (1.27) has been derived for sufficiently small τ with a dimension-independent contraction rate using an approximation of a componentwise reflection coupling.

We now show that a similar strategy as in [51] can be applied directly to the nonlinear equation. We assume that the interaction coefficient $\vartheta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a globally Lipschitz continuous function:

Assumption 14. *There exists a constant $L \in (0, \infty)$ such that*

$$|\vartheta(x, x') - \vartheta(y, y')| \;\le\; L \cdot (|x - y| + |x' - y'|) \quad \textit{for any } x, x', y, y' \in \mathbb{R}^d.$$

In our first theorem, we assume the contractivity at infinity condition (1.28) instead of a Lyapunov condition. Existence and uniqueness of solutions of the nonlinear SDE can then be proven as in [22]. In that case we can obtain contractivity w.r.t. an underlying metric of type

$$\rho_0(x,y) \;=\; f(|x-y|)$$

where f is an appropriately chosen concave function. Let $\mathcal{W}^1$ denote the standard L^1 Wasserstein distance defined w.r.t. the Euclidean metric on $\mathbb{R}^d$. Notice that in the upcoming theorem, we allow the function κ from Assumption 9 to take negative values. We obtain the following counterpart to Corollary 3.4 in [51]:

Theorem 3 (Contraction rates for nonlinear diffusions I)**.** *Suppose that Assumptions 9 and 14 hold true with a function $\kappa : (0,\infty) \to \mathbb{R}$ satisfying*

$$\limsup_{r\to\infty} \kappa(r) \;<\; 0. \tag{1.28}$$

For $x \in \mathbb{R}^d$, let (μ_t^x) denote the marginal laws of a strong solution (X_t) of Equation (1.26) with initial condition $X_0 = x$. Then there exist a concave and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and constants $c, K, A \in (0,\infty)$ such that for any $\tau \in \mathbb{R}$ and $x, y \in \mathbb{R}^d$,

$$\begin{aligned} \mathcal{W}_{\rho_0}(\mu_t^x, \mu_t^y) \;&\le\; \exp\left((|\tau|\, K - c)\, t\right) \rho_0(x,y), \qquad \textit{and} \qquad &(1.29)\\ \mathcal{W}^1(\mu_t^x, \mu_t^y) \;&\le\; 2\,A\, \exp\left((|\tau|\, K - c)\, t\,\right) |x-y|. &(1.30)\end{aligned}$$

The constants are explicitly given by

$$\begin{aligned} c^{-1} \;&=\; \int_0^{R_2}\int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa^+(u),\, du\right) dr\, ds,\\ A \;&=\; \exp\left(\frac{1}{2}\int_0^{R_1} s\,\kappa^+(s)\,ds\right),\\ K \;&=\; 4\,L\,\exp\left(\frac{1}{2}\int_0^{R_1} s\kappa^+(s)\,ds\right), \end{aligned} \tag{1.31}$$

where

$$\begin{aligned} R_1 \;&=\; \inf\{R \ge 0 : \kappa(r) \le 0 \;\; \textit{for all}\; r \ge R\}, \qquad \textit{and} &(1.32)\\ R_2 \;&=\; \inf\{R \ge R_1 : \kappa(r)\, R\,(R - R_1) \le -4 \;\; \textit{for all}\; r \ge R\}. &(1.33)\end{aligned}$$

The function f is linear for $r \ge R_2$, and

$$\frac{1}{2} \le f'(r)\,\exp\left(\frac{1}{2}\int_0^{r\wedge R_1} s\,\kappa^+(s)\,ds\right) \le 1 \qquad \textit{for } 0 < r < R_2.$$

The precise definition of the function f is given in the proof in Section 1.5.3.

Our next goal is to replace (1.28) by the following dissipativity condition:

Assumption 15 (Drift condition). *There exist constants $D, \lambda \in (0,\infty)$ such that*

$$\langle x, b(x)\rangle \;\leq\; -\lambda\,|x|^2 \qquad \textit{for any } x \in \mathbb{R}^d \textit{ with } |x| \geq D.$$

Let $V(x) = 1 + |x|^2$. Assumption 15 implies that V is a Lyapunov function for the nonlinear diffusion (1.26), cf. Lemma 3 below.

A major difficulty in the McKean-Vlasov case is that solutions X_t and Y_t with different starting points follow dynamics with different drifts. Therefore, it is not clear how to construct a coupling (X_t, Y_t) such that $X_t = Y_t$ for $t > T$ holds for an almost surely finite stopping time T. Using the multiplicative semimetric we are still able to retrieve a local contraction:

Theorem 4 (Contraction rates for nonlinear diffusions II). *Suppose that Assumptions 9, 14 and 15 hold true. For $x \in \mathbb{R}^d$, let (μ_t^x) denote the marginal laws of a strong solution (X_t) of Equation (1.26) with initial condition $X_0 = x$. Then there exist a concave, bounded and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and constants $c, \epsilon, K_0, K_1 \in (0,\infty)$ such that:*

(i) For any $R \in (0,\infty)$ there is $\tau_0 \in (0,\infty)$ such that for any $\tau \in \mathbb{R}$ and $x, y \in \mathbb{R}^d$ with $|\tau| \leq \tau_0$ and $|x|, |y| \leq R$,

$$\mathcal{W}_{\rho_2}(\mu_t^x, \mu_t^y) \;\leq\; \exp(-c\,t)\,\rho_2(x,y), \qquad \textit{and} \tag{1.34}$$

$$\mathcal{W}^1(\mu_t^x, \mu_t^y) \;\leq\; K_0\,\exp(-c\,t)\;|x-y|\,(1 + \epsilon\,V(x) + \epsilon\,V(y))\,. \tag{1.35}$$

(ii) There is $\tau_0 \in (0,\infty)$ s.t. for any $\tau \in \mathbb{R}$ with $|\tau| \leq \tau_0$ and $x, y \in \mathbb{R}^d$,

$$\mathcal{W}_{\rho_2}(\mu_t^x, \mu_t^y) \;\leq\; \exp(-c\,t)\,\left(\rho_2(x,y) + K_1\,[\epsilon V(x) + \epsilon V(y)]^2\right). \tag{1.36}$$

The function ρ_2 is given by (1.16). For the explicit definition of the function f and the constants $c, \epsilon, \tau_0, K_0, K_1$ see the proof in Section 1.5.3.

The assumption that τ is sufficiently small is natural, since for large τ, Equation (1.26) can have several distinct stationary solutions. Nevertheless, we do not claim that our bound on τ is sharp.

1.2.4 Subgeometric ergodicity with explicit constants

We now consider the case where the drift is not strong enough to provide a Kantorovich contraction like (1.11). Instead of Assumption 10, we only assume a subgeometric drift condition as it has been used for example in [43].

Assumption 16 (Subgeometric drift condition). *There are a function $V \in C^2(\mathbb{R}^d)$ with $\inf_{x\in\mathbb{R}^d} V(x) > 0$, a strictly positive, increasing and concave C^1 function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\eta(V(x)) \to \infty$ as $|x| \to \infty$, as well as a constant $C \in (0,\infty)$ such that*

$$\mathcal{L}V(x) \;\leq\; C - \eta(V(x)) \qquad \textit{for any } x \in \mathbb{R}^d.$$

The following example shows how V and η can be chosen explicitly, cf. also [43].

Example 6 (Choice of V and η)**.** *Suppose that*

$$\langle b(x), x\rangle \quad\leq\quad -\gamma\,|x|^q$$

holds for $|x| \geq R$ with constants $R, \gamma \in (0, \infty)$ and $q \in (0, 1)$. Let $V \in C^2(\mathbb{R}^d)$ be a strictly positive function such that outside a compact set, $V(x) = \exp(\alpha\,|x|^q)$ for some $\alpha \in (0, 2\gamma/q)$, and fix $\beta \in (0, \gamma - \alpha q/2)$. Then Assumption 16 is satisfied with

$$\eta(r) = \begin{cases} \alpha^{\frac{2}{q}-1} q\beta r \log(r)^{2-\frac{2}{q}} & \text{for } r \geq e^{\frac{2}{q}-1}, \\ \alpha^{\frac{2}{q}-1}\beta \left(\frac{2}{q}-1\right)^{1-\frac{2}{q}} \left(2e^{1-\frac{2}{q}}(q-1)r^2 + (4-3q)r\right) & \text{for } r < e^{\frac{2}{q}-1}. \end{cases}$$

From now on we assume that Assumption 9 holds true, and we define the functions φ and Φ as in (1.5) above. Let $R_1 = \sup\{|x-y| : (x,y) \in S_1\}$, where

$$S_1 \;=\; \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : \eta(V(x)) + \eta(V(y)) \leq 4C\right\}.$$

The set S_1 is chosen such that for $(x,y) \notin S_1$,

$$\mathcal{L}\,V(x) + \mathcal{L}\,V(y) \leq -\,(\,\eta(V(x)) \,+\, \eta(V(y))\,)\,/2.$$

Notice that R_1 is finite, since $\eta(\,V(x)\,) \to \infty$ as $|x| \to \infty$. Moreover, S_1 is recurrent for any Markovian coupling (X_t, Y_t) of solutions of (1.1). Let

$$\epsilon^{-1} = \max\left(1, 4C\int_0^{R_1} \phi(r)^{-1}\,dr\right) = \max\left(1, 4C\int_0^{R_1} e^{\frac{1}{2}\int_0^r t\,\kappa(t)\,dt}\,dr\right). \tag{1.37}$$

The following growth condition on the Lyapunov function replaces Assumption 11:

Assumption 17 (Growth condition in subgeometric case)**.** *There exist a constant $\alpha > 0$ and a bounded set $S_2 \supseteq S_1$ such that for any $(x,y) \in \mathbb{R}^{2d} \setminus S_2$, we have*

$$\eta(V(x)) + \eta(V(y)) \geq 4C\left(1 + \alpha\int_0^{R_1} \phi(r)^{-1}\,dr\;\eta(\Phi(|x-y|))\right).$$

Notice that Φ grows at most linearly. Let $R_2 = \sup\{|x-y| : (x,y) \in S_2\}$. We state our main result for the subgeometric case.

Theorem 5 (Subgeometric decay rates)**.** *Suppose that Assumptions 9, 16 and 17 hold true. Then there exist a concave, bounded and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and constants $c, \epsilon \in (0, \infty)$ s.t.*

$$\|p_t(x,\cdot) - p_t(y,\cdot)\|_{\mathrm{TV}} \quad\leq\quad \frac{\rho_1(x,y)}{H^{-1}(c\,t)} \qquad \textit{for any } x, y \in \mathbb{R}^d \textit{ and } t \geq 0. \tag{1.38}$$

Here the distance ρ_1 *is defined by* (1.4) *and* (1.37)*, the function* $H : [l, \infty) \to [0, \infty)$ *is given by*

$$H(t) \;=\; \int_l^t \frac{1}{\eta(s)}\, ds \qquad \textit{with} \qquad l \;=\; 2\,\epsilon \inf_{x \in \mathbb{R}^d} V(x),$$

$c = \min(\alpha, \beta, \gamma)/2$ *where* β *is given by* (1.12)*, and*

$$\gamma \;=\; \inf\{\, \epsilon\, \eta(r)/\eta(\epsilon r) \,:\, r \geq l/\epsilon\}\,.$$

The function f *is constant for* $r \geq R_2$*, and*

$$\frac{1}{2} \;\leq\; f'(r)\, \exp\left(\frac{1}{2}\int_0^r t\, \kappa(t)\, dt\right) \;\leq\; 1 \qquad \textit{for any } r \in (0, R_2).$$

The precise definition of the function f *is given in the proof in Section 1.5.4.*

The crucial difference in comparison to Theorem 1 is, that we do not provide upper bounds on $\mathcal{W}_{\rho_1}$, but use the additive distance to derive moment bounds for coupling times instead. These bounds are partially based on a technique from [74], see Section 1.3.6 further below.

Remark 4. *Since* $\eta(s)$ *is concave, it is growing at most linearly as* $s \to \infty$*. In particular,* $\int_l^\infty (1/\eta(s))\, ds = \infty$*, and thus the inverse function* H^{-1} *maps* $[0, \infty)$ *to* $[l, \infty)$*. Since* $\epsilon \leq 1$ *and* η *is increasing, we always have* $\gamma \geq \epsilon$ *. If* $\eta(r) = r^a$ *for some* $a \in (0, 1)$*, then* $\gamma \geq \epsilon^{1-a}$*.*

It is well-known that the local Lipschitz assumption on b together with Assumption 16 implies the existence of a unique invariant probability measure π satisfying $\int \eta(V(x))\, \pi(dx) \leq C$, see e.g. [74, Section 4]. Theorem 5 can be used to quantify the speed of convergence towards the invariant measure using cut-off arguments. Following [74, Section 4], we obtain:

Corollary 5. *Under the Assumptions of Theorem 5,*

$$\|p_t(x, \cdot) - \pi\|_{\mathrm{TV}} \;\leq\; \frac{R_2 + \epsilon\, V(x)}{H^{-1}(\,c\, t\,)} + \frac{(2\,\epsilon\, b + 1)\, C}{\eta(b\, H^{-1}(\,c\, t\,))} \quad \textit{for any } x \in \mathbb{R}^d \textit{ and } t \geq 0,$$

where $b = \eta^{-1}(2C)/l$*.*

The proofs are given in Section 1.5.4.

1.3 Discussion

1.3.1 Comparison to Meyn-Tweedie approach

The classical Harris theorem, as propagated by Meyn-Tweedie, allows to derive geometric ergodicity for a large class of Markov chains under conditions which are easy to verify. The approach is very generally applicable, but it is usually not trivial to make the results quantitative. The first assumption is that the Markov chain at hand is recurrent w.r.t. some bounded subset S of the state space and that one has some kind of control over the average length of excursions from this set. The second assumption which is typically imposed is a minorization condition which often takes the following form: There are constants $t, \epsilon \in (0, \infty)$ and a probability measure Q such that

$$p_t(x, \cdot) \quad \geq \quad \epsilon\, Q(\cdot) \tag{1.39}$$

holds for all $x \in S$, where p_t denotes the transition kernel of the chain.

The recurrence condition can be quantified performing direct computations with the generator of the Markov chain via Lyapunov techniques. The minorization condition is usually much harder to quantify. In the context of diffusions of the form (1.1) there are abstract methods available which allow to conclude that the condition (1.39) can indeed be satisfied, cf. e.g. [101, Discussion after Remark 1.29]. Nevertheless, using such methods, it is not clear how the resulting constant ϵ depends on the drift coefficient b and how a perturbation of b translates to a change of ϵ. In the diffusion setting, Roberts and Rosenthal developed in [135] a method to provide explicit bounds for ϵ that are closely connected to the drift coefficient b. Their method is based on reflection coupling and an application of the Bachelier-Lévy formula. In comparison to their results, we establish contractions of the transition kernels, and our contraction rates are based only on *one-sided* Lipschitz bounds for the drift coefficient. This often leads to much more precise bounds.

1.3.2 Relation to functional inequalities

Functional inequalities are now a common tool to get rates for convergence to equilibrium in L^2 distance or in entropy. For the class of diffusion processes considered here, the Poincaré inequality takes the form

$$\mathrm{Var}_\pi(f) \quad \leq \quad \frac{1}{2}\, C_P \int |\nabla f|^2 \, d\pi \tag{1.40}$$

for smooth functions f, where π is the stationary distribution. Equation (1.40) is equivalent to L^2 convergence to equilibrium (and in fact L^2 contractivity) with rate C_P^{-1}. It turns out to be quite difficult to prove a Poincaré inequality for a general non-reversible diffusion such as (1.1), as usual criteria rely on the explicit knowledge of the invariant probability measure π. If we assume that $b(x) = -\nabla V(x)/2$, then the diffusion is reversible with respect to $d\pi \propto e^{-V} dx$ and plenty criterias are available to prove Poincaré inequalities. In particular, it is shown in [4], that if there exists a set B, constants $\lambda, C \in (0,\infty)$, and a positive twice continuously differentiable function V such that

$$\mathcal{L}V \;\leq\; -\lambda V \,+\, C\, I_B,$$

and a local Poincaré inequality of the form

$$\int_B (f - \pi(f\, I_B))^2\, d\pi \;\leq\; \frac{1}{2}\,\kappa_B \int |\nabla f|^2\, d\pi$$

holds, then a global Poincaré inequality holds with constant $C_P = \lambda^{-1}(1 + C\,\kappa_B)$. Notice that a Poincaré inequality implies back the Lyapunov condition. Using the additive metric and Corollary 1, one has that a Poincaré inequality holds, but the identification of the constant is a hard task in general. However, using the multiplicative semimetric and the gradient bounds of Corollary 4, one may prove that in the reversible case a Poincaré inequality holds with the same constant c than in Corollary 4. Here the reflection coupling serves as an alternative to a local Poincaré inequality. The latter is usually established via Holley-Stroock's perturbation argument [82] which may lead to quite poor estimates.

Notice also that, by a result of Sturm and von Renesse [150], for a reversible diffusion with stationary distribution given by $e^{-V}\, dx$, a strict contraction in L^p Wasserstein distance is *equivalent* to a lower bound on the Hessian of V. The latter condition is a special case of the Bakry-Emery criterion and usually linked to logarithmic Sobolev inequalities, cf. [3]. In [23], a reinforced Lyapunov condition has been used to prove stronger functional inequalities than Poincaré inequalities (namely super Poincaré inequalities, including logarithmic Sobolev inequalities). In a similar spirit, one can replace the *global* curvature condition (9) by a *local* one using a reinforced Lyapunov condition, cf. Section 1.6 further below. Note however that, although our results are sufficient to prove back some Poincaré inequality, it does not seem possible to get stronger inequalities starting from our contractions.

1.3.3 Dimension dependence

In our results above, dependence on the dimension d usually enters through the value of the constant C in the Lyapunov condition, which affects the size of R_2. For example, in Theorem 1, the contraction rate is $c = \min(\alpha, \beta, \lambda)/2$, where α and λ are given by Assumptions 11 and 10 respectively, and the constant β defined in (1.12) depends both on R_2 and on the function κ defined in Assumption 9. In order to illustrate the dependence on the dimension of R_2, let us assume that there are constants $A, \gamma \in (0,\infty)$ and $q \geq 1$ such that

$$\langle x, b(x)\rangle \;\leq\; -\gamma\, |x|^q \qquad \text{for all } |x| \geq A.$$

Suppose first that $q = 2$. Then $V(x) = 1 + |x|^2$ satisfies the Lyapunov condition in Assumption 10 with constants $C = O(d)$ and $\lambda = \Omega(1)$. In this case, the set S_2 in Assumption 11 can be chosen such that $R_2 = O(\sqrt{d})$. Hence, assuming a one-sided Lipschitz condition with constant κ as in Example 4, the lower bound c for the contraction rate in Theorem 1 is of order $\Omega(1/d)$ if $\kappa = 0$ (convex case), or, more generally, if $\kappa = O(1/d)$. On the other hand, for $\kappa = \Omega(1)$, c is exponentially small in the dimension. By Example 5, similar statements hold true for the lower bound on the contraction rate w.r.t. the multiplicative semimetric derived in Corollary 3.

Now assume more generally $q \geq 1$. In this case, a Lyapunov function with polynomial growth does not necessarily exist. Instead, by Remark 1, one can choose a Lyapunov function V with constant $\lambda = 1$ such that outside of a compact set, $V(x) = \exp\left(a\,|x|^q\right)$ for some $a < 2\gamma/q$. In this case, $C = O(\exp(\eta d))$ for some finite constant $\eta > 0$, and one can choose R_2 of order $O(d^{1/q})$. Again, assuming a one-sided Lipschitz condition, the constant c in Theorem 1 is of polynomial order $\Omega(d^{-2/q})$ if $\kappa = 0$ (convex case), or, more generally, if $\kappa = O(d^{-2/q})$. For the multiplicative semimetric, we are not able to prove a polynomial order in the dimension in this case. Nevertheless, for $q \in (1, 2)$, an application of Corollary 3 with a Lyapunov function satisfying $V(x) = \exp(|x|^\alpha)$ for large $|x|$ for some $\alpha \in (2 - q, 1)$ yields at least a sub-exponential order in d. For $\kappa = \Omega(1)$, the values of c decay exponentially in the dimension.

We finally remark that, in some situations, it is possible to combine the techniques presented here with additional arguments to derive explicit and dimension-free contraction rates for diffusions, see for example the results presented in Chapter 2 of this thesis.

1.3.4 Extensions of the results

Similarly as in [51], the results presented above can easily be generalized to diffusions with a constant and nondegenerate diffusion matrix σ. In the case of non-constant and non-degenerate diffusion coefficients $\sigma(x)$, it should still be possible to retrieve related results replacing reflection coupling by the *Kendall-Cranston coupling* w.r.t. the intrinsic Riemannian metric induced by the diffusion coefficients, cf. [95, 32].

The main contraction results, Theorem 1 and Theorem 2, are based on Assumption 9, a *global* generalized one-sided Lipschitz condition. It is possible to relax this condition to a *local* bound which, up to some technical details, holds only on the set for which the coupling (X_t, Y_t) is recurrent. A corresponding generalization is given in Section 1.6.

In the recent work [108], Majka extends the results from [51] to stochastic differential equations driven by Lévy jump processes with rotationally invariant jump measures, deriving Kantorovich contractions for the transition semigroups with explicit constants, see also [153]. One of the key assumptions in [108] is the “contractivity at infinity” condition (1.28). Using an additive distance similar to (1.4), it should be possible to extend the results presented there, replacing the latter assumption by a more general geometric drift condition.

An extension of the theory presented in this article to a class of degenerate and infinite-dimensional diffusions is considered in Chapter 2 combining asymptotic couplings with the multiplicative distance (1.16).

In this work, we derive explicit contraction rates for diffusion processes. An important question is whether similar results can be obtained for time-discrete approximations. There are at least two different approaches to tackle this question. The first approach, which is considered in forthcoming work by one of the authors, is to establish related coupling approaches directly for Markov chains. Another possibility is to consider time discretization as a perturbation of the diffusion process, and to apply directly the contraction results for the diffusion, cf. [36, 45, 46] and also [134, 138, 121, 128, 58, 137].

1.3.5 McKean-Vlasov equations

For the class of nonlinear diffusions considered above, Theorems 3 and 4 considerably relax assumptions in previous works. Both, the PDE approach in [20] and the approach based on synchronous coupling in [109, 22] require global positive curvature bounds. In the case where the curvature is strictly positive with degeneracy at a finite number of points, algebraic contraction rates have been derived by synchronous coupling. The "dissipation of W^2 approach" in [14] yields exponential decay to equilibrium for sufficiently small τ provided the confinement and interaction forces both derive from a potential, the confinement force satisfies condition (1.28), and the interaction potential is bounded with a lower bound on the curvature. The approach in [18] yields exponential convergence to equilibrium in total variation distance in the small and bounded interaction case. Theorem 3 above relaxes these assumptions on the interaction potential while requiring only a "strict convexity at infinity" condition on the confinement potential. Moreover, Theorem 4 replaces the latter condition on the confinement potential by the dissipativity condition in Assumption 15. With additional technicalities, it should be possible to relax this dissipativity condition to $\langle x, b(x)\rangle \leq -\lambda\, |x|$ and generalize Lemma 3.

1.3.6 Subgeometric ergodicity

Our results in the subgeometric case can be interpreted as a variation of statements from the lecture notes [74, Section 4]. There, M. Hairer derives subgeometric ergodicity for diffusions, estimating hitting times of recurrent sets and combining these with a minorization condition. While the principle result from [74, Section 4] is already contained in [5, 43], the method of proof shows new and interesting aspects avoiding discrete-time approximations. The main tool used is the following statement, which gives an elegant proof for the integrability of hitting times:

Lemma 2 ([74], reformulated)**.** *Let $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly positive, increasing and concave C^1 function and denote by (Z_t) a continuous semimartingale, i.e. $Z_t = Z_0 + A_t + M_t$, where (A_t) is of finite variation, (M_t) is a local martingale, $E[Z_0] < \infty$ and $A_0 = M_0 = 0$. Let T be a stopping time. If there are constants $l, c \in (0, \infty)$ such that*

$$Z_t \;\geq\; l \quad \text{and} \quad A_t \leq -c \int_0^t \eta(\, Z_s\,)\, ds \quad \text{almost surely for } t < T, \tag{1.41}$$

then T is almost surely finite and satisfies the inequality

$$E\left[\, H^{-1}(\, c\,T\,)\,\right] \leq E\left[\, H^{-1}(\, H(\, Z_T\,) + c\,T\,)\right] \leq E[\, Z_0\,],$$

where $H : [l, \infty) \to [0, \infty)$ is given by $H(t) = \int_l^t \frac{1}{\eta(s)}\, ds$.

Our result for the subgeometric case, Theorem 5, relies on the above tool. The main difference to [74] is that we do not impose any kind of minorization condition or renewal theory. Instead, we consider a reflection coupling (X_t, Y_t) of the diffusions, defined in Section 1.4.2, and we directly establish bounds on the integrability of the coupling time $T = \inf\{t \geq 0 : X_t = Y_t\}$ using Lemma 2 and the additive distance (1.4). For the reader's convenience, a proof of Lemma 2 is included in Section 1.5.4. It should be mentioned that subgeometric ergodicity of Markov processes has been studied by many others authors in various settings, see [44, 59, 122, 145, 146, 19, 110, 83] and the references therein.

1.4 Couplings

1.4.1 Synchronous coupling for diffusions

Given initial values $(x_0, y_0) \in \mathbb{R}^{2d}$ and a d-dimensional Brownian motion (B_t), we define a *synchronous coupling* of two solutions of (1.1) as a diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$ solving

$$\begin{aligned} dX_t &= b(X_t)\; dt + dB_t, \qquad & X_0 &= x_0, \\ dY_t &= b(Y_t)\; dt \; + dB_t, \qquad & Y_0 &= y_0. \end{aligned}$$

1.4.2 Reflection coupling for diffusions

Reflection coupling goes back to [107]. Given initial values $(x_0, y_0) \in \mathbb{R}^{2d}$ and a d-dimensional Brownian motion (B_t), we define a *reflection coupling* of two solutions of (1.1) as a diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$ satisfying

$$\begin{aligned} dX_t &= b(X_t)\, dt + dB_t, & \\ dY_t &= b(Y_t)\, dt + (I - 2\, e_t\, \langle e_t, \cdot \rangle)\, dB_t \qquad & \text{for } t < T, \\ Y_t &= X_t & \text{for } t \geq T, \\ (X_0, Y_0) &= (x_0, y_0), & \end{aligned}$$

where $T = \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time. Here, for $t < T$, e_t is the unit vector given by $e_t = (X_t - Y_t)/|X_t - Y_t|$.

1.4.3 Coupling for McKean-Vlasov processes

We construct a coupling for two solutions of (1.26). The coupling will be realized as a process (X_t, Y_t) with values in $\mathbb{R}^{2d}$. We first describe the coupling in words: We fix a parameter $\delta > 0$ and use a reflection coupling of the driving Brownian motions whenever $|X_t - Y_t| \geq \delta$. If, on the other hand, $|X_t - Y_t| \leq \delta/2$ we use a synchronous coupling. Inbetween there is a transition region, where a mixture of both couplings is used. One should think of δ being close to zero.

The technical realization of the coupling is near to [51]. In order to implement the above coupling, we introduce Lipschitz functions $\mathrm{rc} : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ and $\mathrm{sc} : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ satisfying

$$\mathrm{sc}^2(x,y) + \mathrm{rc}^2(x,y) = 1. \tag{1.42}$$

We impose that $\mathrm{rc}(x,y) = 1$ holds whenever $|x - y| \geq \delta$ and $\mathrm{rc}(x,y) = 0$ holds if $|x - y| \leq \delta/2$. The functions rc and sc can be constructed using standard cut-off techniques. Notice that in the case where the drift coefficient b and the nonlinearity ϑ are Lipschitz, equation (1.26) admits a unique, strong and non-explosive solution (X_t) for any initial value $x_0 \in \mathbb{R}^d$. The uniqueness holds pathwise and in law. Moreover, the law $\mu_t^{x_0}$ of X_t has finite second moments, i.e. $\int |y|^2 \, \mu_t^{x_0}(dy) < \infty$, see [117, Theorem 2.2] and [144]. For a fixed $x_0 \in \mathbb{R}^d$ we define

$$b^{x_0}(t,y) \quad = \quad b(y) + \theta \int \vartheta(y,z) \, \mu_t^{x_0}(dz).$$

The results from [117, Theorem 2.2] imply that the function $b^{x_0} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is continuous. It is easy to see that Assumption 14, in combination with a Lipschitz bound on b, implies that there is $M > 0$ such that

$$\sup_{t \geq 0} | \, b^{x_0}(t,y) - b^{x_0}(t,z) \, | \quad \leq \quad M \cdot |y - z| \qquad \text{for any } y, z \in \mathbb{R}^d. \tag{1.43}$$

Fix now initial values $(x_0, y_0) \in \mathbb{R}^{2d}$, the parameter $\delta > 0$ and two independent Brownian motions (B_t^1) and (B_t^2). For the given x_0 and y_0, we construct diffusion coefficients b^{x_0} and b^{y_0}, as above, and define the coupling $(U_t) = (X_t, Y_t)$ as the solution of the standard diffusion

$$\begin{aligned} dX_t &= b^{x_0}(t, X_t)\, dt && + \mathrm{rc}(U_t) \; dB_t^1 + \mathrm{sc}(U_t) \; dB_t^2, \\ dY_t &= b^{y_0}(t, Y_t)\, dt && + \mathrm{rc}(U_t) \; (I - 2 e_t \langle e_t, \cdot \rangle) \; dB_t^1 + \mathrm{sc}(U_t) \; dB_t^2, \end{aligned}$$

with $(X_0, Y_0) = (x_0, y_0)$ and

$$e_t \;=\; \frac{X_t - Y_t}{|X_t - Y_t|} \quad \text{for } X_t \neq Y_t, \qquad e_t \;=\; u \quad \text{for } X_t = Y_t,$$

where $u \in \mathbb{R}^d$ is some arbitrary fixed unit vector. Note that the concrete choice of u is irrelevant for the dynamic, since $\mathrm{rc}(x,x) = 0$. Inequality (1.43) implies that the above diffusion process admits a unique, strong and non-explosive solution. Using Levy's characterization of Brownian motion and (1.42), one can verify that the marginal processes (X_t) and (Y_t) solve the standard equations

$$dX_t = b^{x_0}(t, X_t)\, dt + B_t, \qquad X_0 = x_0 \tag{1.44}$$

$$dY_t = b^{y_0}(t, Y_t)\, dt + \hat{B}_t, \qquad Y_0 = y_0. \tag{1.45}$$

with respect to the Brownian motions

$$\begin{aligned} B_t &= \int_0^t \mathrm{rc}(U_s)\, dB_s^1 + \int_0^t \mathrm{sc}(U_s)\, dB_s^2 \qquad \text{and} \\ \hat{B}_t &= \int_0^t \mathrm{rc}(U_s)\, (I - 2e_s \langle e_s, \cdot \rangle)\, dB_s^1 + \int_0^t \mathrm{sc}(U_s)\, dB_s^2. \end{aligned} \tag{1.46}$$

Since the solutions (X_t) and (Y_t) of (1.44) and (1.45) are pathwise unique, they coincide a.s. with the strong solutions of (1.26) w.r.t. the Brownian motions (B_t) and $(\hat{B}_t)$ and initial values x_0 and y_0, respectively. Hence (X_t, Y_t) is indeed a coupling for (1.26).

1.5 Proofs

Let us start with a crucial tool which will be used throughout our proofs: A general construction of the function f appearing in the main theorems, characterized by a differential inequality. We define a concave function $f : [0, \infty) \to [0, \infty)$ depending on various parameters. Fix constants $R_1, R_2 \in \mathbb{R}_+$ such that $R_1 \leq R_2$, and let functions

$$h : [0, R_2] \to [0, \infty), \qquad j : [0, R_2] \to [0, \infty) \qquad \text{and} \qquad i : [0, R_1] \to [0, \infty)$$

be given. We suppose that i and j are continuous, j is non-decreasing and h is continuously differentiable with $h' \geq 0$. The function f is given by

$$f(r) \;=\; \int_0^{r \wedge R_2} \phi(s)\, g(s)\, ds,$$

where ϕ and g are defined as

$$\begin{aligned} &\phi(r) = \exp(-h(r)) \qquad \text{and} \\ &g(r) = 1 - \frac{\beta}{4} \int_0^{r \wedge R_2} j(\,\Phi(s)\,)\, \phi(s)^{-1}\, ds - \frac{\xi}{4} \int_0^{r \wedge R_1} i(s)\, \phi(s)^{-1}\, ds. \end{aligned}$$

Here the function Φ and the constants β and ξ are given by

$$\Phi(r) = \int_0^r \phi(s)\, ds, \quad \beta^{-1} = \int_0^{R_2} j(\,\Phi(s)\,)\, \phi(s)^{-1}\, ds, \quad \xi^{-1} = \int_0^{R_1} i(s)\, \phi(s)^{-1}\, ds. \tag{1.47}$$

The function f is a generalization of the concave distance function constructed in [51]. It is continuously differentiable on $(0, R_2)$ and constant on $[R_2, \infty)$. The derivative f' on $(0, R_2)$ is given by the product ϕg, where ϕ and g are positive and non-increasing functions. Hence f is a concave and non-decreasing function. Notice that g maps the interval $[0, R_2]$ into $[1/2, 1]$, which implies that the following inequalities hold for any $r \in [0, R_2]$:

$$r\, \phi(R_2) \leq \Phi(r) \leq 2\, f(r) \leq 2\, \Phi(r) \leq 2\, r. \tag{1.48}$$

The crucial property of the function f is that it is twice continuously differentiable on $(0, R_1) \cup (R_1, R_2)$ and that it satisfies on this set the (in)equality

$$\begin{aligned} f''(r) &= -h'(r)\, f'(r) - \frac{\beta}{4}\, j(\,\Phi(r)\,) - \frac{\xi}{4}\, i(r)\, I_{r<R_1} \\ &\leq -h'(r)\, f'(r) - \frac{\beta}{4}\, j(\,f(r)\,) - \frac{\xi}{4}\, i(r)\, I_{r<R_1}. \end{aligned} \tag{1.49}$$

Observe that f is not continuously differentiable at the point R_2 and thus we sometimes work with the left-derivative f'_- which exists everywhere. The function f can formally be extended to a concave function on $\mathbb{R}$ by setting $f(r) = -r$ for $r < 0$. We can associate with f a signed measure μ_f on $\mathbb{R}$, which takes the role of a generalized second derivative. For $x < y$ the measure is defined by $\mu_f(\,[x, y)\,) = f'_-(y) - f'_-(x)$. On the set $(0, R_1) \cup (R_1, R_2)$ the measure satisfies

$$\mu_f(dx) = f''(x)\, dx,$$

since f is twice continuously differentiable. Furthermore,

$$\mu_f\left((-\infty, 0] \cup (R_2, \infty)\right) = 0 \qquad \text{and} \qquad \mu_f\left(\{R_1, R_2\}\right) \leq 0.$$

1.5.1 Proofs of results in Section 1.2.1

Proof of Lemma 1. Let $(x, y) \in \mathbb{R}^{2d}$ such that $(x, y) \notin S_2$. Assume w.l.o.g. that $\max(|x|, |y|) = |x| \geq \mathcal{R}$. Using our assumption, the triangle inequality and the estimate $\Phi(r) \leq r$, we get

$$V(x) \geq 4C\lambda^{-1}\,(1 + 2\,|x|) \geq 4C\lambda^{-1}\,(1 + |x - y|) \geq 4C\lambda^{-1}\,(1 + \Phi(|x - y|)).$$

□

Proof of Theorem 1. We use the function f defined at the beginning of Section 1.5 with the following parameters: The constants R_1 and R_2 are specified by (1.7) and (1.10) respectively. For $r \geq 0$ we set $i(r) = 1$, $j(r) = r$ and

$$h(r) = \frac{1}{2}\int_0^r s\,\kappa(s)\,ds, \quad \text{where } \kappa \text{ is defined in Assumption 9.} \tag{1.50}$$

We fix initial values $(x, y) \in \mathbb{R}^{2d}$ and prove (1.11) for Dirac measures δ_x and δ_y. This is sufficient, since for general $\mu, \nu \in \mathcal{P}_V(\mathbb{R}^d)$ one can show, arguing similarly to [149, Theorem 4.8], that for any coupling γ of μ and ν we have

$$\mathcal{W}_{\rho_1}(\mu p_t, \nu p_t) \quad \leq \quad \int \mathcal{W}_{\rho_1}(\delta_x p_t, \delta_y p_t)\,\gamma(dx\,dy).$$

Let $U_t = (X_t, Y_t)$ be a reflection coupling with initial values (x, y), as defined in Section 1.4.2. We will argue that $E\left[e^{ct}\rho_1(X_t, Y_t)\right] \leq \rho_1(x, y)$ holds for any $t \geq 0$. Denote by $T = \inf\{t \geq 0 : X_t = Y_t\}$ the coupling time. Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. The process (Z_t) satisfies the SDE

$$\begin{aligned} dZ_t &= (b(X_t) - b(Y_t))\,dt + 2\,e_t\,\langle e_t, dB_t\rangle \quad \text{for } t < T, \\ dZ_t &= 0 \quad \text{for } t \geq T, \quad \text{where } e_t = Z_t/r_t. \end{aligned}$$

Until the end of the proof, all Itô equations and differential inequalities hold almost surely for $t < T$, even though we do not mention it every time. An application of Itô's formula shows that (r_t) satisfies the equation

$$dr_t \quad = \quad \langle e_t, b(X_t) - b(Y_t)\rangle\,dt + 2\,\langle e_t, dB_t\rangle \quad \text{for } t < T. \tag{1.51}$$

Let (L_t^x) denote the right-continuous local time of the semimartingale (r_t). Since f is a concave function, we can apply the general Itô-Tanaka formula of Meyer and Wang (cf. e.g. [93, Thm. 22.5] or [133, Ch. VI]) to conclude

$$\begin{aligned} f(r_t) - f(r_0) &= \int_0^t f_-'(r_s)\,\langle e_s, b(X_s) - b(Y_s)\rangle\,ds + 2\int_0^t f_-'(r_s)\,\langle e_s, dB_s\rangle \\ &\quad + \frac{1}{2}\int_{-\infty}^{\infty} L_t^x\,\mu_f(dx) \qquad \text{for } t < T, \end{aligned} \tag{1.52}$$

where f_-' denotes the left-derivative of f and μ_f is the non-positive measure representing the second derivative of f, i.e., $\mu_f(\,[x, y)\,) = f_-'(y) - f_-'(x)$ for $x \leq y$. Moreover, the generalized Itô formula implies for every measurable function $v : \mathbb{R} \to [0, \infty)$ the equality

$$\int_0^t v(r_s)\,d[r]_s \quad = \quad \int_{-\infty}^{\infty} v(x)\,L_t^x\,dx \quad \text{for any } t < T. \tag{1.53}$$

Observe that (1.53) implies that the Lebesgue measure of the set $\{0 \leq s \leq T : r_s \in \{R_1, R_2\}\}$, i.e., the time that (r_s) spends at the points R_1 and R_2 before coupling, is almost surely zero. Our function f is twice continuously differentiable on $(0,\infty)\setminus\{R_1,R_2\}$. The measure $\mu_f(dy)$ is non-positive and thus (1.53) implies

$$\int_{-\infty}^{\infty} L_t^x \mu_f(dx) \quad\le\quad \int_{-\infty}^{\infty} I_{\mathbb{R}\setminus\{R_1,R_2\}}(x) f''(x) L_t^x dx = 4\int_0^t f''(r_s)\, ds, \quad t<T.$$

We can conclude that a.s. the following differential inequalities hold for $t<T$:

$$\begin{aligned} df(r_t) &\le (f'(r_t)\,\langle e_t, b(X_t)-b(Y_t)\rangle + 2\,f''(r_t))\,dt + 2\,f'(r_t)\,\langle e_t, dB_t\rangle \quad (1.54)\\ &\le (-(\beta/2)\,f(r_t)\,I_{r_t<R_2} - (\xi/2)\,I_{r_t<R_1})\,dt + 2\,f'(r_t)\,\langle e_t, dB_t\rangle\,. \end{aligned}$$

For the second inequality, we have used that f is constant on $[R_2,\infty)$ and that inequality (1.49) holds on $(0,R_2)\setminus\{R_1\}$ with h given by (1.50). Moreover, using Assumption 9, we estimated

$$\langle e_t, b(X_t)-b(Y_t)\rangle = \langle Z_t/r_t, b(X_t)-b(Y_t)\rangle \quad\le\quad \kappa(r_t)\,r_t.$$

We now turn to the Lyapunov functions. Assumption 10 implies that a.s.

$$d\,(\epsilon\,V(X_t)+\epsilon\,V(Y_t)) \quad\le\quad 2\,C\,\epsilon - \lambda\,(\epsilon\,V(X_t)+\epsilon\,V(Y_t))\,dt + dM_t,$$

where (M_t) denotes a local martingale. If $r_t\ge R_1$, the definition of S_1 implies

$$2\,C\,\epsilon - \lambda\,(\epsilon\,V(X_t)+\epsilon\,V(Y_t)) \quad\le\quad -(\lambda/2)\,(\epsilon\,V(X_t)+\epsilon\,V(Y_t))\,.$$

If $r_t\ge R_2$, then by Assumption 11,

$$2C\epsilon - \lambda\,(\epsilon V(X_t)+\epsilon V(Y_t)) \quad\le\quad -(\alpha/2)\,f(r_t) - (\lambda/2)\,(\epsilon\,V(X_t)+\epsilon\,V(Y_t))\,,$$

where we have used that by (1.8) and (1.47), $\epsilon=\xi/(4C)$ and $\Phi(r)\ge f(r)$. We can conclude that a.s. ,

$$\begin{aligned} & d(\epsilon V(X_t)+\epsilon V(Y_t))\\ \le\ & ((\xi/2)I_{r_t<R_1} - (\alpha/2)f(r_t)I_{r_t\ge R_2} - (\lambda/2)(\epsilon V(X_t)+\epsilon V(Y_t)))\,dt + dM_t. \end{aligned}$$

Summarizing the above results, we can conclude that a.s., for $t<T$,

$$d\rho_1(X_t,Y_t) = df(r_t) + d\,(\epsilon\,V(X_t)+\epsilon\,V(Y_t)) \;\le\; -c\,\rho_1(X_t,Y_t)\,dt + dM'_t,$$

where (M'_t) denotes a local martingale and $c=\min(\alpha,\beta,\lambda)/2$. The product rule for semimartingales then implies a.s. for $t<T$:

$$d(e^{ct}\rho_1(X_t,Y_t)) \quad=\quad c\,e^{ct}\rho_1(X_t,Y_t)\,dt + e^{ct}\,d\rho_1(X_t,Y_t) \;\le\; e^{ct}\,dM'_t.$$

We introduce a sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ given by

$$T_n = \inf\{t\ge 0 : |X_t-Y_t|\le 1/n \quad\text{or}\quad \max(|X_t|\,,|Y_t|)\ge n\}.$$

We have $T_n\uparrow T$ a.s. by non-explosiveness. Therefore we finally obtain:

$$\begin{aligned} &\mathcal{W}_{\rho_1}(\delta_x p_t, \delta_y p_t) \le E\left[\rho_1(X_t,Y_t)I_{t<T}\right] = \lim_{n\to\infty} E\left[\rho_1(X_t,Y_t)I_{t<T_n}\right]\\ &\le \mathrm{e}^{-ct}\liminf_{n\to\infty} E\left[\mathrm{e}^{c(t\wedge T_n)}\,\rho_1(X_{t\wedge T_n},Y_{t\wedge T_n})\right] \le \mathrm{e}^{-ct}\,\mathcal{W}_{\rho_1}(\delta_x,\delta_y) \end{aligned}$$

□

Bounds for Example 4. The statement is a special case of Theorem 1. The only thing to verify is the lower bound

$$\beta \;\geq\; \sqrt{\kappa/\pi}\left(\int_0^{R_2} \exp\left(\kappa r^2/4\right)\,dr\right)^{-1}.$$

As $\int_0^\infty \exp(\kappa r^2/4)\,dr = \sqrt{\pi/\kappa}$, this follows from the definitions of Φ and ϕ:

$$\beta^{-1} \;=\; \int_0^{R_2} \phi(r)^{-1}\,\Phi(r)\,dr \;\leq\; \sqrt{\kappa/\pi}\int_0^{R_2} \exp\left(\kappa r^2/4\right)\,dr.$$

□

Proof of Corollary 1. It is well-known that, in our setup, the Markov transition kernels (p_t) admit a unique invariant measure π satisfying $\pi p_t = \pi$ for any $t \geq 0$ and $\int V(x)\,\pi(dx) \leq C/\lambda$, see e.g. [73]. In [75, Lemma 2.1] it is proven that for any probability measures ν_1 and ν_2 we have

$$\int_{\mathbb{R}^d} V(x)\;|\nu_1 - \nu_2|\,(dx) \;=\; \inf_\gamma \int \left[\,V(x) + V(y)\,\right]\; I_{x\neq y}\,\gamma(dx\,dy),$$

where the infimum is taken over all couplings γ with marginals ν_1 and ν_2 respectively. In our setup, this implies that for any $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ and $t \geq 0$,

$$\int_{\mathbb{R}^d} V(z)\;|\mu p_t - \pi|\,(dz) \;\leq\; \epsilon^{-1}\mathcal{W}_{\rho_1}(\mu p_t, \pi p_t) \;\leq\; \epsilon^{-1}\,e^{-ct}\,\mathcal{W}_{\rho_1}(\mu,\pi).$$

This implies the bound on the mixing time, since

$$\mathcal{W}_{\rho_1}(\delta_x,\pi) \leq \int \left[f(|x-y|) + \epsilon\,V(x) + \epsilon\,V(y)\,\right]\pi(dy) \leq R_2 + \epsilon\,V(x) + \epsilon\,C/\lambda.$$

□

Proof of Corollary 2. The proof relies on arguments from [51]. Let $x \in \mathbb{R}^d$. Assumption 10 implies that $\delta_x p_t \in \mathcal{P}_V(\mathbb{R}^d)$ for any $t \geq 0$ and hence $p_t g(x) = E_x[g(X_t)]$ is well defined and finite for any measurable g which is Lipschitz w.r.t. ρ_1. Fix $(x,y) \in \mathbb{R}^{2d}$ and $t \geq 0$, and let (X_t, Y_t) be an arbitrary coupling of $\delta_x p_t$ and $\delta_y p_t$. We bound the Lipschitz norm of $x \mapsto p_t g(x)$:

$$|p_t g(x) - p_t g(y)| \;\leq\; E_{(x,y)}[|g(X_t) - g(Y_t)|] \;\leq\; |g|_{\mathrm{Lip}(\rho_1)}\;E_{(x,y)}[\rho_1(X_t,Y_t)].$$

Since the above inequality holds for any coupling, Theorem 1 implies

$$|p_t g|_{\mathrm{Lip}(\rho_1)} \;\leq\; |g|_{\mathrm{Lip}(\rho_1)}\;e^{-ct}.$$

This estimate implies bounds on the bias of ergodic averages:

$$\begin{aligned}
&\left| E_x\left[\frac{1}{t}\int_0^t g(X_s)\,ds - \int g\,d\pi\right]\right| \leq \frac{1}{t}\int_0^t \int |p_s g(x) - p_s g(y)|\;\pi(dy)\,ds \\
&\leq \frac{1-e^{-ct}}{c\,t}\,|g|_{\mathrm{Lip}(\rho_1)} \int \rho_1(x,y)\,\pi(dy) \\
&\leq \frac{1-e^{-ct}}{c\,t}\,|g|_{\mathrm{Lip}(\rho_1)} \left(R_2 + \epsilon V(x) + \epsilon\frac{C}{\lambda}\right),
\end{aligned}$$

where we have used that f is bounded by R_2.

