Centre de Mathématiques Appliquées, École Polytechnique

Intership report

Coupling by reflection and exponential ergodicity of stochastic differential equations

Anna Bahrii

supervised by
Professor Giovanni Conforti

September 16, 2022

Abstract

One of the most important questions in the studying stochastic differential equations involves studying the long-time behavior of a system. We want to know if the solution of the SDE converges to equilibrium and if yes, then the question of convergence speed arises. The objective of this report is to show that coupling by reflection applied to a suitable modification of the Wasserstein distance leads to exponential convergence of a system for any initial distribution under mild assumptions on the drift field.

Contents

1 Introduction

The goal of this report is to provide a consistent approach for addressing questions about exponential convergence of random dynamical systems. Stochastic differential equations can be viewed as differential equations in which some randomness is allowed. SDEs are widely used for describing various processes in natural sciences. Long-time properties of such equations are of special interest, in particular, we are interested in the exponential stability of stochastic processes. The main focus of the report is explaining how coupling by reflection w.r.t. Wasserstein distance with a modified underlying distance function allows us to profit from randomness in dynamical systems and get exponential convergence to equilibrium. The report is organised as follows:

In section 2 we give some introduction to stochastic processes. We prove the existence of Brownian motion and list some properties that naturally lead us to stochastic calculus. We describe some core objects of Itô stochastic calculus, including Itô formula that plays a key role in the analysis of coupling by reflection.

In section 3 we provide a step-by-step approach for studying the stability of dynamical systems. First, we explore the long-time behaviour of deterministic systems. Then, we arrive at the question of stability of randomly perturbed systems, and we describe two types of coupling: synchronous coupling and coupling by reflection. The idea of reflection coupling together with the change of the underlying distance function in the Wasserstein metric is described carefully. A special effort has been put into providing detailed calculations from the paper [2].

2 Brownian motion & Stochastic differential equations

In this section, we provide some preliminaries that are required for the understanding of the main results. In particular, we introduce Brownian motion and stochastic differential equations. The material here is mainly adapted from [4] and [3].

2.1 Brownian motion

Definition 2.1.1. *A* ***stochastic process*** *is a parameterised collection of random variables* $\{X_t\}_{t\geq 0}$*, that are defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with values in* $\mathbb{R}^d$*.*

Definition 2.1.2. ***Brownian motion*** *is a stochastic process* $(B_t)_{t\geq 0}$ *that is defined by the following properties:*

1. $B_0 = x,\ x \in \mathbb{R}$,
2. *The increments are independent, i.e for* $\forall t, s, r \geq 0$ *s.t.* $t < s < r$ $B_r - B_s \perp\!\!\!\perp B_s - B_t$,
3. *The increments* $B_t - B_s$ *are normally distributed with mean zero and variance* $t - s$ $\forall t, s \geq 0$,
4. $(B_t)_{t\geq 0}$ *has almost surely continuous paths.*

If $x = 0$*, then* $(B_t)_{t\geq 0}$ *is called a* ***standard Brownian motion****.*

Definition 2.1.3. *A d-dimensional stochastic process* $B_t = (B_t^1, ..., B_t^d)$*, where* $B_t^i, 1 \leq i \leq d$*, is a one-dimensional Brownian motion, and* $B_t^i \perp\!\!\!\perp B_t^j$*,* $1 \leq i < j \leq d$*, is called a* ***d-dimensional Brownian motion****.*

Theorem 2.1.1 (Wiener 1923)**.** *The standard Brownian motion exists.*

(Proof sketch). The following proof of the existence is known as Paul Lévy's construction of Brownian motion. The idea is to construct Brownian motion as a uniform limit of continuous functions that are built on the set of dyadic points $\mathcal{D} = \cup_{n\in\mathbb{N}}\mathcal{D}_n,\ \mathcal{D}_n = \{\frac{k}{2^n} : 0 \leq k \leq 2^n\}$. It suffices to apply this approach for the interval $[0, 1]$ and the Brownian motion on the interval $[0, \infty)$ can be properly constructed "by gluing the parts".

We start with defining a collection of i.i.d. random variables $\{Z_t : t \in \mathcal{D}\}$ which are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the standard normal distribution. Then, we define random variables B_k, $k \in \mathcal{D}_n$ for $n \in \mathbb{N}$ such that

1. $B_0 := 0,\ B_1 := Z_1$ for $n = 0$,
2. The increment $B_t - B_s$ are normally distributed with mean zero and variance $t - s$ for $\forall t, s \in \mathcal{D}_n$,
3. The increments are independent, i.e. $B_r - B_s \perp\!\!\!\perp B_s - B_t$ for $\forall t < s < r \in \mathcal{D}_n$,
4. The vectors $(B_k)_{k\in\mathcal{D}_n}$ and $(Z_k)_{k\in\mathcal{D}\setminus\mathcal{D}_n}$ are independent.

We continue to build these random variables by induction. Thus, for $k \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ we have

$$B_k = \frac{B_{k-2^{-n}} + B_{k+2^{-n}}}{2} + \frac{Z_k}{2^{(n+1)/2}}.$$

It can be verified that B_k satisfies the specified properties. Now, we interpolate between the values on the dyadic points. We define the function F_t^n for $n = 0$

$$F_t^0 = \begin{cases} Z_1 \text{ for } t = 0, \\ 0 \text{ for } t = 1, \\ \textit{linear} \text{ for } 0 \le t \le 1, \end{cases}$$

and for $n \ge 1$

$$F_t^n = \begin{cases} 2^{-(n+1)/2} Z_t \text{ for } t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}, \\ 0 \text{ for } t \in \mathcal{D}_{n-1}, \\ \textit{linear} \text{ between consecutive points in } \mathcal{D}_n. \end{cases}$$

The functions F_t^n are continuous on $[0,1]$ and it can be shown by induction on $n \ge 0$ that

$$B_k = \sum_{i=0}^{n} F_k^i = \sum_{i=0}^{\infty} F_k^i, \quad n \in \mathbb{N},\ k \in \mathcal{D}_n.$$

Using Borel-Cantelli lemma we can find a bound for $\|F^n\|_\infty$ and, hence, show that , almost surely, this series is uniformly convergent on $[0,1]$ and the limit is exactly Brownian motion. We denote the limit of the series as $\{B_t : t \in [0,1]\}$. The last step is to show that the limit has independent increments with the right distribution. Indeed, it follows from the properties of B on the dense set $\mathcal{D} \subset [0,1]$ and the fact that it has continuous paths.

To extend the construction Brownian motion $B : [0,1] \to \mathbb{R}$ to $[0,\infty)$ we consider a sequence of independent random variables $B^0, B^1, ...$ that are continuous on $[0,1]$ and define $\{B_t : t \ge 0\}$ as

$$B_t = B_{t-\lfloor t \rfloor}^{\lfloor t \rfloor} + \sum_{i=0}^{\lfloor t \rfloor -1} B_1^i, \quad t \ge 0.$$

□

Remark 2.1.1. *Another approach for constructing Brownian motion is viewing it as a scaling limit of a random walk. This result is known as a Donsker's theorem and is proven in [4].*

The following theorem describes one of the pathological properties of Brownian motion.

Theorem 2.1.2. *Almost surely, Brownian motion is nowhere differentiable. Moreover, the upper and lower right derivatives of Brownian motion are equal to $+\infty$ and $-\infty$ respectively:*

$$\limsup_{h\to 0} = \frac{B_{t+h} - B_t}{h} = +\infty,$$

$$\liminf_{h\to 0} = \frac{B_{t+h} - B_t}{h} = -\infty,$$

almost surely.

It follows from theorem 2.1.2 that the paths of Brownian motion fail to satisfy the requirements for applying the rules of classical analysis. That is why a new approach for stochastic processes is needed. Here, we will be using Itô calculus, the central concept of which are Itô integral and Itô formula. This integral can be viewed as a stochastic generalisation of the Riemann–Stieltjes integral in classical analysis.