We now turn to the variance bound. Integrating (1.15) implies

$$E_x[V(X_t)^2] \;\leq\; C^*/\lambda^* + e^{-\lambda^* t}\,V(x)^2 \quad \text{for any } t \geq 0.$$

For reals a, b, c, the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ holds true. Hence

$$\begin{aligned}
\int\int \rho_1(y,z)^2\,p_t(x,dy)\,p_t(x,dz) \;&\leq\; 3\left(R_2^2 + 2\,\epsilon^2 \int V(y)^2\,p_t(x,dy)\right) \\
&\leq\; 3\left(R_2^2 + 2\,\epsilon^2\left[C^*/\lambda^* + e^{-\lambda^* t}\,V(x)^2\right]\right).
\end{aligned}$$

Let $A = 3\left(R_2^2 + 2\,\epsilon^2\left(C^*/\lambda^* + e^{-\lambda^* t}\,V(x)^2\right)\right)$. For $t \geq 0$ and $h \geq 0$,

$$\begin{aligned}
\mathrm{Var}_x\left[g(X_t)\right] &= \frac{1}{2}\int\int (\,g(y) - g(z)\,)^2\,p_t(x,dy)\,p_t(x,dz) \leq \frac{A}{2}\;|g|^2_{\mathrm{Lip}(\rho_1)}, \\
\mathrm{Var}_x\left[(p_h g)(X_t)\right] &\leq \frac{A}{2}\;|p_h g|^2_{\mathrm{Lip}(\rho_1)} \;\leq\; \frac{A}{2}\;|g|^2_{\mathrm{Lip}(\rho_1)}\;e^{-2\,c\,h}.
\end{aligned}$$

We get an estimate on the decay of correlations by Cauchy-Schwarz:

$$\begin{aligned}
&\mathrm{Cov}_x\left[g(X_t), g(X_{t+h})\right] = \mathrm{Cov}_x\left[g(X_t), (p_h g)(X_t)\right] \\
&\leq \mathrm{Var}_x\left[g(X_t)\right]^{1/2}\,\mathrm{Var}_x\left[(p_h g)(X_t)\right]^{1/2} \;\leq\; \frac{A}{2}\;|g|^2_{\mathrm{Lip}(\rho_1)}\;e^{-c\,h}.
\end{aligned}$$

Finally, we obtain the variance bound

$$\begin{aligned}
\mathrm{Var}_x\left[\frac{1}{t}\int_0^t g(X_s)ds\right] &= \frac{2}{t^2}\int_0^t\int_r^t \mathrm{Cov}_x\left[\,g(X_s)\,, g(X_r)\,\right]\,ds\,dr \\
&= \frac{A}{t^2}\;|g|^2_{\mathrm{Lip}(\rho_1)}\int_0^t\int_r^t e^{-c\,(s-r)}\,ds\,dr \;=\; \frac{A}{c\,t}\,|g|^2_{\mathrm{Lip}(\rho_1)}.
\end{aligned}$$

□

1.5.2 Proofs of results in Section 1.2.2

Proof of Theorem 2. We use the function f defined at the beginning of Section 1.5 with the following parameters: The constants R_1 and R_2 are specified by (1.22), we fix $\epsilon \in (0,\infty)$ satisfying (1.18), set $i(r) = \Phi(r)$ and $j(r) = r$ and define

$$h(r) \;=\; \frac{1}{2}\int_0^r s\,\kappa(s)\,ds + 2\,Q(\epsilon)\,r, \tag{1.55}$$

where Φ, κ and Q are given by (1.19), Assumption 9 and (1.17) respectively.

Fix initial values $(x, y) \in \mathbb{R}^{2d}$. It is enough to prove (1.23) for Dirac measures δ_x and δ_y, see the proof of Theorem 1 for details. Let $U_t = (X_t, Y_t)$ be a reflection coupling with initial values (x, y), as defined in Section 1.4.2. We will argue that $E\left[e^{ct}\rho_2(X_t, Y_t)\right] \leq \rho_2(x, y)$ holds for any $t \geq 0$. Denote by $T = \inf\{t \geq 0 : X_t = Y_t\}$ the coupling time. Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. The proof of Theorem 1 shows that $f(r_t)$ satisfies a.s.

$$df(r_t) \leq (f'(r_t)\,\langle e_t, b(X_t) - b(Y_t)\rangle + 2f''(r_t))\;dt + 2\,f'(r_t)\,\langle e_t, dB_t\rangle \tag{1.56}$$

for $t < T$, where $e_t = Z_t/r_t$. As in the proof of Theorem 1, the Lebesgue measure of the set $\{0 \leq s \leq T : r_s \in \{R_1, R_2\}\}$, i.e. the time that (r_t) spends at the points R_1 and R_2 before coupling, is almost surely zero. This justifies to write f' and f'' in the above inequality. Observe that Assumption 9 implies the upper bound

$$\langle e_t, b(X_s) - b(Y_s)\rangle = \langle Z_t/r_t, b(X_t) - b(Y_t)\rangle \leq \kappa(r_t)\,r_t.$$

The function f is constant on $[R_2, \infty)$, and $f(r) \leq \Phi(r)$. Moreover, on $(0, R_2)\setminus\{R_1\}$ the function f satisfies inequality (1.49). By (1.56), (1.49) and (1.55), we can conclude that a.s. for $t < T$,

$$\begin{aligned} df(r_t) \leq\; & (-4\,Q(\epsilon)\,f'(r_t) - \beta/2f(r_t)\,I_{r_t<R_2} - \xi/2f(r_t)\,I_{r_t<R_1})\,dt \\ & + 2\,f'(r_t)\,\langle e_t, dB_t\rangle\,. \end{aligned} \tag{1.57}$$

We now turn to the Lyapunov functions and set $G(x, y) = 1 + \epsilon\,V(x) + \epsilon\,V(y)$. By definition of the coupling in Section 1.4.2, we have a.s. for $t < T$:

$$\begin{aligned} dG(X_t, Y_t) \;=\; & (\epsilon\,\mathcal{L}V(X_t) + \epsilon\,\mathcal{L}V(Y_t))\;dt \\ & +\epsilon\,\langle\nabla V(X_t) + \nabla V(Y_t), dB_t\rangle - 2\,\epsilon\,\langle e_t, \nabla V(Y_t)\rangle\,\langle e_t, dB_t\rangle\,. \end{aligned} \tag{1.58}$$

Assumption 10 implies $\mathcal{L}V(X_t) + \mathcal{L}V(Y_t) \leq 2\,C - \lambda\,(V(X_t) + V(Y_t))$. Notice that by (1.18), (1.19) and (1.47) with $i(r) = \Phi(r)$,

$$2C\epsilon \leq \left(2\int_0^{R_1}\Phi(r)\,\phi(r)^{-1}\,dr\right)^{-1} = \xi/2.$$

Recall that $c = \min\left(\beta, \lambda, 4C\epsilon\lambda\right)/2$. Using the definitions (1.20) and (1.21) of the sets S_1 and S_2 respectively, we can conclude that a.s. for $t < T$:

$$\begin{aligned} d(\epsilon V(X_t) + \epsilon V(Y_t)) \leq\; & (\xi/2I_{r_t<R_1} - c\,G(X_t, Y_t)I_{r_t\geq R_2})\;dt \\ & + \epsilon\,\langle\nabla V(X_t) + \nabla V(Y_t), dB_t\rangle - 2\epsilon\,\langle e_t, \nabla V(Y_t)\rangle\,\langle e_t, dB_t\rangle\,. \end{aligned} \tag{1.59}$$

By (1.52) and (1.58), the covariation of $f(r_t)$ and $\epsilon\,V(X_t) + \epsilon\,V(Y_t)$ is, almost surely for $t < T$, given by:

$$d[\,f(r)\,, \epsilon\,V(X) + \epsilon\,V(Y)\,]_t \;=\; 2\,f'(r_t)\,\epsilon\,\langle\nabla V(X_t) - \nabla V(Y_t), e_t\rangle\;dt.$$

Using Cauchy-Schwarz and (1.17), we can derive the following bound for any $x, y \in \mathbb{R}^d$ with $x \neq y$:

$$\begin{aligned}\epsilon \left\langle \nabla V(x) - \nabla V(y), \frac{x-y}{|x-y|} \right\rangle &\leq (1 + \epsilon V(x) + \epsilon V(y)) \frac{\epsilon |\nabla V(x)| + \epsilon |\nabla V(y)|}{(1 + \epsilon V(x) + \epsilon V(y))} \\ &\leq 2\, Q(\epsilon)\, G(x,y).\end{aligned}$$

Hence, almost surely for $t < T$:

$$d[\, f(r)\,, \epsilon V(X) + \epsilon V(Y)\,]_t \;\leq\; 4\, Q(\epsilon)\, f'(r_t)\, G(X_t, Y_t)\, dt. \tag{1.60}$$

The product rule for semimartingales implies almost surely for $t < T$:

$$d\left(f(r_t)G(X_t, Y_t)\right) = G(X_t, Y_t)\, df(r_t) + f(r_t)\, dG(X_t, Y_t) + [f(r), G(X, Y)]_t.$$

By (1.57), we have

$$\begin{aligned}G(X_t, Y_t) df(r_t) \;\leq\; & (-\beta/2\, \rho_2(X_t, Y_t)\, I_{r_t < R_2} - \xi/2\, \rho_2(X_t, Y_t)\, I_{r_t < R_1})\, dt \\ & -4\, Q(\epsilon)\, f'(r_t) G(X_t, Y_t)\, dt + dM_t^1,\end{aligned} \tag{1.61}$$

where (M_t^1) is a local martingale. Moreover, (1.59) implies

$$f(r_t)\, dG(X_t, Y_t) \;\leq\; [\xi/2\, f(r_t)\, I_{r_t < R_1} - c\, \rho_2(X_t, Y_t)\, I_{r_t \geq R_2}]\, dt + dM_t^2, \tag{1.62}$$

where (M_t^2) is again a local martingale. Observe that $G \geq 1$. Combining (1.60), (1.61) and (1.62) we can conclude a.s. for $t < T$:

$$\begin{aligned}d\rho_2(X_t, Y_t) \;&\leq\; -c\, \rho_2(X_t, Y_t) + dM_t, \\ d\left(e^{ct} \rho_2(X_t, Y_t)\right) \;&=\; c\, e^{ct}\, \rho_2(X_t, Y_t)\, dt + e^{ct}\, d\rho_2(X_t, Y_t) \;\leq\; e^{ct}\, dM_t,\end{aligned}$$

where (M_t) is a local martingale. We can finish the proof of (1.23) using a stopping argument, see the end of the proof of Theorem 1 for details. □

Proof of Corollary 4. Analogously to the proof of Corollary 2, we can conclude that $p_t g(x)$ is finite for any function g which is Lipschitz w.r.t. ρ_2, any $x \in \mathbb{R}^d$ and $t \geq 0$. Moreover,

$$|p_t g|_{\mathrm{Lip}(\rho_2)} \;\leq\; |g|_{\mathrm{Lip}(\rho_2)}\, e^{-ct} \qquad \text{holds for any } t \geq 0.$$

In particular, for any $x, y \in \mathbb{R}^d$ we can conclude that

$$\begin{aligned}|p_t g(x) - p_t g(y)| \;&\leq\; |g|_{\mathrm{Lip}(\rho_2)}\, e^{-ct} \rho_2(x, y) \\ &\leq\; |g|_{\mathrm{Lip}(\rho_2)}\, e^{-ct}\, |x - y|\, (1 + \epsilon V(x) + \epsilon V(y))\,,\end{aligned}$$

where we used $f(r) \leq r$. If the map $x \mapsto p_t g(x)$ is differentiable at $x \in \mathbb{R}^d$, we can deduce the gradient bound (1.25). □

1.5.3 Proofs of results in Section 1.2.3

Proof of Theorem 3. In contrast to the proofs above, we do not use the function f defined in the beginning of Section 1.5, but the one constructed in [51], i.e., we set

$$f(r) = \int_0^r \phi(s)\, g(s \wedge R_2)\, ds,$$

where ϕ and g are defined as

$$\phi(r) = \exp\left(-\frac{1}{2}\int_0^r u\,\kappa^+(u)\,du\right) \qquad \text{and} \qquad g(r) = 1 - \frac{c}{2}\int_0^r \Phi(s)\,\phi(s)^{-1}\,ds.$$

The function Φ and the constant c are given by

$$\Phi(r) = \int_0^r \phi(s)\,ds \qquad\qquad \text{and} \qquad\qquad c^{-1} = \int_0^{R_2} \Phi(s)\,\phi(s)^{-1}\,ds.$$

The constants R_1 and R_2 are defined in (1.32) and (1.33) respectively. Notice that by definition, $\kappa^+(r) = 0$ for any $r \geq R_1$ and thus f is linear on the interval $[R_2, \infty)$. The function f is twice continuously differentiable on $(0, R_2)$, and

$$2\, f''(r) = -\, r\,\kappa^+(r)\, f'(r) -\, c\,\Phi(r) \leq -\, r\,\kappa^+(r)\, f'(r) - c\, f(r). \tag{1.63}$$

We now prove (1.29) and fix initial values $(x, y) \in \mathbb{R}^{2d}$ as well as a small constant $\delta > 0$. The coupling $U_t = (X_t, Y_t)$, defined in Section 1.4.3, yields the upper bound

$$\mathcal{W}_{\rho_0}(\mu_t^x, \mu_t^y) \;\leq\; E[\rho_0(X_t, Y_t)] \;=\; E[f(|X_t - Y_t|)].$$

Let $\gamma = c - |\tau|\, K$. Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. We will argue that there is a constant $C > 0$, independent of δ, such that

$$e^{\gamma\, t} E[f(r_t)] \leq f(r_0) + e^{\gamma\, t}\, C\,\delta \qquad \text{holds true for any } t \geq 0. \tag{1.64}$$

From this inequality one can then conclude, that for any $t \geq 0$ we have

$$\mathcal{W}_{\rho_0}(\mu_t^x, \mu_t^y) \leq e^{-\gamma\, t}\, \rho_0(x, y) + C\,\delta,$$

which finishes the proof of (1.29) since $\delta > 0$ can be chosen arbitrarily small. Moreover (1.30) directly follows from (1.29) and the inequality

$$r\,\phi(R_1) \leq \Phi(r) \leq 2\, f(r) \leq 2\,\Phi(r) \leq 2\, r.$$

We now show (1.64). By definition of the coupling in Section 1.4.3,

$$dZ_t \;=\; \left(b^x(t, X_t) - b^y(t, Y_t)\right)\, dt + 2\ \mathrm{rc}(U_t)\, e_t\, dW_t,$$

where $W_t = \int_0^t \langle e_s, dB_s^1\rangle$ is a one dimensional Brownian motion. Notice that whenever $r_t < \delta/2$, we have $\mathrm{rc}(U_t) = 0$ by definition. Using an approximation argument, cf.

the proof of Lemma 8 further below or arguing similarly to [51, Lemma 6.2], one can show that r_t satisfies almost surely the equation

$$dr_t = \langle \tilde{e}_t, b^x(t, X_t) - b^y(t, Y_t)\rangle \; dt + 2 \; \mathrm{rc}(U_t) \, dW_t, \tag{1.65}$$

where $\tilde{e}_t = Z_t/r_t$ for $r_t \neq 0$, $\tilde{e}_t = (b^x(t, X_t) - b^y(t, Y_t))/ \left| b^x(t, X_t) - b^y(t, Y_t) \right|$ if $r_t = 0$ and $|b^x(t, X_t) - b^y(t, Y_t)| > 0$, and $\tilde{e}_t$ is an arbitrary unit vector otherwise. Similarly as in the proof in Section 1.5.1, we now apply the Itô-Tanaka formula for semimartingales to conclude that almost surely,

$$\begin{aligned} f(r_t) - f(r_0) \;\; &= \;\; \int_0^t f'_-(r_s) \; \langle \tilde{e}_s, b^x(s, X_s) - b^y(s, Y_s)\rangle \; ds \\ &\quad + 2 \int_0^t \mathrm{rc}(U_s) \, f'_-(r_s) \, dW_s + \frac{1}{2} \int_{-\infty}^{\infty} L_t^x \, \mu_f(dx), \end{aligned}$$

where L_t^x is the right-continuous local time of (r_t) and μ_f is the non-positive measure representing the second derivative of f. By (1.53), the Lebesgue measure of the set $\{0 \leq s \leq t : r_s \in \{R_1, R_2\}\}$ is almost surely zero. Since f is twice continuously differentiable, except possibly at R_1 and R_2, we can replace f'_- by f' in the equation above. Moreover, since f is concave, the measure of the points R_1 and R_2 w.r.t. μ_f is non-positive. Hence by (1.53),

$$\int_{-\infty}^{\infty} L_t^x \, \mu_f(dx) \;\; \leq \;\; \int_0^t f''(r_s) \, d[r]_s \;\; = \;\; 4 \int_0^t \mathrm{rc}(U_s)^2 \, f''(r_s) \, ds \quad \text{a.s., and thus}$$

$$\begin{aligned} f(r_t) \;\; &= \;\; f(r_0) \; + \; M_t \; + \; \int_0^t H_s \, ds, \qquad \text{where} && (1.66) \\ M_t \;\; &= \;\; 2 \int_0^t \mathrm{rc}(U_s) f'(r_s) dW_s, \qquad \text{and} && (1.67) \\ H_s \;\; &\leq \;\; f'(r_s) \; \langle \tilde{e}_s, b^x(t, X_s) - b^y(t, Y_s)\rangle \; + \; 2 \; \mathrm{rc}(U_s)^2 \, f''(r_s). && (1.68) \end{aligned}$$

We can bound the inner product using the definitions of b^x, b^y and κ, as well as the Lipschitz bounds on b and ϑ:

$$\begin{aligned} &\langle \tilde{e}_t, b^x(t, X_t) - b^y(t, Y_t)\rangle \qquad\qquad (1.69) \\ &= \langle \tilde{e}_t, b(X_t) - b(Y_t)\rangle + \tau \left\langle \tilde{e}_t, \int \vartheta(X_t, z) \mu_t^x(dz) - \int \vartheta(Y_t, z) \mu_t^y(dz) \right\rangle \\ &\leq I_{r_t \geq \delta} r_t \kappa(r_t) + I_{r_t < \delta} \, |b|_{\mathrm{Lip}} \, \delta + |\tau| \, L(r_t + \mathcal{W}^1(\mu_t^x, \mu_t^y)). \end{aligned}$$

Notice that $\mathcal{W}^1(\mu_t^x, \mu_t^y) \leq E[r_t]$. Remembering that by (1.31), $K = \frac{4L}{\phi(R_1)}$ and combining (1.69) with the inequality $r \leq 2 \, f(r)/\phi(R_1)$, we obtain

$$\begin{aligned} &\langle \tilde{e}_t, b^x(t, X_t) - b^y(t, Y_t)\rangle \\ &\quad \leq \;\; I_{r_t \geq \delta} \, r_t \, \kappa(r_t) + I_{r_t < \delta} \; |b|_{\mathrm{Lip}} \; \delta + |\tau| \; K/2 \, (f(r_t) + E[f(r_t)]). \end{aligned}$$

The product rule for semimartingales shows that

$$d(e^{\gamma t} f(r_t)) \;=\; e^{\gamma t} dM_t \;+\; e^{\gamma t} (\gamma f(r_t) \;+\; H_t)\, dt.$$

Using that $\gamma = c - |\tau|\, K$ and the bound $f' \leq 1$, we can conclude that

$$\begin{aligned} d(e^{\gamma t} f(r_t)) \;\leq\; & e^{\gamma t} dM_t + e^{\gamma t}\, |\tau|\, K/2\, (E[f(r_t)] - f(r_t))\, dt \\ & + e^{\gamma t} I_{r_t<\delta} \left(c\, f(r_t) + |b|_{\mathrm{Lip}}\, \delta \right) dt \\ & + e^{\gamma t} I_{r_t \geq \delta} (c\, f(r_t) \;+ r_t\, \kappa(r_t)\, f'(r_t) \;+ 2 f''(r_t))\, dt. \end{aligned} \tag{1.70}$$

Here we used that $f'' \leq 0$ and $\mathrm{rc}(U_t) = 1$ whenever $r_t \geq \delta$. We now argue that for any $r \in (0, \infty) \setminus \{R_2\}$ we have

$$c\, f(r) + r\, \kappa(r)\, f'(r) + 2\, f''(r) \leq 0. \tag{1.71}$$

For $r \in (0, R_2)$ this inequality follows directly from the definition of f, see (1.63). For $r > R_2$ we have $f''(r) = 0$, but $\kappa(r)$ is sufficiently negative instead: First notice that for $r \geq R_1$, $\phi(r)$ is constant and hence $\Phi(r) = \Phi(R_1) + \phi(R_1)(r - R_1)$. Analogously to [51, Theorem 2.2.] we get

$$\begin{aligned} \int_{R_1}^{R_2} \Phi(s)\phi(s)^{-1} ds &= \int_{R_1}^{R_2} (\Phi(R_1) + \phi(R_1)(s - R_1))\phi(R_1)^{-1} ds \\ &= \Phi(R_1)\phi(R_1)^{-1}(R_2 - R_1) + (R_2 - R_1)^2/2 \\ &\geq (R_2 - R_1)(\Phi(R_1) + \phi(R_1)(R_2 - R_1))\phi(R_1)^{-1}/2 \\ &= (R_2 - R_1)\Phi(R_2)\phi(R_1)^{-1}/2. \end{aligned}$$

For $r \geq R_2$ we have $f'(r) = \phi(R_1)/2$, and thus we get

$$f'(r)\, r\, \kappa(r) \leq -2 \frac{\phi(R_1)}{R_2 - R_1} \frac{r}{R_2} \leq -2 \frac{\phi(R_1)}{R_2 - R_1} \frac{\Phi(r)}{\Phi(R_2)} \leq -c\, \Phi(r) \leq -c\, f(r),$$

where we used the definition of R_2 in (1.33) and the fact that $c^{-1} = \int_0^{R_2} \Phi(s)/\phi(s)\, ds$. Hence (1.71) holds for any $r \in (0, \infty) \setminus \{R_2\}$. By (1.70), we conclude that

$$E[e^{\gamma t} f(r_t) - f(r_0)] \leq \delta \left(|b|_{\mathrm{Lip}} + c \right) \int_0^t e^{\gamma s}\, ds,$$

where we used that $f(r) \leq r$. □

We now prepare the proof of Theorem 4 by providing a priori bounds. Notice that Assumption 14 implies that there are constants $A, B > 0$ s.t.

$$|\vartheta(x, y)| \leq A + B\, (\, |x| + |y|\,) \qquad \text{for any } x, y \in \mathbb{R}^d. \tag{1.72}$$

Lemma 3 (A priori bounds)**.** *Let $V(x) = 1 + |x|^2$. Suppose that Assumptions 14 and 15 hold true. Then there is a constant $C \in (0, \infty)$ such that for any $\tau \in \mathbb{R}$ with $|\tau| \leq \lambda/(8B)$, $x \in \mathbb{R}^d$ and $t \geq 0$, a solution (X_t) of (1.26) with $X_0 = x$ satisfies*

$$dV(X_t) \;\; \leq \;\; \left[(C - \lambda\, V(X_t)) + \left(2\,|\tau|\, B\, |X_t|\, E[|X_t|] - \frac{\lambda}{4}\, |X_t|^2\right)\right]\, dt + 2\,\langle X_t, dB_t\rangle\, .$$

In particular, $\quad E\,[V(X_t)] \;\leq\; C/\lambda + e^{-\lambda\, t}\, V(x).$

Proof of Lemma 3. Let $M_t = \int_0^t \langle X_s, dB_s\rangle$. By Itô's formula,

$$\frac{1}{2}\, dV(X_t) = \langle X_t, b(X_t)\rangle\; dt + \tau\langle X_t, \int \vartheta(X_t, y)\, \mu_t^x(dy)\rangle\, dt + d\, dt + dM_t.$$

Using Assumption 15, inequality (1.72) and $|\tau| \leq \lambda/(8B)$, we conclude

$$\begin{aligned}\frac{1}{2} dV(X_t) &\leq [C_1 - \lambda\, |X_t|^2 + |\tau|\, (A\, |X_t| + B(|X_t|^2 + |X_t|\, E_x[|X_t|]))]dt + dM_t \\ &\leq \left[C_2 - \frac{5}{8}\lambda\, |X_t|^2 + |\tau|\, B\, |X_t|\, E_x[|X_t|]\right] dt + dM_t,\end{aligned}$$

with $C_1 = \sup_{|x| \leq D} \left|\langle x, b(x)\rangle + \lambda\, |x|^2 + d\right|$ and a constant $C_2 > C_1$ s.t.

$$-\lambda\, r^2/4 + |\tau|\; A\, r \;\;\leq\;\; C_2 - C_1 \quad \text{for any } r \in \mathbb{R}_+.$$

It follows that we can find a constant $C > 0$ such that

$$dV(X_t) \leq \left[C - \lambda V(X_t) + 2\, |\tau|\;\, B\;\, |X_t|\;\, E[|X_t|]\; - \frac{\lambda}{4}\; |X_t|^2\right] dt + 2dM_t.$$

Applying the product rule for semimartingales we get

$$d(e^{\lambda\, t}\, V(X_t)) \leq e^{\lambda\, t}\; \left[C + 2\;\, |\tau|\;\, B\;\, |X_t|\;\, E[|X_t|]\; - \frac{\lambda}{4}\; |X_t|^2\right]\; dt + 2\, e^{\lambda\, t}\, dM_t.$$

We introduce the stopping times $T_n = \inf\{t \geq 0 : |X_t| \geq n\}$ and remark that almost surely, $T_n \uparrow \infty$, since the solution (X_t) is non-explosive. Using Fatou's Lemma and monotone convergence, we can conclude that

$$\begin{aligned}E_x[e^{\lambda\, t}\, V(X_t)] &\leq \liminf_{n\to\infty} E_x[e^{\lambda\, (t \wedge T_n)}\, V(X_{t\wedge T_n})] \\ &\leq V(x) + \int_0^t e^{\lambda\, s}\; \left[C + 2\;\, |\tau|\;\, B\, E[|X_s|]^2\; - \frac{\lambda}{4}\, E[|X_s|^2]\right]\; ds.\end{aligned}$$

This concludes the proof, since by assumption, $|\tau| \leq \lambda/(8B)$. □

Proof of Theorem 4. We use the Lyapunov function $V(x) = 1 + |x|^2$. Assumption 15 provides a rate λ and Lemma 3 a constant C. We follow Section 1.2.2 defining

$$\begin{aligned} S_1 &= \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : V(x) + V(y) \le 2C/\lambda\right\}, \\ S_2 &= \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : V(x) + V(y) \le 8C/\lambda\right\}, \\ R_i &= \sup\left\{|x-y| : (x,y) \in S_i\right\}, \quad i = 1,2. \end{aligned}$$

We define f as in the beginning of Section 1.5 w.r.t. the following parameters:

$$h(s) = \frac{1}{2}\int_0^s r\kappa(r)dr + 2s, j(s) = s, i(s) = \Phi(s), c = \frac{1}{4}\min(\beta,\lambda), \epsilon = \frac{\xi}{4C},$$

with κ given by Assumption 9. We assume $|\tau| < \lambda/(8B)$ so that Lemma 3 applies.

Fix initial values $(x,y) \in \mathbb{R}^{2d}$, as well as a small constant $\delta > 0$. The coupling $U_t = (X_t, Y_t)$ defined in Section 1.4.3 yields the upper bound

$$\mathcal{W}_{\rho_2}(\mu_t^x, \mu_t^y) \le E[\rho_2(X_t, Y_t)].$$

Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. Equations (1.65), (1.66), (1.67) and (1.68) are still valid in our setup. By (1.69) we can conclude that

$$H_t \le \left(I_{r_t \ge \delta} f'(r_t)\kappa(r_t)r_t + I_{r_t<\delta}\,|b|_{\mathrm{Lip}}\,\delta\right) + 2\,\mathrm{rc}(U_t)^2 f''(r_t) + |\tau|\,L(r_t + E[r_t]).$$

By definition, f is constant on $[R_2, \infty)$, and, for $r \in (0, R_2) \setminus \{R_1\}$,

$$2\,f''(r) \le -f'(r)\,[\kappa(r)\,r\, + 4] - (\beta/2)\,f(r) - (\xi/2)\,f(r)\,I_{r<R_1}.$$

Using that f is concave with $f(r) \le r$ and $\mathrm{rc}(U_t) = 1$ for $r_t \ge \delta$, we obtain

$$\begin{aligned} df(r_t) \le &\left[-\frac{\beta}{2}\,f(r_t)I_{r_t<R_2} - \frac{\xi}{2}\,f(r_t)\,I_{r_t<R_1} - 4\,\mathrm{rc}(U_t)^2\,f'(r_t)\right]\,dt \\ &+ \left(|b|_{\mathrm{Lip}} + \frac{\beta}{2} + \frac{\xi}{2}\right)\,\delta\,dt + |\tau|\;L\,(r_t + E[r_t])\,dt + 2\;\mathrm{rc}(U_t)\,f'(r_t)\,dW_t. \end{aligned} \tag{1.73}$$

Next, we observe that Lemma 3 implies

$$\begin{aligned} dV(X_t) &\le [C - \lambda\,V(X_t)]\;dt + 2\,|\tau|\;B\,V(X_t)\,E[V(X_t)]\,dt + 2\,\langle X_t, dB_t\rangle\,, \\ dV(Y_t) &\le [C - \lambda\,V(Y_t)]\;dt + 2\,|\tau|\;B\,V(Y_t)\,E[V(Y_t)]\,dt + 2\left\langle Y_t, d\hat{B}_t\right\rangle, \end{aligned}$$

where (B_t) and $(\hat{B}_t)$ are the Brownian motions defined in (1.46). Let

$$G(x,y) = 1 + \epsilon\,V(x) + \epsilon\,V(y).$$

The set S_1 is chosen such that $2\,C\,\epsilon - \lambda\,\epsilon\,V(X_t) + \lambda\,\epsilon\,V(Y_t) \le 0$ whenever $r_t \ge R_1$. For $r_t \ge R_2$ we have

$$\begin{aligned} 2\,C\,\epsilon - \lambda\,\epsilon\,V(X_t) + \lambda\,\epsilon\,V(Y_t) &\le -2\,C\,\epsilon - (\lambda/2)\,\epsilon\,V(X_t) - (\lambda/2)\,\epsilon\,V(Y_t) \\ &\le -\min(\beta/2, \lambda/2)\,G(X_t, Y_t), \end{aligned}$$

since $\epsilon = \xi/(4C)$ and $\xi \geq \beta$. We conclude that

$$dG(X_t, Y_t) \leq I_{r_t<R_1} 2\, C\, \epsilon - I_{r_t \geq R_2} \min(\beta/2, \lambda/2)\, G(X_t, Y_t)\, dt + \epsilon\, 2\, \langle X_t, dB_t\rangle$$
$$+ 2\epsilon\, |\tau|\, B\, [V(X_t)\, E[V(X_t)] + V(Y_t)\, E[V(Y_t)]]\, dt + \epsilon\, 2 \left\langle Y_t, d\hat{B}_t \right\rangle. \quad (1.74)$$

Note that $|\nabla V(x)| = 2|x| \leq V(x)$. Therefore, and by (1.66) and (1.67), we obtain similarly to (1.60):

$$\begin{aligned} d[f(r), G(X,Y)]_t &= 2\, \mathrm{rc}(U_t)^2\, f'(r_t)\, \epsilon \langle \nabla V(X_t) - \nabla V(Y_t),\, e_t\rangle\, dt \\ &\leq 2\, \mathrm{rc}(U_t)^2\, f'(r_t)\, G(X_t, Y_t)\, dt. \end{aligned} \quad (1.75)$$

Using the product rule together with (1.73),(1.74) and (1.75), we see that

$$\begin{aligned} d\rho_2(X_t, Y_t) &= d\, (f(r_t) G(X_t, Y_t)) \\ &= G(X_t, Y_t) df(r_t) + f(r_t) dG(X_t, Y_t) + d[f(r), G(X,Y)]_t \\ &\leq -\min(\beta/2, \lambda/2)\, f(r_t)\, G(X_t, Y_t)\, dt + |\tau|\ L\, G(X_t, Y_t)\, (r_t + E[r_t])\, dt \\ &\quad +2\, \epsilon\ |\tau|\ B\, f(r_t)\, [V(X_t)\, E[V(X_t)] + V(Y_t)\, E[V(Y_t)]]\ dt \\ &\quad +G(X_t, Y_t) \left(|b|_{\mathrm{Lip}} + (\beta+\xi)/2\right)\, \delta\, dt + d\tilde{M}_t, \end{aligned}$$

where $(\tilde{M}_t)$ denotes a local martingale. We further bound the perturbation terms originating from the nonlinearity. For $r < R_2$, inequality (1.48) holds true and thus there is a constant $K_0 \in (0, \infty)$ s.t. for any $x, y \in \mathbb{R}^d$ we have

$$|x-y| \ \leq\ K_0\, f(|x-y|)\, (1 + \epsilon\, V(x) + \epsilon\, V(y)) \ =\ K_0\, \rho_2(x,y).$$

Hence, we can bound

$$|\tau|\ L\, G(X_t, Y_t)\, (r_t + E[r_t]) \leq |\tau|\ L\, K_0\, (\rho_2(X_t, Y_t) + G(X_t, Y_t) E[\rho_2(X_t, Y_t)]).$$

Moreover,

$$\epsilon\, V(X_t)\, E[V(X_t)] + \epsilon\, V(Y_t)\, E[V(Y_t)] \ \leq\ \epsilon^{-1}\, E[G(X_t, Y_t)]\, G(X_t, Y_t).$$

Recall that $2c = \min(\beta/2, \lambda/2)$. Hence by the bounds above,

$$d(e^{ct}\rho_2(X_t, Y_t)) \ =\ c\, \rho_2(X_t, Y_t)\, e^{ct}\, dt + e^{ct} d\rho_2(X_t, Y_t) \ \leq\ e^{ct} J_t\, dt \ +\ e^{ct} d\tilde{M}_t,$$

where

$$\begin{aligned} J_t = &-c\rho_2(X_t, Y_t)\ +\ |\tau|\ L\, K_0\, (\rho_2(X_t, Y_t) + G(X_t, Y_t) E[\rho_2(X_t, Y_t)]) \\ &+\ 2\ |\tau|\ B\, \epsilon^{-1}\, E[G(X_t, Y_t)]\, \rho_2(X_t, Y_t)\ + G(X_t, Y_t) \left(|b|_{\mathrm{Lip}} + (\beta+\xi)/2\right) \delta. \end{aligned}$$

Optional stopping and Fatou's lemma now shows that

$$E[e^{ct}\rho_2(X_t, Y_t)] \ \leq\ \rho_2(x,y) + \int_0^t e^{cs}\, E[J_s]\, ds.$$

Using the a priori bounds from Lemma 3, we see that there is a constant $C_1 \in (0, \infty)$, not depending on δ, such that

$$\Big(|b|_{\mathrm{Lip}} + (\beta + \xi)/2\Big) \int_0^t e^{c\,s}\, E[G(X_s, Y_s)]\, ds \;\leq\; C_1.$$

Since $G \geq 1$, we can conclude that

$$\begin{aligned} |\tau|\; L\, K_0 \int_0^t \left(E[\rho_2(X_s, Y_s)] + E[G(X_s, Y_s)] E[\rho_2(X_s, Y_s)]\right)\, e^{c\,s}\, ds \\ +2\; |\tau|\; B\, \epsilon^{-1} \int_0^t E[G(X_s, Y_s)]\, E[\rho_2(X_s, Y_s)]\, e^{c\,s}\, ds \\ \leq\;\; |\tau|\, C_2 \int_0^t E[G(X_s, Y_s)]\, E[\rho_2(X_s, Y_s)]\, e^{c\,s}\, ds, \end{aligned}$$

where $C_2 = 2\,(L\, K_0 + B/\epsilon)$. Moreover, the a priori estimates imply

$$\begin{aligned} &\int_0^t e^{c\,s}\; E[G(X_s, Y_s)]\, E[\rho_2(X_s, Y_s)]\, ds \\ &\leq C_3 \int_0^t e^{c\,s}\, E[\rho_2(X_s, Y_s)]\, ds + C_4(x, y) \int_0^t e^{(c-\lambda)\,s}\, E[\rho_2(X_s, Y_s)]\, ds, \end{aligned}$$

where $C_3 = 1 + \epsilon\, 2\, C/\lambda$ and $C_4(x, y) = \epsilon\, V(x) + \epsilon\, V(y)$. If τ is sufficiently small, i.e., if $|\tau|\, C_2(C_3 + C_4(x, y)) \leq c$, we can conclude that for any $\delta > 0$,

$$\mathcal{W}_{\rho_2}(\mu_t^x, \mu_t^y) \leq E[\rho_2(X_t, Y_t)] \leq e^{-c\,t}\, \rho_2(x_0, y_0) + C_1\, \delta.$$

However, observe that C_4 depends on the initial values x and y, i.e. we get a local contraction in the sense that for a given $R > 0$, we can find a constant $\tau_0 \in (0, \infty)$, such that (1.34) holds for all $|\tau| \leq \tau_0$ and initial values $x, y \in \mathbb{R}^d$ with $|x|\,, |y| \leq R$. Inequality (1.35) follows readily from (1.34) and the definition of K_0.

In order to obtain a related statement which is valid for any initial condition, see (1.36), we assume $|\tau|\, C_2 C_3 < c$. Similarly as above, we obtain

$$\begin{aligned} &E[\rho_2(X_t, Y_t)] \\ &\leq e^{-c\,t}\, \rho_2(x_0, y_0) + C_1\, \delta + e^{-c\,t}\, |\tau|\; C_2\, C_4(x, y) \int_0^t e^{(c-\lambda)\,s}\, E[\rho_2(X_s, Y_s)]\, ds \end{aligned}$$

Using once again the apriori estimates and the bound $f \leq R_2$, we see that

$$\int_0^t e^{(c-\lambda)\,s} E[\rho_2(X_s, Y_s)] ds \leq R_2(1 + 2\,\epsilon\, C/\lambda + \epsilon V(x) + \epsilon V(y)) \int_0^t e^{(c-\lambda)\,s} ds.$$

Since $\lambda > c$, there is a constant $K_1 \in (0, \infty)$, neither depending on the initial values (x, y) nor on δ, such that

$$E[\rho_2(X_t, Y_t)] \leq e^{-c\,t}\, \rho_2(x_0, y_0) + C_1\, \delta + e^{-c\,t}\, K_1\, (\epsilon V(x) + \epsilon V(y))^2.$$

Since $\delta > 0$ is arbitrary, we have shown (1.36). $\square$

1.5.4 Proofs of results in Section 1.2.4

Before proving Theorem 5, we include a proof of Lemma 2 from Section 1.3.6 that is based on [74, Section 4].

Proof of Lemma 2. The function H is C^2 with strictly positive first derivative, and thus the inverse function $H^{-1} : [0, \infty] \to [l, \infty]$ is also strictly increasing and C^2. We define a function $G : [l, \infty) \times [0, \infty) \to [0, \infty)$ by

$$G(x, t) = H^{-1}(\, H(x) \, + c\, t\,).$$

Observe that for any fixed $t \geq 0$ the map $x \mapsto G(x, t)$ is a concave C^2 function on (l, ∞), which can be seen by the following computation:

$$\partial_x^2 G \; = \partial_x \left(\frac{\eta \circ G}{\eta} \right) \; = \; \frac{(\eta\eta') \circ G}{\eta^2} - \frac{(\eta \circ G)\, \eta'}{\eta^2} \; \leq \; 0.$$

Since $x \mapsto G(x, t)$ is concave, Itô's formula shows that almost surely,

$$dG(Z_t, t) \; \leq \; \partial_t G(Z_t, t)\, dt + \partial_x G(Z_t, t)\, dA_t + dW_t,$$

where (W_t) denotes a local martingale. Observe that $\partial_t\, G = c\, \eta \circ G > 0$ and $\partial_x\, G = \frac{\eta \circ G}{\eta} > 0$. Using our Assumption (1.41), we can conclude that a.s.

$$dG(Z_t, t) \leq dW_t \qquad \text{for } t < T.$$

Let $(T_n)_{n \in \mathbb{N}}$ be a localizing sequence for (W_t) with $T_n \uparrow \infty$. We see

$$\begin{aligned} E[\, G(\, Z_{t \wedge T}\, , t \wedge T\,)\,] &= E[\, \liminf_{n \to \infty} \, G(\, Z_{t \wedge T \wedge T_n}\, , t \wedge T \wedge T_n\,)\,] \\ &\leq \liminf_{n \to \infty} E\, [\, G(\, Z_{t \wedge T \wedge T_n},\, t \wedge T \wedge T_n\,)\,] \leq E[\, G(\, Z_0,\, 0\,)\,] = E[\, Z_0\,]. \end{aligned}$$

Since H is nonnegative and H^{-1} is increasing, we get

$$E[\, H^{-1}(\, c\, (t \wedge T)\,)\,] \leq E[\, G(\, Z_{t \wedge T},\, t \wedge T\,)\,] \leq E[\, Z_0\,] < \infty.$$

Since the inequality holds for any $t \geq 0$ and $H^{-1}(t) \to \infty$ as $t \to \infty$, the time T is a.s. finite, and we can finish the proof using Fatou's lemma:

$$E[\, H^{-1}(\, c\, T\,)\,] \leq E[\, G(\, Z_T,\, T\;)] \leq \liminf_{t \to \infty} E[\, G(\, Z_{t \wedge T},\, t \wedge T\,)\,] \leq E[\, Z_0\,].$$

□

Proof of Theorem 5. We use the function f defined in the beginning of Section 1.5 with the following parameters: $i \equiv 1$ constant, $j = \eta$, and $h(r) = \frac{1}{2} \int_0^r s\, \kappa(s)\, ds$, where κ is defined in Assumption 9.