Definition 2.1.4. *Let $\mathcal{V} = \mathcal{V}(S,T)$ be the class of functions $f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$ such that*

- $(t,\omega) \longmapsto f(t,\omega)$ *is $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B}$ is Borel σ-algebra on $\mathbb{R}^+$,*
- $f(t,\omega)$ *is $\mathcal{F}_t$-adapted for $\forall t \geq 0$,*
- $\mathbb{E}[\int_S^T f(t,\omega)^2 dt] < \infty, \ 0 < S < T.$

Definition 2.1.5. *A function $\phi \in \mathcal{V}(S,T)$ is called* ***elementary*** *if it has the form*

$$\phi(t,\omega) = \sum_{j=0}^{d} e_j(\omega)\mathcal{X}_{[t_j,t_{j+1})}(t), \quad S = t_0 < ... < t_d = T.$$

Definition 2.1.6. *Let $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ be a filtered probability space and $(B_t)_{t\geq 0}$ be a $\mathcal{F}_t$-adapted Brownian motion. Let $f \in \mathcal{V}(S,t)$. Then the* ***Itô integral*** *of the function f from S to T is defined by*

$$\int_S^T f(t,\omega)dB_t = \lim_{n\to\infty} \phi_n(t,\omega)dB_t,$$

where $\{\phi_n\}_{n\in\mathbb{N}}$ is a sequence of elementary functions such that

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (f(t,\omega) - \phi_n(t,\omega))^2 dt\Big] = 0.$$

It is important to note that in the context of Itô integrals there is no differentiation theory, only integration theory. But it is possible to establish an Itô integral version of the chain rule, which is called the Itô formula.

Definition 2.1.7 (Multi-dimensional Itô process)**.** *Let $B_t = (B_t^1, ..., B_t^m)$ be a m-dimensional Brownian motion on $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ and adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. If $u(t,\omega) = (u_1(t,\omega), ..., u_d(t,\omega))^T$ and $v(t,\omega) = (v_{ij}) \in \mathbb{R}^{d\times m}$ are such that $u_i, v_i, i = 1, ..., d$, are adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ and $\mathbb{E}[\int_0^t |v|^2 ds] = \mathbb{E}[\int_0^t \sum_{ij} |v_{ij}|^2 ds] < \infty$, then*

$$X_t^i = X^0 + \int\limits_0^t u_i(s,\omega)ds + \int\limits_0^t v_{ij}(s,\omega)dB_s^0 + ... + \int\limits_0^t v_{ij}(s,\omega)dB_s^m,$$
$$1 \le i \le m,\ 1 \le j \le d.$$

*is a d-**dimensional Itô process**. Or, in matrix notation*

$$dX_t = udt + vdB_t.$$

Theorem 2.1.3 (Multi-dimensional Itô formula)**.** *Let*

$$dX_t = udt + vdB_t$$

be a d-dimensional Itô process. Let $g \in C^2([0,\infty) \times \mathbb{R}^d, \mathbb{R}^n), n \in \mathbb{N}$. Then $g(t, X_t)$ is an Itô process and

$$dg(t,X_t) = \frac{\partial g}{\partial t}(t,X_t)dt + \nabla_X g(t,X_t)^T dX_t + \frac{1}{2}\operatorname{Tr}\left(v^T \nabla_X^2 g(t,X_t) v\right)dt. \tag{2.1}$$

2.2 Stochastic differential equations

Consider the differential equation of type

$$\frac{dX_t}{dt} = r(t)X_t,$$

where $r(t), t \ge 0$, is some given deterministic function for which the solution of the equation exists. Stochastic differential equation arises naturally when the coefficient $r(t)$ is perturbed by some random noise:

$$\frac{dX_t}{dt} = (r(t) + \text{"noise"})X_t.$$

Considering the "noise" term as a Brownian motion, we obtain the following *stochastic differential equation*:

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \tag{2.2}$$

where $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ is the *drift vector field* of the system, $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ is a *diffusion matrix* and $(B_t)_{t\in[0,T]}$ is a d-dimensional Brownian motion. A solution to 2.2 is known as a *diffusion process*. To convince ourselves that this equation is well-defined, we consider a discrete version of 2.2 for $0 = t_0 < t_1 < ... < t_m = t$:

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)\Delta B_k, \quad 0 \le k \le m,$$

where

$$X_k = X_{t_k},\ B_k = B_{t_k},\ \Delta t_k = t_{k+1} - t_k,\ \Delta B_k = B_k \Delta t_k.$$

For $k \le m$ we have

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_k + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta B_j.$$

If the limit of the right hand side exists when $\Delta t_j \to 0$, then by applying the integration notation we will obtain

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \ t \geq 0. \tag{2.3}$$

But if the function $\sigma(t, X_t(\omega))$ satisfies the conditions in the definition 2.1.6 in notions $f(t, \omega) = \sigma(t, X_t(\omega))$ and $S = 0, T = t$, then the term $\int_0^t \sigma(s, X_s)dB_s$ is an Itô integral and thus the equation 2.3 is well-defined.

Remark 2.2.1. *Equation 2.2 can intuitively be understood as follows: a system follows a deterministic drift defined by the function $b(\cdot)$ and then gets a random movement that is described by the the term $\sigma(t, \cdot)dB_t$, $t \geq 0$.*

Consider now the *initial value problem*:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = \xi. \tag{2.4}$$

Let Brownian motion $(B_t)_{t\geq 0}$ and random vector ξ in 2.4 are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_t = \sigma(\xi, (B_s)_{s\leq t})$ be the *natural filtration* generated by $(B_s)_{s\geq 0}$ and ξ by the time $t \geq 0$.

Definition 2.2.1 (Strong solution of SDE)**.** *A stochastic process $(X_t)_{t\geq 0}$ with continuous paths is called a **strong solution** of 2.4 for the given Brownian motion $(B_t)_{t\geq 0}$ and the initial condition ξ if*

1. $\mathbb{P}(X_0 = \xi) = 1$,
2. *X_t is $\mathcal{F}_t$-adapted for $\forall t \geq 0$,*
3. *The integral equation 2.3 holds $\mathbb{P}$-almost surely for $\forall t \geq 0$.*

Theorem 2.2.1 (Existence and uniqueness of the solution of SDE)**.** *Let $T > 0$ and $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable functions such that*

$$\exists D \in \mathbb{R} : \ |b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y| \quad \forall x, y \in \mathbb{R}^n, \ \forall t \in [0, T] \tag{2.5}$$

and

$$\exists C \in \mathbb{R} : \ |b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|) \quad \forall x \in \mathbb{R}^n, \ \forall t \in [0, T], \tag{2.6}$$

for $t \in [0, T]$. Let ξ be a random variable that is independent from the σ-algebra $\mathcal{F}_\infty$ generated by $B_s, s \geq 0$, and such that $\mathbb{E}|\xi|^2 < \infty$. Then the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T,$$

with the initial condition

$$X_0 = \xi$$

has a unique t-continuous strong solution $X_t(\omega)$ such that $X_t(\omega)$ is adapted to the filtration $\mathcal{F}_t$ generated by ξ and $(B_s)_{s\leq t}$, and $\mathbb{E}\Big[\int_0^T |X_t|^2 dt\Big] < \infty$.

Idea of proof. The uniqueness of solutions of the SDE follows from the Itô isometry and the Lipschitz condition 2.5, and as usually is proved from the opposite – by assuming that there exist two distinct solutions and arriving at a contradiction.

The proof of existence of solutions is quite similar to the existence proof for ordinary differential equations: with the use of Picard iteration, we define an approximating sequence, show that the sequence converges to a limit and that the limit solves the SDE.