We now prove (1.38). Let $U_t = (X_t, Y_t)$ be a reflection coupling with initial values (x, y), as defined in Section 1.4.2. Denote by $T = \inf\{t \geq 0 : X_t = Y_t\}$ the coupling time. We will argue that the stochastic process $(\rho_1(X_t, Y_t))$ satisfies the conditions of Lemma 2, except that the map $t \mapsto \rho_1(X_t, Y_t)$ is not continuous at $t = T$. Nevertheless, this obstacle can be overcome by a stopping argument. Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. Following the lines of the proof of Theorem 1, one can show that a.s. for $t < T$, $f(r_t)$ satisfies

$$\begin{aligned} df(r_t) &\leq [f'(r_t)\, \langle e_t, b(X_t) - b(Y_t)\rangle + 2f''(r_t)]\, dt + 2\, f'(r_t)\, \langle e_t, dB_t\rangle \\ &\leq [-\beta/2\, \eta(\, f(r_t)\,)\, I_{r_t<R_2} - \xi/2\, I_{r_t<R_1}]\, dt + 2\, f'(r_t)\, \langle e_t, dB_t\rangle\,. \end{aligned}$$

We turn to the Lyapunov functions. Assumption 16 implies that a.s.,

$$d(\epsilon\, V(X_t) + \epsilon\, V(Y_t)) \;\leq\; (2C\epsilon - (\epsilon\, \eta(\, V(X_t)\,) + \epsilon\, \eta(\, V(Y_t)\,)))\; dt + dM_t,$$

where (M_t) is a local martingale. Observe that by definition of γ in Theorem 5, and by concavity of η, we have

$$\epsilon(\, \eta(\, V(X_t)\,) + \eta(\, V(Y_t)\,)) \geq \epsilon\, \eta(\, V(X_t) + V(Y_t)\,) \geq \gamma\, \eta(\, \epsilon\, V(X_t) + \epsilon\, V(Y_t)\,).$$

If $r_t \geq R_1$, we know by definition of S_1 that

$$2C\epsilon - \;(\epsilon\, \eta(\, V(X_t)\,) + \epsilon\, \eta(\, V(Y_t)\,)) \leq -\,\gamma/2\, \eta(\, \epsilon\, V(X_t) + \epsilon\, V(Y_t)\,).$$

If $r_t \geq R_2$, then by Assumption 17,

$$2C\epsilon - (\epsilon\, V(X_t) + \epsilon\, V(Y_t)) \leq -\alpha/2\, \eta(\, f(r_t)\,) \,-\, \gamma/2\, \eta(\, \epsilon\, V(X_t) + \epsilon\, V(Y_t)\,),$$

where we have used that η is increasing, $\epsilon = \xi/(4C)$, and $\Phi \geq f$. Thus a.s.,

$$\begin{aligned} d(\epsilon V(X_t) + \epsilon V(Y_t)) \;\leq\; & (\xi/2 I_{r_t<R_1} - \alpha/2\, \eta(\, f(r_t)\,)\, I_{r_t\geq R_2}) \\ & -\gamma/2\, \eta(\, \epsilon\, V(X_t) + \epsilon\, V(Y_t)\,)\, dt + dM_t. \end{aligned}$$

Summarizing the above results, we can conclude that almost surely, the following differential inequality holds for $t < T$:

$$\begin{aligned} d\rho_1(X_t, Y_t) &= df(r_t) + d(\epsilon\, V(X_t) + \epsilon\, V(Y_t)) \\ &\leq -\min(\alpha, \beta)/2\;\, \eta(\, f(r_t)\,) - \gamma/2\, \eta(\epsilon\, V(X_t) + \epsilon\, V(Y_t))\, dt + dM'_t \\ &\leq -\min(\alpha, \beta, \gamma)/2\, \eta(\, \rho_1(X_t, Y_t)\,)\, dt + dM'_t, \end{aligned}$$

where (M'_t) denotes a local martingale and $\min(\alpha, \beta, \gamma)/2 = c$.

Now let $T_n = \inf\{t \geq 0 : |X_t - Y_t| \leq \frac{1}{n}\}$. By non-explosiveness we have $T_n \uparrow T$. We have shown that the semimartingale $R_t = \rho_1(X_{t\wedge T_n}, Y_{t\wedge T_n})$ satisfies the assumptions of Lemma 2 for the stopping time T_n. Thus

$$\begin{aligned} E[\, H^{-1}(\, c\, T\,)\,] &\leq \liminf_{n\to\infty} E[\, H^{-1}(\, c\, T_n\,)\,] \\ &\leq \liminf_{n\to\infty} E\left[\, H^{-1}(\, H(\, R_{T_n}\,) + c\, T_n\,)\right] \leq E[\, R_0\,] = \rho_1(x, y). \end{aligned}$$

Inequality (1.38) now follows from an application of the Markov inequality, and by the fact that H^{-1} is strictly increasing:

$$P[T>t] = P[\,H^{-1}(cT) > H^{-1}(c\,t)\,] \leq \frac{E[H^{-1}(cT)]}{H^{-1}(c\,t)} \leq \frac{\rho_1(x,y)}{H^{-1}(c\,t)}.$$

□

Proof of Corollary 5. The proof is similar to the one of [74, Theorem 4.1]. Consider the probability measure $\pi_R(\cdot) = \pi(\cdot \cap A_R)/\pi(A_R)$ where $A_R = \{x \in \mathbb{R}^d : V(x) \leq R\}$ for some constant $R \in (0,\infty)$ to be determined below. Since $\pi p_t = \pi$,

$$\begin{aligned} \|p_t(x,\cdot) - \pi\|_{\mathrm{TV}} \;&\leq\; \int \|p_t(x,\cdot) - p_t(y,\cdot)\|_{\mathrm{TV}}\, \pi_R(dy) + \|\pi_R\, p_t - \pi\, p_t\|_{\mathrm{TV}} \\ &\leq \frac{\int \rho_1(x,y)\,\pi_R(dy)}{H^{-1}(c\,t)} + \pi(A_R^c) \;\leq\; \frac{R_2 + \epsilon\, V(x) + \epsilon \int V(y)\,\pi_R(dy)}{H^{-1}(c\,t)} + \pi(A_R^c), \end{aligned}$$

where we have used that $f \leq R_2$. Similarly to [19, Lemma 4.1], one can see that Assumption 16 implies that the invariant measure π satisfies $\int \eta(\,V(y)\,)\,\pi(dy) \leq C$. Hence, the Markov inequality implies $\pi(A_R^c) \leq C/\eta(R)$. Since $x \mapsto \eta(x)/x$ is non-increasing we have

$$V(x) \leq \eta(\,V(x)\,)\,R\,/\,\eta(\,R\,)$$

for any $x \in \mathbb{R}^d$ such that $V(x) \leq R$. This yields the upper bound

$$\int_{V \leq R} V\, d\pi \;\leq\; C\,R/\eta(R).$$

We can conclude that

$$\|p_t(x,\cdot) - \pi\|_{\mathrm{TV}} \;\leq\; \frac{R_2 + \epsilon\, V(x)}{H^{-1}(c\,t)} + \frac{\epsilon\, C\, R}{\pi(A_R)\,\eta(R)\,H^{-1}(ct)} + \frac{C}{\eta(R)}.$$

We now choose R. Set $b = \eta^{-1}(\,2\,C\,)/l$ and define $R = b\,H^{-1}(c\,t)$. Since $\eta(b\,H^{-1}(0)) = \eta(\,b\,l\,) = 2\,C$, we also have a lower bound for $\pi(A_R)$:

$$\pi(A_R) \;=\; 1 - \pi(A_R^c) \;\geq\; 1 - C/\eta(R) \;\geq\; 1/2.$$

Combining the bounds, we obtain the assertion

$$\|p_t(x,\cdot) - \pi\|_{\mathrm{TV}} \;\leq\; \frac{R_2 + \epsilon\, V(x)}{H^{-1}(\,c\,t\,)} + \frac{C\,(\,2\,\epsilon\, b + 1\,)}{\eta(\,b\,H^{-1}(c\,t))}.$$

□

1.6 Extensions

This section is not part of the original article [48] and has been added by the third author. We extend Theorem 1 and Theorem 2 from Section 1.2 replacing the *global* one-sided Lipschitz condition (1.2) by a *local* bound.

We assume that the drift coefficient b is *locally* Lipschitz. Let $|b|_{\mathrm{Lip}(S)}$ be the local Lipschitz constant of b on a bounded set $S \subset \mathbb{R}^d \times \mathbb{R}^d$, i.e.

$$|b|_{\mathrm{Lip}(S)} := \sup \left\{ \frac{|b(x) - b(y)|}{|x-y|} : (x,y) \in S,\, x \neq y \right\} < \infty.$$

In particular, for any bounded set $S \subset \mathbb{R}^d \times \mathbb{R}^d$, there is a continuous function $\kappa(S, \cdot) : (0, \infty) \to [0, \infty)$ such that

$$\int_0^1 r\, \kappa(S, r)\, dr \quad < \quad \infty \quad \text{and} \tag{1.76}$$

$$\langle x - y, b(x) - b(y) \rangle \quad \leq \quad \kappa(S, |x-y|) \cdot |x-y|^2 \quad \text{for any } (x,y) \in S. \tag{1.77}$$

Clearly, the constant function $\kappa(S) \equiv |b|_{\mathrm{Lip}(S)}$ satisfies (1.76) and (1.77).

1.6.1 Extension of Theorem 1

We derive Kantorovich contractions w.r.t. the additive distance (1.4). Suppose that Assumption 10 holds true and fix a function V with corresponding constants $C, \lambda \in (0, \infty)$. Exactly as in Section 1.2.1, we define a set S_1 and its diameter R_1 by (1.6) and (1.7) respectively. Assumption 11 is replaced by the slightly more complicated growth condition:

Assumption 18. *There exist a constant $\alpha > 0$ and a bounded set $S_2 \supseteq S_1$ such that for any $(x,y) \in \mathbb{R}^{2d} \setminus S_2$, we have*

$$V(x) + V(y) \;\geq\; \frac{4C}{\lambda} \left(1 + \frac{1}{\xi(S_2)} \left(2\, |b(x) - b(y)| + \alpha\, |x-y| \right) \right),$$

where $1/\xi(S_2) = \int_0^{R_1} \exp\left(\frac{1}{2} \int_0^r t\, \kappa(S_2, t)\, dt \right) dr$ and $\kappa(S_2, \cdot)$ is a continuous function satisfying (1.76) *and* (1.77)*.*

Suppose that Assumption 18 holds true and fix α, S_2 and $\kappa(S_2)$. As before, we define

$$R_2 \quad = \quad \sup\left\{ |x-y| : (x,y) \in S_2 \right\}. \tag{1.78}$$

Since the set S_2 is fixed, we just write $\kappa(t)$ instead of $\kappa(S_2, t)$ and use this function to define the maps ϕ and Φ by (1.5), completely analogous to Section 1.2.1.

Theorem 6 (Contraction rates for additive metric). *Suppose that Assumptions 10 and 18 hold true. Then there exist a concave, bounded and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and constants $c, \epsilon \in (0, \infty)$ s.t.*

$$\mathcal{W}_{\rho_1}(\mu p_t, \nu p_t) \le e^{-ct}\, \mathcal{W}_{\rho_1}(\mu, \nu) \quad \text{for any } \mu, \nu \in \mathcal{P}(\mathbb{R}^d) \text{ and } t \ge 0. \tag{1.79}$$

Here the underlying distance ρ_1 is defined by (1.4) with ϵ determined by (1.8). The contraction rate is given by $c = \min(\beta, \alpha, \lambda)/2$, where β is given by (1.12). The function f is constant for $r \ge R_2$, and

$$\frac{1}{2} \;\le\; f'(r)\, \exp\left(\frac{1}{2}\int_0^r t\,\kappa(t)\,dt\right) \;\le\; 1 \qquad \text{for any } r \in (0, R_2).$$

The precise definition of the function f is given in the proof.

Before we prove the theorem, we briefly demonstrate the applicability. We write $B_R := \{x \in \mathbb{R}^d : |x| < R\}$ for a ball in $\mathbb{R}^d$ with radius R.

Lemma 4. *If there exists constants $\alpha, \mathcal{R} \in (0, \infty)$ s.t. for any $x \in \mathbb{R}^d$ with $|x| \ge \mathcal{R}$,*

$$V(x) \;\ge\; \frac{4C}{\lambda}\left(1 + \frac{2}{\xi(B_{\mathcal{R}} \times B_{\mathcal{R}})}\left(2 \sup_{|y| \le |x|} |b(y)| + \alpha\, |x|\right)\right),$$

then Assumption 18 holds true.

The statement can be proven similarly to Lemma 1 in Section 1.2.1.

Example 7. *Assume that there are constants $A, D, \gamma \in (0, \infty)$, $q \ge 1$ and $0 \le p < q$ such that*

$$\begin{aligned} \langle b(x), x\rangle &\le -\gamma\, |x|^q && \text{for all } |x| \ge D \text{ and} \\ |b(x)| &\le A\,(1 + |x|^p) && \text{for all } x \in \mathbb{R}^d. \end{aligned}$$

According to Remark 1 there are constants $a, \lambda, C \in (0, \infty)$ and a function V satisfying $V(x) = \exp(a\,|x|^q)$ for large $|x|$, such that Assumption 10 is satisfied. For any $R > 0$ and $(x, y) \in B_R \times B_R$ with $|x - y| = r > 0$, we have that

$$\left\langle \frac{x - y}{|x - y|}, \frac{b(x) - b(y)}{|x - y|} \right\rangle \;\le\; 2\,A\,(1 + R^p)/r$$

and thus (1.76) and (1.77) are satisfied for $\kappa(B_R \times B_R, t) := 2\,A\,(1 + R^p)/t$. Since $p < q$, Lemma 4 is applicable.

Proof of Theorem 6. The proof is analogous to the proof of Theorem 1 and only small changes have to be made. We use the function f defined at the beginning of Section 1.5 with the following parameters: The constants R_1 and R_2 are the diameter of the sets S_1 and S_2, and are given by (1.6) and (1.78) respectively. We set $i(r) \equiv 1$, $j(r) = r$ and define

$$h(r) = \frac{1}{2}\int_0^r s\,\kappa(s)\,ds, \quad \text{with } \kappa(s) = \kappa(S_2, s) \text{ as in Assumption 18.}$$

We fix initial values $(x, y) \in \mathbb{R}^{2d}$ and prove (1.79) for Dirac measures δ_x and δ_y. Let $U_t = (X_t, Y_t)$ be a reflection coupling with initial values (x, y), as defined in Section 1.4.2. We will argue that $E\,[e^{c\,t}\rho_1(X_t, Y_t)] \leq \rho_1(x, y)$ holds for any $t \geq 0$. Denote by $T = \inf\{t \geq 0 : X_t = Y_t\}$ the coupling time. Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. Notice that the differential (in)equalities (1.51) and (1.54) still hold true. Clearly, if $(X_t, Y_t) \in S_1$, then $r_t \leq R_1$ and if $(X_t, Y_t) \in S_2$, then $r_t \leq R_2$. Moreover, the Lebesgue measure of the time (r_t) spends at the points R_1 and R_2 before coupling is almost surely zero, see the proof of Theorem 1 for details. By (1.77), we have for $(X_t, Y_t) \in S_2$ and $t < T$,

$$\langle e_t, b(X_t) - b(Y_t)\rangle \;\leq\; \kappa(S_2, r_t)\, r_t.$$

Using inequality (1.49), (1.54) and the facts that $|f'| \leq 1$ and $f'' \leq 0$, we can conclude that almost surely for $t < T$,

$$\begin{aligned} df(r_t) \;&\leq\; (f'(r_t)\,\langle e_t, b(X_t) - b(Y_t)\rangle + 2\,f''(r_t))\;dt + 2\,f'(r_t)\,\langle e_t, dB_t\rangle \\ &\leq\; (-(\beta/2)\,f(r_t)\,I_{(X_t,Y_t)\in S_2} - (\xi/2)\,I_{(X_t,Y_t)\in S_1})\,dt \\ &+\; |b(X_t) - b(Y_t)|\,I_{(X_t,Y_t)\notin S_2}\,dt \;+\; 2\,f'(r_t)\,\langle e_t, dB_t\rangle\,. \end{aligned} \tag{1.80}$$

We now turn to the Lyapunov functions. Assumption 10 implies that a.s.

$$d\,(\epsilon\,V(X_t) + \epsilon\,V(Y_t)) \;\leq\; 2\,C\,\epsilon - \lambda\,(\epsilon\,V(X_t) + \epsilon\,V(Y_t))\;dt + dM_t,$$

where (M_t) denotes a local martingale. If $(X_t, Y_t) \notin S_1$, then by definition

$$2\,C\,\epsilon - \lambda\,(\epsilon\,V(X_t) + \epsilon\,V(Y_t)) \;\leq\; -(\lambda/2)\,(\epsilon\,V(X_t) + \epsilon\,V(Y_t))\,.$$

In the case $(X_t, Y_t) \notin S_2$, Assumption 18 implies

$$\begin{aligned} &2C\epsilon - \lambda\,(\epsilon V(X_t) + \epsilon V(Y_t)) \\ \leq\;& -(\alpha/2)\,f(r_t)\; - |b(X_t) - b(Y_t)| - (\lambda/2)\,(\epsilon\,V(X_t) + \epsilon\,V(Y_t))\,, \end{aligned}$$

where we have used that by (1.8) and (1.47), $\epsilon = \xi(S_2)/(4C)$ and that $f(r) \leq r$. We can conclude that almost surely,

$$\begin{aligned} d(\epsilon V(X_t) + \epsilon V(Y_t)) \;\leq\;& ((\xi/2)I_{(X_t,Y_s)\in S_1} - (\lambda/2)(\epsilon V(X_t) + \epsilon V(Y_t)))\,dt \\ &- ((\alpha/2)f(r_t) \;+\; |b(X_t) - b(Y_t)|)\,I_{(X_t,Y_t)\notin S_2}\,dt \;+\; dM_t. \end{aligned} \tag{1.81}$$

Combining (1.80) and (1.81), we can conclude that a.s. for $t < T$,

$$d\rho_1(X_t, Y_t) = df(r_t) + d\,(\epsilon\,V(X_t) + \epsilon\,V(Y_t)) \;\leq\; -c\,\rho_1(X_t, Y_t)\,dt + dM'_t,$$

where (M'_t) denotes a local martingale and $c = \min(\alpha, \beta, \lambda)/2$. The proof can now be finished in the same way as the proof of Theorem 1. $\square$

1.6.2 Extension of Theorem 2

We consider Kantorovich contractions w.r.t. the multiplicative semimetric (1.16). As before, the aim is to replace the global bounds by local ones. To this end, we need a slightly more general geometric drift condition.

Assumption 19. *There is a C^2 function $V : \mathbb{R}^d \to \mathbb{R}_+$, a constant $C \in (0, \infty)$ and a continuous function $\Psi : \mathbb{R}^d \to \mathbb{R}_+$ with $\lambda = \inf_x \Psi(x) > 0$ such that $V(x) \to \infty$ as $|x| \to \infty$, and*

$$\mathcal{L}V(x) \;\leq\; C - \Psi(x)V(x) \qquad \text{for any } x \in \mathbb{R}^d.$$

Similar drift conditions occur in [23]. The idea to use this type of drift condition in the context of the multiplicative semimetric is due to A. Guillin.

Remark 5. *Assume that there are constants $\mathcal{R} > 0$, $\gamma > 0$ and $q \geq 1$ such that*

$$\langle b(x), x\rangle \;\;\leq\;\; -\gamma\, |x|^q \qquad \text{for any } |x| \;\geq\; \mathcal{R}.$$

Let V be a C^2 function with $V(x) = \exp(a\,|x|^q)$ outside of a compact set. If $a, \lambda \in (0, \infty)$ satisfy $\frac{aq}{2} + \frac{\lambda}{aq} < \gamma$, then there is $C \in (0, \infty)$ such that Assumption 19 is satisfied with $\Psi(x) = \lambda \max(1, |x|^{2q-2})$.

We proceed similarly to Section 1.2.2, but replace global bounds by local versions. The set $S_1 \subset \mathbb{R}^{2d}$ and its diameter R_1 are now defined by

$$\begin{aligned} S_1 &= \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : [\Psi V](x) + [\Psi V](y) \leq 2C\right\} \quad \text{and} && (1.82)\\ R_1 &= \sup\left\{|x - y| : (x, y) \in S_1\right\}, && (1.83)\end{aligned}$$

where $[\Psi V](x) = \Psi(x)V(x)$. Similarly to Assumption 12, we impose the following condition.

Assumption 20. *The logarithm of V is locally Lipschitz continuous, i.e., for any bounded set $S \subset \mathbb{R}^d$, we have that*

$$\sup_{x\in S} \frac{|\nabla V(x)|}{V(x)} \;\;<\;\; \infty.$$

For bounded sets $S \subset \mathbb{R}^d \times \mathbb{R}^d$, we define

$$Q(S) \;\;=\;\; \sup_{(x,y)\in S} \left(\frac{|\nabla V(x)|}{V(x)} + \frac{|\nabla V(y)|}{V(y)}\right) < \infty. \qquad (1.84)$$

We impose the following growth condition:

Assumption 21. *There exist constants $\alpha, R \in (0,\infty)$ such that for any $x, y \in \mathbb{R}^d$ with $x \neq y$ and $(x,y) \notin B_R^2 := B_R \times B_R$, we have*

$$[\Psi V](x) + [\Psi V](y) \;\geq\; 2C \,+\, \left(\frac{4C}{\xi(B_R^2)} \,+\, V(x) \,+\, V(y)\right)\left(\frac{|b(x)-b(y)|}{|x-y|} \,+\, \frac{\alpha}{2}\right),$$

where $1/\xi(B_R^2) = \int_0^{R_1}\int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa(B_R^2,u)\,du \,+\, Q(B_R^2)(s-r)\right) dr\, ds$ and $\kappa(B_R^2,\cdot)$ is a continuous function satisfying (1.76) *and* (1.77).

Fix α, $S_2 = B_R^2$ and $\kappa(t) = \kappa(S_2,t)$. We write $Q = Q(S_2)$, $\xi = \xi(S_2)$ and set

$$R_2 = \sup\{|x-y| : (x,y) \in S_2\}. \tag{1.85}$$

The reason for choosing S_2 as a product of balls is only to reduce technical complexity in the proofs. As before, we define functions ϕ and Φ by

$$\phi(r) \;=\; \exp\left(-\frac{1}{2}\int_0^r t\,\kappa(t)\,dt \,-\, Q\,r\right), \quad \Phi(r) \;=\; \int_0^r \phi(t)\,dt \tag{1.86}$$

and set $\epsilon = \xi/(4C)$.

Theorem 7 (Contraction rates for multiplicative semimetric)**.** *Suppose that Assumptions 19, 20 and 21 hold true. Then there exist a concave, bounded and non-decreasing continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$ and a constant $c \in (0,\infty)$ such that*

$$\mathcal{W}_{\rho_2}(\mu p_t, \nu p_t) \;\leq\; e^{-c\,t}\,\mathcal{W}_{\rho_2}(\mu,\nu) \qquad \text{for any } \mu,\nu \in \mathcal{P}(\mathbb{R}^d) \text{ and } t \geq 0. \tag{1.87}$$

Here $c = \min(\beta,\alpha)/2$ where

$$\begin{aligned}\beta^{-1} \;&=\; \int_0^{R_2} \Phi(r)\phi(r)^{-1}\,dr \\ &=\; \int_0^{R_2}\int_0^s \exp\left(\frac{1}{2}\int_r^s u\,\kappa(u)\,du + Q\,(s-r)\right) dr\, ds,\end{aligned}$$

the distance ρ_2 is defined by (1.16)*, and f is constant for $r \geq R_2$ and satisfies*

$$\frac{1}{2} \;\leq\; f'(r)\,\exp\left(\frac{1}{2}\int_0^r u\,\kappa(u)\,du \,+\, Q\,r\right) \;\leq\; 1 \qquad \text{for } r \in (0,R_2).$$

The precise definition of the function f is given in the proof.

Before we prove the statement, we make brief remarks on applicability.

Lemma 5. *Suppose that Assumption 19 holds true. Moreover, we assume that the function V is of type $V(x) = V(|x|)$ and that it is non-decreasing in $|x|$. If there are constants $\alpha, \mathcal{R} \in (0,\infty)$ and such that for all $|x| \geq \mathcal{R}$,*

$$V(x) \;\geq\; 4C\,\max\left(\frac{1}{\lambda}, \frac{1}{\xi(B_{\mathcal{R}}^2)}\right) \qquad \text{and} \tag{1.88}$$

$$\Psi(x) \;\geq\; 6\left(|b|_{\mathrm{Lip}\left(B_{|x|}^2\right)} + \frac{\alpha}{2}\right), \tag{1.89}$$

then Assumption 21 holds true.

Proof of Lemma 5. Let $(x,y) \notin B^2_{\mathcal{R}}$ and assume w.l.o.g. that $|x| = \max(|x|, |y|) \geq \mathcal{R}$. Using (1.88), (1.89), the monotonicity of V and that $\Psi \geq \lambda$, we can conclude that

$$\begin{aligned}
[\Psi V](x) &\geq \frac{\lambda}{2} V(x) + \frac{1}{2}[\Psi V](x) \\
&\geq 2C + 3\,V(x) \left(|b|_{\mathrm{Lip}\left(B^2_{|x|}\right)} + \frac{\alpha}{2} \right) \\
&\geq 2C + \left(\frac{4C}{\xi(B^2_{\mathcal{R}})} + V(x) + V(y) \right) \left(\frac{|b(x) - b(y)|}{|x-y|} + \frac{\alpha}{2} \right).
\end{aligned}$$

□

Example 8. *Assume that there are constants $D, \gamma \in (0,\infty)$ and $q > 1$ such that*

$$\langle b(x), x \rangle \;\leq\; -\gamma\, |x|^q \qquad \text{for all } |x| \geq D.$$

Then, there are constants $a, \lambda, C \in (0,\infty)$, a function V satisfying $V(x) = \exp(a\,|x|^q)$ outside of a compact set, and $\Psi(x) = \lambda \max(1, |x|^{2q-2})$ such that Assumption 19 is satisfied. Moreover, for this choice, there are constants $A, B \in (0,\infty)$ such that

$$Q(B_R \times B_R) \leq A + B\,R^{q-1}.$$

Assume that there are constants $0 \leq p \leq q-1$ and $D, E \in (0,\infty)$ such that for any $R > 0$,

$$|b|_{\mathrm{Lip}\left(B^2_R\right)} \leq D + E\,R^p.$$

Then, the constant function $\kappa(B_R \times B_R) \equiv D + E\,R^p$ satisfies (1.76) and (1.77) and there are constants $F, m \in (0,\infty)$ such that $1/\xi(B_R \times B_R) \leq F\, \exp(m\,R^{q-1})$ with this choice of κ. In particular, there is $\mathcal{R} \in (0,\infty)$ such that (1.88) is satisfied. Moreover, since $2q-2 > q-1 \geq p$, (1.89) is satisfied for large enough $\mathcal{R}$ and arbitrary α.

Proof of Theorem 7. The proof is similar to the proof of Theorem 2. We use the function f defined at the beginning of Section 1.5 with the following parameters: The set S_1 and its diameter R_1 are specified by (1.82) and (1.83) respectively. Assumption 21 yields a set $S_2 = B^2_R$ and the corresponding diameter R_2 is defined by (1.85). We set $j(r) = r$, $i(r) = \Phi(r)$, $\kappa(t) = \kappa(S_2, t)$ and $Q = Q(S_2)$, where $\Phi(r)$ is determined by (1.86). We define

$$h(r) \;=\; \frac{1}{2} \int_0^r s\,\kappa(s)\,ds \;+ Q\,r.$$

Coupling: We construct a coupling of solutions to (1.1). It is realized as a standard diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$. We use a mixture of synchronous and reflection coupling, similar to Section 1.4.3. The technical realization of the coupling is near to [51]. Fix small $\delta > 0$. One should think of δ being close to zero. We introduce Lipschitz functions $\mathrm{rc}, \mathrm{sc} : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ satisfying (1.42) and

$$\begin{array}{llll} \mathrm{rc}(x,y) = 1 & \text{whenever} \quad (x,y) \in B^2_{R-\delta} & \text{and} & |x-y| \geq \delta, \\ \mathrm{rc}(x,y) = 0 & \text{whenever} \quad (x,y) \notin B^2_R & \text{or} & |x-y| \leq \delta/2. \end{array}$$

The functions can be constructed using standard cut-off techniques. Fix initial values $(x,y) \in \mathbb{R}^{2d}$ and two independent Brownian motions (B^1_t) and (B^2_t). We define our coupling $U_t = (X_t, Y_t)$ as the solution of the SDE

$$\begin{aligned} dX_t &= b(X_t)\,dt \;+\; \mathrm{rc}(U_t)\,dB^1_t \;+\; \mathrm{sc}(U_t)\,dB^2_t \\ dY_t &= b(Y_t)\,dt \;+\; \mathrm{rc}(U_t)\,(I - 2\,e_t\,\langle e_t, \cdot\rangle)\,dB^1_t \;+\; \mathrm{sc}(U_t)\,dB^2_t, \end{aligned}$$

with $(X_0, Y_0) = (x,y)$ and

$$e_t \;=\; \frac{X_t - Y_t}{|X_t - Y_t|} \quad \text{for } X_t \neq Y_t, \qquad e_t \;=\; u \quad \text{for } X_t = Y_t,$$

where $u \in \mathbb{R}^d$ is some arbitrary fixed unit vector. Notice that the concrete choice of u is irrelevant for the dynamic, since $\mathrm{rc}(x,x) = 0$. The equation is a standard SDE with locally Lipschitz coefficients satisfying a non-explosive criteria and thus there is a unique, strong and global solution. Using Levy's characterization of Brownian motion and (1.42), one can verify that (X_t, Y_t) is indeed a coupling.

Calculations: We now prove (1.87) and fix initial values $(x,y) \in \mathbb{R}^{2d}$ as well as small $\delta > 0$. It is enough to prove the statement for Dirac measures δ_x and δ_y. The coupling $U_t = (X_t, Y_t)$, defined in the last paragraph, yields the upper bound

$$\mathcal{W}_{\rho_2}(\delta_x p_t, \delta_y p_t) \;\leq\; E[\rho_2(X_t, Y_t)].$$

Set $Z_t = X_t - Y_t$ and $r_t = |Z_t|$. We will argue that for each $t \geq 0$ and $\delta > 0$, there is a constant $\tilde{C}(t,\delta) \in (0,\infty)$ with the property $\tilde{C}(t,\delta) \to 0$ for $\delta \to 0$ and fixed t, such that

$$e^{ct} E[\rho_2(X_t, Y_t)] \leq \rho_2(x,y) + \tilde{C}(t,\delta) \qquad \text{holds true.} \tag{1.90}$$

From this inequality one can then conclude, that for any $t \geq 0$ we have

$$\mathcal{W}_{\rho_2}(\delta_x p_t, \delta_y p_t) \leq e^{-ct}\,\rho_2(x,y) + \tilde{C}(t,\delta),$$

which then finishes the proof of (1.87) since $\delta > 0$ can be chosen arbitrarily small.

We now argue (1.90). Observe that $W_t = \int_0^t \langle e_s, dB^1_s\rangle$ is a one-dimensional Brownian motion. The process (Z_t) satisfies almost surely the equation

$$dZ_t \;=\; (b(X_t) - b(Y_t))\;dt \;+\; 2\;\mathrm{rc}(U_t)\,e_t\,dW_t.$$

Using an approximation argument, similarly to [159, Lemma 3], one can show that (r_t) satisfies almost surely the equation

$$dr_t \;=\; \langle e_t, b(X_t) - b(Y_t)\rangle \; dt + \; 2\, \mathrm{rc}(U_t)\, dW_t.$$

Similarly as in the proof in Section 1.5.1, we apply the Itô-Tanaka formula for semimartingales to conclude that almost surely,

$$\begin{aligned} f(r_t) - f(r_0) \;&=\; \int_0^t f'_-(r_s)\; \langle e_s, b(X_s) - b(Y_s)\rangle \; ds \\ &\quad + 2 \int_0^t \mathrm{rc}(U_s)\, f'_-(r_s)\, dW_s + \frac{1}{2}\int_{-\infty}^{\infty} L_t^x\, \mu_f(dx), \end{aligned} \tag{1.91}$$

where L_t^x is the right-continuous local time of (r_t), f'_- is the left derivative of f and μ_f is the non-positive measure representing the second derivative of f. By (1.53), we conclude that, almost surely for all $t > 0$,

$$\mathcal{L}(\{0 \le s \le t : r_s \in \{R_1, R_2\} \text{ and } \mathrm{rc}(U_s) > 0\}) = 0, \tag{1.92}$$

i.e. the Lebesgue measure of the time r_s spends at the points R_1 and R_2 while $\mathrm{rc}(U_s) > 0$, is almost surely zero. Notice that f is continuous differentiable on the set $(0, \infty) \setminus \{R_2\}$ and twice continuously differentiable on $(0, \infty) \setminus \{R_1, R_2\}$. Since μ_f is non-positive and by (1.53), we can conclude that, almost surely for all $t > 0$,

$$\begin{aligned} \int_{-\infty}^{\infty} L_t^x\, \mu_f(dx) \;&\le\; \int_{-\infty}^{\infty} I_{x \notin \{R_1,R_2\}}\, L_t^x\, \mu_f(dx) \;=\; \int_0^t I_{r_s \notin \{R_1,R_2\}}\, f''(r_s)\, d[r]_s \\ &= 4 \int_0^t I_{r_s \notin \{R_1,R_2\}}\, \mathrm{rc}(U_s)^2\, f''(r_s)\, ds, \quad \text{i.e. }, \end{aligned}$$

$$df(r_t) \le \big(f'_-(r_t)\; \langle e_t, b(X_t) - b(Y_t)\rangle \; + \; I_{r_s \notin \{R_1,R_2\}}\, 2\, \mathrm{rc}(U_t)^2\, f''(r_t)\big)\; dt + dM_t^1,$$

where $M_t^1 = 2\int_0^t \mathrm{rc}(U_s)\, f'_-(r_s)\, dW_s$. We now turn to the Lyapunov functions and set $G(x, y) = 1 + \epsilon\, V(x) + \epsilon\, V(y)$. By definition of the coupling, we have a.s.

$$\begin{aligned} dG(X_t, Y_t) \;&=\; (\epsilon\, \mathcal{L}V(X_t) + \epsilon\, \mathcal{L}V(Y_t))\; dt \\ &\quad +\; \epsilon\; \mathrm{sc}(U_t)\, \big\langle \nabla V(X_t) + \nabla V(Y_t), dB_t^2 \big\rangle \\ &\quad +\; \epsilon\; \mathrm{rc}(U_t)\, \big\langle \nabla V(X_t) + \nabla V(Y_t), dB_t^1 \big\rangle \\ &\quad -\; 2\, \epsilon\; \mathrm{rc}(U_t)\; \langle \nabla V(Y_t), e_t\rangle\, \big\langle e_t, dB_t^1 \big\rangle\,. \end{aligned} \tag{1.93}$$

By Assumption 19, we can conclude that

$$dG(X_t, Y_t) \;\le\; \epsilon\, (2\, C \; - [\Psi V](X_t) - [\Psi V](Y_t))\; dt + dM_t^2,$$

where $[\Psi V](x) = \Psi(x)V(x)$ and M_t^2 is a local martingale. By (1.91) and (1.93), the covariation of $f(r_t)$ and $G(X_t, Y_t)$ is given by:

$$d[\, f(r)\, , G(X, Y)\,]_t \;=\; 2\; \mathrm{rc}(U_t)^2\, f'_-(r_t)\, \epsilon\; \langle \nabla V(X_t) - \nabla V(Y_t), e_t\rangle \; dt.$$

Whenever $(X_t,Y_t)\notin S_2$, we have $\mathrm{rc}(U_t)=0$ by definition. Moreover, using Cauchy-Schwarz and (1.84), we can derive the following bound for $(x,y)\in S_2$ with $x\neq y$:

$$\begin{aligned}\epsilon\left\langle \nabla V(x)-\nabla V(y),\frac{x-y}{|x-y|}\right\rangle &\leq (1+\epsilon\,V(x)+\epsilon\,V(y))\frac{\epsilon\,|\nabla V(x)|+\epsilon\,|\nabla V(y)|}{(1+\epsilon\,V(x)+\epsilon\,V(y))}\\ &\leq Q\,G(x,y).\end{aligned}$$

Hence, almost surely,

$$d[\,f(r)\,,G(X,Y)\,]_t \quad\leq\quad 2\;\mathrm{rc}(U_t)^2\,Q\,f'_-(r_t)\,G(X_t,Y_t)\,dt. \tag{1.94}$$

The product rule for semimartingales, (1.91), (1.93) and (1.94) imply that, almost surely,

$$\begin{aligned}d\,(f(r_t)G(X_t,Y_t)) &= G(X_t,Y_t)\,df(r_t)+f(r_t)\,dG(X_t,Y_t)+[f(r),G(X,Y)]_t\\ &\leq H_t\;dt\;+\;dM^3_t,\end{aligned}$$

where M^3_t is a local martingale and

$$\begin{aligned}H_t \;=\;& G(X_t,Y_t)\,\big(f'_-(r_t)\,\langle e_t,b(X_t)-b(Y_t)\rangle\;+\;I_{r_t\notin\{R_1,R_2\}}\,2\;\mathrm{rc}(U_t)^2\,f''(r_t)\big)\\ &+\epsilon\,f(r_t)\,\left(2\,C\;-[\Psi V](X_t)-[\Psi V](Y_t)\right)\\ &+2\;\mathrm{rc}(U_t)^2\,Q\,f'_-(r_t)\,G(X_t,Y_t).\end{aligned}$$

Define

$$J_t \;=\; f'_-(r_t)\,G(X_t,Y_t)\;|b(X_t)-b(Y_t)|\;+\;\epsilon\,f(r_t)\;\left(2\,C\;-[\Psi V](X_t)-[\Psi V](Y_t)\right). \tag{1.95}$$

We argue that there is $\tilde{C}(\delta)\in(0,\infty)$ with the property $\tilde{C}(\delta)\to 0$ for $\delta\to 0$ such that almost surely for all $t\geq 0$,

$$\int_0^t H_s\,ds\leq -c\int_0^t \rho_2(X_s,Y_s)\,ds+C(\delta)\,t \tag{1.96}$$

We do a case distinction to derive this upper bound.

Case 1: $(X_s,Y_s)\notin S_2$**.** In this case we have $\mathrm{rc}(U_s)=0$ and thus $H_s\leq J_s$. By Assumption 21, we can conclude that

$$\begin{aligned}J_s &\leq \left(f'_-(r_t)-\frac{f(r_t)}{r_t}\right)\,G(X_t,Y_t)\;|b(X_t)-b(Y_t)|\;-\;\alpha/2\,\rho_2(X_t,Y_t)\\ &\leq -\,c\,\rho_2(X_t,Y_t),\end{aligned}$$

where we have used that $\epsilon=\xi/(4C)$, that $f(r)\geq f'_-(r)\,r$ for $0<r\leq R_2$ and that $f'_-(r)=0$ for $r>R_2$.

Case 2: $(X_s,Y_s)\in B^2_R$ **and** $(X_s,Y_s)\notin B^2_{R-\delta}$**.** Assume w.l.o.g. that $\max(|X_s|\,,|Y_s|)=|X_s|$ and let X'_s be the point on the boundary of the ball $B_R=\{x\in\mathbb{R}^d:|x|<R\}$

which is on the line passing through 0 and X_s and is closer to X_s. Observe that $|X_s - X'_s| \le \delta$. We define J'_s similarly to (1.95), replacing X_s by X'_s and r_s by $r'_s = |X'_s - Y_s|$. Observe that $(X'_s, Y_s) \notin S_2^R$ and thus by Assumption 21, we have similarly as in Case 1, $J'_s \le -c\,\rho_2(X'_s, Y_s)$. We define

$$K_s \;=\; \mathrm{rc}(U_s)^2\, G(X_s,Y_s)\,\big(I_{r_s \notin \{R_1,R_2\}}\, 2\, f''(r_s) \,+\, 2\, f'_-(r_s)\, Q\big)$$

Notice that $H_s \le J_s \,+\, K_s$. Observe that $K_s = 0$ if $\mathrm{rc}(U_s) = 0$. If $\mathrm{rc}(U_s) > 0$, we may assume by (1.92) that $r_s \notin \{R_1, R_2\}$ and by (1.49) that $K_s \le 0$. Moreover,

$$J_s \;\le\; J'_s + |J'_s - J_s| \;\le -\,c\,\rho_2(X_s,Y_s) \,+\, c\,|\rho_2(X'_s,Y_s) - \rho_2(X_s,Y_s)| \,+\, |J'_s - J_s|\,.$$

Observe that the functions f and f'_- are uniformly continuous and bounded on the set $[0, R_2]$. Moreover, the functions b, Ψ and V are uniformly continuous and bounded on the set B_R. Therefore, for any $\delta > 0$ there is $C^1(\delta) \in (0,\infty)$ with the property $C^1(\delta) \to 0$ for $\delta \to 0$ such that

$$J_s \le -c\,\rho_2(X_s,Y_s) + C^1(\delta).$$

Case 3: $(X_s,Y_s) \in B^2_{R-\delta}$ **and** $|X_s - Y_s| < \delta$**.** Similarly to the arguments in Case 2, we can conclude that $H_s \le J_s + K_s$ and may assume that $K_s \le 0$. Notice that on the ball B_R, V is bounded and b is Lipschitz. Moreover, $f(r) \le r$ for all $r \ge 0$. We can conclude that there is a constant $C^2(\delta) \in (0,\infty)$ which converges to zero for $\delta \to 0$ such that $J_s \le -c\,\rho_2(X_s,Y_s) + C^2(\delta)$.

Case 4: $(X_s,Y_s) \in B^2_{R-\delta}$ **and** $|X_s - Y_s| \ge \delta$**.** In this case we have $r_s < R_2$ and $\mathrm{rc}(U_s) = 1$. In particular, we may assume by (1.92) that $r_s \ne R_1$. Recall the definition of H_s. By (1.77), we have the bound

$$G(X_s,Y_s)\, f'(r_s)\, \langle e_s, b(X_s) - b(Y_s)\rangle \le G(X_s,Y_s)\, f'(r_s)\, \kappa(r_s)\, r_s.$$

Inequality (1.49) implies

$$\begin{aligned} G(X_s,Y_s)\, 2\, f''(r_s) \;&\le\; -\,G(X_s,Y_s)\, f'(r_s)\,(\kappa(r_s)\, r_s + 2\,Q) \\ &\quad -\frac{\beta}{2}\,\rho_2(X_s,Y_s) - I_{r_s < R_1}\,\frac{\xi}{2}\,\rho_2(X_s,Y_s). \end{aligned}$$

Notice that if $r_s > R_1$, then $(X_s,Y_s) \notin S_1$ and thus

$$\epsilon\, f(r_t)\;(2\,C\; - [\Psi V](X_t) - [\Psi V](Y_t)) \le 0.$$

Moreover, since $\epsilon = \xi/(4C)$ and since $G \ge 1$,

$$\epsilon\, f(r_t)\, 2\, C \;\;\le\;\; \frac{\xi}{2}\,\rho_2(X_s,Y_s).$$

We see that $H_s \le -(\beta/2)\,\rho_2(X_s,Y_s)$.

Combining the arguments from the four cases, we can conclude (1.96). Applying the product rule for semimartingales and using the latter mentioned inequality, we see that

$$\begin{aligned} d(e^{ct}\,\rho_2(X_t,Y_t)) &\le e^{ct}\,dM_t^3 + e^{ct}\,c\,\rho_2(X_t,Y_t)\,dt + e^{ct}\,d\rho_2(X_t,Y_t) \\ &\le e^{ct}\,dM_t^3 + e^{ct}\,C(\delta)\,dt. \end{aligned}$$

Using a stopping argument, we can conclude (1.90) for an appropriate constant $\tilde{C}(t,\delta)$ satisfying $\tilde{C}(t,\delta) \to 0$ for $\delta \to 0$. $\square$

2 Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions

Given a separable and real Hilbert space $\mathbb{H}$ and a trace-class, symmetric and nonnegative operator $\mathcal{G} : \mathbb{H} \to \mathbb{H}$, we examine the equation

$$dX_t = -X_t\,dt + b(X_t)\,dt + \sqrt{2}\,dW_t, \qquad X_0 = x \in \mathbb{H},$$

where (W_t) is a $\mathcal{G}$-Wiener process on $\mathbb{H}$ and $b : \mathbb{H} \to \mathbb{H}$ is Lipschitz. We assume there is a splitting of $\mathbb{H}$ into a finite-dimensional space $\mathbb{H}^l$ and its orthogonal complement $\mathbb{H}^h$ such that $\mathcal{G}$ is strictly positive definite on $\mathbb{H}^l$ and the nonlinearity b admits a contraction property on $\mathbb{H}^h$. Assuming a geometric drift condition, we derive a Kantorovich (L^1 Wasserstein) contraction with an *explicit* contraction rate for the corresponding Markov kernels. Our bounds on the rate are based on the eigenvalues of $\mathcal{G}$ on the space $\mathbb{H}^l$, a Lipschitz bound on b and a geometric drift condition. The results are derived using coupling methods.

R. Zimmer. Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. *ArXiv e-print 1605.07863*, May 2016

Financial support from the German Science foundation through the *Hausdorff Center for Mathematics* is gratefully acknowledged.

2.1 Introduction

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle, |\cdot|)$ be a separable and real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $|\cdot|$. Suppose that a trace-class, symmetric and nonnegative operator $\mathcal{G} : \mathbb{H} \to \mathbb{H}$ is given. Let $(\mathbf{e}_k)_{k\in\mathbb{N}_+}$ be an orthonormal basis of $\mathbb{H}$ such that for nonnegative real numbers (λ_k), we have $\mathcal{G}\mathbf{e}_k = \lambda_k \mathbf{e}_k$ and $\sum_{k=1}^{\infty} \lambda_k < \infty$, see e.g. [132] for the existence of such a basis. Denote by (W_t) a $\mathcal{G}$-Wiener process on $\mathbb{H}$, i.e. $W_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, B_t^k\, \mathbf{e}_k$ for independent Brownian motions (B_t^k). We consider the stochastic differential equation

$$dX_t = -X_t\,dt + b(X_t)\,dt + \sqrt{2}\,dW_t, \qquad X_0 = x \in \mathbb{H}, \tag{2.1}$$

on the space $\mathbb{H}$ and assume that the nonlinearity $b : \mathbb{H} \to \mathbb{H}$ is Lipschitz. In particular, there is a strong, non-explosive and continuous solution (X_t) taking values in $\mathbb{H}$, see e.g. [103]. Moreover, (X_t) is a Feller process and we denote the

Markov transition kernels by (p_t). Given a probability measure μ on $\mathbb{H}$, we write $\mu p_t(dx) = \int p_t(y, dx)\, \mu(dy)$.

Equation (2.1) has a natural appearance in the domain of sampling problems and acts as a diffusion limit for Markov chain Monte Carlo (MCMC) methods, see [71, 114, 38, 143, 70] and the references therein. In particular, if $U : \mathbb{H} \to \mathbb{R}_+$ is a smooth function, if $\mathcal{G}$ is positive definite and if we choose the nonlinearity $b(x) = -\mathcal{G}\nabla_{\mathbb{H}} U(x)$ in (2.1), then the results from [71] imply that the Markov kernels (p_t) admit a unique invariant probability measure π satisfying $\pi p_t = \pi$ for any $t \geq 0$. The measure π is given by

$$\pi(dx) \quad \propto \quad \exp(-U(x))\; \mathcal{N}(0, \mathcal{G})(dx), \tag{2.2}$$

where $\mathcal{N}(0, \mathcal{G})$ denotes a centered normal distribution on $\mathbb{H}$ with covariance operator $\mathcal{G}$. Such measures appear for example in the area of diffusion bridges, cf. [71].