The full proof of the theorem can be found in [3]. □

3 Long-time behaviour of stochastic differential equations

The goal of this section is to explore the long-time stability of the following SDE

$$dX_t = b(X_t)dt + \sigma dB_t, \quad X_0 \sim \mu, \tag{3.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a given drift vector, σ is a constant diffusion matrix with values in $\mathbb{R}^{d\times d}$, μ is a probability distribution and $(B_t)_{t\geq 0}$ is Brownian motion. Studying the long-time stability of the SDE means addressing of the following questions:

– What conditions on the drift field $b(\cdot)$ guarantee the existence of a stationary probability measure μ_∞ such that $\mathcal{L}(X_t) \longrightarrow \mu_\infty$ as $t \to \infty$?

– If such a measure exists, then is the convergence exponentially fast?

– What is the right metric to choose on the space of probability measures to answer the questions about convergence?

To make our approach to addressing mentioned questions clear and more consistent, we start with studying the long-time behaviour of ordinary differential equations.

3.1 Ordinary differential equations

As an appetiser, consider the ODE with drift vector $b(\cdot)$ of gradient type:

$$\dot{X}_t = -\frac{1}{2}\nabla U(X_t), \quad X_0 = x, \tag{3.2}$$

where $U \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R})$ is a *potential function*, $x \in \mathbb{R}^d$. $X_t \in \mathrm{C}^1(\mathbb{R}, \mathbb{R}^d), t \geq 0$ is known as the *gradient flow* of the potential function U started at point x. It can be seen as a deterministic counterpart of 3.1 when $b(X_t) = -\frac{1}{2}\nabla U(X_t)$ and $\sigma = 0$. We assume that U satisfies the conditions for existence of a unique solution for 3.2. The following statements demonstrate that having some specific conditions on U, we can make two solutions of 3.2 with different initial conditions meet exponentially fast.

Proposition 3.1.1. *For $\forall K > 0$ the inequality*

$$\forall x, y \in \mathbb{R}^d \quad (x - y) \cdot (\nabla U(x) - \nabla U(y)) \geq K|x - y|^2 \tag{3.3}$$

holds if and only if

$$\forall x, y \in \mathbb{R}^d \quad |X_t - Y_t| \leq e^{-\frac{Kt}{2}}|x - y|, \tag{3.4}$$

where $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are strong solutions of 3.2 with initial conditions $X_0 = x$ and $Y_0 = y$.

Proof. ($\Rightarrow$)

To prove inequality 3.4 we can use Grönwall's inequality, i.e. if we can show that for a well-chosen function $v \in \mathrm{C}^1(\mathbb{R}, \mathbb{R})$

$$v'(t) \leq \beta(t)v(t),$$

then we will get

$$v(t) \leq v(0)e^{\int_0^t \beta(s)ds}.$$

Let $v(t) = |X_t - Y_t|^2$. Then $v'(t) = 2(X_t - Y_t) \cdot (\dot{X}_t - \dot{Y}_t)$. From 3.2 and 3.3 we have $2(X_t - Y_t) \cdot (\dot{X}_t - \dot{Y}_t) \leq -K|X_t - Y_t|^2$ and by the Grönwall's inequality we have

$$|X_t - Y_t|^2 \leq e^{-\frac{Kt}{2}}|x-y|^2 \Rightarrow |X_t - Y_t| \leq e^{-\frac{Kt}{2}}|x-y|.$$

($\Leftarrow$)

$$\frac{d(|X_t - Y_t|)}{dt} \leq \frac{d(e^{-\frac{Kt}{2}}|x-y|)}{dt} \implies \frac{(X_t - Y_t) \cdot (\dot{X}_t - \dot{Y}_t)}{|X_t - Y_t|} \leq -\frac{K}{2}e^{-\frac{Kt}{2}}|x-y|,$$
$$(X_t - Y_t) \cdot (\nabla U(X_t) - \nabla U(Y_t)) \geq Ke^{-\frac{Kt}{2}}|x-y||X_t - Y_t| \geq K|X_t - Y_t|^2.$$

□

The next proposition demonstrates that condition 3.3 is equivalent to the strong convexity of the potential function.