Let μ be a given initial distribution. Given the outlined connection to sampling problems, an important question is whether the measure μp_t converges towards π for $t \to \infty$ in some reasonable distance and how one can obtain explicit rates for the speed of convergence. We give conditions under which the convergence takes place in Kantorovich and L^p Wasserstein distances at an exponential rate and focus on establishing concrete bounds on the speed of convergence. Inspired by the sampling setup, we work in the following setting: Fix $n \in \mathbb{N}_+$. We consider a splitting of the Hilbert space $\mathbb{H}$ into a space $\mathbb{H}^l = \langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$, spanned by the first n basis vectors, and its orthogonal complement $\mathbb{H}^h$, i.e. $\mathbb{H} = \mathbb{H}^l \oplus \mathbb{H}^h$. We call $\mathbb{H}^l$ *low*-dimensional and $\mathbb{H}^h$ *high*-dimensional space. Given $x \in \mathbb{H}$, we denote by x^l and x^h the orthogonal projections onto $\mathbb{H}^l$ and $\mathbb{H}^h$ respectively. Our main assumptions are:

Assumption 22. *There are constants* $0 \leq H_h < 1$ *and* $L_l, L_h, H_l \geq 0$ *such that*

$$\left|b^h(x) - b^h(y)\right| \;\leq\; H_l \left|x^l - y^l\right| + H_h \left|x^h - y^h\right| \quad \textit{and} \tag{2.3}$$

$$\left|b^l(x) - b^l(y)\right| \;\leq\; L_l \left|x^l - y^l\right| + L_h \left|x^h - y^h\right| \quad \textit{for any } x, y \in \mathbb{H}. \tag{2.4}$$

Assumption 23. $\mathcal{G}$ *is strictly positive definite on* $\mathbb{H}^l$*, i.e. for any* $k \in \mathbb{N}$ *with* $1 \leq k \leq n$*, we have* $\lambda_k > 0$.

In the sampling setup described above, assuming that the map $x \mapsto \nabla U(x)$ is Lipschitz on $\mathbb{H}$, it is always possible to find a splitting $\mathbb{H} = \mathbb{H}^l \oplus \mathbb{H}^h$ such that Assumptions 22 and 23 are satisfied, cf. Section 2.3. In addition to the above assumptions, we need some kind of localization argument, i.e. we assume either that b is vanishing outside of a ball or that a geometric drift condition holds, cf. Assumptions 24 and 25 respectively. Based on these assumptions we derive quantitative Kantorovich contractions for the Markov kernels using coupling methods. The resulting contraction rates are given explicitly in terms of the eigenvalues of $\mathcal{G}$ on the space $\mathbb{H}^l$, the constants from Assumption 22 and the localization argument.

Outline. The main results are presented in Section 2.2.1. The key statements are Theorem 8 and Theorem 9. The couplings are specified in Section 2.2.2 and the proofs are given in Section 2.2.3. Applications are considered in Section 2.3. In the remaining part of the introduction we present additional motivation and references.

The ergodicity of degenerate and infinite-dimensional models has been extensively studied in the last two decades and by now there exists a comprehensive theory [67, 69, 68, 34]. Huge parts of the theory have been developed trying to answer the question, whether the $2D$ stochastic Navier-Stokes equation is uniquely ergodic in a hypoelleptic setting, where only a few dimensions are stimulated directly with noise, cf. [66, 65]. As an intermediate step to tackle the truly hypoelleptic setting, many authors [157, 115, 113, 100, 99, 158, 17, 16] worked in an intermedium setting: They considered a splitting of the underlying Hilbert space into a finite-dimensional space $\mathbb{H}^u$ of "unstable modes", where the dynamics is forced directly with noise, and an infinite-dimensional complement $\mathbb{H}^s$ of "stable modes", where the driving noise can be degenerate. Stable and unstable modes means in this context that the long time behavior of the dynamics is determined by the behavior on the space $\mathbb{H}^u$, cf. [72]. In this context, J.C. Mattingly proposed in [115] a coupling approach to conclude exponential mixing properties for the $2D$ stochastic Navier-Stokes equation. In a related context, M. Hairer demonstrated in [72] the strength of *asymptotic couplings* to show mixing properties of degenerate systems. Finally, J.C. Mattingly and M. Hairer were able to proof the unique ergodicity of the $2D$ Navier-Stokes equation in a hypoelleptic setting, which was a milestone in the development of ergodic theory for degenerate and infinite-dimensional systems [65, 67, 68]. Embedding some of the key concepts of the theory into a uniformly applicable framework, Mattingly, Hairer and Scheutzow developed the weak Harris theorem [69]. It can be interpreted as a generalization of classical Harris type theorems [77, 118, 96, 135, 75] which have become standard tools for proving geometric ergodicity of finite-dimensional Markov processes. The weak Harris theorem further extends the range of possible applications and allows to establish geometric ergodicity under verifiable conditions. Nevertheless, being a uniform framework, applicable for a large class of Markov processes, the (weak) Harris theorem usually does not provide sharp constants for specific models and the resulting constants are often not connected to the structure of the model in a transparent way. This is due to the fact that the corresponding Kantorovich distance is usually chosen in a somehow ad hoc way, cf. [48].

In this work we do not have the aim of developing a uniform framework for various models. We focus on the very specific model (2.1) and establish Kantorovich contractions with explicit constants by adapting the underlying Kantorovich distance in a very specific way to the structure of the model. The approach is based on a technique from [50, 51]. Here, A. Eberle establishes Kantorovich contractions with explicit constants for finite-dimensional and nondegenerate diffusions using a combination of reflection coupling [107] and concave distance functions. While the principle idea to study Kantorovich distances w.r.t. concave underlying distances occurred at other places in the literature before [28, 69], it is noteworthy that [50, 51] presents a technique, how one can construct an explicit concave distance function which, under some reasonable assumptions, maximizes the resulting contraction rate under the reflection coupling up to constant factors. Eberle's results are based on the assumption that the underlying deterministic system of the diffusion is contractive for "large distances". In the recent work [48] this assumption is replaced by a more general Lyapunov drift condition combining Lyapunov functions with concave distance functions and reflection coupling, partially motivated by [69, 75]. In this work, we use the main ideas from [50, 51, 48] and extend them to the infinite-dimensional and possibly degenerate process (2.1) by constructing an explicit *asymptotic coupling* (X_t, Y_t) of solutions to (2.1) in the sense of [72, 115], i.e. a coupling for which the processes X_t and Y_t converge to each other but do not necessarily meet in finite time. The Kantorovich contraction of the Markov kernels is then established by adapting the underlying cost function in a very specific way to the chosen coupling and model.

Up to the author's knowledge there are currently two works which use a reflection coupling to conclude exponential mixing properties of infinite-dimensional systems. In [33] a reflection coupling is used to prove exponential convergence for a reaction-diffusion and Burgers equation driven by space-time white noise. The article [151] makes use of an "approximated reflection coupling" to derive gradient estimates and exponential mixing for a class of nonlinear monotone SPDES, where the driving noise is a $\mathcal{G}$-Wiener process, $\mathcal{G}$ being trace-class and satisfying $\langle x, \mathcal{G}x\rangle > 0$ for any $x \in \mathbb{H}$. Moreover, it is assumed that the solution of the SPDE lies in the image of $\mathcal{G}$, i.e. that the equation has some kind of smoothing properties. In both articles exponential convergence in total variation norms is concluded. In contrast to these settings, we allow the operator $\mathcal{G}$ to be degenerate on the infinite-dimensional space $\mathbb{H}^h$ and equation (2.1) does not provide the additional smoothing assumed in [151]. Moreover, in our setting it is in general not true that for arbitrary $x, y \in \mathbb{H}$ we have $\|\delta_x p_t - \delta_y p_t\|_{\mathrm{TV}} \to 0$ for $t \to \infty$, see e.g. [66, Example 3.14].

2.2 Main results

We present our main results. In Section 2.2.1 we formulate the main statements. The coupling approach leading to those statements is explained in Section 2.2.2. Finally, the proofs are provided in Section 2.2.3.

2.2.1 Results

We now formulate our contraction results. As a preparation, we first introduce a norm $|\cdot|_\alpha$ on $\mathbb{H}$ which is equivalent to the Hilbert space norm, but has the advantage that it puts additional weight on the components in the space $\mathbb{H}^h$. This enables us to exploit the contraction property provided by Assumption 22. We then formulate three Kantorovich contractions with an increasing level of difficulty: In Proposition 1 we assume that the map b is a contraction w.r.t. $|\cdot|_\alpha$ and thus a Kantorovich contraction can be established with ease. In Theorem 8 we assume that b is a contraction w.r.t. $|\cdot|_\alpha$ only for "large distances" and adapt the underlying metric of the Kantorovich distance accordingly by involving a concave function. Finally, in Theorem

9 we replace the contraction property for large distances by a more general geometric drift condition and combine the metric considered in Theorem 8 with Lyapunov functions.

Suppose that Assumption 22 holds true. Denote

$$\alpha \;=\; \frac{1+L_h}{1-H_h} \geq 1 \qquad \text{and} \qquad \beta \;=\; \alpha H_l + L_l - 1. \tag{2.5}$$

We define a norm $|\cdot|_\alpha$ on $\mathbb{H}$, where the $\mathbb{H}^h$ component is weighted by α:

$$|x|_\alpha \;=\; \left|x^l\right| \;+\; \alpha \;\left|x^h\right|.$$

Observe that $|\cdot|_\alpha$ is equivalent to $|\cdot|$, i.e. for any $x \in \mathbb{H}$,

$$|x| \;\leq\; |x|_\alpha \;\leq\; \sqrt{2}\,\alpha\,|x|\,. \tag{2.6}$$

Assumption 22 implies that the nonlinearity b is a contraction w.r.t. $|\cdot|_\alpha$ in "certain regions of $\mathbb{H}$". More precisely, we have the following statement:

Lemma 6. *Assumption 22 implies the inequality*

$$|b(x)-b(y)|_\alpha \leq (1+\beta)\left|x^l-y^l\right|_\alpha + \left(1-\frac{1}{\alpha}\right)\left|x^h-y^h\right|_\alpha \quad \textit{for any } x,y \in \mathbb{H}. \tag{2.7}$$

Moreover, if $x, y \in \mathbb{H}$ satisfy

$$(1+\beta)\left|x^l-y^l\right| \;\leq\; \frac{1}{2}\left|x^h-y^h\right|, \tag{2.8}$$

then it follows

$$|b(x)-b(y)|_\alpha \;\leq\; \left(1-\frac{1}{2\alpha}\right)|x-y|_\alpha\,. \tag{2.9}$$

Proof. Assumption 22 implies the inequalities

$$\begin{aligned}
|b(x)-b(y)|_\alpha &= \left|b^l(x)-b^l(y)\right| \;+\; \alpha\;\left|b^h(x)-b^h(y)\right| \\
&\leq (\alpha\; H_l+L_l)\;\left|x^l-y^l\right| \;+\; (H_h+L_h/\alpha)\,\alpha\left|x^h-y^h\right| \\
&= (1+\beta)\;\left|x^l-y^l\right|_\alpha \;+\; (1-1/\alpha)\;\left|x^h-y^h\right|_\alpha.
\end{aligned}$$

If (2.8) holds true, then we can further estimate:

$$\begin{aligned}
|b(x)-b(y)|_\alpha &\leq \left|x^h-y^h\right|_\alpha - 1/2\left|x^h-y^h\right| \\
&\leq |x-y|_\alpha - \left|x^l-y^l\right| - 1/(2\alpha)\left|x^h-y^h\right|_\alpha \\
&\leq (1-\min\{1/(2\alpha),1\})\,|x-y|_\alpha\,.
\end{aligned}$$

Since $\alpha \geq 1$, we conclude that $\min\{1/(2\alpha),1\} = 1/(2\alpha)$. □

Given a continuous function $d : \mathbb{H} \times \mathbb{H} \to \mathbb{R}_+$, the L^1 transportation cost of two Borel probability measures μ and ν on $\mathbb{H}$ w.r.t. the cost function d is defined by

$$\mathcal{W}_d(\nu, \mu) \quad = \quad \inf_{\gamma} \int d(x, y)\, \gamma(dx\, dy),$$

where the infimum is taken over all couplings γ with marginals ν and μ respectively. If the function d is a metric, then $\mathcal{W}_d$ is called *Kantorovich distance*. Let $\mathcal{P}$ be the set of Borel probability measures on $\mathbb{H}$ with finite first moment, i.e. $\int |x|\, \mu(dx) < \infty$ for $\mu \in \mathcal{P}$.

If $\beta < 0$, then (2.7) reveals that b is a *contraction* on $\mathbb{H}$ w.r.t. $|\cdot|_\alpha$ which implies the following trivial result.

Proposition 1. *Let Assumption 22 be true and $\beta < 0$, then*

$$\mathcal{W}_{d_1}(\mu p_t, \nu p_t) \quad \leq \quad e^{-ct}\, \mathcal{W}_{d_1}(\mu, \nu) \quad \text{for any } \mu, \nu \in \mathcal{P} \text{ and } t \geq 0, \tag{2.10}$$

where the distance d_1 and the rate c are given by

$$d_1(x, y) \quad = \quad |x - y|_\alpha \quad \text{and} \quad c \ = \ \min\left\{\alpha^{-1}, |\beta|\right\}.$$

The assumption $\beta < 0$ implies that the underlying deterministic system of (2.1) is contractive and thus the statement even holds in the case $\mathcal{G} \equiv 0$. A proof using synchronous coupling is given in Section 2.2.3 for the readers convenience.

In order to tackle the case $\beta \geq 0$, we demand that the noise in the space $\mathbb{H}^l$ is nondegenerate, i.e. that Assumption 23 holds true. Moreover, we assume that b is a contraction w.r.t. $|\cdot|_\alpha$ for "large distances". More precisely, we assume :

Assumption 24. *There are $R \in (0, \infty)$ and $0 \leq M < 1$ such that*

$$|b(x) - b(y)|_\alpha \quad \leq \quad M\ |x - y|_\alpha \quad \text{for any } x, y \in \mathbb{H} \text{ with } |x - y|_\alpha \geq R.$$

The assumption is for example satisfied, if b vanishes outside of a ball. Subsequently, we will replace Assumption 24 by a more general geometric drift condition, cf. Assumption 25. Denote by $\lambda_\star = \min\{\lambda_k : k \in \mathbb{N}, 1 \leq k \leq n\}$ the smallest eigenvalue of $\mathcal{G}$ on $\mathbb{H}^l$. We present our first main statement.

Theorem 8. *Let Assumption 22, 23, and 24 be true and assume $\beta \geq 0$. There is a distance d_2 and a constant $c \in (0, \infty)$ such that*

$$\mathcal{W}_{d_2}(\mu p_t, \nu p_t) \quad \leq \quad e^{-ct}\, \mathcal{W}_{d_2}(\mu, \nu) \quad \text{for any } \mu, \nu \in \mathcal{P} \text{ and } t \geq 0. \tag{2.11}$$

The rate c is given explicitly in (2.41). If $\beta > 0$, a lower bound is given by

$$c \quad \geq \quad \frac{1}{2} \exp\left(-\frac{\beta}{8\lambda_\star} R^2\right) \min\left\{\beta, 1 - M, \frac{1}{2\alpha}\right\}. \tag{2.12}$$

The distance d_2 is equivalent to $|\cdot|$ and is given by

$$d_2(x,y) \;=\; f\left(|x-y|_\alpha\right),$$

where $f:\mathbb{R}_+\to\mathbb{R}_+$ is a strictly increasing, concave and continuous function with $f(0)=0$. The function is explicitly defined in (2.38)*. It satisfies the relations*

$$\begin{aligned} \frac{1}{2} \;&\le\; f'(r)\,\exp\left(\frac{\beta}{8\lambda_\star}r^2\right) \;\le\; 1 \quad \text{for } 0<r<R \text{ and} \\ f(r) \;&=\; f(R)+\frac{1}{2}\exp\left(-\frac{\beta}{8\lambda_\star}R^2\right)(r-R) \quad \text{for } r\ge R. \end{aligned}$$

Theorem 8 extends ideas from [50, 51] to an infinite-dimensional and degenerate setting using asymptotic couplings in a similar spirit as [72, 115]. The proof is given in Section 2.2.3. The occurring factors 1/2 and 1/8 are, to some extend, arbitrary. Notice that the degenerate case $\mathcal{G}|_{\mathbb{H}^h}\equiv 0$ is covered by the statement. Given $p\ge 1$, we write

$$\mathcal{W}^p(\mu,\nu) \;=\; \left(\inf_\gamma \int |x-y|^p\;\gamma(dx\,dy)\right)^{1/p}$$

for the L^p *Wasserstein distance* of two measures μ and ν. The Kantorovich contraction (2.11) has several consequences. Following [51], we present some applications.

Corollary 6. *There is a unique invariant probability measure $\pi\in\mathcal{P}$ such that*

$$\mathcal{W}^1(\delta_x p_t,\pi) \;\le\; 4\,\alpha\, e^{\frac{\beta R^2}{8\lambda_\star}}\, e^{-c\,t}\,\mathcal{W}^1(\delta_x,\pi) \quad \text{for any } x\in\mathbb{H} \text{ and } t\ge 0. \tag{2.13}$$

For measurable $g:\mathbb{H}\to\mathbb{R}$, we denote the Lipschitz constant w.r.t. d_2 by

$$|g|_{\mathrm{Lip}(d_2)} = \sup\left\{|g(x)-g(y)|/d_2(x,y) : x,y\in\mathbb{H}, x\ne y\right\}. \tag{2.14}$$

Corollary 7. *For any Lipschitz function g and $t\ge 0$,*

$$\sup\left\{\frac{|(p_tg)(x)-(p_tg)(y)|}{|x-y|} : x,y\in\mathbb{H},\; x\ne y\right\} \;\le\; \sqrt{2}\,\alpha\,|g|_{\mathrm{Lip}(d_2)}\;e^{-c\,t}.$$

Further consequences are discussed after Theorem 9.

We now generalize Theorem 8 and replace Assumption 24 by a geometric drift condition using arguments related to the recent work [48]. Lyapunov drift conditions are widely used to study ergodicity and stability of Markov processes, see e.g. [118, 96, 69] and the references therein. Suppose that a continuous function $V:\mathbb{H}\to[1,\infty)$ is given for which the Fréchet derivatives $\mathcal{D}V$ and $\mathcal{D}^2V$ exist, are continuous and bounded in bounded subsets of $\mathbb{H}$. Let

$$\mathcal{L}V(x) \;=\; \langle\mathcal{D}V(x),-x+b(x)\rangle \;+\; \frac{1}{2}\sum_{k=1}^{\infty}\lambda_k\,\mathcal{D}^2V(x)\,[\mathbf{e}_k,\mathbf{e}_k]. \tag{2.15}$$

Assumption 25. *There are constants $C, \eta \in (0,\infty)$ such that for any $x \in \mathbb{H}$,*

$$\mathcal{L}V(x) \;\leq\; C - \eta\, V(x). \tag{2.16}$$

Moreover, we assume that

$$\lim_{r\to\infty} \inf_{|x|=r} V(x) \;=\; \infty \qquad\qquad \textit{and} \qquad\qquad \theta \;=\; \sup_{x\in\mathbb{H}} \frac{|\mathcal{D}V(x)|}{V(x)} \;<\; \infty.$$

The condition $\theta < \infty$ is imposed for simplicity and can be weakened. We call a function V satisfying the above conditions a *Lyapunov function.* A typical candidate for a Lyapunov function is $V(x) = 1 + |x|^2$. Let

$$S = \{(x,y) \in \mathbb{H}\times\mathbb{H} : V(x) + V(y) < 8C/\eta\} \quad \text{and} \quad R = \sup_{(x,y)\in S} |x-y|_\alpha\,. \tag{2.17}$$

The set is chosen such that for any $(x,y) \notin S$,

$$\mathcal{L}V(x) + \mathcal{L}V(y) \;\leq\; -(\eta/2)\;(\,V(x) + V(y)\,)\;-\;2\,C. \tag{2.18}$$

Since V is bounded from below, the set S cannot be empty and by continuity of V, $R > 0$. Moreover, Assumption 25 implies that $R < \infty$.

Let $\mathcal{P}_V$ be the set of probability measures μ on $\mathbb{H}$ satisfying $\int V(x)\,\mu(dx) < \infty$ and write $\lambda^\star = \max\{\lambda_k : k \in \mathbb{N}, 1 \leq k \leq n\}$ for the largest eigenvalue of $\mathcal{G}$ on $\mathbb{H}^l$. We call a continuous function $d : \mathbb{H}\times\mathbb{H} \to [0,\infty)$ a *semimetric*, if it is symmetric and satisfies $d(x,y) = 0$ if and only if $x = y$. We present our main result.

Theorem 9. *Let Assumptions 22, 23 and 25 be true and assume $\beta \geq 0$. There is a semimetric d_3 and a constant $c \in (0,\infty)$ such that*

$$\mathcal{W}_{d_3}(\mu p_t, \nu p_t) \;\leq\; e^{-ct}\,\mathcal{W}_{d_3}(\mu,\nu) \quad \textit{for any } \mu,\nu \in \mathcal{P}_V \textit{ and } t \geq 0. \tag{2.19}$$

The rate c is given explicitly in (2.61)*. If $\beta > 0$, then a lower bound is given by*

$$c \;\geq\; \frac{1}{2}\,\min\left\{\,\exp\left(-\frac{\beta}{8\lambda_\star}R^2 - 2\theta\frac{\lambda^\star}{\lambda_\star}R\right)\,\min\left\{\,\frac{\beta}{2}\,,\,\frac{1}{4\alpha}\right\}\,,\,\eta\,\right\}. \tag{2.20}$$

The semimetric d_3 is given by

$$d_3(x,y) \;=\; f\left(|x-y|_\alpha\right)\left(1 + \epsilon\,V(x) + \epsilon\,V(y)\right), \tag{2.21}$$

where $\epsilon \in (0,\infty)$ is a small constant. The function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is non-decreasing, concave and continuous with $f(0) = 0$. It is constant for $r \geq R$ and satisfies for $0 < r < R$ the inequality

$$\frac{1}{2} \;\leq\; f'(r)\,\exp\left(\frac{\beta}{8\lambda_\star}r^2 + 2\theta\frac{\lambda^\star}{\lambda_\star}r\right) \;\leq\; 1.$$

The explicit definitions of f and ϵ are given in (2.59) *and* (2.61) *further below.*

The extension of Theorem 8 to the case of a geometric drift condition is in the same spirit as in the related work [48]. The multiplicative structure of d_3 is inspired by [69]. A proof of the theorem is given in Section 2.2.3. Notice that the function d_3 is in general not a metric, since the triangle inequality might be violated. Nevertheless, as pointed out in [69, Lemma 4.14], one can show that if the Lyapunov function V growths at most exponentially in $|x|$, then d_3 satisfies a *weak* triangle inequality, i.e. there is $K \in (0, \infty)$ s.t. for all $x, y, z \in \mathbb{H}$ we have $d_3(x, y) \leq K[d_3(x, z) + d_3(z, y)]$. This is sufficient for several applications, as we discuss now. The applications are well-known in the literature.

Corollary 8. *Suppose that the assumptions of Theorem 9 hold true. Let $p \geq 1$ and assume there is a constant $K \in (0, \infty)$ such that $|x - y|^p \leq K\,(V(x) + V(y))$ for any $x, y \in \mathbb{H}$. Then, the Markov kernels (p_t) admit a unique invariant probability measure $\pi \in \mathcal{P}_V$ such that for any $\mu \in \mathcal{P}_V$ and $t \geq 0$,*

$$\mathcal{W}^p(\mu p_t, \pi)^p \quad \leq \quad 2 \,\exp\left(\frac{\beta}{8\lambda_\star} + 2\theta\,\frac{\lambda^\star}{\lambda_\star}\right) \max\left\{1, \frac{K}{\epsilon\ \min\{1, R\}}\right\} e^{-c\,t}\ \mathcal{W}_{d_3}(\mu, \pi).$$

If π is symmetric w.r.t. (p_t), which is for example the case in the setting considered in Section 2.3.1 further below, then Corollary 8 implies a $L^2(\pi)$ spectral gap, cf. [70, Proposition 2.8 and Theorem 2.15] for a precise statement. A Kantorovich contraction as in Theorem 9 has further remarkable consequences: For example it allows to make statements about Markov processes which are perturbations of (X_t), cf. e.g. [69, Section 4.1: Stability of invariant measures]. Moreover, it allows for quantifications of bias and variances of ergodic averages, cf. [91, 51, 48]. Since the latter sources do not provide statements which are directly applicable in the setting of Theorem 9, we formulate slightly adapted versions. Notice that similarly to (2.14) we can define $|\cdot|_{\mathrm{Lip}(d_3)}$ for the semimetric d_3.

Corollary 9. *Under the assumptions of Theorem 9, it holds*

$$\sup\left\{\frac{|(p_t g)(x) - (p_t g)(y)|}{|x - y|} : x \neq y\right\} \quad \leq \quad \sqrt{2}\alpha\,|g|_{\mathrm{Lip}(d_2)}\,(1 + \epsilon V(x) + \epsilon V(y))\,e^{-c\,t}$$

for any measurable function g satisfying $|g|_{\mathrm{Lip}(d_3)} < \infty$ and any $t \geq 0$.

In particular, if $x \mapsto p_t g(x)$ is Fréchet differentiable at some point $x \in \mathbb{H}$, then Corollary 9 provides a bound on $|\nabla p_t g(x)|$.

Corollary 10. *Under the assumptions of Corollary 8, we have for any measurable function $g : \mathbb{H} \to \mathbb{R}$ with $|g|_{\mathrm{Lip}(d_3)} < \infty$, any $x \in \mathbb{H}$ and $t > 0$,*

$$\left| E_x\left[\frac{1}{t}\int_0^t g(X_s)\,ds - \int g d\pi\right]\right| \quad \leq \quad \frac{1 - e^{-c\,t}}{c\,t}\,|g|_{\mathrm{Lip}(d_3)}\,R\,(1 + \epsilon\,V(x) + \epsilon\,C/\eta).$$

Corollary 11. *Suppose that the assumptions of Theorem 9 hold true. Moreover, we assume that the function $x \mapsto V(x)^2$ satisfies the geometric drift condition*

$$(\mathcal{L}V^2)(x) \quad \leq \quad C^\star - \eta^\star V(x)^2 \qquad \textit{for any } x \in \mathbb{R}^d,$$

with constants $C^\star, \eta^\star \in (0,\infty)$. It follows that

$$|\mathrm{Cov}_x[g(X_t), g(X_{t+h})]| \quad \leq \quad \frac{3R^2}{2}|g|^2_{\mathrm{Lip}(d_3)}(1 + 2\,\epsilon^2[C^\star/\eta^\star + e^{-\eta^\star t}V(x)^2])e^{-ch} \tag{2.22}$$

for any measurable function g satisfying $||g||_{\mathrm{Lip}(d_3)} < \infty$ and any $t \geq 0$. In particular,

$$\mathrm{Var}_x\left[\frac{1}{t}\int_0^t g(X_s)\,ds\right] \quad \leq \quad \frac{3\,R^2}{c\,t}||g||^2_{\mathrm{Lip}(d_3)}\,\left(1 + 2\,\epsilon^2\left[C^\star/\eta^\star \,+\, e^{-\eta^\star t}\,V(x)^2\right]\right).$$

The proofs of Corollaries 9, 10 and 11 are nearly identical to the ones given in [51, 48] and are not repeated here. We remark that Theorem 9 can also be used to make statements about the existence of solutions for the Poisson equation $-\mathcal{L}u = g$ associated with (2.1) for a certain class of functions g. For a precise statement regarding this topic, we refer the reader to [147, Theorem 3.1].

2.2.2 Couplings

We introduce the couplings used to derive upper bounds on the Kantorovich distances occurring in Proposition 1, Theorem 8 and Theorem 9.

Synchronous coupling

Fix initial values $(x_0, y_0) \in \mathbb{H} \times \mathbb{H}$. We call (X_t, Y_t) a *synchronous coupling*, if it is a solution of the equation

$$\begin{aligned} dX_t \;&=\; -X_t\,dt \;+\; b(X_t)\,dt \;+\; \sqrt{2}\,dW_t, \\ dY_t \;&=\; -Y_t\,dt \;+\; b(Y_t)\,dt \;+\; \sqrt{2}\,dW_t, \qquad (X_0, Y_0) \;=\; (x_0, y_0), \end{aligned}$$

on the space $\mathbb{H} \oplus \mathbb{H}$, where (W_t) is a $\mathcal{G}$-Wiener process on $\mathbb{H}$. The coupling is well-known and used to prove Proposition 1.

Reflection coupling for nondegenerate and finite-dimensional diffusions

In order to explain the coupling leading to Theorem 8 and Theorem 9, we shortly recall reflection coupling for nondegenerate and finite-dimensional diffusions, which goes back to [107]. We consider the following SDE in $\mathbb{R}^d$:

$$dR_t \;=\; a(R_t)\,dt \;+\; \sigma\,dB_t, \tag{2.23}$$

where $a : \mathbb{R}^d \to \mathbb{R}^d$ is (say) Lipschitz, $\sigma \in \mathbb{R}^{d\times d}$ satisfies $\det(\sigma) > 0$ and (B_t) is a d-dimensional Brownian motion. A reflection coupling (R_t, S_t) starting at $(r_0, s_0) \in \mathbb{R}^{2d}$ is a solution of the equation

$$\begin{aligned} dR_t &= a(R_t)\, dt \;+\; \sigma\, dB_t, \quad (R_0, S_0) \;=\; (r_0, s_0), \\ dS_t &= a(S_t)\, dt \;+\; \sigma \left(\mathrm{I}_d - 2\frac{\sigma^{-1}(R_t - S_t)}{|\sigma^{-1}(R_t - S_t)|} \left\langle \frac{\sigma^{-1}(R_t - S_t)}{|\sigma^{-1}(R_t - S_t)|}, \cdot \right\rangle \right) dB_t, \quad t < T \\ S_t &= R_t, \quad t \geq T, \end{aligned}$$

where $T = \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time. One of the crucial properties of reflection coupling is that $r_t = |R_t - S_t|$ satisfies almost surley the equation

$$dr_t \;=\; r_t^{-1} \langle R_t - S_t, a(R_t) - a(S_t) \rangle\, dt + 2 \left|\sigma^{-1}(R_t - S_t)\right|^{-1} r_t\, dW_t, \quad t < T,$$

where (W_t) is a one-dimensional Brownian motion. We see that the driving noise (W_t) has a direct impact on $|R_t - S_t|$, see [51] for details.

Switching between reflection and synchronous coupling

We present the coupling used to prove Theorem 8 and Theorem 9. Before we introduce the coupling in a rigorous way, we shortly explain the strategy: Let (X_t, Y_t) be a *synchronous coupling* of solutions to (2.1), i.e. let the processes (X_t) and (Y_t) be driven by the same noise. We argue pathwise. Assume that $X_t - Y_t$ satisfies for some $t \geq 0$ the inequality

$$H_l \left|X_t^l - Y_t^l\right| \;\leq\; (1 - H_h) \left|X_t^h - Y_t^h\right|/2, \tag{2.24}$$

then Assumption 22 implies that

$$\left|b^h(X_t) - b^h(Y_t)\right| \;\leq\; H_l \left|X_t^l - Y_t^l\right| \;+\; H_h \left|X_t^h - Y_t^h\right| \;\leq\; (1 + H_h) \left|X_t^h - Y_t^h\right|/2,$$

where $(1+H_h)/2 < 1$ by assumption. In particular, as long as $X_t - Y_t$ satisfies (2.24), $\left|X_t^h - Y_t^h\right|$ decreases exponentially fast, while $\left|X_t^l - Y_t^l\right|$ might increase at the same time. At some point, as time increases, $X_t - Y_t$ might not satisfy (2.24) any more. The idea is now to use a reflection coupling of X_t^l and Y_t^l in the space $\mathbb{H}^l$ with the aim of decreasing $\left|X_t^l - Y_t^l\right|$ relative to $\left|X_t^h - Y_t^h\right|$. As soon as $X_t - Y_t$ satisfies again (2.24), we switch the coupling to a *synchronous coupling* and wait for a decrease of $\left|X_t^h - Y_t^h\right|$. If $\left|X_t^h - Y_t^h\right|$ gets again "small" compared to $\left|X_t^l - Y_t^l\right|$, we switch to a *reflection coupling* in $\mathbb{H}^l$ and so an and so forth. The coupling is visualized in Figure 2.1. As remarked above, during the phases $X_t - Y_t$ satisfies (2.24), $\left|X_t^l - Y_t^l\right|$ might increase. In order to see a contraction of $X_t - Y_t$, we measure the distance with the weighted norm $|\cdot|_\alpha$ and replace the sector condition (2.24) by (2.8) provided by Lemma 6. Indeed, as long as $X_t - Y_t$ satisfies (2.8), $|X_t - Y_t|_\alpha$ decreases exponentially fast. This is of course not true, if $X_t - Y_t$ fails to satisfy (2.8). Nevertheless, in the setting of Theorem 8, an exponential decay of $f(|X_t - Y_t|_\alpha)$ still holds *on average*,

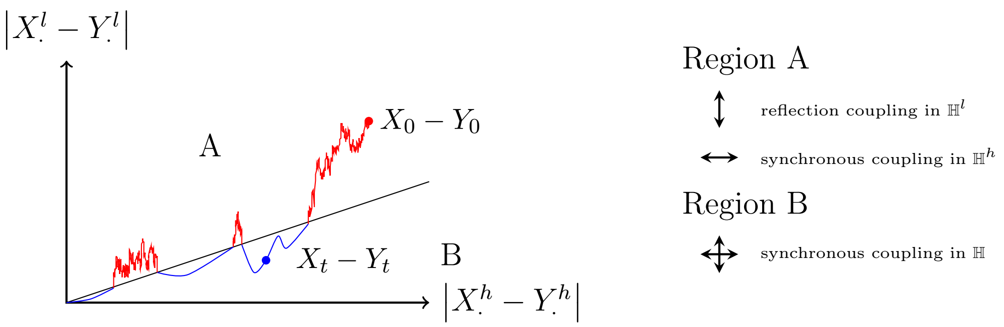

Figure 2.1: Asymptotic coupling for a degenerate diffusion

if we use an appropriate concave function f following [51]. The coupling strategy is similar to the ones from [72, 115]: We identify a region where the deterministic system corresponding to (2.1) has a contraction property and then use the available noise to drive the coupling into those regions.

We now define the coupling in a rigorous way. Fix small $\delta > 0$ and denote

$$\begin{aligned} \mathcal{S}_{SC} &= \left\{x \in \mathbb{H} : 4\,(\beta+1)\left|x^l\right| \le \left|x^h\right|\right\} \cup \left\{x \in \mathbb{H} : |x|_\alpha \le \delta/2\right\}, \\ \mathcal{S}_{RC} &= \left\{x \in \mathbb{H} : 2\,(\beta+1)\left|x^l\right| \ge \left|x^h\right|\right\} \cap \left\{x \in \mathbb{H} : |x|_\alpha \ge \delta\right\}. \end{aligned} \tag{2.25}$$

In comparison to the informal description further above, we add transition regions to realize transitions between the different coupling types. We describe the coupling first in words: The driving noise in the subspace $\mathbb{H}^h$ is always coupled *synchronously*, i.e. the same noise is used to drive X_t^h and Y_t^h. In the finite-dimensional subspace $\mathbb{H}^l$ we use a *reflection coupling* of the driving noise if $X_t - Y_t \in \mathcal{S}_{RC}$ and a *synchronous coupling* if $X_t - Y_t \in \mathcal{S}_{SC}$. The definition of the above sets is motivated by Lemma 6. The two sets $\mathcal{S}_{RC}$ and $\mathcal{S}_{SC}$ are closed, disjoint and $\inf_{x\in\mathcal{S}_{RC},y\in\mathcal{S}_{SC}} |x-y| > 0$. The region "in between", i.e. $\mathbb{H} \setminus (\mathcal{S}_{RC} \cup \mathcal{S}_{SC})$, is a transition region where a mixture of both couplings is used. The parameter δ occurs only for technical reasons and one should think of δ being close to 0.

We now specify the technical realization of the coupling which follows [51, Section 6]. For given $(x, y) \in \mathbb{H} \times \mathbb{H}$, we define linear operators $R(x, y) : \mathbb{H} \to \mathbb{H}$ and $S(x, y) : \mathbb{H} \to \mathbb{H}$ by

$$\begin{aligned} S(x,y)z &= z^h + \mathrm{sc}(x,y)\, z^l \qquad \text{and} \\ R(x,y)z &= \mathrm{rc}(x,y)\, z^l. \end{aligned}$$

Here $\mathrm{sc}, \mathrm{rc} : \mathbb{H} \oplus \mathbb{H} \to [0, 1]$ are Lipschitz functions, satisfying for any $x, y \in \mathbb{H}$,

$$\mathrm{sc}^2(x,y) + \mathrm{rc}^2(x,y) = 1 \quad \text{and} \quad \mathrm{rc}(x,y) = \begin{cases} 1 & \text{if } (x-y) \in \mathcal{S}_{RC}. \\ 0 & \text{if } (x-y) \in \mathcal{S}_{SC}. \end{cases} \tag{2.26}$$

Regarding the existence of the above functions, we remark that it is enough to construct a suitable function $h : \mathbb{R}_+ \times \mathbb{R}_+ \to [0,1]$ such that

$$\mathrm{rc}(x,y) \quad = \quad h\left(\left|x^h - y^h\right|, \left|x^l - y^l\right|\right) \quad \text{and} \quad \mathrm{sc}(x,y) \quad = \quad \sqrt{1 - \mathrm{rc}^2(x,y)}$$

satisfy the above conditions. This can be done using standard cut-off techniques. Let now $\mathbb{W}^1$ and $\mathbb{W}^2$ be independent $\mathcal{G}$-Wiener processes on $\mathbb{H}$ and fix some arbitrary unit vector $u \in \mathbb{H}^l$. Given starting points $(x_0, y_0) \in \mathbb{H} \times \mathbb{H}$ we define $(X_t, Y_t)_{t \geq 0}$ as a strong solution of

$$\begin{aligned} dX_t &= -X_t\, dt + b(X_t)\, dt + \sqrt{2}\, R(U_t)\, d\mathbb{W}^1_t + \sqrt{2}\, S(U_t)\, d\mathbb{W}^2_t, \\ dY_t &= -Y_t\, dt + b(Y_t)\, dt + \sqrt{2}\mathcal{G}^{1/2}\left(I - 2e_t \langle e_t, \cdot \rangle\right) \mathcal{G}^{-1/2}\, R(U_t)\, d\mathbb{W}^1_t + \sqrt{2} S(U_t)\, d\mathbb{W}^2_t, \end{aligned}$$

on $\mathbb{H} \oplus \mathbb{H}$, where $(X_0, Y_0) = (x_0, y_0)$, $U_t = (X_t, Y_t)$ and

$$e_t = \begin{cases} \left|\mathcal{G}^{-1/2}(X^l_t - Y^l_t)\right|^{-1} \mathcal{G}^{-1/2}(X^l_t - Y^l_t) & \text{if } \left|X^l_t - Y^l_t\right| > 0, \\ u & \text{if } \left|X^l_t - Y^l_t\right| = 0. \end{cases} \tag{2.27}$$

Notice that $\left|X^l_t - Y^l_t\right| = 0$ implies $\mathrm{rc}(X_t, Y_t) = 0$ and thus the arbitrary value u in (2.27) is not relevant for the dynamic. The operator $\mathcal{G}^{-1/2}$ is well defined on the space $\mathbb{H}^l$ due to Assumption 23. Furthermore, by assumption, the maps $(x,y) \mapsto (b(x), b(y))$, $(x,y) \mapsto R(x,y)$ and $(x,y) \mapsto S(x,y)$ are Lipschitz on $\mathbb{H} \oplus \mathbb{H}$. Observe that $(\mathbb{W}_t)$ defined by $\mathbb{W}_t = (\mathbb{W}^1_t, \mathbb{W}^2_t)$ is a $\mathbb{G}$-Wiener process on $\mathbb{H} \oplus \mathbb{H}$ with $\mathbb{G}(x,y) = (\mathcal{G}x, \mathcal{G}y)$. Therefore, the above equation is a standard SDE with Lipschitz coefficients on the Hilbert space $\mathbb{H} \oplus \mathbb{H}$. The existence of a continuous, unique and non-explosive solution is well-known, see e.g. [103, Theorem 3.3]. Using the infinite-dimensional analog of Levy's characterization of Brownian motion, see e.g. [35, Theorem 4.4], and (2.26) one can check that

$$\begin{aligned} t \quad &\mapsto \quad \int_0^t R(U_s)\, d\mathbb{W}^1_s \;+\; \int_0^t S(U_s)\, d\mathbb{W}^2_s \quad \text{and} \\ t \quad &\mapsto \quad \int_0^t \mathcal{G}^{1/2}\left(I - 2e_s \langle e_s, \cdot \rangle\right) \mathcal{G}^{-1/2}\, R(U_s)\, d\mathbb{W}^1_s \;+\; \int_0^t S(U_s)\, d\mathbb{W}^2_s \end{aligned}$$

are $\mathcal{G}$-Wiener processes on $\mathbb{H}$ and hence (X_t, Y_t) is indeed a coupling.

2.2.3 Proofs

Proof of Proposition 1. Fix initial values $x_0, y_0 \in \mathbb{H}$. We first show that (2.10) holds for Dirac measures $\mu = \delta_{x_0}$ and $\nu = \delta_{y_0}$. Let (X_t, Y_t) be a synchronous coupling as defined in Section 2.2.2. In the following, all Itô differential (in)equalities hold almost surely for all $t \geq 0$ without further mentioning.

Observe that the difference process $Z_t = X_t - Y_t$ satisfies the equation

$$dZ_t \quad = \quad (\, -Z_t + b(X_t) - b(Y_t)\,)\; ds. \tag{2.28}$$

As before, we write Z^l_t and Z^h_t for the orthogonal projections of Z_t onto $\mathbb{H}^l$ and $\mathbb{H}^h$ respectively.

Lemma 7. *The processes* $(|Z_t^l|)$ *and* $(|Z_t^h|)$ *satisfy the equations*

$$d\left|Z_t^l\right| = I_{Z_t^l\neq 0}\left\langle \frac{Z_t^l}{\left|Z_t^l\right|}, -Z_t + b(X_t) - b(Y_t)\right\rangle dt, \tag{2.29}$$

$$d\left|Z_t^h\right| = I_{Z_t^h\neq 0}\left\langle \frac{Z_t^h}{\left|Z_t^h\right|}, -Z_t + b(X_t) - b(Y_t)\right\rangle dt. \tag{2.30}$$

We proof Lemma 7 further below and continue, assuming that it holds true. The coupling (X_t, Y_t) yields an upper bound for the Kantorovich distance:

$$\mathcal{W}_{d_1}(\delta_{x_0}p_t, \delta_{y_0}p_t) \leq E\left[|Z_t|_\alpha\right] = e^{-ct}\, E\left[e^{ct}\,|Z_t|_\alpha - |Z_0|_\alpha\right] + e^{-ct}\, E\left[|Z_0|_\alpha\right].$$

The product rule for semimartingales implies

$$d(e^{ct}\,|Z_t|_\alpha) = c\, e^{ct}\,|Z_t|_\alpha\, dt + e^{ct}\, d\,|Z_t|_\alpha\,. \tag{2.31}$$

Combining Lemma 7 and (2.7), we conclude that

$$d\,|Z_t|_\alpha \leq \left(\beta\left|Z_t^l\right| - \alpha^{-1}\left|Z_t^h\right|_\alpha\right) dt \leq -c\,|Z_t|_\alpha\, dt. \tag{2.32}$$

By (2.31) and (2.32), $E\left[e^{ct}\,|Z_t|_\alpha - |Z_0|_\alpha\right] \leq 0$ and therefore Proposition 1 holds for Dirac measures. For the general case, let $\mu, \nu \in \mathcal{P}$. With arguments similar to [149, Theorem 4.8] one can show that for any coupling γ of μ and ν, it holds

$$\mathcal{W}_{d_1}(\mu p_t, \nu p_t) \leq \int \mathcal{W}_{d_1}(\delta_x p_t, \delta_y p_t)\,\gamma(dx\,dy) \leq e^{-ct}\int d_1(x,y)\,\gamma(dx\,dy).$$

Taking the infimum over all couplings γ, we finish the proof of Proposition 1. □

Proof of Lemma 7. We argue pathwise. The chain rule combined with (2.28) yields

$$d\left|Z_t^l\right|^2 = 2\left\langle Z_t^l, -Z_t + b(X_t) - b(Y_t)\right\rangle dt, \tag{2.33}$$

$$d\left|Z_t^h\right|^2 = 2\left\langle Z_t^h, -Z_t + b(X_t) - b(Y_t)\right\rangle dt. \tag{2.34}$$

We introduce a C^2 approximation of the map $t \mapsto \sqrt{t}$. Given $\epsilon > 0$, we define

$$s(r) = \begin{cases} -(1/8)\,\epsilon^{-3/2}\, r^2 + (3/4)\,\epsilon^{-1/2}\, r + (3/8)\,\epsilon^{1/2} & r < \epsilon \\ \sqrt{r} & r \geq \epsilon. \end{cases} \tag{2.35}$$

For any $r \in [0,\infty)$, $s(r) \to \sqrt{r}$ for $\epsilon \downarrow 0$. Let $r_t^l = \left|Z_t^l\right|^2$. Using (2.33) and the chain rule, we see that for any $t \geq 0$,

$$s(r_t^l) - s(r_0^l) = \int_0^t I_{r_u^l\geq\epsilon}\left\langle \frac{Z_u^l}{|Z_u^l|}, -Z_u + b(X_u) - b(Y_u)\right\rangle du \tag{2.36}$$

$$+ \int_0^t I_{0<r_u^l<\epsilon}\, s'(r_u^l)\, 2\left\langle Z_u^l, -Z_u + b(X_u) - b(Y_u)\right\rangle du. \tag{2.37}$$

Observe that for $0 < r_u^l < \epsilon$,

$$\left|\left\langle X_u^l - Y_u^l, -(X_u - Y_u) + b(X_u) - b(Y_u)\right\rangle\right| \;\leq\; \epsilon \;+\; \sqrt{\epsilon}\,\left|b(X_u) - b(Y_u)\right|.$$

Moreover, $\sup_{u\in[0,t]} |b(X_u) - b(Y_u)|$ can be bounded by a constant, since (X_u) and (Y_u) are continuous and b is Lipschitz. Observe that $\sup_{0\leq r\leq\epsilon} |s'(r)| \lesssim \epsilon^{-1/2}$. The Lebesgue dominated convergence theorem yields that the integral (2.37) vanishes in the limit as $\epsilon \downarrow 0$. Arguing similarly for the integral on the r.h.s. of (2.36), we retrieve (2.29). The proof of (2.30) is analogous. □

Proof of Theorem 8. We first define the function f explicitly. The function is constructed using a technique from [51, 50]. Related constructions can be found in [28, 29, 33, 48]. For real numbers a and b, we write $a \wedge b = \min\{a, b\}$.

$$\begin{aligned} f(r) &= \int_0^r \phi(s\wedge R)\, g(s\wedge R)\, ds, & \Phi(r) &= \int_0^r \phi(s\wedge R)\, ds, \\ \phi(r) &= \exp\left(-\frac{\beta}{8\lambda_\star} r^2\right), & \gamma^{-1} &= \int_0^R \phi(s)^{-1}\,\Phi(s)\, ds, \\ g(r) &= 1 - \frac{\gamma}{2}\int_0^r \phi(s)^{-1}\,\Phi(s)\, ds. \end{aligned} \tag{2.38}$$

We summarize important properties. The derivative of the function f at $r \in (0, \infty)$ is given by the product $\phi(r\wedge R)\, g(r\wedge R)$. The functions ϕ and g are strictly positive and non-increasing on $(0, R)$ and thus f is strictly increasing and concave. Notice that $g(R) = 1/2$. On the interval $[R, \infty)$ the function f is linear with slope $\phi(R)/2$. Moreover, for any $r \in (0, \infty)$,

$$r \;\leq\; \phi(R)^{-1}\Phi(r), \qquad \Phi(r) \;\leq\; r, \quad \text{and} \quad \Phi(r)/2 \;\leq\; f(r) \;\leq\; \Phi(r), \tag{2.39}$$

which follows directly from the above definitions. Notice that $f(r)$ is twice continuously differentiable at $r \in (0, R)$ and that it satisfies for such r the (in)equality

$$4\,\lambda_\star\, f''(r) \;=\; -\beta\, f'(r)\, r - 2\,\lambda_\star\,\gamma\,\Phi(r) \;\leq\; -\beta\, f'(r)\, r - 2\,\lambda_\star\,\gamma\, f(r). \tag{2.40}$$

We define the rate c by

$$c \;=\; \min\left\{f'(R)\,(1-M), f'(R)/(2\alpha), 2\,\lambda_\star\,\gamma\right\}. \tag{2.41}$$

In order to see (2.12), observe that $f'(R) = \phi(R)/2$ and

$$\gamma^{-1} \;=\; \int_0^R \phi(s)^{-1}\Phi(s)ds \;\leq\; \int_0^R \exp\left(\frac{\beta}{8\lambda_\star} s^2\right) s\, ds \;=\; 4\,\lambda_\star\, \frac{\exp\left(\frac{\beta}{8\lambda_\star} R^2\right) - 1}{\beta}.$$

Fix $(x_0, y_0) \in \mathbb{H}\times\mathbb{H}$. We argue that (2.11) holds for Dirac measures $\mu = \delta_{x_0}$ and $\nu = \delta_{y_0}$. Fix small $\delta > 0$ and let $U_t = (X_t, Y_t)$ be the coupling with initial values

(x_0, y_0) defined in Section 2.2.2. We use the notation $Z_t = X_t - Y_t$ and $r_t = |Z_t|_\alpha$. The coupling yields an upper bound for the Kantorovich distance:

$$\mathcal{W}_{d_2}(\delta_{x_0} p_t, \delta_{y_0} p_t) \;\le\; E\left[f(r_t)\right] \;=\; e^{-ct} E\left[e^{ct} f(r_t) - f(r_0)\right] + e^{-ct} E\left[f(r_0)\right]. \quad (2.42)$$

We now establish bounds on $E\left[e^{ct} f(r_t) - f(r_0)\right]$. All Itô differential (in)equalities hold almost surely for all $t \ge 0$ without further mentioning.

Lemma 8. *The process* (r_t) *satisfies*

$$\begin{aligned} dr_t \;&=\; I_{Z_t^l \neq 0} \left\langle \frac{Z_t^l}{\left|Z_t^l\right|}, -Z_t + b(X_t) - b(Y_t) \right\rangle dt \;+\; 2\sqrt{2}\, \mathrm{rc}(U_t)\, \frac{\left|Z_t^l\right|}{\left|\mathcal{G}^{-1/2} Z_t^l\right|}\, dB_t \\ &+\; \alpha\, I_{Z_t^h \neq 0} \left\langle \frac{Z_t^h}{\left|Z_t^h\right|}, -Z_t + b(X_t) - b(Y_t) \right\rangle dt, \end{aligned}$$

where $B_t = \int_0^t \left\langle \mathcal{G}^{-1/2} e_t, d\mathbb{W}_t^{1,l} \right\rangle$ *is a one-dimensional Brownian motion.*

Observe that by (2.26) and (2.25), $Z_s^l = 0$ implies $\mathrm{rc}(U_s) = 0$.

Lemma 9. *The process* $(f(r_t))$ *satisfies*

$$df(r_t) \;=\; f'(r_t)\, dr_t \;+\; 4\, I_{r_t \neq R}\, f''(r_t)\, \mathrm{rc}(U_t)^2 \left|Z_t^l\right|^2 \left|\mathcal{G}^{-1/2} Z_t^l\right|^{-2} dt.$$

Assuming that Lemma 9 holds true, we can apply the product rule for semimartingales to conclude

$$d\left(e^{ct} f(r_t)\right) \;=\; c\, e^{ct}\, f(r_t)\, dt \;+\; e^{ct}\, df(r_t). \quad (2.43)$$

Lemma 10. *There is a function* $h : \mathbb{R}_+ \to \mathbb{R}_+$ *with* $\lim_{r \downarrow 0} h(r) = 0$ *such that*

$$df(r_t) \;\le\; -c\, f(r_t)\, dt \;+\; h(\delta)\, dt \;+\; f'(r_t)\, dM_t^1, \quad (2.44)$$

where $M_t^1 = \int_0^t 2\sqrt{2}\, \mathrm{rc}(U_t) \left|Z_t^l\right| \left|\mathcal{G}^{-1/2} Z_t^l\right|^{-1} dB_t$.

Notice that M_t^1 is a martingale and $f' \le 1$. By Lemma 10, (2.42) and (2.43),

$$\mathcal{W}_{d_2}(\delta_{x_0} p_t, \delta_{y_0} p_t) \le h(\delta)/c + e^{-ct}\, \mathcal{W}_{d_2}(\delta_{x_0}, \delta_{y_0}).$$

Passing to the limit $\delta \to 0$, we see that (2.11) holds for Dirac measures. The general case can be concluded with the same argument used at the end of the proof of Proposition 1. □

Proof of Lemma 8. We first consider the projection of Z_t onto $\mathbb{H}^h$. From the definition of the coupling in Section 2.2.2, we see that

$$dZ_t^h \;=\; \left(-Z_t^h \;+\; b^h(X_t) - b^h(Y_t) \right)\, dt.$$

Using the same approximation argument as in the proof Lemma 7, we conclude

$$d\left|Z_t^h\right| \;=\; I_{Z_t^h\neq 0}\left|Z_t^h\right|^{-1}\left\langle Z_t^h, -Z_t \;+\; b(X_t)-b(Y_t)\right\rangle\; dt.$$

Using the definition of the coupling in Section 2.2.2, we see that the projection of Z_t onto $\mathbb{H}^l$ satisfies

$$dZ_t^l \;=\; \left(-Z_t^l+b^l(X_t)-b^l(Y_t)\right)\; dt \;+\; 2\sqrt{2}\;\mathrm{rc}(U_t)\,\mathcal{G}^{1/2}\,e_t\,\left\langle \mathcal{G}^{-1/2}e_t, d\mathbb{W}_t^{1,l}\right\rangle,$$

Notice that $B_t=\int_0^t\left\langle \mathcal{G}^{-1/2}e_s, d\mathbb{W}_s^{1,l}\right\rangle$ is a one-dimensional Brownian motion, which follows from Levy's characterization of Brownian motion. A Hilbert space version of Itô's formula, see e.g. [60, Theorem 2.9], allows to conclude

$$\begin{aligned} d\left|Z_t^l\right|^2 \;&=\; 2\left\langle Z_t^l,-Z_t \;+\; b(X_t)-b(Y_t)\right\rangle dt+8\;\mathrm{rc}(U_t)^2\;\left|\mathcal{G}^{1/2}e_t\right|^2 dt\\ &+\; 4\sqrt{2}\;\mathrm{rc}(U_t)\;\left\langle Z_t^l,\mathcal{G}^{1/2}e_t\right\rangle\; dB_t. \end{aligned}$$

Given $\epsilon>0$, let $s(t)$ be the C^2 approximation of $t\mapsto\sqrt{t}$ defined in (2.35). Itô's formula shows

$$\begin{aligned} s\left(\left|Z_t^l\right|^2\right)-s\left(\left|Z_0^l\right|^2\right) \;&=\; \int_0^t s'(\left|Z_v^l\right|^2)\,2\,\left\langle Z_v^l,-Z_v \;+\; b(X_v)-b(Y_v)\right\rangle dv \qquad (2.45)\\ &+\; \int_0^t s'(\left|Z_v^l\right|^2)\,8\;\mathrm{rc}(U_v)^2\;\left|\mathcal{G}^{1/2}e_v\right|^2 dv\\ &+\; \int_0^t s''(\left|Z_v^l\right|^2)\,16\,\mathrm{rc}(U_v)^2\,(\left\langle Z_v^l,\mathcal{G}^{1/2}e_v\right\rangle)^2 dv\\ &+\; \int_0^t s'(\left|Z_v^l\right|^2)\,4\sqrt{2}\;\mathrm{rc}(U_v)\;\left\langle Z_v^l,\mathcal{G}^{1/2}e_v\right\rangle\; dB_v. \end{aligned}$$

We now pass to the limit $\epsilon\downarrow 0$. The integral on the r.h.s. of (2.45) converges to

$$\int_0^t I_{Z_u^l\neq 0}\;\left|Z_u^l\right|^{-1}\;\left\langle Z_u^l,-Z_u+b(X_u)-b(Y_u)\right\rangle\; du,$$

which can be argued similarly as in the proof of Lemma 7. Regarding the limits of the remaining integrals, notice that by (2.25) and (2.26),

$$\left|Z_t^l\right| \;<\; \frac{\delta}{4}\;\min\left\{1,\frac{1}{4\;\alpha\;(\beta+1)}\right\} \tag{2.46}$$

implies $\mathrm{rc}(U_t)=0$. Indeed, if $\left|Z_t^h\right|\leq\delta/(4\alpha)$ and (2.46) holds, then $\left|Z_t\right|_\alpha<\delta/2$ and thus $Z_t\in\mathcal{S}_{\mathrm{SC}}$. If $\left|Z_t^h\right|>\delta/(4\alpha)$ and (2.46) holds, then

$$4\;(\beta+1)\left|Z_t^l\right| \;<\; \delta/(4\alpha) \;<\; \left|Z_t^h\right|$$

and thus again $Z_t\in\mathcal{S}_{\mathrm{SC}}$. On the other hand, $s(t)=\sqrt{t}$ for $t\geq\epsilon$, which concludes the lemma. □

Proof of Lemma 9. The function f can be continued to a concave function on $\mathbb{R}$ by setting $f(x) = x$ for $x < 0$. The generalized Itô formula for concave functions, see e.g. [93, Thm. 22.5], implies that $(f(r_t))$ satisfies the equation

$$f(r_t) - f(r_0) \;=\; \int_0^t f'_-(r_s)\, dr_s \;+\; \frac{1}{2}\int_{-\infty}^{\infty} L_t^x\, \mu_f(dx), \tag{2.47}$$

where f'_- denotes the left-derivative of f, μ_f is the signed measure induced by f'_-, i.e. $\mu_f[x,y) = f'_-(y) - f'_-(x)$ for $x \le y$, and L_t^x denotes the right-continuous local time of (r_t). A further consequence of the generalized Itô formula is that, outside of a fixed null set, we retrieve for any measurable and nonnegative function $v : \mathbb{R} \to \mathbb{R}_+$ the equality

$$\int_{\mathbb{R}} L_t^x\, v(x)\, dx \;=\; \int_0^t v(r_s)\, d[r]_s \qquad \forall t \ge 0, \tag{2.48}$$

see e.g. [93, Thm. 22.5]. Since f' exists everywhere and is continuous, we have $\mu_f[\{R\}] = 0$. Observe that f is twice continuously differentiable except at the point R. Hence by (2.47), (2.48) and Lemma 8, we can conclude that $(f(r_t))$ satisfies the equations

$$\begin{aligned}
f(r_t) - f(r_0) \;&=\; \int_0^t f'(r_s)\, dr_s \;+\; \frac{1}{2}\int_{-\infty}^{\infty} I_{x\neq R}\, L_t^x\, \mu_f(dx), \\
\int_{-\infty}^{\infty} I_{x\neq R}\, L_t^x\, \mu_f(dx) \;&=\; \int_{-\infty}^{\infty} I_{x\neq R}\, L_t^x\, f''(x)\, dx \;=\; \int_0^t I_{r_s\neq R}\, f''(r_s)\, d[r]_s \\
&=\; \int_0^t I_{r_s\neq R}\, f''(r_s)\, 8\, \mathrm{rc}(U_s)^2\, \frac{|Z_s^l|^2}{|\mathcal{G}^{-1/2} Z_s^l|^2}\, ds.
\end{aligned}$$

□

Proof of Lemma 10. Let

$$\begin{aligned}
w(U_t) \;&=\; I_{Z_t^l\neq 0}\, |Z_t^l|^{-1} \left\langle Z_t^l, -Z_t + b(X_t) - b(Y_t)\right\rangle \\
&+\; \alpha\, I_{Z_t^h\neq 0}\, |Z_t^h|^{-1} \left\langle Z_t^h, -Z_t + b(X_t) - b(Y_t)\right\rangle .
\end{aligned}$$

Combining Lemma 8 and 9, we conclude

$$df(r_t) \;=\; \left(f'(r_t)\, w(U_t) \;+ 4\, I_{r_t\neq R}\, f''(r_t)\, \mathrm{rc}(U_t)^2\, \frac{|Z_t^l|^2}{|\mathcal{G}^{-1/2} Z_t^l|^2} \right) dt + dM_t^2. \tag{2.49}$$

with $M_t^2 = \int_0^t f'(r_t)\, dM_t^1$. Notice that for any $t \ge 0$,

$$|\mathcal{G}^{-1/2} Z_t^l| \;\le\; \lambda_\star^{-1/2}\, |Z_t^l|. \tag{2.50}$$

Moreover, Lemma 6 implies that

$$w(U_t) \;\le\; -r_t \;+\; |b(X_t) - b(Y_t)|_\alpha \;\le\; -|Z_t^h| + \beta\, |Z_t^l|. \tag{2.51}$$

Recall that f is concave and non-decreasing. By (2.49), (2.50) and (2.51),

$$\begin{aligned} df(r_t) \;&\le\; f'(r_t)\,\left(-r_t \;+\; |b(X_t) - b(Y_t)|_\alpha\right)\, dt \\ &\quad +\; 4\,\lambda_\star\, I_{r_t\neq R}\, f''(r_t)\ \mathrm{rc}(U_t)^2\ dt \;+\; dM_t^2. \end{aligned} \tag{2.52}$$

If $r_t \ge R$, then Assumption 24 and (2.39) imply

$$-r_t \;+\; |b(X_t) - b(Y_t)|_\alpha \;\le\; -(1-M)\, r_t \;\le\; -(1-M)\, f(r_t). \tag{2.53}$$

If $Z_t \notin \mathcal{S}_{\mathrm{RC}}$ and $r_t \ge \delta$, then by (2.25) and Lemma 6

$$-r_t \;+\; |b(X_t) - b(Y_t)|_\alpha \;\le\; -1/(2\alpha)\, r_t \;\le\; -1/(2\alpha)\, f(r_t). \tag{2.54}$$

If $Z_t \in \mathcal{S}_{\mathrm{RC}}$ and $\delta \le r_t < R$, we argue as follows: By definition we have $\mathrm{rc}(U_t) = 1$. Lemma 6 implies the bound

$$-r_t \;+\; |b(X_t) - b(Y_t)|_\alpha \;\le\; \beta\, r_t. \tag{2.55}$$

Observe that for $r \in (0, R)$, inequality (2.40) holds true and therefore

$$f'(r_t)\,\beta\, r_t \;+ 4\,\lambda_\star\, f''(r_t) \;\le\; -2\,\lambda_\star\,\gamma\, f(r_t). \tag{2.56}$$

Recall (2.38) and (2.39) to see that if $r_t \le \delta$ holds, then we can estimate

$$f'(r_t)\,\beta\, r_t \;\le\; \beta\,\delta \qquad \text{and} \qquad f(r_t) \;\le\; r_t \;\le\; \delta. \tag{2.57}$$

Combining (2.52), (2.53), (2.54), (2.55), (2.56), (2.57) and (2.41), we conclude

$$df(r_t) \;\le\; -c\, f(r_t)\ dt \;+\; (c+\beta)\,\delta\ dt \;+\; dM_t^2. \tag{2.58}$$

The claim follows by setting $h(\delta) \;=\; (c \;+\; \beta)\ \delta$. ☐

Proof of Corollary 6. By (2.39), (2.11) and (2.6), we conclude for any $x, y \in \mathbb{H}$,

$$\mathcal{W}^1(\delta_x p_t, \delta_y p_t) \;\le\; 2\,\phi(R)^{-1}\mathcal{W}_{d_2}(\delta_x p_t, \delta_y p_t) \;\le\; 4\,\alpha\,\phi(R)^{-1}\, e^{-c\,t}\, \mathcal{W}^1(\delta_x, \delta_y).$$

The fact that the Markov kernels (p_t) admit a unique invariant measure π satisfying $\pi p_t = \pi$ for any $t \ge 0$ now follows by standard arguments, see e.g. [51, Corollary 2.5]. ☐

Proof of Corollary 7. The proof follows [51, Section 4]. Let (X_t) be a solution of (2.1) with $X_0 = x$. Assumption 24 implies that the first moments of X_t are uniformly bounded in time, i.e. $\sup_{t\ge 0} E_x[|X_t|] < \infty$. In particular, for any $x \in \mathbb{H}$, $t \ge 0$ and any Lipschitz function g, $\int g(y)\, p_t(x, dy) < \infty$. Fix $x, y \in \mathbb{H}$ and let (X_t, Y_t) be *any* coupling of $\delta_x p_t$ and $\delta_y p_t$. It follows

$$|(p_t g)(x) - (p_t g)(y)| \le E[|g(X_t) - g(Y_t)|] \le |g|_{\mathrm{Lip}(d_2)}\, E[d_2(X_t, Y_t)],$$

and hence by (2.11), (2.6) and (2.39),

$$|(p_t g)(x) - (p_t g)(y)| \le |g|_{\mathrm{Lip}(d_2)}\, e^{-ct} f\left(|x-y|_\alpha\right) \le \sqrt{2}\,\alpha\ |g|_{\mathrm{Lip}(d_2)}\ e^{-ct}\ |x-y|\,.$$

☐

Proof of Theorem 9. The proof is close to the proof of Theorem 8. We use again the coupling from Section 2.2.2, but use a slightly different function f.

$$\begin{aligned} f(r) &= \int_0^{r\wedge R} \phi(s)\, g(s)\, ds & \Phi(r) &= \int_0^{r\wedge R} \phi(s)\, ds \\ \phi(r) &= \exp\left(-\frac{\beta}{8\lambda_\star} r^2 - 2\theta \frac{\lambda^\star}{\lambda_\star} r\right) & \gamma^{-1} &= \int_0^{R} \Phi(s)\, \phi(s)^{-1}\, ds \\ g(r) &= 1 - \frac{\gamma}{2} \int_0^{r\wedge R} \Phi(s)\, \phi(s)^{-1}\, ds \end{aligned} \tag{2.59}$$

We highlight the differences to the situation in Theorem 8. This time, f is constant on $[R, \infty)$ and it is not differentiable at the point R. Nevertheless, it is concave and the left-derivative f'_- exists everywhere. Observe that the inequalities (2.39) still hold true on the interval $[0, R]$. Moreover, the function f is twice continuously differentiable on $(0, R)$ and satisfies there the (in)equality

$$\begin{aligned} 4\,\lambda_\star\, f''(r) &= -f'(r)\, (\beta\, r + 8\,\theta\,\lambda^\star) - 2\,\lambda_\star\,\gamma\,\Phi(r) \\ &\le -f'(r)\, (\beta\, r + 8\,\theta\,\lambda^\star\,) - 2\,\lambda_\star\,\gamma\, f(r). \end{aligned} \tag{2.60}$$

The contraction rate c in (2.19) and the constant ϵ in (2.21) are given by

$$c \;=\; \min\left\{\lambda_\star\gamma, \frac{\phi(R)}{8\,\alpha}, \frac{\eta}{2}\right\} \quad \text{and} \quad 2\,C\,\epsilon \;=\; \min\left\{\lambda_\star\gamma, \frac{\phi(R)}{8\,\alpha}\right\} \ge c. \tag{2.61}$$

The lower bound (2.20) can be derived similarly as in the proof of Theorem 8.

Fix small $\delta > 0$, initial conditions $(x_0, y_0) \in \mathbb{H} \times \mathbb{H}$ and let $U_t = (X_t, Y_t)$ be the coupling defined in Section 2.2.2. We use the notation

$$\begin{aligned} Z_t &= X_t - Y_t, & r_t &= |Z_t|_\alpha\,, \\ G(x,y) &= 1 \;+\; \epsilon\, V(x) + \epsilon\, V(y) & Q_t &= f(r_t)\, G(X_t, Y_t). \end{aligned}$$

The coupling yields an upper bound for the Kantorovich distance:

$$\mathcal{W}_{d_3}(\delta_{x_0} p_t, \delta_{y_0} p_t) \;\le\; E\left[Q_t\right] \;=\; e^{-ct} E\left[e^{ct} Q_t - Q_0\right] \;+\; e^{-ct} E\left[Q_0\right]. \tag{2.62}$$

We now estimate $E\left[e^{ct} Q_t - Q_0\right]$ and proceed similarly to the proof of Theorem 8. Observe that Lemma 8 still holds true, since we use the same coupling as in the proof of Theorem 8.

Lemma 11. *The process* $(f(r_t))$ *satisfies*

$$\begin{aligned} df(r_t) &= f'_-(r_t)\; dr_t \;+\; \frac{1}{2}\int_{-\infty}^{\infty} L_t^x\; \mu_f(dx) \\ &\le f'_-(r_t)\; dr_t \;+\; 4\, I_{r_t \ne R}\, f''(r_t)\, \mathrm{rc}(U_t)^2 \left|Z_t^l\right|^2 \left|\mathcal{G}^{-1/2} Z_t^l\right|^{-2} dt. \end{aligned}$$

The notation μ_f and L_t^x is defined in the proof of Lemma 9.

Lemma 12. *The process $(G(X_t, Y_t))$ satisfies*

$$dG(X_t, Y_t) \;=\; \epsilon\;(\;\mathcal{L}V(X_t) + \mathcal{L}V(Y_t)\;)\;dt \;+\; dM_t^3, \tag{2.63}$$

where (M_t^3) is a local martingale given by

$$\begin{aligned} dM_t^3 \;&=\; \sqrt{2}\,\epsilon\,\left\langle \mathcal{D}V(X_t) + \mathcal{D}V(Y_t), d\mathbb{W}_t^{2,h} \right\rangle \\ &+\; \sqrt{2}\,\epsilon\;\mathrm{sc}(U_t)\left\langle \mathcal{D}V(X_t) + \mathcal{D}V(Y_t), d\mathbb{W}_t^{2,l} \right\rangle \\ &+\; \sqrt{2}\,\epsilon\;\mathrm{rc}(U_t)\left\langle \mathcal{D}V(X_t) + \mathcal{D}V(Y_t), d\mathbb{W}_t^{1,l} \right\rangle \\ &-\; 2\sqrt{2}\,\epsilon\;\mathrm{rc}(U_t)\left\langle \mathcal{D}V(Y_t), \mathcal{G}^{1/2}e_t \right\rangle \left\langle \mathcal{G}^{-1/2}e_t, d\mathbb{W}_t^{1,l} \right\rangle. \end{aligned}$$

The product rule for semimartingales implies

$$d\left(e^{ct}Q_t\right) \;=\; c\,e^{ct}\,Q_t\,dt \;+\; e^{ct}\,dQ_t. \tag{2.64}$$

Lemma 13. *There is $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{r\downarrow 0} h(r) = 0$ such that*

$$dQ_t \;\le\; -c\,Q_t\;dt \;+\; (1 + \epsilon\,V(X_t) + \epsilon\,V(Y_t))\,h(\delta)\;dt \;+\; dM_t^4, \tag{2.65}$$

where $M_t^4 = \int_0^t f(r_s)\;dM_s^3 + \int_0^t G(X_s, Y_s)\;f'_-(r_s)\;dM_s^1$ is a local martingale.

The martingale (M_t^1) is defined in Lemma 10.

Lemma 14. *For any $t \ge 0$, there is $K_t \in (0, \infty)$, not depending on δ, such that*

$$E\left[e^{ct}Q_t - Q_0\right] \;\le\; K_t\,h(\delta).$$

Combining Lemma 14 with (2.62) yields

$$\mathcal{W}_{d_3}(\delta_{x_0}p_t, \delta_{y_0}p_t) \;\le\; K_t\,h(\delta) \;+\; e^{-ct}\,\mathcal{W}_{d_3}(\delta_{x_0}, \delta_{y_0}).$$

Passing to the limit $\delta \to 0$, we see that (2.19) holds for Dirac measures. The general case can be concluded with the same argument used at the end of the proof of Proposition 1. □

Proof of Lemma 11. The proof is analogous to the proof of Lemma 9, except that now f is not continuously differentiable everywhere. In particular, f'_- has a discontinuity at the point R and therefore we do not have $\mu_f[\{R\}] = 0$. Nevertheless, since f is concave we know that $\mu_f[\{R\}] < 0$. □

Proof of Lemma 12. The assumptions imposed on V allow to apply Itô's formula in a Hilbert space setting, see e.g. [60, Theorem 2.9]. Recalling the definition of the coupling from Section 2.2.2, we see that (2.63) holds true. □

Proof of Lemma 13. The product rule for semimartingales implies that (Q_t) satisfies

$$dQ_t \;=\; G(X_t, Y_t)\; df(r_t) \;+\; f(r_t)\; dG(X_t, Y_t) \;+\; d\left[f(r.), G(X., Y.)\right]_t\,, \quad (2.66)$$

where $[\cdot, \cdot]$ denotes the quadratic variation. By Lemma 11, (2.50) and (2.51),

$$\begin{aligned} G(X_t, Y_t)\; df(r_t) \;&\leq\; G(X_t, Y_t)\, f'_-(r_t)\, \left(-\left|Z^h_t\right| + \beta \left|Z^l_t\right|\right)\, dt \\ &+\; G(X_t, Y_t)\, 4\, \lambda_\star\, I_{r_t \neq R}\, f''(r_t)\, \mathrm{rc}(U_t)^2 dt + dM^5_t, \end{aligned} \quad (2.67)$$

where $M^5_t = \int_0^t G(X_s, Y_s) f'_-(r_s)\; dM^1_s$ is a local martingale.

Lemma 12 and Assumption 25 imply that

$$f(r_t)\; dG(X_t, Y_t) \;\leq\; f(r_t)\, \epsilon\, \left(2\, C - \eta\, \left(V(X_t) + V(Y_t)\right)\right) dt \;+\; dM^6_t, \quad (2.68)$$

with $M^6_t = \int_0^t f(r_t)\; dM^3_t$. Using Lemma 8, 11 and 12, we establish the bound

$$\begin{aligned} [f(r.), G(X., Y.)]_t &= 4\epsilon \int_0^t f'_-(r_s)\, \mathrm{rc}(U_s)^2 \frac{\left|Z^l_s\right|}{\left|\mathcal{G}^{-1/2} Z^l_s\right|^2} \left\langle \mathcal{D}V(X_s) - \mathcal{D}V(Y_s), Z^l_s \right\rangle ds \\ &\leq 4\, \epsilon\, \lambda^\star \int_0^t f'_-(r_s)\; \mathrm{rc}(U_s)^2 (\; |\mathcal{D}V(X_s)| + |\mathcal{D}V(Y_s)|\;)\, ds, \end{aligned}$$

where $\lambda^\star$ is the largest eigenvalue of $\mathcal{G}$ on $\mathbb{H}^l$. Assumption 25 implies

$$[f(r.), G(X., Y.)]_t \;\leq\; 8\, \theta\, \lambda^\star \int_0^t f'_-(r_s)\; \mathrm{rc}(U_s)^2\, G(X_s, Y_s)\; ds. \quad (2.69)$$

Combining (2.66), (2.67), (2.68) and (2.69), we conclude that

$$\begin{aligned} dQ_t \;&\leq\; G(X_t, Y_t) f'_-(r_t)\, \left(-\left|Z^h_t\right| + \beta \left|Z^l_t\right| + 8\, \theta\, \lambda^\star\, \mathrm{rc}(U_t)^2\right)\; dt \\ &+\; G(X_t, Y_t)\, 4\, \lambda_\star\, I_{r_t \neq R}\, f''(r_t)\; \mathrm{rc}(U_t)^2\; dt \\ &+\; f(r_t)\, \epsilon\, \left(2\, C - \eta\, \left(V(X_t) + V(Y_t)\right)\right)\; dt \;+\; dM^5_t \;+\; dM^6_t. \end{aligned}$$

We are now in a position to argue (2.65) and do a pathwise case distinction:

If $r_t > R$, then $(X_t, Y_t) \notin S$ by (2.17). By (2.18) and (2.61),

$$\begin{aligned} f(r_t)\, \epsilon\, \left(2\, C - \eta\, \left(V(X_t) + V(Y_t)\right)\right) \;&\leq\; f(r_t)\, \left(-2\, C\, \epsilon - \eta/2\, \left(\epsilon\, V(X_t) + \epsilon\, V(Y_t)\right)\right) \\ &\leq\; -c\; f(r_t)\, G(X_t, Y_t) \;=\; -c\; Q_t. \end{aligned}$$

Moreover, f is constant on (R, ∞) and thus $f'(r_t) = f''(r_t) = 0$.

Now assume that $\delta \leq r_t \leq R$ and $Z_t \in \mathcal{S}_{\mathrm{RC}}$. By (2.26), we have that $\mathrm{rc}(U_t) = 1$. Notice that equality (2.48) implies for any fixed $t \geq 0$,

$$\lambda_{\mathrm{Leb}}\left(\{0 \leq s \leq t : r_s = R \text{ and } \mathrm{rc}(U_s) > 0\}\right) = 0, \quad (2.70)$$

i.e. the Lebesgue measure of the time (r_s) spends at the point R up to time t, while $\mathrm{rc}(U_s) > 0$ is almost surely zero. This allows us to neglect the case $r_t = R$.

Moreover, f is twice continuously differentiable on $(0, R)$ and fulfils inequality (2.60). We conclude that for $\delta \le r_t < R$ with $Z_t \in \mathcal{S}_{\mathrm{RC}}$,

$$\begin{aligned} & G(X_t, Y_t)\, f'(r_t)\, \left(\beta \left|Z_t^l\right| \ + \ 8\,\theta\,\lambda^\star\right) + G(X_t, Y_t)\, 4\,\lambda_\star\, f''(r_t) + f(r_t)\,\epsilon\, 2\,C \\ \le \quad & -2\,\lambda_\star\,\gamma\, G(X_t, Y_t)\, f(r_t) \ + \ f(r_t)\,\epsilon\, 2\,C \ \le \ -c\, Q_t, \end{aligned}$$

where we used (2.61) and $G \ge 1$.

If $\delta \le r_t \le R$ and $Z_t \notin \mathcal{S}_{\mathrm{RC}}$, then by (2.25), $2\,(\beta+1)\left|Z_t^l\right| \le \left|Z_t^h\right|$, but we do not necessarily have $\mathrm{rc}(U_t) = 1$. Nevertheless, (2.70) is still true and (2.60) implies

$$f'_-(r_t)\, 8\,\theta\,\lambda^\star\, \mathrm{rc}(U_t)^2 \ + \ 4\,\lambda_\star\, f''(r_t)\, \mathrm{rc}(U_t)^2 \ \le \ 0 \qquad \text{for } 0 < r_t < R. \tag{2.71}$$

Lemma 6 shows that

$$-\left|Z_t^h\right| \ + \ \beta\left|Z_t^l\right| \ \le \ -1/(2\alpha)\, r_t \ \le \ -1/(2\alpha)\, f(r_t) \tag{2.72}$$

and thus

$$\begin{aligned} & G(X_t, Y_t)\, f'_-(r_t)\, \left(-\left|Z_t^h\right| + \beta\left|Z_t^l\right|\right) \ + \ f(r_t)\,\epsilon\, 2\,C \\ \le \quad & -\phi(R)/(4\alpha)\, Q_t \ + \ f(r_t)\,\epsilon\, 2\,C \le -c\, Q_t, \end{aligned}$$

where we used (2.61) and the fact that f' is nonnegative and decreasing on $(0, R)$ with $f'_-(R) = \phi(R)/2$.

Now assume $r_t \le \delta$. Similarly to the last case, (2.71) holds true. Since $f'_- \le 1$ and $f(r) \le r$, we can estimate

$$G(X_t, Y_t)\, f'_-(r_t)\, \beta\left|Z_t^l\right| \ + \ f(r_t)\,\epsilon\, 2\,C \ \ \le \ \ G(X_t, Y_t)\, (\beta + 2\,C\,\epsilon)\,\delta$$

We conclude the lemma setting $h(\delta) = (c + \beta + 2\,C\,\epsilon)\,\delta$. □

Proof of Lemma 14. We introduce stopping times

$$\begin{aligned} T \ &= \ \inf\{t \ge 0 : X_t = Y_t\} \quad \text{and} \\ T_m \ &= \ \inf\{t \ge 0 : |X_t - Y_t| \le 1/m \text{ or } \max\{|X_t|, |Y_t|\} \ge m\}. \end{aligned}$$

Since the process (X_t, Y_t) is non-explosive, we have $T_m \uparrow T$ for $m \uparrow \infty$. We get

$$\begin{aligned} E\left[e^{c\,t}\, Q_t\right] \ &= \ E\left[e^{c\,t}\, Q_t\, I_{t<T}\right] \ = \ \lim_{m\to\infty} E\left[e^{c\,t\wedge T_m}\, Q_{t\wedge T_m}\, I_{t<T_m}\right] \\ &\le \ \liminf_{m\to\infty} E\left[e^{c\,t\wedge T_m}\, Q_{t\wedge T_m}\right]. \end{aligned}$$

Fix $m \in \mathbb{N}$ and notice that the stopped process $(M^4_{t\wedge T_m})$ defined in Lemma 13 is a martingale. Using (2.64) and Lemma 13, we conclude

$$\begin{aligned} E\left[e^{c\,(t\wedge T_m)} Q_{t\wedge T_m} - Q_0\right] \ &\le \ E\left[\int_0^{t\wedge T_m} e^{c\,s}\, G(X_s, Y_s)\, ds\right]\, h(\delta) \\ &\le \ E\left[\int_0^{t} e^{c\,s}\, G(X_s, Y_s)\, ds\right]\, h(\delta). \end{aligned}$$

Assumption 25 implies that there is a constant $A \in (0, \infty)$ such that

$$\sup_{t\in[0,\infty)} \left(\, E[V(X_t)] \; + \; E[V(Y_t)] \, \right) \; < \; A.$$

□

Proof of Corollary 8. Let $x, y \in \mathbb{H}$ with $|x-y|_\alpha \le \min\{1, R\}$. By (2.6) and (2.39),

$$|x-y|^p \;\; \le \;\; |x-y|_\alpha \;\; \le \;\; 2\,\phi^{-1}(\min\{1,R\})\; f(|x-y|_\alpha)\;(1+\epsilon\, V(x)+\epsilon\, V(y)).$$

On the other hand, if $|x-y|_\alpha > \min\{1, R\}$, then we get

$$|x-y|^p \le K(V(x)+V(y)) \le \frac{K}{\epsilon f(\min\{1,R\})} f(|x-y|_\alpha)(1+\epsilon\, V(x)+\epsilon\, V(y)).$$

By (2.39), $f(\min\{1,R\}) \;\ge\; \Phi(\min\{1,R\})/2 \;\ge\; \min\{1,R\}\;\phi(\min\{1,R\})/2$. Combining the bounds, we get for any $x, y \in \mathbb{H}$,

$$|x-y|^p \;\; \le \;\; 2\,\phi^{-1}(\min\{1,R\})\; \max\left\{1, \frac{K}{\epsilon\; \min\{1,R\}}\right\}\; d_3(x,y). \tag{2.73}$$

Using (2.73) and Theorem 9, we can conclude that

$$\mathcal{W}^p(\mu p_t, \nu p_t)^p \;\le\; 2\,\phi^{-1}(\min\{1,R\})\; \max\left\{1, \frac{K}{\epsilon\; \min\{1,R\}}\right\}\; e^{-ct}\mathcal{W}_{d_3}(\mu,\nu) \tag{2.74}$$

for any $\mu, \nu \in \mathcal{P}_V$ and $t \ge 0$. Notice that Assumption 25 implies that

$$\sup_{t\ge 0} \int V(y)\,(\delta_x p_t)(dy) < \infty \qquad \text{for any } x \in \mathbb{H}. \tag{2.75}$$

In particular, (2.74) and (2.75) together imply that there is a constant $C \in (0, \infty)$ such that

$$\mathcal{W}^p(\delta_x p_m, \delta_x p_n)^p \;\; = \;\; \mathcal{W}^p(\delta_x p_{m-n} p_n, \delta_x p_n)^p \;\le\; C\, e^{-cn} \tag{2.76}$$

for any integers $m > n > 0$. We see that $(\delta_x p_n)_{n\in\mathbb{N}}$ is a Cauchy sequence w.r.t. $\mathcal{W}^p$. Moreover, the L^p Wasserstein space is Polish and convergence w.r.t. $\mathcal{W}^p$ implies weak convergence. The Krylov-Bogolioubov criteria thus implies that there is a measure π_0 such that $\pi_0 p_1 = \pi_0$, cf. e.g. [74, Theorem 1.10]. It is straightforward to check that $\pi = \int_0^1 \pi_0 p_s\, ds$ is invariant w.r.t. (p_t), cf. e.g. [98, Section 3]. Moreover, Assumption 25 implies that any invariant probability measure $\pi^\star$ satisfies $\pi^\star \in \mathcal{P}_V$, cf. e.g. [73, Proposition 4.24], and thus (2.74) implies that π is the only invariant measure. □

2.3 Applications

We demonstrate the applicability of the results from Section 2.2.

2.3.1 Absolutely continuous measures w.r.t. a normal distribution

General setup

Suppose that $\mathcal{G}$ is the covariance operator of a non-degenerate and centered normal distribution $\mathcal{N}(0, \mathcal{G})$ on a separable Hilbert space $(\mathbb{H}, \langle \cdot, \cdot \rangle, |\cdot|)$, i.e. $\mathcal{G}$ is trace-class, symmetric and positive-definite. Define a probability measure π by (2.2), where $U : \mathbb{H} \to \mathbb{R}$ is a given potential which is bounded from below, Fréchet differentiable and for which $x \mapsto \nabla U(x)$ is Lipschitz. We define b in equation (2.1) as $b(x) = -\mathcal{G}\nabla U(x)$. The results from [71] imply that π is an invariant measure for (p_t), i.e. $\pi p_t = \pi$ for any $t \geq 0$. We now give sufficient conditions under which the results from Section 2.2 are applicable in this context.

Remark 6. *The article [71] by Hairer, Stuart and Voss considers two different SPDEs for which π is a stationary distribution and which can both be used for sampling purposes. The first one is given by*

$$d\tilde{X}_t \;=\; \Delta \tilde{X}_t \, dt \;-\; \nabla U(\tilde{X}_t)\, dt \;+\; \sqrt{2}\, d\tilde{W}_t, \tag{2.77}$$

where $(\tilde{W}_t)$ is a cylindrical Wiener process over $\mathbb{H}$. The second one is given by (2.1) *and this is the one we study in this article. Formally, the latter equation is obtained from* (2.77) *by "preconditioning". The solutions for the equations behave quite differently: While* (2.77) *only admits mild solutions in general, strong solutions are possible for* (2.1). *Moreover, as pointed out in [71] under reasonable assumptions, the process $(\tilde{X}_t)$ is strong Feller and it is possible to apply classical Harris' theorems to study ergodic properties. In contrast to this, the process (X_t) solving* (2.1) *is not strong Feller in general and the study of ergodic properties is more involved.*

Fix an orthonormal basis $(\mathbf{e}_k)_{k \in \mathbb{N}_+}$ of $\mathbb{H}$ such that $\mathcal{G}\mathbf{e}_k = \lambda_k \mathbf{e}_k$ holds for a sequence (λ_k) of positive reals satisfying $\sum_{k=1}^{\infty} \lambda_k < \infty$. For the sake of simplicity, we assume $\lambda_k \downarrow 0$ and $\lambda_1 = 1$. Observe that Theorem 8 and 9 still hold true, if we replace Assumptions 22 and 24 by the slightly more general Assumptions 26 and 27 respectively.

Assumption 26. *There are constants $0 \leq H_h < 1$ and $L_l, L_h, H_l \geq 0$ such that*

$$\left\langle \frac{x^h - y^h}{|x^h - y^h|}, b(x) - b(y) \right\rangle \;\leq\; H_l \left|x^l - y^l\right| \;+\; H_h \left|x^h - y^h\right| \tag{2.78}$$

for any $x, y \in \mathbb{H}$ with $x^h \neq y^h$ and in the case $x^l \neq y^l$, we have

$$\left\langle \frac{x^l - y^l}{|x^l - y^l|}, b(x) - b(y) \right\rangle \;\leq\; L_l \left|x^l - y^l\right| \;+\; L_h \left|x^h - y^h\right|. \tag{2.79}$$

Assumption 27. *There are $R \in (0, \infty)$ and $0 \le M < 1$ such that*

$$I_{x^l \neq y^l} \left\langle \frac{x^l - y^l}{|x^l - y^l|}, b(x) - b(y) \right\rangle + I_{x^h \neq y^h} \alpha \left\langle \frac{x^h - y^h}{|x^h - y^h|}, b(x) - b(y) \right\rangle \le M \, |x - y|_\alpha$$

for any $x, y \in \mathbb{H}$ with $|x - y|_\alpha \ge R$.

In the following we focus on potentials $U : \mathbb{H} \to \mathbb{R}$ of the form

$$U(x) \;=\; \frac{a}{2} |x|^2 \;+\; m(x), \tag{2.80}$$

where $a \ge 0$ and $m : \mathbb{H} \to \mathbb{R}$ satisfies:

Assumption 28. *The function m is bounded from below and Fréchet differentiable. There is $L \ge 1$ such that*

$$|\nabla m(x) - \nabla m(y)| \;\le\; L \; |x - y| \qquad \text{holds true for any } x, y \in \mathbb{H}.$$

Lemma 15. *Let Assumption 28 be true and define $n = \min\left\{ k \in \mathbb{N}_+ : \lambda_{k+1} < \frac{1}{2L} \right\}$. We consider the splitting $\mathbb{H} = \mathbb{H}^l \oplus \mathbb{H}^h$ with $\mathbb{H}^l = \langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_n \rangle$. In this setting, Assumption 26 is satisfied with*

$$H_l \;=\; H_h \;=\; 1/2, \quad L_l \;=\; L_h \;=\; L, \quad \alpha \;=\; 2\,(1 + L) \quad \text{and} \quad \beta \;=\; 2\,L.$$

Proof. Let $x, y \in \mathbb{H}$ with $x^h - y^h \neq 0$. We have

$$\begin{aligned} -\left\langle x^h - y^h, \mathcal{G}(\nabla U(x) - \nabla U(y)) \right\rangle \;&=\; -a \left\langle x^h - y^h, \mathcal{G}(x - y) \right\rangle \\ &\quad\; - \left\langle x^h - y^h, \mathcal{G}(\nabla m(x) - \nabla m(y)) \right\rangle. \end{aligned}$$

Observe that $-a \left\langle x^h - y^h, \mathcal{G}(x - y) \right\rangle \le 0$. Using Cauchy-Schwarz, we get

$$\begin{aligned} \left| \left\langle x^h - y^h, \mathcal{G}(\nabla m(x) - \nabla m(y)) \right\rangle \right| \;&\le\; \lambda_{n+1} \, L \, \left| x^h - y^h \right| \, |x - y| \\ &\le\; 1/2 \, \left| x^h - y^h \right| \, |x - y| \,. \end{aligned}$$

This implies (2.78) with $H_l = H_h = 1/2$. Inequality (2.79) can be argued similarly. □

Lemma 16. *Assume that there is $\mathcal{R} > 0$ such that $\nabla m(x) = 0$ for any $|x| \ge \mathcal{R}$. Then Assumption 27 can be satisfied with $M = 3/4$ and $R = 8\,L\,\mathcal{R}$.*

Proof. Let $x, y \in \mathbb{H}$ with $|x - y|_\alpha \ge R$. The statement is clear if $\min\{|x|, |y|\} \ge \mathcal{R}$. Assume w.l.o.g. that $|x| < \mathcal{R}$, $|y| \ge \mathcal{R}$ and let $z \in \mathbb{H}$ with $|z| = \mathcal{R}$, then

$$\begin{aligned} & I_{x^l \neq y^l} \left\langle \frac{x^l - y^l}{|x^l - y^l|}, b(x) - b(y) \right\rangle \;+\; I_{x^h \neq y^h} \; \alpha \; \left\langle \frac{x^h - y^h}{|x^h - y^h|}, b(x) - b(y) \right\rangle \\ \le\;& \left| (\mathcal{G}(\nabla m(x) - \nabla m(z)))^l \right| \;+\; \alpha \; \left| (\mathcal{G}(\nabla m(x) - \nabla m(z)))^h \right| \\ \le\;& L \, |x - z| \;+\; \alpha \, \lambda_{n+1} \, L \, |x - z| \le 2\,L\,\mathcal{R} + 2\,(1 + L)\,\mathcal{R} \;\le\; 3/4 \, |x - y|_\alpha \,. \end{aligned}$$

□

Corollary 12. *If the assumptions of Lemma 15 and 16 are satisfied, then Theorem 8 holds with* $4\,c \;\geq\; \exp\left(-32\,L^4\,\mathcal{R}^2\right)/(1+L)$.

We give a sufficient condition for the existence of a Lyapunov function.

Lemma 17. *Let Assumption 28 be true and set* $V(x) = 1+|x|^2$. *If there are constants* $b \in (0,\infty)$ *and* $0 \leq c < 1$ *such that*

$$|\nabla m(x)| \;\leq\; b \;+\; c\,|x| \qquad \textit{holds for any } x \in \mathbb{H}, \tag{2.81}$$

then for any $0 < \eta < 1-c$ *there is* $C \in (0,\infty)$ *such that Assumption 25 is satisfied with* (V, C, η).

Proof. We have to find C such that (2.16) holds true for all $x \in \mathbb{H}$. Observe that,

$$\begin{aligned} \mathcal{L}V(x) \;&=\; 2\;\langle x, -x - a\,\mathcal{G}\,x - \mathcal{G}\nabla m(x)\rangle \;+\; \mathrm{trace}(\mathcal{G}) \\ &\leq\; 2\left(-\,|x|^2 + b\,|x| + c\,|x|^2\right) \;+\; \mathrm{trace}(\mathcal{G}). \end{aligned}$$

The claim follows since $c < 1$. □

We see that Theorem 9 is applicable if the assumptions of Lemma 17 are satisfied.

Transition path sampling

We present a concrete sampling context for which the results from the last subsection are applicable. We follow here [71] and consider the $\mathbb{R}^d$-valued SDE

$$dX_t \;=\; -\nabla_{\mathbb{R}^d} W(X_t)\,dt \;+\; dB_t, \qquad X_0 = 0, \tag{2.82}$$

where (B_t) is a d-dimensional Brownian motion.

Assumption 29. *The potential* $W : \mathbb{R}^d \to \mathbb{R}$ *is given by*

$$W(x) \;=\; \frac{a}{2}\,|x|^2 \;+\; H(x),$$

with $a > 0$ *and* $H : \mathbb{R}^d \to \mathbb{R}$ *is a* C^4 *function for which all* k*-fold partial derivatives,* $k \in \{1,2,3,4\}$, *satisfy*

$$\left|\partial^k H(x)\right|\,|x|^{k-2} \to 0 \textit{ for } |x| \to \infty.$$

Suppose that we are interested in the law π of $(X_t)_{t\in[0,1]}$ conditioned on the event $X_1 = 0$. We describe a typical setup for the above situation. Set $\mathbb{H} = L^2([0,1],\mathbb{R}^d)$ and let (Δ_0, D_0) be the self-adjoint Laplacian with Dirichlet boundary condition, i.e. the domain D_0 is given by all differentiable functions f, such that f' is absolutely continuous with $f'' \in L^2$ and such that $f(0) = f(1) = 0$. Let $\mathcal{G} = -\Delta_0^{-1}$. Observe that $\mathbf{e}_k = \sqrt{2}\,\sin(\pi k t)$, $k \in \mathbb{N}$, is an orthonormal basis of $\mathbb{H}$ satisfying $\mathcal{G}\mathbf{e}_k = \lambda_k \mathbf{e}_k$ with $\lambda_k = (\pi k)^{-2}$. In particular, the operator $\mathcal{G}$ is trace-class, symmetric and positive definite on $\mathbb{H}$. It is well-known, that the distribution of a standard Brownian Bridge on $\mathbb{H}$ is a centered normal distribution with covariance operator $\mathcal{G}$. Under Assumption 29 one can argue, using Girsanov's theorem and Itô's formula, that the law π of $(X_t)_{t\in[0,1]}$ conditioned on $X_1 = 0$ is given by (2.2), with $U : \mathbb{H} \to \mathbb{R}$ defined by

$$U(x) \;=\; \frac{1}{2}\int_0^1 \Phi(x_s)\, ds \quad \text{and} \quad \Phi(x) \;=\; |\nabla_{\mathbb{R}^d} W(x)|^2 + \Delta_{\mathbb{R}^d} W(x). \tag{2.83}$$

We refer the reader to [71] for a detailed exposition. It is now a straightforward calculation to check that Lemma 17 is applicable, if Assumption 29 is satisfied.

2.3.2 Finite-dimensional approximations

In this work we focus on explicit contraction rates for the process (2.1). In the light of sampling applications one might ask, if it is possible to make related statements about finite-dimensional approximations. We shortly argue, that this is indeed the case.

Suppose that Assumptions 22, 23, and 24 are true. Let $\mathbb{H}^l$ be of dimension $n \in \mathbb{N}_+$. Fix some $d > n$ and write $\mathbb{H}^d = \langle \mathbf{e}_1, \ldots, \mathbf{e}_d \rangle$ for the subspace spanned by the first d basis vectors. Given $x \in \mathbb{H}$, we write x^d for the orthogonal projection onto $\mathbb{H}^d$. Let (X_t) be a solution of (2.1) with $X_0 = x_0$. A straightforward d-dimensional approximation $(\tilde{X}_t)$ is given by the solution of the SDE

$$d\tilde{X}_t \;=\; -\tilde{X}_t\, dt \;+\; b^d(\tilde{X}_t)\, dt \;+\; \sqrt{2}\, dW_t^d, \qquad \tilde{X}_0 = x_0^d. \tag{2.84}$$

A similar approximation is e.g. considered in [38]. Observe that the nonlinearity $x \mapsto b^d(x)$ satisfies Assumptions 22 and 24 on the space $\mathbb{H}^d$ with the *same constants* as $x \mapsto b(x)$ on $\mathbb{H}$. In particular, we can apply Theorem 8 to equation (2.84) and see that the corresponding Markov kernels satisfy a Kantorovich contraction with a *dimension-independent* and explicit contraction rate. A related statement holds true for Theorem 9. We remark that the unique invariant measure π^d for (2.84) does in general not agree with the invariant measure π of (2.1). A study of the approximation error can be found in [38].

It might also be possible to use the presented results to make statements about the speed of convergence of time-discrete approximations of (2.84), e.g. Euler approximations. There are at least two different approaches to this question: The first possibility is to implement a similar coupling strategy directly for Markov chains. We refer in this context to the forthcoming work [47]. The second possibility is to interpret the approximation as a *perturbation* of the original equation, see [138, 121, 128, 137, 36, 45] and the references therein. Nevertheless, the last question goes beyond the scope of this work.

Acknowledgement: I want to thank my supervisor A. Eberle for suggesting the topic, for his support and helpful advice. I am grateful for several fruitful discussions with A. Guillin on related topics. Moreover, I want to thank M. Hairer for his hospitality during a stay at the University of Warwick and for pointing out the connection of my work to the 2D Navier-Stokes equation. I also want to thank an anonymous referee who encouraged me to extend the discussion of consequences after Theorem 9 and to add Remark 6.

3 Sticky couplings of multidimensional diffusions with different drifts

We present a novel approach of coupling two multidimensional and nondegenerate Itô processes (X_t) and (Y_t) which follow dynamics with different drifts. Our coupling is *sticky* in the sense that there is a stochastic process (r_t), which solves a one-dimensional stochastic differential equation with a *sticky boundary* behavior at zero, such that almost surely $|X_t - Y_t| \leq r_t$ for all $t \geq 0$. The coupling is constructed as a weak limit of Markovian couplings. We provide explicit, non-asymptotic and longtime stable bounds for the probability of the event $\{X_t = Y_t\}$.

A. Eberle and R. Zimmer. Sticky couplings of multidimensional diffusions with different drifts. *ArXiv e-print 1612.06125*, December 2016

Financial support from the German Science foundation through the *Hausdorff Center for Mathematics* is gratefully acknowledged.

3.1 Introduction

Let (B_t) and $(\tilde{B}_t)$ be d-dimensional Brownian motions. We consider two diffusion processes with values in $\mathbb{R}^d$ which follow dynamics with different drifts, i.e.

$$dX_t = b(t, X_t)\, dt + dB_t, \qquad X_0 = x, \tag{3.1}$$

$$dY_t = \tilde{b}(t, Y_t)\, dt + d\tilde{B}_t, \qquad Y_0 = y. \tag{3.2}$$

We assume that the drift coefficients $b, \tilde{b} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ are locally Lipschitz. Moreover, we impose assumptions which imply that a geometric Lyapunov drift condition holds for (3.1) and that there is a constant $M > 0$ such that uniformly $|b - \tilde{b}| \leq M$.

Diffusions with different drifts occur in many application areas. For example, one could consider a Langevin diffusion (X_t) and a perturbation or approximation (Y_t) of the latter. Other natural examples are McKean-Vlasov processes, where the drift coefficients depend not only on the current position of the process but also on the corresponding law. A natural question arising is how to obtain explicit bounds for the distance of X_t and Y_t in Kantorovich distances, e.g. in total variation norm. There are a few articles which try to answer this question in a general setting: Using Girsanov's theorem and coupling on the path space, the works [92, 104, 105] establish bounds on the total variation norm of such diffusions. In [12] bounds for the distance between transition probabilities of diffusions with different drifts are derived using analytic arguments, see also the related work [112]. The drawback of these approaches is that the derived bounds are typically only useful for small time horizons and are not longtime stable. The article [9] provides bounds for the distance between stationary measures of diffusions with different drifts. Coupling methods are used in [46] to provide longtime stable bounds on the distance between a Langevin diffusion and its Euler approximation. Howitt constructs in [84] a *sticky coupling* of two one-dimensional Brownian motions with different drifts using time-change arguments which are restricted to the one-dimensional setting.

In this article, we discuss a novel approach of constructing couplings (X_t, Y_t) of solutions to (3.1) and (3.2) in a multidimensional setting. Consider for example the case where $\tilde{b}$ differs from b by a non-zero constant m, i.e., $\tilde{b}(t,x) = b(t,x) + m$ for some $m \in \mathbb{R}^d$, and let (X_t) and (Y_t) be solutions of (3.1) and (3.2) respectively. In this case, whenever X_t and Y_t meet, the drift forces the processes to immediately move apart from each other. It is clear that, regardless of how the processes are coupled, one cannot hope for the existence of an almost surely finite stopping time T such that $P[X_t = Y_t \, \forall t \geq T] = 1$. Nevertheless, we construct a coupling such that for any given $t > 0$, we have $P[X_t = Y_t] > 0$ and the coupling is *sticky* in the sense that there is a continuous semimartingale (r_t) which solves a one-dimensional stochastic differential equation with a *sticky boundary* behavior at zero such that almost surely $|X_t - Y_t| \leq r_t$ for all $t \geq 0$. This allows us to establish explicit, non-asymptotic and longtime stable bounds for the probability of the event $\{X_t = Y_t\}$. The coupling is constructed as a weak limit of Markovian couplings. The idea for the coupling is based on [51, 48] where coupling approaches for particle systems and nonlinear McKean-Vlasov processes are discussed, cf. Section 3.2.2 for a comprehensive comparison. We show that sticky couplings can be applied effectively to provide total variation bounds between the laws of both linear and nonlinear diffusions with varying drifts.

Outline: The main results are presented in Section 3.2. In Section 3.3 we recall results on the existence and uniqueness of one-dimensional SDEs with sticky boundary, we establish an approximation result for the latter, and we study the longtime behavior of solutions to such equations using coupling methods. Based on these results, the proof of our main theorem and the construction of the *sticky coupling* are presented in Section 3.4.

3.2 Main results

3.2.1 Sticky couplings

We impose the following assumptions:

Assumption 30. *There is a constant $M \in [0, \infty)$ such that*

$$\left| b(t,x) - \tilde{b}(t,x) \right| \leq M \qquad \text{for any } x \in \mathbb{R}^d \text{ and } t \geq 0.$$

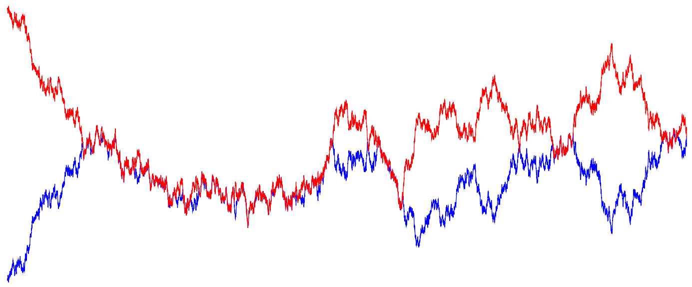

Figure 3.1: Sticky coupling of one-dimensional diffusions with different drifts

Assumption 31. *There is a Lipschitz function* $\kappa : [0,\infty) \to \mathbb{R}$ *such that*

$$\langle x-y, b(t,x)-b(t,y)\rangle \le \kappa(|x-y|)\cdot |x-y|^2 \quad \textit{for any } x,y\in\mathbb{R}^d \textit{ and } t\ge 0.$$

Outside of a bounded interval, the function κ *is constant and strictly negative.*

The assumptions imply in particular that the unique strong solutions (X_t) and (Y_t) of (3.1) and (3.2) respectively are non-explosive. We present our main result:

Theorem 10 (Sticky coupling)**.** *Suppose that Assumptions 30 and 31 hold true. Then for any initial values* $x, y \in \mathbb{R}^d$*, there is a coupling* (X_t, Y_t) *of solutions to* (3.1) *and* (3.2)*, respectively, such that* $X_t - Y_t$ *is* sticky at zero *in the sense that the difference is controlled by a solution of a one-dimensional SDE with a sticky boundary behavior at zero. More precisely, there is a real-valued process* (r_t) *solving the SDE*

$$dr_t \;=\; (M \;+\; \kappa(r_t)\, r_t)\; dt \;+\; 2\, I(r_t>0)\; dW_t, \qquad r_0 \;=\; |x-y|\,, \tag{3.3}$$

driven by a one-dimensional Brownian motion (W_t)*, such that almost surely,*

$$|X_t - Y_t| \;\le\; r_t \qquad \textit{for any } t\ge 0. \tag{3.4}$$

The process (r_t) *is* sticky *at zero in the sense that almost surely,*

$$2M \int_0^t I(r_s = 0)\; ds \;=\; \ell_t^0(r), \qquad 0 \;\le\; t \;<\; \infty, \tag{3.5}$$

where $\ell_t^0(r)$ *is the right local time at* 0 *of* (r_t)*, i.e.,*

$$\ell_t^0(r) \;=\; \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}\int_0^t I(0\le r_s<\epsilon)\, d[r]_s \;=\; 4 \lim_{\epsilon\downarrow 0}\frac{1}{\epsilon}\int_0^t I(0<r_s<\epsilon)\, ds.$$

Equation (3.3) *admits an invariant probability measure* π*. For* $M=0$*,* $\pi=\delta_0$*, and for* $M>0$*,* π *is determined by*

$$\pi(dx) \;\propto\; \left(\frac{2}{M}\,\delta_0(dx) \;+\; \exp\left(\frac{1}{2}\int_0^x (M+\kappa(y)\, y)\, dy\right)\lambda_{(0,\infty)}(dx)\right). \tag{3.6}$$

If the initial conditions coincide, i.e., if $x = y$, then for any $t \geq 0$,

$$P[X_t = Y_t] \;\geq\; \pi[\{0\}] \;=\; \left(1 + \frac{M}{2}\int_0^\infty \exp\left(\frac{1}{2}\int_0^x (M + \kappa(y)\, y)\, dy\right) dx\right)^{-1} \tag{3.7}$$

In general, there are constants $c, \epsilon \in (0, \infty)$, depending only on M and κ, such that for any $t > 0$ and any initial values $x, y \in \mathbb{R}^d$,

$$P[X_t \neq Y_t] \;\leq\; \frac{1}{\epsilon}\,\frac{c}{e^{c\,t} - 1}\;|x - y| \;+\; \pi[(0, \infty)]. \tag{3.8}$$

The constants c and ϵ are given by

$$c = \left(2\int_0^{R_1} \frac{\Phi(s)}{\phi(s)}\, ds\right)^{-1} \quad \text{and} \quad \epsilon = \min\left\{\left(2\int_0^{R_1} \frac{1}{\phi(s)}\, ds\right)^{-1},\; c\,\Phi(R_1)\right\},$$

where $\phi(r) = \exp\left(-\frac{1}{2}\int_0^r (M \;+\; \kappa(s)\, s)^+\, ds\right)$, $\Phi(r) = \int_0^r \phi(s)\, ds$,

$$R_0 \;=\; \inf\{R \geq 0 : (M \;+\; \kappa(r)\, r) \leq 0 \quad \text{for any } r \geq R\}, \quad \text{and} \tag{3.9}$$

$$R_1 \;=\; \inf\{R \geq R_0 : R(R - R_0)\; (M/r \;+\; \kappa(r)) \leq -4 \quad \text{for any } r \geq R\} \tag{3.10}$$

In Section 3.3 we also provide explicit bounds on the expected values $E[|X_t - Y_t|]$, cf. Theorem 14 further below.

The coupling (X_t, Y_t) in Theorem 10 is constructed as a weak limit of Markovian couplings. The construction of the coupling and the proof of the theorem are given in Section 3.4.

Remark 7 (Reflection coupling)**.** *The classical reflection coupling of Lindvall and Rogers [107] occurs as a special case of the coupling in Theorem 10 when the drift coefficients coincide, i.e., $b = \tilde{b}$. In this case we can choose $M = 0$ so that 0 is an absorbing boundary for the diffusion process (r_t). The equation* (3.11) *reduces to*

$$P[X_t \neq Y_t] \;\leq\; \frac{1}{\epsilon}\,\frac{c}{e^{c\,t} - 1}\;|x - y|\,, \tag{3.11}$$

which is a well-known bound for reflection coupling [107, 27].

In the two special cases $M = 0$ and $x = y$, the bound in (3.11) takes a very simple and intuitive form. In general, however, the rate c depends on M. This dependence can be avoided by considering a modified coupling.

Theorem 11. *There is a coupling $(\tilde{X}_t, \tilde{Y}_t)$ of solutions to* (3.1) *and* (3.2) *such that*

$$P[\tilde{X}_t \neq \tilde{Y}_t] \;\leq\; \frac{1}{\tilde{\epsilon}}\,\frac{\tilde{c}}{e^{\tilde{c}t} - 1}\;|x - y| \;+\; \pi[(0, \infty)] \quad \text{for any } t \geq 0, \tag{3.12}$$

where $\tilde{c}, \tilde{\epsilon}$ are defined analogously to c and ϵ but with $M = 0$.

Proof of Theorem 11. Consider a process (Z_t) satisfying

$$dZ_t \;=\; b(t,Z_t)\,dt \,+\, dB_t, \qquad Z_0 = y.$$

Let $(\tilde{X}_t, \tilde{Z}_t)$ be a standard reflection coupling of (X_t) and (Z_t), i.e., a sticky coupling in the case where the drifts coincide. Then we can glue this coupling with a sticky coupling of (Z_t) and (Y_t), i.e., there are processes $(\tilde{X}_t, \tilde{Z}_t, \tilde{Y}_t)$ defined on a joint probability space such that $(\tilde{X}_t, \tilde{Z}_t)$ is a sticky coupling of (X_t, Z_t), and $(\tilde{Z}_t, \tilde{Y}_t)$ is a sticky coupling of (Z_t, Y_t), see e.g. the "glueing lemma" in [149]. For $t \geq 0$, we obtain by Theorem 10:

$$P[\tilde{X}_t \neq \tilde{Y}_t] \;\leq\; P[\tilde{X}_t \neq \tilde{Z}_t] \,+\, P[\tilde{Z}_t \neq \tilde{Y}_t] \;\leq\; \frac{1}{\tilde{\epsilon}}\, \frac{\tilde{c}}{e^{\tilde{c}t}-1}\; |x-y| \,+\, \pi[(0,\infty)].$$

□

To make the bounds in the theorems more explicit, we now assume that we are given constants $\mathcal{R}, L \in [0,\infty)$ and $K \in (0,\infty)$ such that for any $t \geq 0$,

$$\langle x-y, b(t,x)-b(t,y)\rangle \;\leq\; \begin{cases} L\,|x-y|^2 & \text{for any } x,y \in \mathbb{R}^d, \\ -K\,|x-y|^2 & \text{for } x,y \in \mathbb{R}^d \text{ s.t. } |x-y| \geq \mathcal{R}. \end{cases} \tag{3.13}$$

Hence Assumption 31 is satisfied with $\kappa(r) = L\,I(r<\mathcal{R}) - K\,I(r \geq \mathcal{R})$. In this case, the exponential decay rate $\tilde{c}$ in Theorem 11 is bounded from below by

$$\tilde{c}^{-1} \;\leq\; \begin{cases} 4\,\max(\mathcal{R}^2,\, K^{-1}) & \text{if } L = 0, \\ 3e\,\max(\mathcal{R}^2,\, 4K^{-1}) & \text{if } L\mathcal{R}^2 \leq 4, \\ 8\sqrt{\pi}L^{-1/2}(L^{-1}+K^{-1})\mathcal{R}^{-1}\exp\left(L\mathcal{R}^2/4\right) + 16K^{-2}\mathcal{R}^{-2} & \text{if } L\mathcal{R}^2 > 4, \end{cases}$$

see Lemma 1 in [51] (Note that the definitions of the function κ and the constant c in [51] differ from the definitions above by a factor -2, 2, respectively). The following lemma provides explicit upper bounds on the longtime asymptotics of the probabilities in (3.11) and (3.12). The proof is included in Section 3.4.

Lemma 18. *Suppose that Condition* (3.13) *is satisfied. Then* $\pi[(0,\infty)] = \alpha/(1+\alpha)$ *where* α *is a nonnegative constant such that for* $M \leq K\mathcal{R}$*,*

$$\alpha \;\leq\; \left(\pi^{1/2}e^{1/2}K^{-1/2} + 2\mathcal{R}\max(4, L\mathcal{R}^2+2M\mathcal{R})^{-1}\right)\; M\; \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right),$$

and for $M \geq K\mathcal{R}$*,*

$$\alpha \leq \left(\sqrt{\frac{\pi}{K}} + \frac{2\mathcal{R}}{\max(4, 2M\mathcal{R}+L\mathcal{R}^2)}\right) M \exp\left(\frac{M^2}{4K} + \frac{L+K}{4}\mathcal{R}^2\right).$$

The theorems imply bounds on the total variation distance between the laws of X_t and Y_t for any time $t \geq 0$. We now verify that in two simple examples, the bound in (3.12) is of the correct order:

Example 9 (Ornstein-Uhlenbeck processes)**.** *Fix $m \in \mathbb{R}^d \setminus \{0\}$. We consider Ornstein-Uhlenbeck processes on $\mathbb{R}^d$, given by*

$$dX_t = -X_t/2 \, dt + dB_t, \qquad X_0 = x, \tag{3.14}$$

$$dY_t = -(Y_t - m)/2 \, dt + d\tilde{B}_t, \qquad Y_0 = y, \tag{3.15}$$

where (B_t) and $(\tilde{B}_t)$ are d-dimensional Brownian motions. Let $d(t)$ denote the total variation distance between the laws of X_t and Y_t at time t. It is well-known that X_t and Y_t are normally distributed with

$$\begin{aligned} \mathrm{Law}(X_t) &= \mathcal{N}\left(e^{-t/2}\, x,\ (1-e^{-t})\, \mathrm{I}_d\right), \\ \mathrm{Law}(Y_t) &= \mathcal{N}\left(e^{-t/2}\, y \ +\ (1-e^{-t/2})\, m,\ (1-e^{-t})\, \mathrm{I}_d\right). \end{aligned}$$

The total variation distance between d-dimensional normal distributions $\mathcal{N}(a, bI_d)$ and $\mathcal{N}(\tilde{a}, bI_d)$ with $a, \tilde{a} \in \mathbb{R}^d$ and $b \in (0,\infty)$ is given by $\Phi_1(|a-\tilde{a}|/(2\sqrt{b}))$ where

$$\Phi_1(r) \ := \ \sqrt{2/\pi} \int_0^r \exp(-x^2/2)\, dx,$$

cf. e.g. [37, Exercise 15.12]. Hence for any $t > 0$,

$$d(t) \ = \ ||\, \mathrm{Law}(X_t) - \mathrm{Law}(Y_t)||_{TV} \ = \ \Phi_1\left(\frac{\left|m + e^{-t/2}(y-m-x)\right|}{2\sqrt{1-e^{-t}}}\right). \tag{3.16}$$

We now compare the upper bound (3.12) *for the total variation distance that has been derived by sticky couplings to the exact expression* (3.16)*. Observe that Assumptions 30 and 31 are satisfied with $M = |m|/2$ and the constant function $\kappa(r) = -1/2$ respectively. By a straightforward computation we obtain*

$$\pi[(0,\infty)] \ = \ 1 - \left(1 + \sqrt{\pi/8}\, |m|\, e^{m^2/8}\, (1 + \Phi_1(|m|/2))\right)^{-1}. \tag{3.17}$$

Asymptotically as $t \to \infty$, the upper bound for $P[\tilde{X}_t \neq \tilde{Y}_t]$ in (3.12) *approaches* (3.17)*, whereas the total variation distance $d(t)$ converges to $\Phi_1(|m|/2)$. Comparing both expressions for small and large values of $|m|$, we see that as $|m| \to 0$,*

$$\pi[(0,\infty)] \sim \sqrt{\pi/8}\, |m|\,, \quad \textit{whereas} \quad \Phi_1(|m|/2) \sim |m|/\sqrt{2\pi},$$

and as $|m| \to \infty$,

$$1 - \pi[(0,\infty)] \ \sim \ \frac{2}{\sqrt{2\pi}\,|m|} e^{-|m|^2/8}, \quad \textit{whereas} \quad 1 - \Phi_1(|m|/2) \ \sim \ \frac{4}{\sqrt{2\pi}\,|m|} e^{-|m|^2/2}.$$

Hence as $m \downarrow 0$, the bounds for the long time limit of the total variation distance provided by sticky couplings are of the correct order up to a multiplicative constant, whereas for $m \to \infty$, we loose a factor 4 in the exponential.

Furthermore, we can compare the decay rate $\tilde{c}$ in (3.12) *with the rate of convergence of $d(t)$ to its limit $\Phi_1(|m|/2)$. Asymptotically as $t \uparrow \infty$,* (3.16) *implies*

$$\begin{aligned} |d(t) - \Phi_1(|m|/2)| &\sim \Phi_1'(|m|/2)e^{-t/2}\,|y-m-x|/2 \\ &= (2\pi)^{-1/2}e^{-m^2/8}e^{-t/2}\,|y-m-x|\,. \end{aligned} \tag{3.18}$$

On the other hand, in this case $\tilde{c} = 1/8$ and $\tilde{\epsilon} = 1/(2\sqrt{8})$, so by (3.12),

$$P[\tilde{X}_t \neq \tilde{Y}_t] - \pi[(0,\infty)] \;\leq\; 2^{-1/2}(e^{t/8}-1)^{-1}\,|x-y|\,. \tag{3.19}$$

We see that the exponential rate of decay in our bound differs from the optimal rate only by a factor 4.

Example 10 (Confined Brownian motion)**.** *Fix $R, k, m \in (0,\infty)$, and let*

$$b(x) = 0 \quad \text{for } |x| \leq R, \qquad \text{and} \qquad b(x) = -k(x - R\,sgn(x))/2 \quad \text{otherwise.}$$

Moreover, let $\tilde{b}(x) = b(x) + m/2$. In this case, Condition (3.13) *is satisfied with $L = 0$, $K = k/6$ and $\mathcal{R} = 3R$, and Assumption 30 holds with $M = m/2$. Assuming $m \leq kR$ and $mR \leq 4/3$, Theorem 11 and the first bound in Lemma 18 show that there is a coupling $(\tilde{X}_t, \tilde{Y}_t)$ of the corresponding solutions to* (3.1) *and* (3.2) *with arbitrary initial values x and y such that*

$$\limsup_{t\to\infty} P[\tilde{X}_t \neq \tilde{Y}_t] \;\leq\; \left(\frac{3e}{4}R + \left(3\pi e^3/2\right)^{1/2} k^{-1/2}\right) m. \tag{3.20}$$

On the other hand, the unique invariant probability measures for (3.1) *and* (3.2) *are given explicitly by $\nu(dx) = Z_f^{-1} f(x)\,dx$, $\mu(dx) = Z_g^{-1} g(x)\,dx$, respectively, where $f(x) = \exp(-k\max(|x|-R,0)^2/2)$, $g(x) = \exp(mx)f(x)$, $Z_f = \int_{-\infty}^{\infty} f(x)\,dx$ and $Z_g = \int_{-\infty}^{\infty} g(x)\,dx$. Noting that $Z_g \geq Z_f$, an explicit computation yields the lower bounds*

$$\|\mu - \nu\|_{TV} \;\geq\; (\exp(-mR) - 1 + mR)/(mR),$$

and, for $Rk^{1/2} \leq 1$,

$$\|\mu - \nu\|_{TV} \;\geq\; \left(1 - \exp(-mR + m^2/(2k)) + 2^{1/2}(\pi k)^{-1/2} m \exp(-mR)\right)/4,$$

see page 134 further below. In particular,

$$\liminf_{m\downarrow 0} \|\mu - \nu\|_{TV}/m \;\geq\; \frac{1}{4}\left(R + (2/\pi)^{1/2} k^{-1/2}\right).$$

Hence for small m, the bound in (3.20) *is sharp up to a constant factor.*

Remark 8 (Comparison with Girsanov couplings)**.** *An alternative approach to construct couplings of solutions to* (3.1) *and* (3.2) *is by Girsanov's Theorem. If the initial conditions X_0 and Y_0 coincide and $T \in [0,\infty)$ is a fixed constant, then Girsanov's Theorem can be applied to construct a coupling (X_s, Y_s) such that with positive probability, $X_s = Y_s$ for all $s \in [0,T]$. Moreover, explicit bounds on this probability can be derived via Hellinger integrals [92, 104, 105]. Notice, however, that the corresponding bounds typically degenerate rapidly as $T \to \infty$. Hence Girsanov's Theorem provides a very strong coupling over short time intervals, whereas the sticky couplings introduced above are stable for long times in the sense that* $\liminf_{t\to\infty} P[X_t = Y_t] \geq \pi[\{0\}] > 0$.

3.2.2 McKean-Vlasov processes

We consider nonlinear diffusions on $\mathbb{R}^d$ of type

$$\begin{aligned} dX_t &= \eta(X_t)\,dt \;+\; \tau \int \vartheta(X_t, y)\,\mu_t^x(dy)\,dt \;+\; dB_t, \quad X_0 = x, \qquad (3.21)\\ \mu_t^x &= \mathrm{Law}(X_t), \end{aligned}$$

where (B_t) is a d-dimensional Brownian motion and $\tau \in \mathbb{R}$. The SDE is nonlinear in the sense of McKean, i.e., the future development after time t depends on the current state X_t and on the law of X_t, cf. e.g. [144, 117]. Let $\eta : \mathbb{R}^d \to \mathbb{R}^d$ and $\vartheta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be Lipschitz continuous functions. Then the equation above admits a unique strong solution, cf. [117, Theorem 2.2]. Let us fix initial values $x_0, y_0 \in \mathbb{R}^d$, $x_0 \neq y_0$, and consider solutions (X_t) and (Y_t) of (3.21) with $X_0 = x_0$ and $Y_0 = y_0$ respectively. We define drift coefficients

$$b^{x_0}(t,x) \;=\; \eta(x) \;+\; \tau \int \vartheta(x,y)\,\mu_t^{x_0}(dy), \qquad (3.22)$$

$$b^{y_0}(t,x) \;=\; \eta(x) \;+\; \tau \int \vartheta(x,y)\,\mu_t^{y_0}(dy), \qquad (3.23)$$

which are uniformly Lipschitz in x and continuous in t. Notice that due to pathwise uniqueness, (X_t) and (Y_t) are the unique strong solutions to the equations

$$dX_t \;=\; b^{x_0}(t, X_t)\,dt + dB_t, \qquad X_0 = x_0, \qquad (3.24)$$

$$dY_t \;=\; b^{y_0}(t, Y_t)\,dt + dB_t, \qquad Y_0 = y_0, \qquad (3.25)$$

and hence we can interpret the processes as two diffusions with different drifts.

Assumption 32. *There is a Lipschitz function $\kappa : [0,\infty) \to \mathbb{R}$ such that*

$$\langle x-y, \eta(x) - \eta(y)\rangle \le \kappa(|x-y|)\cdot|x-y|^2 \quad \text{for any } x,y \in \mathbb{R}^d \text{ and } t \ge 0.$$

Outside of a bounded interval, the function κ is constant and strictly negative.

Assuming that Assumption 32 holds, we have shown in [48] that there are constants $A, \lambda, \tau_0 \in (0,\infty)$ such that for $|\tau| \le \tau_0$,

$$\mathcal{W}^1(\mu_t^x, \mu_t^y) \;\le\; A\,e^{-\lambda t}\,|x-y| \quad \text{for any } t \ge 0 \text{ and } x,y \in \mathbb{R}^d, \qquad (3.26)$$

where $\mathcal{W}^1$ denotes the standard L^1 Wasserstein distance. The proof is based on an application of reflection coupling if $|X_t - Y_t| \ge \delta$ and synchronous coupling if $|X_t - Y_t| \le \delta/2$, where δ is a small positive constant. In the intermediate region, a combination of both couplings is applied. The bound in (3.26) is obtained when considering the limit of the resulting bounds as $\delta \downarrow 0$. The couplings considered in [48] now turn out to be approximations of a sticky coupling. By applying directly the sticky coupling and using Corollary 13 further below, we can extend the result in [48] and derive a corresponding exponential decay in total variation norm:

Theorem 12. *Let η and ϑ be Lipschitz and let Assumption 32 be true. There is $\tau_0 \in (0,\infty)$ such that for any $|\tau| \leq \tau_0$ and any $x, y \in \mathbb{R}^d$ there are constants $B, c \in (0,\infty)$ such that,*

$$\|\mu_t^x - \mu_t^y\|_{TV} \quad \leq \quad B\ e^{-ct} \quad \text{for any } t \geq 0. \tag{3.27}$$

The proof is given in Section 3.4.

3.2.3 Outlook[1]

The concept of sticky couplings sheds new light onto several results that have been previously derived using combinations of reflection and synchronous couplings. A first example of this type has been given in Theorem 12. Without carrying out details, we mention three further results that probably can be reinterpreted in terms of sticky couplings:

a) *Componentwise reflection couplings for interacting diffusions.* In [51], Wasserstein bounds for interacting diffusions with small interaction term (for example of mean-field-type) have been derived by coupling each component independently with a reflection coupling if the distance is greater than a given constant $\delta > 0$, and with a synchronous coupling otherwise. Instead, one could now directly consider a componentwise sticky coupling. As time evolves, more and more components in this coupling would get stuck at nearby positions until, after some finite coupling time, all components coincide. We expect that such a coupling could be used to derive total variation bounds similar to those in Theorem 12 for interacting particle systems.

b) *Couplings for infinite-dimensional diffusions.* In [159], Wasserstein contraction rates have been derived for a class of diffusions on a Hilbert space with possibly degenerate noise. Here a reflection coupling has been applied to the projection of the process on a finite dimensional subspace, whereas the remaining (orthogonal) components have been coupled synchronously. Again, because of the interaction between the components, the reflection coupling is switched off when the finite dimensional projections of the two copies are close to each other. Similarly as above, it should be possible to replace the coupling for the finite dimensional projection by a sticky coupling. The resulting infinite dimensional coupling process would then spend a certain amount of time at states where the finite dimensional projections of the two copies coincide. Under the assumptions made in [159], the orthogonal infinite dimensional components would approach each other for large t, and, consequently, the finite dimensional projections would coincide for an increasing proportion of time.

c) *Couplings for Langevin processes.* In a forthcoming paper, we consider couplings for (kinetic) Langevin diffusions $(X_t, V_t)_{t\geq 0}$ with state space $\mathbb{R}^{2d}$ that are given by stochastic differential equations of type

$$\begin{aligned} dX_t &= V_t\, dt, \\ dV_t &= -\gamma V_t\, dt \,-\, u\, \nabla U(X_t)\, dt \,+\, \sqrt{2\gamma u}\, dB_t. \end{aligned} \tag{3.28}$$

[1] This outlook is due to A. Eberle and is *not* a contribution of this thesis.

Here $(B_t)_{t\geq 0}$ is a d dimensional Brownian motion, u and γ are positive constants, and U is a C^1 function on $\mathbb{R}^d$. We apply a reflection coupling that is replaced by a synchronous coupling when the values of $X_t + \gamma^{-1}V_t$ are close to each other for both components. Again, at least informally, this coupling could be replaced by a coupling $((X_t, V_t), (X_t', V_t'))$ that is sticky when $X_t + \gamma^{-1}V_t = X_t' + \gamma^{-1}V_t'$. Under the assumptions that we impose on U, the coupling would be contractive on the corresponding $3d$ dimensional linear subspace of $\mathbb{R}^{4d}$, and as time evolves, it would spend a positive amount of time on this subspace.

We hope that the potential applications listed above show how sticky couplings provide a valuable concept for building intuition about ways to couple diffusion processes in an efficient way. Carrying out carefully the ideas described above would go far beyond the scope of this paper.

3.3 Diffusions on $\mathbb{R}_+$ with a sticky reflecting boundary

In this section we prove some basic results on diffusions on $\mathbb{R}_+$ with a sticky boundary at 0. In particular, we prove the existence of a synchronous coupling of two sticky diffusions and a corresponding comparison theorem, which is then applied to study the long time behavior of the processes. At first, we need to adapt some known facts on existence and uniqueness of weak solutions to our setup. We consider the stochastic differential equation

$$dr_t \;=\; \alpha(t, r_t)\, dt \;+\; 2\, I(r_t > 0)\, dW_t, \qquad \mathrm{Law}(r_0) \;=\; \mu, \tag{3.29}$$

on the positive real line $\mathbb{R}_+ = [0, \infty)$, where (W_t) is a one-dimensional Brownian motion and μ is a probability measure on $\mathbb{R}_+$. Below, we will impose conditions on the drift coefficient $\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ which imply existence and uniqueness of weak solutions. In particular, we will assume that $\alpha(t, 0) > 0$ for any $t \geq 0$. Let us briefly discuss the consequences of this assumption: Suppose that (r_t) is a solution of (3.29). An application of the Itô-Tanaka formula to $f(r_t)$ with the function $f(x) = \max(0, x)$ and a comparison with (3.29) shows that almost surely,

$$\int_0^t \alpha(s, 0)\, I(r_s = 0)\, ds \;=\; \frac{1}{2}\ell_t^0(r), \qquad 0 \;\leq\; t \;<\; \infty, \tag{3.30}$$

where $\ell_t^0(r) = \lim_{\epsilon\downarrow 0} \epsilon^{-1} \int_0^t I(0 \leq r_s \leq \epsilon)\, d\,[r]_s$ is the right local time of (r_t). Equation (3.30) shows that there is *reflection* at zero. Moreover, for almost all trajectories, the Lebesgue measure of the set $\{0 \leq s \leq t : r_s = 0\}$ increases whenever $\ell_t^0(r)$ increases. In this sense (r_t) is *sticky* at zero.

Stochastic differential equations with boundary conditions have a long history. The discovery of a sticky boundary behavior for one-dimensional diffusions seems to go back to Feller [56, 57]. A historical overview is given in [127]. We give references to the most relevant works for our application and some recent developments. Existence and uniqueness results for multidimensional diffusion processes with various boundary behaviors have been established by Ikeda and Watanabe in [85, 155, 156]. These are based on results by Skorokhod and McKean [139, 140, 116]. Martingale problems with boundary conditions have been investigated by Stroock and Varadhan [141], see also the related work [62]. Non-existence of a strong solution to the SDE for sticky Brownian motion has been established in [30]. In [154], Warren identifies the law of a sticky Brownian motion conditioned on the driving Wiener process, see also the related work [76]. A recent publication on existence and uniqueness, which is also a good introduction into the topic, is the work by Engelbert and Peskir [54] and the related work [7]. First steps towards sticky couplings in a one-dimensional setting have been made by Howitt in [84] based on time-changes. The recent articles [63, 64] use Dirichlet forms to investigate sticky diffusions and provide some ergodicity results. Rácz and Shkolnikov [131] construct a multidimensional sticky Brownian motion as a limit of exclusion processes, see also [1] and [78].

3.3.1 Existence, uniqueness and comparison of solutions

We use the concept of weak solutions. Let $(\Omega, \mathcal{A}, (\mathcal{F}_t), P)$ be a filtered probability space satisfying the usual conditions. An $(\mathcal{F}_t)$ adapted process (r_t, W_t) on $(\Omega, \mathcal{A}, P)$ is called a *weak solution* of (3.29) if $P \circ r_0^{-1} = \mu$, (W_t) is a one-dimensional $(\mathcal{F}_t)$-Brownian motion w.r.t. P, and (r_t) is continuous, nonnegative, and P-almost surely,

$$r_t - r_0 \;=\; \int_0^t \alpha(s, r_s)\, ds \;+\; \int_0^t 2\, I(r_s > 0)\, dW_s, \qquad 0 \;\leq\; t \;<\; \infty.$$

We will make the following assumptions on the drift coefficient:

Assumption 33. *For any $R > 0$, $\inf_{t \in [0,R]} \alpha(t, 0) \;>\; 0$.*

Assumption 34. *For any $R > 0$ there is $L_R \in (0, \infty)$ such that*

$$|\alpha(t, x) - \alpha(s, y)| \;\leq\; L_R \;(\; |t - s| \;+\; |x - y| \;) \quad \textit{for any } x, y, s, t \in [0, R].$$

Assumption 35. *There is $C \in (0, \infty)$ such that for any $x \in \mathbb{R}_+$,*

$$\sup_{t \in [0,\infty)} \alpha(t, x) \;\leq\; C \;(\; 1 \;+\; |x| \;)$$

The assumptions above imply existence and uniqueness in law of weak solutions to (3.29). This has been proven by Watanabe in [155, 156] assuming that the maps $(t, x) \mapsto \alpha(t, x)$ and $t \mapsto 1/\alpha(t, 0)$ are bounded and Lipschitz. Using localization techniques for martingale problems, following the work of Stroock and Varadhan [142], Watanabe's results can be transferred to our slightly more general setup:

Uniqueness in law

Let $\mathbb{W} = C(\mathbb{R}_+, \mathbb{R})$ be the space of continuous functions endowed with the topology of uniform convergence on compacts, and let $\mathcal{B}(\mathbb{W})$ denote the Borel σ-Algebra. Let $\mathcal{F}_t = \sigma(\boldsymbol{r}_s : 0 \leq s \leq t)$ be the natural filtration generated by the canonical process $\boldsymbol{r}_t(\omega) = \omega(t)$. Given a solution (r_t) of (3.29), defined on a probability space $(\Omega, \mathcal{A}, P)$, we write $\mathbb{P} = P \circ r^{-1}$ for the law of r on $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$. We say that solutions to (3.29) are *unique in law*, if any two solutions (r_t^1) and (r_t^2) with coinciding initial law have the same law on the space $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$.

In order to apply existing localization techniques for martingale problems, we interpret equation (3.29) as an equation on $\mathbb{R}$, instead of $\mathbb{R}_+$, setting $\alpha(t,x) = \alpha(t,0)$ for $x < 0$. This does not cause any problems since, under the assumptions imposed above, any solution (r_t) with initial law supported on $\mathbb{R}_+$ satisfies almost surely $r_t \geq 0$ for all $t \geq 0$, see e.g. the argument in [54, Proof of Theorem 5].

We follow [142, 94] and define a family of second order differential operators

$$(\mathcal{L}_t f)(x) \quad = \quad \alpha(t,x)\ f'(x)\ +\ (1/2)\ I(x>0)\ f''(x).$$

A probability measure $\mathbb{P}$ on $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$ is called a *solution to the martingale problem w.r.t.* $(\mathcal{L}_t)$ iff for any $f \in C_0^2(\mathbb{R})$,

$$M_t^f \quad = \quad f(\boldsymbol{r}_t) - f(\boldsymbol{r}_0) - \int_0^t (\mathcal{L}_u f)(\boldsymbol{r}_u)\ du$$

is a continuous $(\mathcal{F}_t)$-martingale under $\mathbb{P}$. The solution to the martingale problem is called *unique*, if any two solutions $\mathbb{P}^1$ and $\mathbb{P}^2$ coincide whenever $\mathbb{P}^1 \circ \boldsymbol{r}_0^{-1} = \mathbb{P}^2 \circ \boldsymbol{r}_0^{-1}$. The next two results are well-known:

Lemma 19. *[142, 94] The following statements are equivalent:*

(i) There is a weak solution of (3.29) *with initial distribution* μ.

(ii) There is a solution $\mathbb{P}$ *to the martingale problem w.r.t.* $(\mathcal{L}_t)$ *s.t.* $\mathbb{P} \circ \boldsymbol{r}_0^{-1} = \mu$.

Moreover, the uniqueness of solutions to the martingale problem w.r.t. $(\mathcal{L}_t)$ *and the uniqueness in law of weak solutions to* (3.29) *are equivalent.*

Lemma 20. *[155, 156] Assume that the maps* $(t,x) \mapsto \alpha(t,x)$ *and* $t \mapsto 1/\alpha(t,0)$ *are bounded and Lipschitz. Then for any initial law* μ *on* $\mathbb{R}_+$*, there is a weak solution to* (3.29) *which is unique in law.*

A detailed proof of Lemma 19 can be found in [94, Chapter 5, Section 4.B]. A proof of Lemma 20 is given in [87, Chapter IV, Section 7].

Lemma 21. *If Assumptions 33 and 34 are satisfied then the solution to the martingale problem w.r.t.* $(\mathcal{L}_t)$ *is unique for a given initial law, and thus uniqueness in law holds for solutions to* (3.29).

Proof. We set $\alpha_n(s,x) = \alpha(s \wedge n, x \wedge n)$ for $n \in \mathbb{N}$. By the assumptions, the maps $(t,x) \mapsto \alpha_n(t,x)$ and $t \mapsto 1/\alpha_n(t,0)$ are bounded and Lipschitz continuous. Hence uniqueness holds for the corresponding martingale problem for any initial law μ on $\mathbb{R}_+$ according to Lemma 20 and 19. The uniqueness for the martingale problem w.r.t. $(\mathcal{L}_t)$ for such initial laws can now be shown by a localization argument, cf. [142, Theorem 10.1.2]. $\square$

Approximation, existence and coupling of solutions

We now consider two equations of the form (3.29) with drift coefficients β and γ that both satisfy Assumptions 33, 34 and 35. We construct a synchronous coupling of solutions to these equations as a weak limit of solutions to approximating equations with locally Lipschitz continuous coefficients. We introduce the family of stochastic differential equations, indexed by $n \in \mathbb{N}$, given by

$$\begin{aligned} d\tilde{r}_t^n &= \beta(t,\tilde{r}_t^n)\,dt \;+\; 2\,\vartheta^n(\tilde{r}_t^n)\,d\tilde{W}_t, \qquad \mathrm{Law}(\tilde{r}_0^n,\tilde{s}_0^n) = \tilde{\mu}^n \otimes \tilde{\nu}^n, \\ d\tilde{s}_t^n &= \gamma(t,\tilde{s}_t^n)\,dt \;+\; 2\,\vartheta^n(\tilde{s}_t^n)\,d\tilde{W}_t, \end{aligned} \tag{3.31}$$

Here $(\tilde{W}_t)$ is a Brownian motion, and we assume that:

Assumption 36. *$(\tilde{\mu}^n)$ and $(\tilde{\nu}^n)$ are sequences of probability measures on $\mathbb{R}_+$ converging weakly towards probability measures $\tilde{\mu}$ and $\tilde{\nu}$, respectively.*

Assumption 37. *For each $n \in \mathbb{N}$, the function $\vartheta^n : \mathbb{R}_+ \to [0,1]$ is Lipschitz continuous with $\vartheta^n(0) = 0$, $\vartheta^n(x) > 0$ for $x > 0$, and $\vartheta^n(x) = 1$ for $x \geq 1/n$.*

Remark 9. *In [54], a sticky Brownian motion (r_t) satisfying*

$$dr_t \;=\; I(r_t \neq 0)\,d\tilde{W}_t, \qquad I(r_t = 0)\,\mu\,dt \;=\; d\ell_t^0(r), \qquad \mu \in (0,\infty),$$

is approximated by solutions of equations

$$dr_t^n \;=\; \Big(\sqrt{2\,\mu/n}\;I(|r_t^n| \leq 1/n) \;+\; I(|r_t^n| > 1/n)\Big)\,d\tilde{W}_t,$$

The approximation is tailored in such a way that it is compliant with the time-changes frequently used to show existence and uniqueness of weak solutions to sticky SDEs, see e.g. [54, 156]. Our approximation result follows a similar spirit but it does not rely on time changes.

Lemma 22. *Suppose that β and γ satisfy Assumptions 33, 34 and 35. Moreover, let Assumptions 36 and 37 be true. Then for each $n \in \mathbb{N}$, there is a strong solution $(\tilde{r}_t^n, \tilde{s}_t^n)$ of Equation (3.31) with values in $\mathbb{R}_+^2$. Moreover, uniqueness in law holds.*

Proof. Fix $n \in \mathbb{N}$. For $x < 0$ we set $\vartheta^n(x) = 0$, $\beta(t,x) = \beta(t,0)$, and $\gamma(t,x) = \gamma(t,0)$. Equation (3.31) is then a standard SDE on $\mathbb{R}^2$ with locally Lipschitz coefficients. Hence there is a strong and pathwise unique solution. Moreover, Assumption 35 implies that the solution is non-explosive. Similarly to [54, Proof of Theorem 5], we can apply the Itô-Tanaka formula to the negative part of $\tilde{r}_t^n$ in order to show that the process is nonnegative. Indeed,

$$(\tilde{r}_t^n)^- - (\tilde{r}_0^n)^- \quad = \quad -\int_0^t I(\tilde{r}_s^n \le 0)\, d\tilde{r}_s^n + \frac{1}{2}\ell_t^0(\tilde{r}^n),$$

where $\ell_t^0(\tilde{r}^n)$ is the right local time of $(\tilde{r}_t^n)$, i.e.,

$$\ell_t^0(\tilde{r}^n) = \lim_{\epsilon\downarrow 0} \epsilon^{-1} \int_0^t I(0 \le \tilde{r}_s^n \le \epsilon)\, d\, [\tilde{r}^n]_s = 4 \lim_{\epsilon\downarrow 0} \epsilon^{-1} \int_0^t I(0 \le \tilde{r}_s^n \le \epsilon)\, \vartheta^n(\tilde{r}_s^n)^2\, ds.$$

Since ϑ^n is Lipschitz with $\vartheta^n(0) = 0$, the local time vanishes. Therefore, and since $\beta(s, 0) > 0$ for any $s \ge 0$, we have $0 \le (\tilde{r}_t^n)^- \le (\tilde{r}_0^n)^- = 0$. A similar argument can be used for $(\tilde{s}_t^n)$. □

For each $n \in \mathbb{N}$, there are a probability space $(\Omega^n, \mathcal{A}^n, P^n)$ and random variables $\tilde{r}^n, \tilde{s}^n : \Omega^n \to \mathbb{W}$ such that $(\tilde{r}_t^n, \tilde{s}_t^n)$ is a solution of (3.31). Let $\mathbb{P}^n = P^n \circ (\tilde{r}^n, \tilde{s}^n)^{-1}$ denote the law on $\mathbb{W} \times \mathbb{W}$. For $w = (w_1, w_2) \in \mathbb{W} \times \mathbb{W}$, we define the coordinate mappings $\boldsymbol{r}(w) = w_1$ and $\boldsymbol{s}(w) = w_2$.

Theorem 13. *Suppose that β and γ satisfy Assumptions 33, 34 and 35, and let $\tilde{\mu}$ and $\tilde{\nu}$ be probability measures on $\mathbb{R}_+$. Suppose that the sequences (ϑ^n), $(\tilde{\mu}^n)$ and $(\tilde{\nu}^n)$ satisfy Assumptions 36 and 37. Then there is a random variable $(\tilde{r}, \tilde{s})$ with values in $\mathbb{W} \times \mathbb{W}$, defined on some probability space $(\Omega, \mathcal{A}, P)$, such that $(\tilde{r}_t, \tilde{s}_t)$ is a weak solution of*

$$\begin{aligned} d\tilde{r}_t &= \beta(t, \tilde{r}_t)\, dt + 2\, I(\tilde{r}_t > 0)\, d\tilde{W}_t, \qquad \mathrm{Law}(\tilde{r}_0, \tilde{s}_0) = \tilde{\mu} \otimes \tilde{\nu}, \\ d\tilde{s}_t &= \gamma(t, \tilde{s}_t)\, dt + 2\, I(\tilde{s}_t > 0)\, d\tilde{W}_t, \end{aligned} \tag{3.32}$$

for some Brownian motion $(\tilde{W}_t)$. Moreover, there is a subsequence (n_k) such that $P^{n_k} \circ (\tilde{r}^{n_k}, \tilde{s}^{n_k})^{-1}$ converges weakly towards $P \circ (\tilde{r}, \tilde{s})^{-1}$. If additionally,

$$\beta(t, x) \;\le\; \gamma(t, x) \qquad \text{for any } x, t \in \mathbb{R}_+, \text{ and} \tag{3.33}$$

$$P^n[\, \tilde{r}_0^n \le \tilde{s}_0^n \,] \;=\; 1 \qquad \text{for any } n \in \mathbb{N}, \tag{3.34}$$

then $P[\, \tilde{r}_t \le \tilde{s}_t \text{ for all } t \ge 0 \,] = 1$.

Proof. We fix sequences of diffusion coefficients (ϑ^n) and initial conditions $(\tilde{\mu}^n)$ and $(\tilde{\nu}^n)$ satisfying Assumptions 36 and 37.

Tightness: We claim that the sequence $(\mathbb{P}^n)_{n\in\mathbb{N}}$ of probability measures on $(\mathbb{W} \times \mathbb{W}, \mathcal{B}(\mathbb{W}) \otimes \mathcal{B}(\mathbb{W}))$ is tight. This can be shown by similar arguments as in [80, 81], so we only explain briefly how to adapt these arguments to our setting. At first, we observe that a uniform Lyapunov condition holds for the Markov processes $(\tilde{r}_t^n, \tilde{s}_t^n)$ defined by (3.31). Indeed, these processes solve a local martingale problem w.r.t. the generators

$$\mathcal{L}_t^n \;=\; \beta(t, \cdot)\, \partial_r \;+\; \gamma(t, \cdot)\, \partial_s \;+\; 2(\vartheta^n)^2\, (\partial_r^2 + \partial_s^2) \tag{3.35}$$

defined on smooth functions on $\mathbb{R}^2$. Let $V(x) := 1 + |x|^2$ for $x \in \mathbb{R}^2$. Recall that the drift coefficients in (3.35) do not depend on n and that they satisfy the linear growth Assumption 35. Moreover, the diffusion coefficients are uniformly bounded by one. It follows that there is a constant $\lambda \in (0, \infty)$, not depending on n, such that $\mathcal{L}_t^n V \leq \lambda V$ for any $n \in \mathbb{N}$. From this one can conclude that for each finite time interval $[0, T]$ and every $\epsilon > 0$, there is a compact set $K \subseteq \mathbb{R}^2$ such that for any $n \in \mathbb{N}$, $P[(\tilde{r}_t^n, \tilde{s}_t^n) \in K \text{ for } t \leq T] \geq 1 - \epsilon$. Moreover, the drift and diffusion coefficients are uniformly bounded on the set K. Combining these arguments, we can conclude tightness of the laws on $\mathbb{W} \times \mathbb{W}$. We refer to [80, 81] for a detailled proof in a similar setting. By Prokhorov's Theorem, we can conclude that there is a subsequence $n_k \to \infty$ and a probability measure $\mathbb{P}$ on $\mathbb{W} \times \mathbb{W}$ such that $\mathbb{P}^{n_k} \to \mathbb{P}$ weakly. To simplify notation we will write in the following n instead of n_k, keeping in mind that we have convergence only along a subsequence.

Identification of the limit: We now characterize the measure $\mathbb{P}$. In principle, we follow well-known strategies for identifying limits of semimartingales, cf. [142, 88, 55]. However, we can not apply those results directly, because the diffusion coefficients in (3.32) are discontinuous.

We know that $\mathbb{P} \circ (\boldsymbol{r}_0, \boldsymbol{s}_0)^{-1} = \mu \otimes \nu$, since $\mathbb{P}^n \circ (\boldsymbol{r}_0, \boldsymbol{s}_0)^{-1} = \mu^n \otimes \nu^n$ converges weakly to $\mu \otimes \nu$ by assumption. We define maps $\boldsymbol{M}, \boldsymbol{N} : \mathbb{W} \times \mathbb{W} \to \mathbb{W}$ by

$$\boldsymbol{M}_t = \boldsymbol{r}_t - \boldsymbol{r}_0 - \int_0^t \beta(u, \boldsymbol{r}_u)\, du \quad \text{and} \quad \boldsymbol{N}_t = \boldsymbol{s}_t - \boldsymbol{s}_0 - \int_0^t \gamma(u, \boldsymbol{s}_u)\, du.$$

We claim that $(\boldsymbol{M}_t, \mathcal{F}_t, \mathbb{P})$ and $(\boldsymbol{N}_t, \mathcal{F}_t, \mathbb{P})$ are martingales w.r.t. the canonical filtration $\mathcal{F}_t = \sigma((\boldsymbol{r}_u, \boldsymbol{s}_u)_{0 \leq u \leq t})$. Indeed, the mappings $\boldsymbol{M}$ and $\boldsymbol{N}$ are continuous on $\mathbb{W}$, so by the continuous mapping theorem, $\mathbb{P}^n \circ (\boldsymbol{r}, \boldsymbol{s}, \boldsymbol{M}, \boldsymbol{N})^{-1}$ converges weakly to $\mathbb{P} \circ (\boldsymbol{r}, \boldsymbol{s}, \boldsymbol{M}, \boldsymbol{N})^{-1}$. Notice that for each $n \in \mathbb{N}$, $(\boldsymbol{M}_t, \mathcal{F}_t, \mathbb{P}^n)$ is a martingale. Moreover, for any fixed $t \geq 0$, the family $(\boldsymbol{M}_t, \mathbb{P}^n)_{n \in \mathbb{N}}$ is uniformly integrable. Hence $(\boldsymbol{M}_t, \mathcal{F}_t, \mathbb{P})$ is a continuous martingale, cf. [88, Chapter IX, Proposition 1.12]. In particular, the quadratic variation $([\boldsymbol{M}]_t)$ exists $\mathbb{P}$-almost surely. Notice that, by (3.31), $[\boldsymbol{M}]_t \leq 4t$ $\mathbb{P}^n$-almost surely for every n. Thus for any $t \geq 0$,

$$\begin{aligned}
\mathbb{E}\left[\sup_{0 \leq s \leq t} |\boldsymbol{M}_s|^2\right] &\leq \liminf_{R \to \infty} \mathbb{E}\left[\sup_{0 \leq s \leq t} |\boldsymbol{M}_s|^2 \wedge R\right] = \liminf_{R \to \infty} \lim_{n \to \infty} \mathbb{E}^n\left[\sup_{0 \leq s \leq t} |\boldsymbol{M}_s|^2 \wedge R\right] \\
&\leq \liminf_{n \to \infty} \mathbb{E}^n\left[\sup_{0 \leq s \leq t} |\boldsymbol{M}_s|^2\right] \;\leq\; 4 \liminf_{n \to \infty} \mathbb{E}^n\left[\, [\boldsymbol{M}]_t \,\right] \;\leq\; 16\, t,
\end{aligned}$$

Hence, under $\mathbb{P}$, $(\boldsymbol{M}_t)$ is a square integrable martingale, and thus $(\boldsymbol{M}_t^2 - [\boldsymbol{M}]_t)$ is a martingale, cf. [97, Theorem 21.70]. Similar statements hold for $(\boldsymbol{N}_t)$.

As a next step, we compute the quadratic variations and covariations of $(\boldsymbol{M}_t)$ and $(\boldsymbol{N}_t)$ under $\mathbb{P}$. Here we follow arguments from [131]. Similarly as above, the family $(\boldsymbol{M}_t^2, \mathbb{P}^n)$ is uniformly integrable for any fixed $t \geq 0$, i.e.,

$$\lim_{\delta \to \infty} \sup_{n \in \mathbb{N}} \mathbb{E}^n[\, |\boldsymbol{M}_t|^2 \,;\, |\boldsymbol{M}_t|^2 > \delta\,] \;=\; 0. \tag{3.36}$$

Indeed, by Burkholder's inequality, there is a constant $C \in (0, \infty)$ such that

$$\mathbb{E}^n \left[\boldsymbol{M}_t^4 \right] \;\le\; C\, \mathbb{E}^n \left[[\boldsymbol{M}]_t^2 \right] \;\le\; 16\, C\, t^2 \qquad \text{for any } n \in \mathbb{N}.$$

Let $G : \mathbb{W} \to \mathbb{R}_+$ be bounded, continuous and nonnegative. Equation (3.36) implies

$$\lim_{\delta\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}^n \left[|G\boldsymbol{M}_t^2 - G(\boldsymbol{M}_t^2 \wedge \delta)| \right] \;\le\; |G|_\infty \liminf_{\delta\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}^n \left[\boldsymbol{M}_t^2; \boldsymbol{M}_t^2 > \delta \right] \;=\; 0.$$

Hence for any such G and any $t \ge 0$,

$$\begin{aligned} \mathbb{E}\left[G\,\boldsymbol{M}_t^2\right] \;&=\; \lim_{\delta\to\infty} \mathbb{E}\left[G\,(\boldsymbol{M}_t^2 \wedge \delta)\right] \;=\; \lim_{\delta\to\infty} \lim_{n\to\infty} \mathbb{E}^n \left[G\,(\boldsymbol{M}_t^2 \wedge \delta)\right] \\ &=\; \lim_{n\to\infty} \lim_{\delta\to\infty} \mathbb{E}^n \left[G\,(\boldsymbol{M}_t^2 \wedge \delta)\right] = \lim_{n\to\infty} \mathbb{E}^n \left[G\,\boldsymbol{M}_t^2\right]. \end{aligned} \tag{3.37}$$

We now show that $(\boldsymbol{M}_t^2 - 4\int_0^t I(\boldsymbol{r}_u > 0)\, du, \mathbb{P})$ is a submartingale. Fix $0 \le s < t$. Then for any continuous, bounded and $\mathcal{F}_s$-measurable function $G : \mathbb{W} \to \mathbb{R}_+$,

$$\lim_{n\to\infty} \mathbb{E}^n \left[G \int_s^t 4\,\vartheta^n(\boldsymbol{r}_u)^2\, du \right] = \lim_{n\to\infty} \mathbb{E}^n \left[G\left(\boldsymbol{M}_t^2 - \boldsymbol{M}_s^2\right)\right] = \mathbb{E}\left[G\left(\boldsymbol{M}_t^2 - \boldsymbol{M}_s^2\right)\right]. \tag{3.38}$$

On the other hand, the map $w \mapsto \int_0^\cdot I(w_s > \epsilon)\, ds$ from $\mathbb{W}$ to $\mathbb{W}$ is lower semicontinuous for any $\epsilon \ge 0$. Fatou's lemma and the Portemanteau theorem imply

$$\begin{aligned} \mathbb{E}\left[G \int_s^t I(\boldsymbol{r}_u > 0)\, du \right] \;&\le\; \liminf_{\epsilon\downarrow 0} \mathbb{E}\left[G \int_s^t I(\boldsymbol{r}_u > \epsilon)\, du \right] \\ &\le\; \liminf_{\epsilon\downarrow 0}\; \liminf_{n\to\infty} \mathbb{E}^n \left[G \int_s^t I(\boldsymbol{r}_u > \epsilon)\, du \right]. \end{aligned} \tag{3.39}$$

Notice that for any fixed $\epsilon > 0$,

$$\liminf_{n\to\infty} \mathbb{E}^n \left[G \left(\int_s^t \vartheta^n(\boldsymbol{r}_u)^2\, du - \int_s^t I(\boldsymbol{r}_u > \epsilon)\, du \right) \right] \ge 0. \tag{3.40}$$

By (3.38), (3.39) and (3.40), we have

$$\mathbb{E}\left[G \left(\boldsymbol{M}_t^2 - \boldsymbol{M}_s^2 - 4 \int_s^t I(\boldsymbol{r}_u > 0)\, du \right) \right] \;\ge\; 0.$$

Invoking a monotone class argument, cf. [130, Theorem 8], we see that $(\boldsymbol{M}_t^2 - 4\int_0^t I(\boldsymbol{r}_s > 0)\, ds, \mathcal{F}_t, \mathbb{P})$ is indeed a submartingale. We show that it is also a supermartingale and hence a martingale. By (3.37), for any function G as above,

$$\mathbb{E}\left[G\left(\boldsymbol{M}_t^2 - \boldsymbol{M}_s^2 - 4\,(t-s)\right)\right] \;=\; \lim_{n\to\infty} \mathbb{E}^n \left[G\left(\boldsymbol{M}_t^2 - \boldsymbol{M}_s^2 - 4\,(t-s)\right)\right] \;\le\; 0$$

Hence, $\boldsymbol{M}_t^2 - 4\,t$ is a supermartingale under $\mathbb{P}$. The uniqueness of the Doob-Meyer decomposition [130, Theorem 16] implies that the map $t \mapsto [\boldsymbol{M}]_t - 4\,t$ is $\mathbb{P}$-almost surely decreasing. Observe that $(\boldsymbol{r}_t, \mathcal{F}_t, \mathbb{P})$ is a continuous semimartingale with $[\boldsymbol{r}] = [\boldsymbol{M}]$. Hence the Itô-Tanaka formula implies that $\mathbb{P}$-almost surely,

$$\int_0^t I(\boldsymbol{r}_u = 0) d\,[\boldsymbol{M}]_u \;=\; \int_0^t I(\boldsymbol{r}_u = 0) d\,[\boldsymbol{r}]_u = \int_{-\infty}^{\infty} I(y=0)\ell_t^y(\boldsymbol{r}) dy = 0. \tag{3.41}$$

We conclude that for any $0 \le s < t$,

$$[\boldsymbol{M}]_t - [\boldsymbol{M}]_s \;=\; \int_s^t I(\boldsymbol{r}_u > 0)\, d[\boldsymbol{M}]_u \;\le\; 4\int_s^t I(\boldsymbol{r}_u > 0)\, du,$$

and hence for any $\mathcal{F}_s$-measurable function $G \in C_b(\mathbb{W})$,

$$\mathbb{E}\left[G\left(\boldsymbol{M}_t^2 - \boldsymbol{M}_s^2 - 4\int_s^t I(\boldsymbol{r}_u > 0)\, du\right)\right] \le 0.$$

As above we conclude by a monotone class argument that $(\boldsymbol{M}_t^2 - 4\int_0^t I(\boldsymbol{r}_u > 0)\, du)$ is a supermartingale, and hence a martingale, i.e.,

$$[\boldsymbol{M}] \;=\; 4\int_0^{\cdot} I(\boldsymbol{r}_u > 0)\, du \qquad \mathbb{P}\text{-almost surely.} \tag{3.42}$$

Similarly, we can show that

$$[\boldsymbol{N}] \;=\; 4\int_0^{\cdot} I(\boldsymbol{s}_u > 0)\, du \qquad \mathbb{P}\text{-almost surely.} \tag{3.43}$$

Moreover, we claim that

$$[\boldsymbol{M}, \boldsymbol{N}] \;=\; 4\int_0^{\cdot} I(\boldsymbol{r}_u > 0, \boldsymbol{s}_u > 0)\, du \qquad \mathbb{P}\text{-almost surely.} \tag{3.44}$$

The proof does not involve new arguments, so we just sketch the main steps: With the same arguments as before, one can conclude that

$$t \mapsto \boldsymbol{M}_t \boldsymbol{N}_t - 4\int_0^t I(\boldsymbol{r}_u > 0, \boldsymbol{s}_u > 0)\, du$$

is a submartingale and that the map $t \mapsto \boldsymbol{M}_t \boldsymbol{N}_t - 4t$ is $\mathbb{P}$-almost surely decreasing. Moreover, by (3.42), (3.43), and the Kunita-Watanabe inequality, we see that $\mathbb{P}$-a.s.,

$$\int_s^t I(\boldsymbol{r}_u = 0 \text{ or } \boldsymbol{s}_u = 0)\, d\,[\boldsymbol{M}, \boldsymbol{N}]_u \;=\; 0 \qquad \text{for } 0 \le s \le t, \qquad \text{and thus}$$

$$\begin{aligned}
[\boldsymbol{M}, \boldsymbol{N}]_t - [\boldsymbol{M}, \boldsymbol{N}]_s \;&=\; \int_s^t I(\boldsymbol{r}_u > 0, \boldsymbol{s}_u > 0)\, d\,[\boldsymbol{M}, \boldsymbol{N}]_u \\
&\le\; 4\int_s^t I(\boldsymbol{r}_u > 0, \boldsymbol{s}_u > 0)\, du \qquad \text{for } 0 \le s \le t.
\end{aligned}$$

This completes the proof of (3.44). Invoking a martingale representation theorem, see e.g. [87, Ch. II, Theorem 7.1'], we conclude that there is a probability space $(\Omega, \mathcal{A}, P)$ supporting a Brownian motion $\tilde{W}$, and random variables $(\tilde{r}, \tilde{s})$ such that $P \circ (\tilde{r}, \tilde{s})^{-1} = \mathbb{P} \circ (\boldsymbol{r}, \boldsymbol{s})^{-1}$, and such that $(\tilde{r}_t, \tilde{s}_t, \tilde{W}_t)$ is a weak solution of (3.32).

It remains to show that (3.33) and (3.34) imply $P[\tilde{r}_t \leq \tilde{s}_t \text{ for all } t \geq 0] = 1$. Applying a comparison theorem [86, Theorem 1] to the approximating diffusions (3.31) shows that $\mathbb{P}^n[\boldsymbol{r}_t \leq \boldsymbol{s}_t \text{ for all } t \geq 0] = 1$ for all n. The monotonicity carries over to the limit, since $\mathbb{P}^n \circ (\boldsymbol{r}, \boldsymbol{s})^{-1}$ converges weakly, along a subsequence, towards $\mathbb{P} \circ (\boldsymbol{r}, \boldsymbol{s})^{-1}$. □

3.3.2 Long time behavior

We now derive bounds for solutions to (3.29) that are stable for long times. We assume that $t \mapsto \alpha(t, x)$ is non-increasing, so that the *stickiness* of solutions to (3.29) is non-decreasing in time.

Assumption 38. *The function $\alpha : [0, \infty) \times [0, \infty) \to \mathbb{R}$ is locally Lipschitz continuous with $\alpha(t, x) \leq \alpha(s, x)$ for any $s \leq t$ and $x \in \mathbb{R}_+$, $\alpha(t, 0) > 0$ for any $t \geq 0$, and*

$$\limsup_{r \to \infty} \left(r^{-1} \alpha(0, r)\right) < 0. \tag{3.45}$$

Notice that Assumption 38 implies Assumptions 33, 34 and 35 from above.

Invariant measure in the time-homogenous case

We first consider drift coefficients which do not depend on time, i.e., functions of the form $\alpha(t, x) = \alpha(x)$.

Lemma 23. *Suppose that Assumption 38 holds true, and $\alpha(t, \cdot) = \alpha$ for a function $\alpha : [0, \infty) \to \mathbb{R}$. Let π be the probability measure on $[0, \infty)$ defined by*

$$\pi(dx) \quad = \quad \frac{1}{Z} \left(\frac{2}{\alpha(0)} \, \delta_0(dx) \; + \; \exp\left(\frac{1}{2} \int_0^x \alpha(y) \, dy \right) \lambda_{(0,\infty)}(dx) \right) \tag{3.46}$$

where $Z = \frac{2}{\alpha(0)} + \int_0^\infty \exp\left(\frac{1}{2} \int_0^x \alpha(y) \, dy\right) dx$. Then π is invariant for (3.29)*, i.e., if (r_t) is a solution with initial law π, then* $\mathrm{Law}(r_t) = \pi$ *for any $t \geq 0$.*

Proof. We use an approximation as in (3.31) with $\beta(t, x) = \alpha(x)$ and a sequence of smooth functions $\vartheta^n : [0, \infty) \to [0, 1]$ satisfying Assumption 37. It is well-known that under our assumptions, for each $n \in \mathbb{N}$, the probability measure $\tilde{\mu}^n$ on $\mathbb{R}_+$ with distribution function

$$\tilde{F}^n(x) = \frac{\int_0^x \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{\alpha(z)}{2\vartheta^n(z)^2} dz \right) dy}{\int_0^\infty \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{\alpha(z)}{2\vartheta^n(z)^2} dz \right) dy} \qquad x \in [0, \infty),$$

is an invariant measure for the process $(\tilde{r}_t^n)$ defined by (3.31), see e.g. [111, Chapter 4.4, Theorem 7]. Note in particular that by Assumptions 38 and 37, the occurring integrals are well defined and finite. Let F denote the distribution function of π. We show that for any $x > 0$, $\tilde{F}^n(x) \to F(x)$ as $n \to \infty$, which implies that $\tilde{\mu}_n \to \pi$ weakly. Indeed, fix $x \in (0, \infty]$. Then for $n > 1/x$,

$$\begin{aligned} &\int_0^x \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{\alpha(z)}{2\vartheta^n(z)^2} dz\right) dy \qquad (3.47) \\ = &\int_{1/n}^x \exp\left(\int_{1/n}^y \frac{1}{2}\alpha(z)dz\right) dy + \int_0^{1/n} \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{\alpha(z)}{2\vartheta^n(z)^2} dz\right) dy. \end{aligned}$$

If $C \in (0, \infty)$ is a constant then

$$\begin{aligned} \int_0^{1/n} \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{C}{\vartheta^n(z)^2} dz\right) dy &= \lim_{\epsilon\downarrow 0} \int_\epsilon^{1/n} \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{C}{\vartheta^n(z)^2} dz\right) dy \\ &= \lim_{\epsilon\downarrow 0} \frac{1}{C}\left(1 - \exp\left(-\int_\epsilon^{1/n} \frac{C}{\vartheta^n(z)^2} dz\right)\right) = \frac{1}{C}. \qquad (3.48) \end{aligned}$$

For $0 < y < 1/n$, we have the bounds

$$\begin{aligned} \exp\left(\max_{u\in[0,1/n]} \alpha(u) \int_{1/n}^y \frac{1}{2\vartheta^n(z)^2} dz\right) &\le \exp\left(\int_{1/n}^y \frac{\alpha(z)}{2\vartheta^n(z)^2} dz\right) \\ &\le \exp\left(\min_{u\in[0,1/n]} \alpha(u) \int_{1/n}^y \frac{1}{2\vartheta^n(z)^2} dz\right). \end{aligned}$$

Using (3.47), the continuity of α, and (3.48), we can conclude that as $n \to \infty$,

$$\int_0^x \frac{1}{\vartheta^n(y)^2} \exp\left(\int_{1/n}^y \frac{\alpha(z)}{2\vartheta^n(z)^2} dz\right) dy \;\to\; \int_0^x \exp\left(\int_0^y \frac{1}{2}\alpha(z)dz\right) dy + \frac{2}{\alpha(0)}.$$

Since this also holds for $x = \infty$, we see that $\tilde{F}^n(x) \to F(x)$ for any $x > 0$, and hence $\tilde{\mu}_n \to \pi$ weakly. Consequently, by Lemma 21 and Theorem 13, the laws of the solutions of (3.31) with initial distributions $\tilde{\mu}^n$ converge weakly to the law of the solution of (3.29) with initial distribution π. Since the approximating processes are stationary, the limit process is stationary, too. Hence π is an invariant measure. □

Long time stability in the time-inhomogeneous case

Let (r_t) be a solution of (3.29) with an arbitrary but fixed initial distribution μ on $\mathbb{R}_+$. Our aim is to provide bounds on $P[r_t > 0]$ and $E[r_t]$ for any fixed $t \ge 0$. To this end we fix a continuous function $a : [0, \infty) \to \mathbb{R}$ such that

$$\alpha(0, x) \;\le\; a(x) \text{ for any } x \in [0, \infty), \quad \text{and} \quad \limsup_{r\to\infty} (r^{-1}a(r)) \;<\; 0. \qquad (3.49)$$

For example, by Assumption 38, we can always choose $a(x) = \alpha(0, x)$. However, sometimes it can be more convenient to choose the function a in a different way. Following [50, 51] (see also [27, 28, 29, 33]), we define constants $R_0, R_1 \in (0, \infty)$ and a concave function $f : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$R_0 = \inf\{R \geq 0 : a(r) \leq 0 \quad \text{for any } r \geq R\}, \tag{3.50}$$

$$R_1 = \inf\{R \geq R_0 : R(R - R_0)\, a(r)/r \leq -4 \quad \text{for any } r \geq R\}, \tag{3.51}$$

$$f(r) = \int_0^r \phi(s)\; g(s)\; ds, \;\text{ where } \phi(r) = \exp\left(-\frac{1}{2}\int_0^r a(s)^+ ds\right) \text{ and} \tag{3.52}$$

$$g(r) = 1 - \frac{1}{4}\int_0^{r\wedge R_1} \frac{\Phi(s)}{\phi(s)}\, ds \Big/ \int_0^{R_1} \frac{\Phi(s)}{\phi(s)}\, ds - \frac{1}{4}\int_0^{r\wedge R_1} \frac{1}{\phi(s)}\, ds \Big/ \int_0^{R_1} \frac{1}{\phi(s)}\, ds$$

with $\Phi(r) = \int_0^r \phi(s)\, ds$. The function f is concave, strictly increasing and continuous. Observe that (3.49) implies that $0 < R_0 < R_1 < \infty$. We define constants

$$c = \left(2\int_0^{R_1} \frac{\Phi(s)}{\phi(s)}\; ds\right)^{-1}, \quad \epsilon = \min\left\{\left(2\int_0^{R_1} \frac{1}{\phi(s)}\; ds\right)^{-1},\; c\,\Phi(R_1)\right\}. \tag{3.53}$$

Notice that $1/2 \leq g \leq 1$, and thus $\Phi(r)/2 \leq f(r) \leq \Phi(r)$. Hence for $0 < r < R_1$,

$$2\, f''(r) \;+\; f'(r)\, a(r)^+ \;\leq\; -\epsilon \;-\; c\;\Phi(r) \;\leq\; -\left(\epsilon \;+\; c\; f(r)\right). \tag{3.54}$$

Lemma 24. *Suppose that Assumption 38 holds true. Let (r_t) be a solution of (3.29), and let $T_0 = \inf\{t \geq 0 : r_t = 0\}$. Then for any $t > 0$,*

$$E[f(r_t)\,;\, t < T_0] \;\leq\; e^{-ct}\; E[f(r_0)], \qquad \textit{and} \tag{3.55}$$

$$P[t < T_0] \;\leq\; \frac{1}{\epsilon}\,\frac{c}{e^{ct} - 1}\; E[f(r_0)]. \tag{3.56}$$

Proof. Notice that the function f can be extended to a concave function on $\mathbb{R}$ by setting $f(x) = x$ for $x < 0$. Since the process (r_t) is a continuous semimartingale, we can apply the Itô-Tanaka formula to conclude that almost surely,

$$df(r_t) \;=\; f'(r_t)\, \alpha(t, r_t)\; dt + 2\, f''(r_t)\, I(r_t > 0)\; dt + dM_t, \tag{3.57}$$

where $M_t = 2\int_0^t f'(r_s)\, I(r_s > 0)\; dW_s$ is a martingale. By Assumption 38 and (3.49), $\alpha(t, r_t) \leq \alpha(0, r_t) \leq a(r_t)$. Therefore, for $0 < r_t < R_1$, we can apply (3.54) to bound the right hand side of (3.57). On the other hand, for $r_t \geq R_1$, we have $f''(r_t) = 0$ and $r_t^{-1}\alpha(r_t) < 0$. Moreover, by definition of f and ϕ, $f'(r_t) = \phi(R_0)/2$, and by (3.51), $R_1(R_1 - R_0)\alpha(r_t)/{r_t}^{-1} \leq -4$. Therefore, we can conclude similarly to [51, Proof of Theorem 2.2] that for $r_t > R_1$,

$$\begin{aligned} f'(r_t)\, \alpha(t, r_t) \;&\leq\; \phi(R_0) a(r_t)/2 \;\leq\; -2\frac{\phi(R_0)}{R_1 - R_0}\frac{r_t}{R_1} \;<\; -2\frac{\phi(R_0)}{R_1 - R_0}\frac{\Phi(r_t)}{\Phi(R_1)} \\ &\leq\; -\Phi(r_t) \Big/ \int_{R_0}^{R_1} \Phi(s)\phi(s)^{-1}\, ds \;\leq\; -2\, c\, \Phi(r_t) \\ &\leq\; -c\, \Phi(R_1) \;-\; c\, f(r_t) \;\leq\; -\left(\epsilon + \; c\, f(r_t)\right). \end{aligned} \tag{3.58}$$

Here we have used that $\int_{R_0}^{R_1} \Phi(s)\phi(s)^{-1}\, ds \geq (R_1 - R_0)\Phi(R_1)\phi(R_0)^{-1}/2$. Combining (3.57), (3.54) and (3.58), we see that almost surely,

$$df(r_t) \;\leq\; -(\epsilon + \; c\, f(r_t))\; dt \;+\; dM_t \qquad \text{for } t < T_0. \tag{3.59}$$

Using Itô's product rule and (3.59), we finally obtain

$$\begin{aligned} e^{ct}\, E[f(r_t);\, t < T_0] \;&\leq\; E[f(r_0)] \;+\; E[e^{c(t\wedge T_0)} f(r_{t\wedge T_0}) - f(r_0)] \\ &\leq\; E[f(r_0)] \;-\; \frac{\epsilon}{c}\, \left(E\left[e^{c\,(t\wedge T_0)}\right] - 1\right), \quad \text{and} \\ P[t < T_0] \;&\leq\; E\left[\frac{e^{c\,(t\wedge T_0)} - 1}{e^{c\,t} - 1}\right] \;\leq\; \frac{1}{\epsilon}\, \frac{c}{e^{c\,t} - 1}\, E[f(r_0)]. \end{aligned}$$

□

For $s \in [0, \infty)$, we denote by π_s the invariant probability measure for the *time-homogeneous* sticky diffusion with drift $\alpha(s,\cdot)$ that is given by (3.46), i.e.,

$$\pi_s(dx) \;\propto\; \frac{2}{\alpha(s,0)}\, \delta_0(dx) \;+\; \exp\left(\frac{1}{2}\int_0^x \alpha(s,y)\, dy\right) \lambda_{(0,\infty)}(dx). \tag{3.60}$$

Theorem 14. *Suppose that Assumption 38 holds true, and let (r_t) be a solution of (3.29) with initial distribution μ on $\mathbb{R}_+$. Then for any $t > 0$,*

$$\begin{aligned} E[f(r_t)] \;&\leq\; e^{-c\,t}\, E[f(r_0)] \;+\; \int f\, d\pi_0, \quad E[r_t] \;\leq\; 2\,\phi(R_0)^{-1} E[f(r_t)], \quad \textit{and} \\ P[r_t > 0] \;&\leq\; \frac{1}{\epsilon}\, \frac{c}{e^{c\,t} - 1}\, E[f(r_0)] \;+\; \pi_0[(0,\infty)]. \end{aligned}$$

Proof. Based on the results of Theorem 13, we can construct a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t), P)$ satisfying the usual conditions and supporting random variables $r, W, \tilde{r}, \tilde{s}, \tilde{W} : \Omega \to \mathbb{W}$ such that w.r.t. $(\Omega, \mathcal{A}, (\mathcal{F}_t), P)$,

- (r, W) and $(\tilde{r}, \tilde{s}, \tilde{W})$ are independent,
- (r_t, W_t) is a weak solution of (3.29) with initial distribution μ, and
- $(\tilde{r}_t, \tilde{s}_t, \tilde{W}_t)$ is a weak solution of (3.32) with $\beta(t,x) = \alpha(t,x)$, $\gamma(t,x) = \alpha(0,x)$, $\tilde{\mu} = \delta_0$, $\tilde{\nu} = \pi_0$, and
$$P[\; \tilde{r}_t \leq \tilde{s}_t \text{ for all } t \geq 0 \;] \;=\; 1. \tag{3.61}$$

Let $T := \inf\{t \geq 0 : r_t = \tilde{r}_t\}$ be the first meeting time of (r_t) and $(\tilde{r}_t)$. We define

$$\bar{r}_t \;:=\; r_t \quad \text{for } t < T, \quad \text{and} \quad \bar{r}_t \;:=\; \tilde{r}_t \quad \text{for } t \geq T. \tag{3.62}$$

Then $(\bar{r}_t)$ solves the martingale problem corresponding to (3.29) with initial law μ, cf. e.g. [129, Section 3.1]. By Lemma 21, this martingale problem has a unique solution. Hence, we can conclude that the laws of $\bar{r}$ and r on $\mathbb{W}$ coincide. Let

$T_0 = \inf\{t \geq 0 : r_t = 0\}$. Observe that since $t \mapsto r_t$ and $t \mapsto \tilde{r}_t$ are continuous with $\tilde{r}_0 = 0 \leq r_0$, we have $T \leq T_0$. In particular, by Lemma 24, (3.61), and since f is increasing,

$$\begin{aligned} E[f(r_t)] &= E[f(\bar{r}_t)] = E[f(r_t);\, t < T] + E[f(\tilde{r}_t);\, t \geq T] \\ &\leq E[f(r_t);\, t < T_0] + E[f(\tilde{s}_t)] \leq e^{-ct}\, E[f(r_0)] + \int f\, d\pi_0. \end{aligned}$$

Here we have used that by Lemma 23, the process $(\tilde{s}_t)$ is stationary. By (3.52), (3.50), and since $g \geq 1/2$, we have $f' \geq \phi(R_0)/2$. Hence the inequality $r \leq 2\,\phi(R_0)^{-1} f(r)$ holds for any $r \geq 0$, and thus, we can conclude that

$$E[r_t] \leq 2\phi(R_0)^{-1} E[f(r_t)].$$

Finally, by the second part of Lemma 24, we see that

$$\begin{aligned} P[r_t > 0] &= P[\bar{r}_t > 0] = P[r_t > 0,\, t < T] + P[\tilde{r}_t > 0,\, t \geq T] \\ &\leq P[t < T_0] + P[\tilde{s}_t > 0] \leq \frac{1}{\epsilon}\, \frac{c}{e^{ct} - 1}\, E[f(r_0)] + \pi_0[(0, \infty)]. \end{aligned}$$

□

By applying Theorem 14 on the time intervals $[s, t]$ and $[0, s]$, we obtain:

Corollary 13. *Suppose that Assumption 38 holds true, and let (r_t) be a solution of* (3.29)*. Then for any $0 \leq s < t$,*

$$\begin{aligned} E[f(r_t)] &\leq e^{-ct} E[f(r_0)] + e^{-c\,(t-s)} \int f\, d\pi_0 + \int f\, d\pi_s, \qquad \textit{and} \\ P[r_t > 0] &\leq \frac{1}{\epsilon}\, \frac{c}{e^{c\,(t-s)} - 1} \left(e^{-cs} E[f(r_0)] + \int f\, d\pi_0 \right) + \pi_s[(0, \infty)]. \end{aligned}$$

where f, c and ϵ are defined as above. Furthermore,

$$\begin{aligned} E[f_s(r_t)] &\leq \frac{2}{\phi(R_0)}\, e^{-c_s(t-s)} \left(e^{-cs} E[f(r_0)] + \int f\, d\pi_0 \right) + \int f_s\, d\pi_s, \qquad \textit{and} \\ P[r_t > 0] &\leq \frac{2}{\phi(R_0)\epsilon_s}\, \frac{c_s}{e^{c_s\,(t-s)} - 1} \left(e^{-cs} E[f(r_0)] + \int f\, d\pi_0 \right) + \pi_s[(0, \infty)], \end{aligned}$$

where f_s, c_s and ϵ_s are defined by (3.52)*,* (3.53) *and* (3.46) *with a replaced by $\alpha(s, \cdot)$.*

Proof. Fix $s \in [0, \infty)$. Then the process $(r_{s+t})_{t \geq 0}$ solves (3.29) with drift coefficient $\alpha_s(t, x) = \alpha(s + t, x)$ and initial distribution $P \circ r_s^{-1}$. Since $\alpha_s(t, x) \leq \alpha(s, x) \leq a(x)$ for any $t, x \geq 0$, we can apply Theorem 14 either with a, f, c and ϵ as above, or with

a, f, c and ϵ replaced by $\alpha(s,\cdot), f_s, c_s$ and ϵ_s. For $t > s$ we obtain

$$
\begin{aligned}
E[f(r_t)] &\leq e^{-c\,(t-s)}\; E[f(r_s)] \;+\; \int f\, d\pi_s,\\
P[r_t > 0] &\leq \frac{1}{\epsilon}\,\frac{c}{e^{c\,(t-s)}-1}\; E[f(r_s)] \;+\; \pi_s[(0,\infty)],\\
E[f_s(r_t)] &\leq e^{-c_s\,(t-s)}\; E[f_s(r_s)] \;+\; \int f_s\, d\pi_s,\\
P[r_t > 0] &\leq \frac{1}{\epsilon_s}\,\frac{c_s}{e^{c_s\,(t-s)}-1}\; E[f_s(r_s)] \;+\; \pi_s[(0,\infty)].
\end{aligned}
$$

Noting that $f_s(r_s) \leq r_s$, the assertion follows by applying Theorem 14 once more. □

3.4 Coupling construction and proofs of the main results

In this section, we prove our main theorems. First of all, we construct the sticky coupling (X_t, Y_t) of solutions to (3.1) and (3.2) respectively, advertised in Theorem 10. The coupling is obtained as a weak limit of Markovian couplings (X_t^δ, Y_t^δ), $\delta > 0$. The couplings (X_t^δ, Y_t^δ) are reflection couplings for $|X_t^\delta - Y_t^\delta| \geq \delta$ and synchronous couplings for $|X_t^\delta - Y_t^\delta| = 0$. Inbetween there is an interpolation between the two types of couplings. We argue that the family of couplings is tight and thus there is a subsequence converging to a coupling $(X_t, Y_t)_{t\geq 0}$. It is then argued that this limiting coupling is sticky and shares the properties stated in Theorem 10.

We now define the couplings (X_t^δ, Y_t^δ) rigorously. The technical realization follows [51]. We introduce Lipschitz functions $\mathrm{rc}^\delta, \mathrm{sc}^\delta : \mathbb{R}_+ \to [0,1]$ such that $\mathrm{rc}^\delta(0) = 0$, $\mathrm{rc}^\delta(r) > 0$ for $0 < r < \delta$, $\mathrm{rc}^\delta(r) = 1$ for $r \geq \delta$, and

$$
\mathrm{rc}^\delta(r)^2 + \mathrm{sc}^\delta(r)^2 \;=\; 1 \qquad \text{for any } r \geq 0. \tag{3.63}
$$

Let (B_t^1) and (B_t^2) be independent d-dimensional Brownian motions, and let $u \in \mathbb{R}^d$ be some arbitrary unit vector. We define the coupling (X_t^δ, Y_t^δ) for (3.1) and (3.2) as a diffusion process in $\mathbb{R}^{2d}$ satisfying the stochastic differential equation

$$
\begin{aligned}
dX_t^\delta &= b(t, X_t^\delta)\, dt + \mathrm{rc}^\delta\left(\tilde r_t^\delta\right)\, dB_t^1 \;+\; \mathrm{sc}^\delta\left(\tilde r_t^\delta\right)\, dB_t^2, && (3.64)\\
dY_t^\delta &= \tilde b(t, Y_t^\delta)\, dt + \mathrm{rc}^\delta\left(\tilde r_t^\delta\right)\, \left(\mathrm{Id}_{\mathbb{R}^d} - 2\, e_t^\delta \left\langle e_t^\delta, \cdot\right\rangle\right)\, dB_t^1 \;+\; \mathrm{sc}^\delta\left(\tilde r_t^\delta\right)\, dB_t^2, && (3.65)
\end{aligned}
$$

with initial condition $(X_0^\delta, Y_0^\delta) = (x, y)$. Here $Z_t^\delta = X_t^\delta - Y_t^\delta$, $\tilde r_t^\delta = \left|Z_t^\delta\right|$, $e_t^\delta = Z_t^\delta / \tilde r_t^\delta$ if $\tilde r_t^\delta \neq 0$, and $e_t^\delta = u$ if $\tilde r_t^\delta = 0$. Since $\mathrm{rc}^\delta(0) = 0$, the arbitrary value u is not relevant for the dynamics. The process (X_t^δ, Y_t^δ) can be realized as a standard diffusion process in $\mathbb{R}^{2d}$ with locally Lipschitz coefficients. Moreover, Assumptions 30 and 31 imply the non-explosiveness of the process. Using Lévy's characterization of Brownian motion and (3.63), one can check that (X_t^δ, Y_t^δ) is indeed a coupling of solutions to Equations (3.1) and (3.2). Notice that the process $W_t^\delta = \int_0^t \left\langle e_s^\delta, dB_s^1\right\rangle$ is a one-dimensional Brownian motion.

Lemma 25. *Suppose that Assumptions 30 and 31 are satisfied. Then, almost surely,*

$$d\tilde r_t^\delta \;=\; \left\langle e_t^\delta, b(t,X_t^\delta) - \tilde b(t,Y_t^\delta)\right\rangle \; dt \;+\; 2\,\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)\; dW_t^\delta \tag{3.66}$$

$$\le \qquad \left(\, M \;+\; \kappa(\tilde r_t^\delta)\,\tilde r_t^\delta \,\right)\; dt \;+\; 2\,\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)\; dW_t^\delta. \tag{3.67}$$

Proof. By (3.64) and (3.65),

$$d(\tilde r_t^\delta)^2 = 2\left\langle Z_t^\delta, b(t,X_t^\delta) - \tilde b(t,Y_t^\delta)\right\rangle dt + 4\,\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)^2 dt + 4\,\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)\left\langle Z_t^\delta, e_t^\delta\right\rangle dW_t^\delta.$$

For $\epsilon > 0$, we define a C^2 approximation of the square root by

$$S_\epsilon(r) \;=\; -(1/8)\,\epsilon^{-3/2}\,r^2 \;+\; (3/4)\,\epsilon^{-1/2}\,r \;+\; (3/8)\,\epsilon^{1/2} \qquad \text{for } r<\epsilon, \tag{3.68}$$

$S_\epsilon(r) = \sqrt{r}$ for $r \ge \epsilon$. By Itô's formula,

$$\begin{aligned} dS_\epsilon((\tilde r_t^\delta)^2) = {} & 2\,S_\epsilon'((\tilde r_t^\delta)^2)\left\langle Z_t^\delta, b(t,X_t^\delta) - \tilde b(t,Y_t^\delta)\right\rangle\; dt + 4S_\epsilon'((\tilde r_t^\delta)^2)\,\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)^2 dt \\ & + 8\,S_\epsilon''((\tilde r_t^\delta)^2)\,\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)^2 (\tilde r_t^\delta)^2\, dt \;+\; 4\,S_\epsilon'((\tilde r_t^\delta)^2)\;\mathrm{rc}^\delta\left(\tilde r_t^\delta\right)\tilde r_t^\delta\, dW_t^\delta. \end{aligned}$$

We can now pass to the limit $\epsilon \downarrow 0$ to obtain (3.66). Notice that $\sup_{0\le r\le\epsilon}|S_\epsilon'(r)| \lesssim \epsilon^{-1/2}$, $\sup_{0\le r\le\epsilon}|S_\epsilon''(r)| \lesssim \epsilon^{-3/2}$ and that rc^δ is Lipschitz with $\mathrm{rc}^\delta(0) = 0$. Hence, one can use Lebesgue's dominated convergence theorem for the convergence of the first three integrals. Moreover, the stochastic integral converges almost surely, along a subsequence, to $\int_0^t 2\;\mathrm{rc}^\delta\left(\tilde r_s^\delta\right)\; dW_s^\delta$. Finally, by Assumptions 30 and 31,

$$\begin{aligned} \left\langle Z_t^\delta, b(t,X_t^\delta) - \tilde b(t,Y_t^\delta)\right\rangle \;&\le\; \left\langle Z_t^\delta, b(t,X_t^\delta) - b(t,Y_t^\delta) + b(t,Y_t^\delta) - \tilde b(t,Y_t^\delta)\right\rangle \\ &\le\; M\;\tilde r_t^\delta \;+\; \kappa(\tilde r_t^\delta)\;(\tilde r_t^\delta)^2. \end{aligned}$$

□

In order to control the distance of X_t^δ and Y_t^δ, we introduce a one-dimensional process (r_t^δ) that is defined as the unique and strong solution to the equation

$$dr_t^\delta \;=\; \left(\, M \;+\; \kappa(r_t^\delta)\cdot r_t^\delta \,\right)\; dt \;+\; 2\,\mathrm{rc}^\delta\left(r_t^\delta\right)\; dW_t^\delta, \quad r_0^\delta = \tilde r_0^\delta, \tag{3.69}$$

with $(\tilde r_t^\delta)$ and (W_t^δ) as above.

Lemma 26. *We have* $\left|X_t^\delta - Y_t^\delta\right| = \tilde r_t^\delta \le r_t^\delta$, *almost surely for all* $t \ge 0$.

Proof. The processes $(\tilde r_t^\delta)$ and (r_t^δ) are driven by the same noise, start at the same position, and, by (3.67), the drift of $(\tilde r_t^\delta)$ is smaller or equal to the one of (r_t^δ). Therefore, the assertion follows by Ikeda-Watanabe's comparison theorem for one-dimensional diffusions, cf. [86, Theorem 1.1]. □

Proof of Theorem 10. We consider the diffusion $U_t^\delta := (X_t^\delta, Y_t^\delta, r_t^\delta)$ on $\mathbb{R}^{2d+1}$. Let $\mathbb{P}^\delta$ denote the law of U^δ on the space $C(\mathbb{R}_+, \mathbb{R}^{2d+1})$. We define $\boldsymbol{X}, \boldsymbol{Y} : C(\mathbb{R}_+, \mathbb{R}^{2d+1}) \to C(\mathbb{R}_+, \mathbb{R}^d)$ and $\boldsymbol{r} : C(\mathbb{R}_+, \mathbb{R}^{2d+1}) \to C(\mathbb{R}_+, \mathbb{R})$ as the canonical projections onto the first d, the second d, and the last coordinate.

Notice that in each of the equations (3.64), (3.65) and (3.69), the drift coefficients do not depend on δ and the diffusion coefficients are uniformly bounded. Moreover, Assumptions 30 and 31 imply that, similarly as in the proof of Theorem 13, the diffusions (U_t^δ) satisfy uniformly a Lyapunov non-explosion criterion, and the drift coefficients are uniformly bounded on compact sets. Therefore, the family $(\mathbb{P}^\delta)$ is tight, cf. [80, 81]. In particular, there is a sequence $\delta_n \downarrow 0$ such that $(\mathbb{P}^{\delta_n})$ converges towards a measure $\mathbb{P}$ on $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^{2d+1})$. For each $\delta > 0$, (X_t^δ) and (Y_t^δ) are solutions to (3.1) and (3.2) respectively. Since those solutions are unique in law, we know that $\mathbb{P}^\delta \circ (X^\delta)^{-1} = \mathbb{P} \circ \boldsymbol{X}^{-1}$ and $\mathbb{P}^\delta \circ (Y^\delta)^{-1} = \mathbb{P} \circ \boldsymbol{Y}^{-1}$ for any $\delta > 0$. Hence, $\mathbb{P} \circ (\boldsymbol{X}, \boldsymbol{Y})^{-1}$ is a coupling of (3.1) and (3.2). Moreover, Lemma 21 and the proof of Theorem 13 reveal that, after extending the underlying probability space, there is a Brownian motion $(\tilde{W}_t)$ such that $(r_t, \tilde{W}_t)$ is a solution of (3.3). The statement from Lemma 26 carries over to the limiting processes, since such inequalities are preserved under weak convergence, and thus (3.4) holds. The inequality (3.8) is implied by Theorem 14 setting $\alpha(t,x) = a(x) = M \ + \ \kappa(x) \cdot x$. □

Proof of Lemma 18. By (3.6), $\pi[(0,\infty)] = \frac{\alpha}{1+\alpha}$ with

$$\alpha \ := \ \frac{M}{2} \int_0^\infty \exp\left(\frac{1}{2}\int_0^x (M + \kappa(y)\, y)\, dy\right) dx.$$

In order to provide upper bounds on α, we decompose $\alpha = M(a+b)/2$ with

$$\begin{aligned} a &= \int_{\mathcal{R}}^\infty \exp\left(\frac{1}{2}\int_0^x (M + \kappa(y)\, y)\, dy\right) dx \qquad \text{and} \\ b &= \int_0^{\mathcal{R}} \exp\left(\frac{1}{2}\int_0^x (M + \kappa(y)\, y)\, dy\right) dx. \end{aligned}$$

By Condition (3.13), we have

$$\begin{aligned} \frac{1}{2}\int_0^x (M + \kappa(y)\, y)\, dy &= \frac{1}{2}\int_0^{\mathcal{R}} (M + L\, y)\, dy + \frac{1}{2}\int_{\mathcal{R}}^x (M - K\, y)\, dy \\ &= Mx/2 - Kx^2/4 + (L+K)\mathcal{R}^2/4 \\ &= -K(x - M/K)^2/4 + M^2/(4K) + (L+K)\mathcal{R}^2/4 \end{aligned}$$

for $x \geq \mathcal{R}$ and

$$\frac{1}{2}\int_0^x (M + \kappa(y)\, y)\, dy \ = \ \frac{1}{2}\int_0^x (M + L\, y)\, dy \ = \ Mx/2 + Lx^2/4$$

for $x \leq \mathcal{R}$. We obtain

$$\begin{aligned}
a &= \exp\left(M^2/(4K) + (L+K)\mathcal{R}^2/4\right) \int_{\mathcal{R}}^{\infty} \exp\left(-K\left(x - M/K\right)^2/4\right)\, dx \\
&= \frac{\sqrt{2}}{\sqrt{K}} \exp\left(M^2/(4K) + (L+K)\mathcal{R}^2/4\right) \int_{(\mathcal{R}-M/K)\sqrt{K/2}}^{\infty} \exp\left(-z^2/2\right)\, dz \quad \text{and} \\
b &= \int_0^{\mathcal{R}} \exp\left(Mx/2 + Lx^2/4\right)\, dx
\end{aligned}$$

and give upper bounds for these quantities:

$$b \leq \mathcal{R} \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \tag{3.70}$$

$$\begin{aligned}
b &= \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \int_0^{\mathcal{R}} \exp\left(M(\mathcal{R}-x)/2 - L(\mathcal{R}^2 - x^2)/4\right)\, dx \\
&= \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \int_0^{\mathcal{R}} \exp\left(-My/2 - Ly\left(2\mathcal{R} - y\right)/4\right)\, dy \\
&\leq \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \int_0^{\mathcal{R}} \exp\left(-My/2 - L\mathcal{R}y/4\right)\, dy \\
&\leq \frac{1}{M/2 + L\mathcal{R}/4} \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right).
\end{aligned} \tag{3.71}$$

Combining (3.70) and (3.71), we conclude that

$$b \leq \frac{4\mathcal{R}}{\max(4, 2M\mathcal{R} + L\mathcal{R}^2)} \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right). \tag{3.72}$$

We use the bound $\int_0^\infty e^{-z^2/2} dz \leq \sqrt{2\pi}$ to conclude that

$$\begin{aligned}
a &\leq 2\sqrt{\pi/K} \exp\left(M^2/(4K) + (L+K)\mathcal{R}^2/4\right) \\
&= 2\sqrt{\pi/K} \exp\left(K(R - M/K)^2/4\right) \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \\
&\leq 2\sqrt{\frac{\pi e}{K}} \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \quad \text{for } K(R - M/K)^2 \leq 2.
\end{aligned} \tag{3.73}$$

On the other hand, $\int_y^\infty e^{-z^2/2}\, dz \leq e^{-y^2/2}/y$ for any $y > 0$ and thus

$$\begin{aligned}
a &\leq \frac{2}{K} \frac{1}{\mathcal{R} - M/K} \exp\left((-K(\mathcal{R} - M/K)^2 + M^2/K + (L+K)\mathcal{R}^2)/4\right) \\
&= \frac{2}{\sqrt{K}} \frac{1}{\sqrt{K(R - M/K)^2}} \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \\
&\leq \frac{\sqrt{2}}{\sqrt{K}} \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right)
\end{aligned} \tag{3.74}$$

provided $\mathcal{R} \geq M/K$ and $K(R - M/K)^2 \geq 2$. Combining (3.72), (3.73) and (3.74), we obtain in the case $\mathcal{R} \geq M/K$ the bound

$$\begin{aligned} \alpha &= M(a+b)/2 \\ &\leq \left(\pi^{1/2} e^{1/2} K^{-1/2} + 2\mathcal{R} \max(4, L\mathcal{R}^2 + 2M\mathcal{R})^{-1}\right) M \exp\left(M\mathcal{R}/2 + L\mathcal{R}^2/4\right) \end{aligned}$$

In the case $\mathcal{R} \leq M/K$, (3.72) implies

$$b \leq \frac{4\mathcal{R}}{\max(4, 2M\mathcal{R} + L\mathcal{R}^2)} \exp\left(M^2/(4K) + (L+K)\mathcal{R}^2/4\right). \tag{3.75}$$

Combining (3.75) and (3.73), we can conclude for $\mathcal{R} \leq M/K$ the bound

$$\alpha \leq \left(\sqrt{\frac{\pi}{K}} + \frac{2\mathcal{R}}{\max(4, 2M\mathcal{R} + L\mathcal{R}^2)}\right) M \exp\left(\frac{M^2}{4K} + \frac{L+K}{4}\mathcal{R}^2\right).$$

□

Proof of Theorem 12. The proof is similar to the proof of Theorem 10. We fix $x_0, y_0 \in \mathbb{R}^d$ and corresponding drifts $b(t,x) = b^{x_0}(t,x)$ and $\tilde{b}(t,x) = b^{y_0}(t,x)$ as in (3.22) and (3.23) respectively. Moreover, we choose $\tau_0 \in (0,\infty)$ such that (3.26) holds for $|\tau| \leq \tau_0$. Since ϑ is Lipschitz, we can conclude by (3.26) that for any $x \in \mathbb{R}^d$,

$$\begin{aligned} |b(t,x) - \tilde{b}(t,x)| &= |\tau| \cdot \left| \int \vartheta(x,y) \mu_t^{x_0}(dy) - \int \vartheta(x,y) \mu_t^{y_0}(dy) \right| \\ &\leq |\tau| L\, \mathcal{W}^1(\mu_t^{x_0}, \mu_t^{y_0}) \leq L\, A\, e^{-\lambda t}\, |x_0 - y_0|, \end{aligned}$$

where L is the corresponding Lipschitz constant. We can now repeat the procedure leading to the proof of Theorem 10, replacing M by $|\tau|\, L A e^{-\lambda t} |x_0 - y_0|$. In particular, we can conclude that there is a coupling (X_t, Y_t) of (3.24) and (3.25) and a solution (r_t, W_t) of (3.29) with $r_0 = |x_0 - y_0|$ and drift

$$\alpha(t,x) = |\tau|\, L A e^{-\lambda t} |x_0 - y_0| + \kappa(x) x \tag{3.76}$$

such that $|X_t - Y_t| \leq r_t$. Notice that Assumption 32 implies Assumption 38 for the drift α. We now want to apply Corollary 13. First, we fix the function a in (3.49) as $a(\cdot) := \alpha(0,\cdot)$. Applying Corollary 13 now yields that for any $0 \leq s < t$,

$$\|\mu_t^{x_0} - \mu_t^{y_0}\|_{TV} \leq \frac{1}{\epsilon} \frac{c}{e^{c(t-s)} - 1} \left(e^{-cs} f(|x_0 - y_0|) + \int f\, d\pi_0\right) + \pi_s[(0,\infty)].$$

By (3.60), Assumption 38, and since $f(r) \leq r$, we have $f(|x_0 - y_0|) \leq |x_0 - y_0|$ and $\int f\, d\pi_0 < \infty$. Moreover, by (3.60), (3.76) and Assumption 38,

$$\pi_s[(0,\infty)] \leq \frac{1}{2}\alpha(s,0) \int_0^\infty \exp\left(\frac{1}{2}\int_0^x \alpha(s,y)\, dy\right) dx \leq C\, e^{-\lambda s}$$

where $C := \frac{1}{2}|\tau| LA \int_0^\infty \exp\left(\frac{1}{2}\int_0^x \alpha(s,y)\,dy\right)dx$ is a finite constant. Thus, there is a constant $A \in (0,\infty)$ such that

$$\|\mu_t^{x_0} - \mu_t^{y_0}\|_{TV} \;\leq\; \frac{A}{e^{c\,(t-s)}-1} + Ce^{-\lambda\, s} \;=\; e^{-c\,(t-s)}\frac{A}{1-e^{-c\,(t-s)}} + Ce^{-\lambda\, s}$$

for any $0 \leq s < t$. We can now set $s = t/2$ and use the boundedness of $\|\cdot\|_{TV}$ to see that there is a constant $B \in (0,\infty)$ such that

$$\|\mu_t^{x_0} - \mu_t^{y_0}\|_{TV} \;\leq\; B\,\exp(-\min(c,\lambda)\,t/2) \qquad \text{for all } t \geq 0.$$

It should be stressed, that the constants B and c depend on the initial conditions. □

Computations for Example 10

We now prove lower bounds on the total variation distance between the probability measures $\nu(dx) = Z_f^{-1} f(x)\,dx$ and $\mu(dx) = Z_g^{-1} g(x)\,dx$ on $\mathbb{R}^1$ that have been considered in Example 10. Noticing that by symmetry of f,

$$\begin{aligned} Z_g &= \int_{-\infty}^{\infty} g(x)\,dx \;=\; \int_{-\infty}^{\infty} e^{mx} f(x)\,dx \;=\; \int_0^{\infty} (e^{mx} + e^{-mx}) f(x)\,dx \\ &\geq 2\int_0^{\infty} f(x)\,dx \;=\; \int_{-\infty}^{\infty} f(x)\,dx \;=\; Z_f, \qquad \text{we obtain} \end{aligned}$$

$$\begin{aligned} \|\mu - \nu\|_{TV} &= \int_{\mathbb{R}} (1 - d\mu/d\nu)^+ \,d\nu \;=\; \int_{\mathbb{R}} (1 - e^{mx} Z_f/Z_g)^+ \,\nu(dx) \\ &\geq \int_{-\infty}^{0} (1 - e^{mx})\,\nu(dx) \;=\; \int_0^{\infty} \left(1 - e^{-mx}\right)\,\nu(dx) \\ &= \nu[(0,R)] \int_0^R (1 - e^{-mx})\,dx \Big/ R \\ &\quad + \nu[(R,\infty)] \int_R^{\infty} (1 - e^{-mx}) e^{-k(x-R)^2/2}\,dx \Big/ \int_R^{\infty} e^{-k(x-R)^2/2}\,dx \qquad (3.77) \\ &= \nu[(0,R)] \left(mR - 1 + e^{-mR}\right)/(mR) \\ &\quad + \nu[(R,\infty)] \int_0^{\infty} (1 - e^{-m(R+t)}) e^{-kt^2/2}\,dt \Big/ \int_0^{\infty} e^{-kt^2/2}\,dt\,. \end{aligned}$$

Using that $(e^{-x} - 1 + x)/x \leq 1 - e^{-x}$ for any $x > 0$, we obtain the lower bound

$$\|\mu - \nu\|_{TV} \;\geq\; (e^{-mR} - 1 + mR)/(mR).$$

We now derive an improved bound for small k. Suppose that $R\sqrt{k} \leq 1$. Then

$$\nu[(R,\infty)]/\nu[(0,R)] \;=\; \int_0^{\infty} e^{-kt^2/2}\,dt \Big/ R \;=\; \sqrt{\pi/(2k)}\,R^{-1}$$

implies $\nu[(R,\infty)] = \frac{1}{2}\left(1 + R\sqrt{2k/\pi}\right)^{-1} \geq \frac{1}{4}$. Hence by (3.77),

$$\begin{aligned}
\|\mu-\nu\|_{TV} &\geq \frac{1}{4}\int_0^\infty (1-e^{-m(R+t)})e^{-kt^2/2}\,dt \Big/ \int_0^\infty e^{-kt^2/2}\,dt \\
&= \frac{1}{4}\left(1-e^{-mR+m^2/(2k)}\left(1-\sqrt{2/\pi}\int_0^{m/\sqrt{k}} e^{-s^2/2}\,ds\right)\right) \\
&\geq \frac{1}{4}\left(1-e^{-mR+m^2/(2k)}+\sqrt{2/(\pi k)}me^{-mR}\right).
\end{aligned}$$

Appendix

Supplements for Chapter 0

Proof of Consequence 1. Fix arbitrary $x \in \mathcal{S}$. Inequality (0.6) implies that $\delta_x p_t \in \mathcal{P}_V(\mathcal{S})$ for any $t \in I$. Moreover, (0.4) and (0.5) show that $\mathcal{P}_V(\mathcal{S}) \subset \mathcal{P}_\rho(\mathcal{S}) \subset \mathcal{P}^1(\mathcal{S})$. Fix $t, s \in I$ with $t > s$. The semigroup property, (0.4) and (0.3) allow to conclude the inequality

$$\mathcal{W}^1(\delta_x p_t, \delta_x p_s) \;=\; \mathcal{W}^1(\delta_x p_{t-s} p_s, \delta_x p_s) \;\le\; C_1\, \mathrm{e}^{-cs}\, \mathcal{W}_\rho(\delta_x p_{t-s}, \delta_x). \tag{3.78}$$

Moreover, inequalities (0.5) and (0.6) show that

$$\sup_{t \in I} \mathcal{W}_\rho(\delta_x p_t, \delta_x) < \infty$$

and hence, $(\delta_x p_t)_{t \in I}$ is a Cauchy sequence in the Polish space $(\mathcal{W}^1, \mathcal{P}^1)$. In particular, there is $\pi \in \mathcal{P}^1$ such that $\mathcal{W}^1(\delta_x p_t, \pi) \to 0$ for $t \to \infty$ and thus $\delta_x p_t \to \pi$ weakly. Using the Feller property, we can conclude that for any $f \in C_b$ and $t \in I$,

$$\begin{aligned}
\int f(x)\, \pi(dx) \;&=\; \lim_{s \to \infty} \int f(x)\, (\delta_x p_s)(dx) \;=\; \lim_{s \to \infty} \int f(x)\, (\delta_x p_{s+t})(dx) \\
&=\; \lim_{s \to \infty} \int (p_t f)(x)\, (\delta_x p_s)(dx) \;=\; \int (p_t f)(x)\, \pi(dx) \\
&=\; \int f(x)\, (\pi p_t)(dx)
\end{aligned}$$

and thus $\pi p_t = \pi$ for any $t \in I$. Let $\tilde{\pi}$ be *some* invariant probability measure for (p_t). For any $n \in \mathbb{N}$ and $t \in I \setminus \{0\}$,

$$\begin{aligned}
\int (V(x) \wedge n)\, \tilde{\pi}(dx) \;&=\; \int (V(x) \wedge n)\, (\tilde{\pi} p_t)(dx) \;=\; \int (p_t(V \wedge n))(x)\, \tilde{\pi}(dx) \\
&\le\; \int n \wedge (p_t V)(x)\, \tilde{\pi}(dx) \;\le\; n \wedge C_3 + \mathrm{e}^{-\lambda t} \int n \wedge V(x)\, \tilde{\pi}(dx)
\end{aligned}$$

and thus $\int (V(x) \wedge n)\, \tilde{\pi}(dx) \;\le\; n \wedge C_3$. Using Fatou's lemma, we can conclude that $\tilde{\pi} \in \mathcal{P}_V(\mathcal{S})$. Now (0.3) implies that the measure π obtained above is actually the unique invariant measure. $\square$

Similar arguments can be found in [73, 74, 69].

Bibliography

[1] Madjid Amir. Sticky Brownian motion as the strong limit of a sequence of random walks. *Stochastic Process. Appl.*, 39(2):221–237, 1991.

[2] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[3] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[4] Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin. A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.*, 13:60–66, 2008.

[5] Dominique Bakry, Patrick Cattiaux, and Arnaud Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

[6] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[7] Richard F. Bass. A stochastic differential equation with a sticky point. *Electron. J. Probab.*, 19:no. 32, 22, 2014.

[8] Peter Baxendale. T. E. Harris's contributions to recurrent Markov processes and stochastic flows. *Ann. Probab.*, 39(2):417–428, 2011.

[9] V. I. Bogachev, A. I. Kirillov, and S. V. Shaposhnikov. The Kantorovich and variation distances between invariant measures of diffusions and nonlinear stationary Fokker-Planck-Kolmogorov equations. *Math. Notes*, 96(5-6):855–863, 2014.

[10] V. I. Bogachev, N. V. Krylov, and M. Röckner. On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions. *Comm. Partial Differential Equations*, 26(11-12):2037–2080, 2001.

[11] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov. Positive densities of transition probabilities of diffusion processes. *Theory of Probability & Its Applications*, 53(2):194–215, 2009.

[12] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov. Distances between transition probabilities of diffusions and applications to nonlinear Fokker-Planck-Kolmogorov equations. *J. Funct. Anal.*, 271(5):1262–1300, 2016.

[13] François Bolley, Ivan Gentil, and Arnaud Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.

[14] François Bolley, Ivan Gentil, and Arnaud Guillin. Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.*, 208(2):429–445, 2013.

[15] N. Bou-Rabee and J. M. Sanz-Serna. Randomized Hamiltonian Monte Carlo. *ArXiv e-print 1511.09382*, November 2015.

[16] J. Bricmont, A. Kupiainen, and R. Lefevere. Ergodicity of the 2D Navier-Stokes equations with random forcing. *Comm. Math. Phys.*, 224(1):65–81, 2001. Dedicated to Joel L. Lebowitz.

[17] J. Bricmont, A. Kupiainen, and R. Lefevere. Exponential mixing of the 2D stochastic Navier-Stokes dynamics. *Comm. Math. Phys.*, 230(1):87–132, 2002.

[18] O. A. Butkovsky. On ergodic properties of nonlinear Markov chains and stochastic McKean-Vlasov equations. *Theory Probab. Appl.*, 58(4):661–674, 2014.

[19] Oleg Butkovsky. Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.*, 24(2):526–552, 2014.

[20] José A. Carrillo, Robert J. McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[21] José A. Carrillo, Robert J. McCann, and Cédric Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[22] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields*, 140(1-2):19–40, 2008.

[23] Patrick Cattiaux, Arnaud Guillin, Feng-Yu Wang, and Liming Wu. Lyapunov conditions for super Poincaré inequalities. *J. Funct. Anal.*, 256(6):1821–1841, 2009.

[24] Patrick Cattiaux, Arnaud Guillin, and Li-Ming Wu. A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields*, 148(1-2):285–304, 2010.

[25] Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.

[26] Mu-Fa Chen. *Eigenvalues, inequalities, and ergodic theory*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005.

[27] Mu Fa Chen and Shao Fu Li. Coupling methods for multidimensional diffusion processes. *Ann. Probab.*, 17(1):151–177, 1989.

[28] Mu Fa Chen and Feng Yu Wang. Estimation of the first eigenvalue of second order elliptic operators. *J. Funct. Anal.*, 131(2):345–363, 1995.

[29] Mu-Fa Chen and Feng-Yu Wang. Estimation of spectral gap for elliptic operators. *Trans. Amer. Math. Soc.*, 349(3):1239–1267, 1997.

[30] R. Chitashvili. On the nonexistence of a strong solution in the boundary problem for a sticky Brownian motion. *Proc. A. Razmadze Math. Inst.*, 115:17–31, 1997.

[31] B. Cloez. Wasserstein decay of one dimensional jump-diffusions. *ArXiv e-print 1202.1259*, February 2012.

[32] M. Cranston. Gradient estimates on manifolds using coupling. *J. Funct. Anal.*, 99(1):110–124, 1991.

[33] G. Da Prato, A. Debussche, and L. Tubaro. Coupling for some partial differential equations driven by white noise. *Stochastic Process. Appl.*, 115(8):1384–1407, 2005.

[34] G. Da Prato and J. Zabczyk. *Ergodicity for infinite-dimensional systems*, volume 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[35] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 2014.

[36] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *ArXiv e-print 1412.7392*, December 2014.

[37] Anirban DasGupta. *Probability for statistics and machine learning*. Springer Texts in Statistics. Springer, New York, 2011. Fundamentals and advanced topics.

[38] M. Dashti and A. M. Stuart. The Bayesian Approach To Inverse Problems. *ArXiv e-print 1302.6989*, February 2013.

[39] P. L. Davies. Rates of convergence to the stationary distribution for k-dimensional diffusion processes. *J. Appl. Probab.*, 23(2):370–384, 1986.

[40] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.

[41] Wolfgang Doeblin. Sur deux problèmes de M. Kolmogoroff concernant les chaînes dénombrables. *Bull. Soc. Math. France*, 66:210–220, 1938.

[42] J. L. Doob. *Stochastic processes*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1990. Reprint of the 1953 original, A Wiley-Interscience Publication.

[43] Randal Douc, Gersende Fort, and Arnaud Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3):897–923, 2009.

[44] Randal Douc, Gersende Fort, Eric Moulines, and Philippe Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.

[45] A. Durmus and E. Moulines. Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm. *ArXiv e-print 1507.05021*, July 2015.

[46] A. Durmus and E. Moulines. Sampling from strongly log-concave distributions with the Unadjusted Langevin Algorithm. *ArXiv e-print 1605.01559*, May 2016.

[47] A. Eberle. Quantitative contraction rates for Markov chains on continuous state spaces. *preprint*, 2016.

[48] A. Eberle, A. Guillin, and R. Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. *ArXiv e-print 1606.06012*, June 2016.

[49] A. Eberle and R. Zimmer. Sticky couplings of multidimensional diffusions with different drifts. *ArXiv e-print 1612.06125*, December 2016.

[50] Andreas Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20):1101–1104, 2011.

[51] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields*, pages 1–36. Published online (DOI) 10.1007/s00440–015–0673–1, 2015.

[52] Andreas Eberle. Markov processes. *Lecture notes, Univ. Bonn. http://wt.iam.uni-bonn.de/eberle*, 2016.

[53] Albert Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, 1905.

[54] Hans-Jürgen Engelbert and Goran Peskir. Stochastic differential equations for sticky Brownian motion. *Stochastics*, 86(6):993–1021, 2014.

[55] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[56] William Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77:1–31, 1954.

[57] William Feller. The general diffusion operator and positivity preserving semigroups in one dimension. *Ann. of Math. (2)*, 60:417–436, 1954.

[58] D. Ferré, L. Hervé, and J. Ledoux. Regular perturbation of V-geometrically ergodic Markov chains. *J. Appl. Probab.*, 50(1):184–194, 2013.

[59] G. Fort and G. O. Roberts. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.*, 15(2):1565–1589, 2005.

[60] L. Gawarecki and V. Mandrekar. *Stochastic differential equations in infinite dimensions with applications to stochastic partial differential equations*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[61] Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[62] Carl Graham. The martingale problem with sticky reflection conditions, and a system of particles interacting at the boundary. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(1):45–72, 1988.

[63] M. Grothaus and R. Voßhall. Construction and analysis of sticky reflected diffusions. *ArXiv e-print 1412.3975*, December 2014.

[64] M. Grothaus and R. Voßhall. Strong Feller properties and uniqueness of sticky reflected distorted Brownian motion. *ArXiv e-print 1410.6040*, October 2014.

[65] M. Hairer and J. C. Mattingly. Ergodic properties of highly degenerate 2D stochastic Navier-Stokes equations. *C. R. Math. Acad. Sci. Paris*, 339(12):879–882, 2004.

[66] M. Hairer and J. C. Mattingly. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)*, 164(3):993–1032, 2006.

[67] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.

[68] M. Hairer and J. C. Mattingly. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.*, 16:no. 23, 658–738, 2011.

[69] M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011.

[70] M. Hairer, A: M. Stuart, and S. J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.

[71] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.*, 17(5-6):1657–1706, 2007.

[72] Martin Hairer. Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probab. Theory Related Fields*, 124(3):345–380, 2002.

[73] Martin Hairer. Ergodic properties of markov processes. *Lecture notes, Univ. Warwick. http://www.hairer.org/notes/Markov.pdf*, 2006.

[74] Martin Hairer. Convergence of Markov processes. *Lecture notes, Univ. Warwick. http://www.hairer.org/notes/Convergence.pdf*, 2016.

[75] Martin Hairer and Jonathan C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.

[76] H. Hajri and M. Arnaudon. Application of stochastic flows to the sticky brownian motion equation. *ArXiv e-print 1603.07456*, March 2016.

[77] T. E. Harris. The existence of stationary measures for certain Markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II*, pages 113–124. University of California Press, Berkeley and Los Angeles, 1956.

[78] J. Michael Harrison and Austin J. Lemoine. Sticky Brownian motion as the limit of storage processes. *J. Appl. Probab.*, 18(1):216–226, 1981.

[79] R. Z. Has′minskiĭ. Ergodic properties of recurrent diffusion processes and stabilization of the solution of the Cauchy problem for parabolic equations. *Teor. Verojatnost. i Primenen.*, 5:196–214, 1960.

[80] Martina Hofmanová and Jan Seidler. On weak solutions of stochastic differential equations. *Stoch. Anal. Appl.*, 30(1):100–121, 2012.

[81] Martina Hofmanová and Jan Seidler. On weak solutions of stochastic differential equations II. *Stoch. Anal. Appl.*, 31(4):663–670, 2013.

[82] Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.*, 46(5-6):1159–1194, 1987.

[83] Zhenting Hou, Yuanyuan Liu, and Hanjun Zhang. Subgeometric rates of convergence for a class of continuous-time Markov process. *J. Appl. Probab.*, 42(3):698–712, 2005.

[84] Christopher John Howitt. Stochastic flows and sticky brownian motion. *PhD Thesis, University of Warwick*, 2007.

[85] Nobuyuki Ikeda. On the construction of two-dimensional diffusion processes satisfying Wentzell's boundary conditions and its application to boundary value problems. *Mem. Coll. Sci. Univ. Kyoto Ser. A Math.*, 33:367–427, 1960/1961.

[86] Nobuyuki Ikeda and Shinzo Watanabe. A comparison theorem for solutions of stochastic differential equations and its applications. *Osaka J. Math.*, 14(3):619–633, 1977.

[87] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989.

[88] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[89] Aldéric Joulin. Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli*, 13(3):782–798, 2007.

[90] Aldéric Joulin. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli*, 15(2):532–549, 2009.

[91] Aldéric Joulin and Yann Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.

[92] Yu. M. Kabanov, R. Sh. Liptser, and A. N. Shiryaev. On the variation distance for probability measures defined on a filtered space. *Probab. Theory Relat. Fields*, 71(1):19–35, 1986.

[93] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[94] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[95] Wilfrid S. Kendall. Coupled Brownian motions and partial domain monotonicity for the Neumann heat kernel. *J. Funct. Anal.*, 86(2):226–236, 1989.

[96] Rafail Khasminskii. *Stochastic stability of differential equations*, volume 66 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, second edition, 2012. With contributions by G. N. Milstein and M. B. Nevelson.

[97] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.

[98] Tomasz Komorowski and Anna Walczuk. Central limit theorem for markov processes with spectral gap in the wasserstein metric. *Stochastic Processes and their Applications*, 122(5):2155–2184, 2012.

[99] S. Kuksin and A. Shirikyan. Stochastic dissipative PDEs and Gibbs measures. *Comm. Math. Phys.*, 213(2):291–330, 2000.

[100] S. Kuksin and A. Shirikyan. A coupling approach to randomly forced nonlinear PDE's. I. *Comm. Math. Phys.*, 221(2):351–366, 2001.

[101] Alexei Kulik. *Introduction to Ergodic rates for Markov chains and processes: with applications to limit theorems*, volume 2. Universitätsverlag Potsdam, 2015.

[102] Paul Langevin. Sur la théorie du mouvement brownien. *CR Acad. Sci. Paris*, 146(530-533):530, 1908.

[103] G. Leha and G. Ritter. On diffusion processes and their semigroups in Hilbert spaces with an application to interacting stochastic systems. *Ann. Probab.*, 12(4):1077–1112, 1984.

[104] F. Liese. Hellinger integrals of diffusion processes. *Statistics*, 17(1):63–78, 1986.

[105] F. Liese and W. Schmidt. On the strong convergence, contiguity and entire separation of diffusion processes. *Stochastics Stochastics Rep.*, 50(3-4):185–203, 1994.

[106] Torgny Lindvall. On coupling of diffusion processes. *J. Appl. Probab.*, 20(1):82–93, 1983.

[107] Torgny Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.

[108] M. Majka. Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes. *ArXiv e-print 1509.08816*, September 2015.

[109] Florent Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560, 2003.

[110] M. N. Malyshkin. Subexponential estimates for the rate of convergence to the invariant measure for stochastic differential equations. *Teor. Veroyatnost. i Primenen.*, 45(3):489–504, 2000.

[111] Petr Mandl. *Analytical treatment of one-dimensional Markov processes*. Die Grundlehren der mathematischen Wissenschaften, Band 151. Academia Publishing House of the Czechoslovak Academy of Sciences, Prague; Springer-Verlag New York Inc., New York, 1968.

[112] O. Manita. Estimates for Kantorovich functionals between solutions to Fokker – Planck – Kolmogorov equations with dissipative drifts. *ArXiv e-print 1507.04014*, July 2015.

[113] N. Masmoudi and L.-S. Young. Ergodic theory of infinite dimensional systems with applications to dissipative parabolic PDEs. *Comm. Math. Phys.*, 227(3):461–481, 2002.

[114] J. C. Mattingly, N. S. Pillai, and A. M. Stuart. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.*, 22(3):881–930, 2012.

[115] Jonathan C. Mattingly. Exponential convergence for the stochastically forced Navier-Stokes equations and other partially dissipative dynamics. *Comm. Math. Phys.*, 230(3):421–462, 2002.

[116] H. P. McKean, Jr. A. Skorohod's stochastic integral equation for a reflecting barrier diffusion. *J. Math. Kyoto Univ.*, 3:85–88, 1963.

[117] Sylvie Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[118] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[119] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.*, 24(3):542–574, 1992.

[120] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.

[121] A. Yu. Mitrophanov. Stability and exponential convergence of continuous-time Markov chains. *J. Appl. Probab.*, 40(4):970–979, 2003.

[122] Esa Nummelin and Pekka Tuominen. The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process. Appl.*, 15(3):295–311, 1983.

[123] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.

[124] Yann Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, volume 57 of *Adv. Stud. Pure Math.*, pages 343–381. Math. Soc. Japan, Tokyo, 2010.

[125] Steven Orey. *Lecture notes on limit theorems for Markov chain transition probabilities.* Van Nostrand Reinhold Co., London-New York-Toronto, Ont., 1971. Van Nostrand Reinhold Mathematical Studies, No. 34.

[126] Grigorios A. Pavliotis. *The Langevin Equation*, pages 181–233. Springer New York, New York, NY, 2014.

[127] Goran Peskir. On boundary behaviour of one-dimensional diffusions: From brown to feller and beyond. In *William Feller - Selected Papers II.* Springer, Berlin, 2015.

[128] N. S. Pillai and A. Smith. Ergodicity of Approximate MCMC Chains with Applications to Large Data Sets. *ArXiv e-print 1405.0182*, May 2014.

[129] Enrico Priola and Feng-Yu Wang. Gradient estimates for diffusion semigroups with singular coefficients. *J. Funct. Anal.*, 236(1):244–264, 2006.

[130] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[131] Miklós Z. Rácz and Mykhaylo Shkolnikov. Multidimensional sticky Brownian motions as limits of exclusion processes. *Ann. Appl. Probab.*, 25(3):1155–1188, 2015.

[132] M. Reed and B. Simon. *Methods of modern mathematical physics. I.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis.

[133] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999.

[134] G. O. Roberts, J. S. Rosenthal, and P. O. Schwartz. Convergence properties of perturbed Markov chains. *J. Appl. Probab.*, 35(1):1–11, 1998.

[135] Gareth O. Roberts and Jeffrey S. Rosenthal. Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.*, 1:no. 9, approx. 21 pp. (electronic), 1996.

[136] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998.

[137] D. Rudolf and N. Schweizer. Perturbation theory for Markov chains via Wasserstein distance. *ArXiv e-print 1503.04123*, March 2015.

[138] T. Shardlow and A. M. Stuart. A perturbation theory for ergodic Markov chains and application to numerical approximations. *SIAM J. Numer. Anal.*, 37(4):1120–1137, 2000.

[139] Anatoliy V Skorokhod. Stochastic equations for diffusion processes in a bounded region. *Theory of Probability & Its Applications*, 6(3):264–274, 1961.

[140] Anatoliy V Skorokhod. Stochastic equations for diffusion processes in a bounded region. ii. *Theory of Probability & Its Applications*, 7(1):3–23, 1962.

[141] Daniel W. Stroock and S. R. S. Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971.

[142] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.

[143] A. M. Stuart, Jochen Voss, and Petter Wiberg. Fast communication conditional path sampling of SDEs and the Langevin MCMC method. *Commun. Math. Sci.*, 2(4):685–697, 2004.

[144] Alain-Sol Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[145] Pekka Tuominen and Richard L. Tweedie. Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. in Appl. Probab.*, 26(3):775–798, 1994.

[146] A. Yu. Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.*, 70(1):115–127, 1997.

[147] A. Yu. Veretennikov and A. M. Kulik. The extended Poisson equation for weakly ergodic Markov processes. *Teor. Ĭmovīr. Mat. Stat.*, (85):22–38, 2011.

[148] Cédric Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[149] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[150] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[151] F.-Y. Wang. Asymptotic couplings by reflection and applications for nonlinear monotone SPDES. *Nonlinear Anal.*, 117:169–188, 2015.

[152] Feng-Yu Wang. Functional inequalities for the decay of sub-Markov semigroups. *Potential Anal.*, 18(1):1–23, 2003.

[153] Jian Wang. L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli*, 22(3):1598–1616, 2016.

[154] Jonathan Warren. Branching processes, the Ray-Knight theorem, and sticky Brownian motion. In *Séminaire de Probabilités, XXXI*, volume 1655 of *Lecture Notes in Math.*, pages 1–15. Springer, Berlin, 1997.

[155] Shinzo Watanabe. On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. *J. Math. Kyoto Univ.*, 11:169–180, 1971.

[156] Shinzo Watanabe. On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. II. *J. Math. Kyoto Univ.*, 11:545–551, 1971.

[157] E. Weinan and D. Liu. Gibbsian dynamics and invariant measures for stochastic dissipative PDEs. *J. Statist. Phys.*, 108(5-6):1125–1156, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[158] E. Weinan, J. C. Mattingly, and Ya. Sinai. Gibbsian dynamics and ergodicity for the stochastically forced Navier-Stokes equation. *Comm. Math. Phys.*, 224(1):83–106, 2001. Dedicated to Joel L. Lebowitz.

[159] R. Zimmer. Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. *ArXiv e-print 1605.07863*, May 2016.