Proposition 3.1.2. *For $\forall K > 0$ the inequality*

$$\forall x, y \in \mathbb{R}^d \quad (x-y) \cdot (\nabla U(x) - \nabla U(y)) \geq K|x-y|^2 \tag{3.5}$$

holds if and only if U is a K-convex function, i.e.

$$\forall x, v \in \mathbb{R}^d \quad v^T \nabla^2 U(x) v \geq K|v|^2, \tag{3.6}$$

Proof. By Taylor theorem there exists $z \in \{tx + (1-t)y, t \in [0,1]\}$ such that

$$U(x) = U(y) + \nabla U(y)^T(x-y) + \frac{1}{2}(x-y)^T \nabla^2 U(z)(x-y)$$
$$\implies \nabla U(y)^T(x-y) = U(x) - U(y) - \frac{1}{2}(x-y)^T \nabla^2 U(z)(x-y). \tag{3.7}$$

Using the fact that

$$U(y) \geq U(x) + \nabla U(x)^T(y-x) + \frac{K}{2}|y-x|^2, \tag{3.8}$$

and equation 3.7 we get

$$
\begin{aligned}
\nabla U(x)^T(x-y) - \nabla U(y)^T(x-y) \geq U(x) - U(y) + \frac{K}{2}|x-y|^2 \\
- \nabla U(y)^T(x-y) = U(x) - U(y) + \frac{K}{2}|x-y|^2 - U(x) + U(y) \\
+ \frac{1}{2}(x-y)^T \nabla^2 U(z)(x-y) = \frac{K}{2}|x-y|^2 \\
+ \frac{1}{2}(x-y)^T \nabla^2 U(z)(x-y) \geq K|x-y|^2.
\end{aligned}
$$

Therefore, setting $v := x - y$ we obtain

$$
v^T \nabla^2 U(z) v \geq K v^2 \quad \forall z, v \in \mathbb{R}^d.
$$

□

Propositions 3.1.1 and 3.1.2 imply an important corollary.

Corollary 3.1.1. *If the condition of the 3.1.1 or 3.1.2 holds, then $\exists! x_\infty \in \mathbb{R}^d$, called an attractor, such that $\forall x \in \mathbb{R}^d$*

$$
|X_t - x_\infty| \leq e^{-\frac{Kt}{2}} |x - x_\infty|. \tag{3.9}
$$

Proof. By 3.4 we see that X_t is a contraction mapping on the Euclidean space. Thus, by the Banach fixed point theorem, there exists a unique fixed point x_∞, and we get 3.9. □

These results show that any ODE with a K-convex potential function converges exponentially fast at rate K to the unique equilibrium point.

Example 3.1.1 (Counterexample)**.** *Consider the potential function $U(x) = |x|^4 - |x|^2,\ x \in \mathbb{R}$. Then $U''(x) = 2(6x^2-1)$, and if $x \in [-1,1]$, then for $\forall K \geq 0$ the inequality $2(x^2-1) \geq K$ cannot hold: for $x \in [-1,1]$ the function is concave. Moreover, if $x \leq -1$ and $y \geq 1$, then for the corresponding solutions the following inequality holds*

$$
|X_t - Y_t| \geq \sqrt{2} \quad \forall t \geq 0.
$$

Indeed, solving the ODE

$$
\frac{1}{x(x^2-1)} dX_t = -2dt \tag{3.10}
$$

we have that $X_t \geq -\frac{e^{2t}}{\sqrt{-1+2e^{4t}}}, x \leq -1$ and $Y_t \geq \frac{e^{2t}}{\sqrt{-1+2e^{4t}}}, y \geq 1$. Thus, $|X_t - Y_t| \geq |\frac{2e^{2t}}{\sqrt{-1+2e^{4t}}}| \xrightarrow[t\to\infty]{} \sqrt{2}$. That means that any two gradient flows started from the points $x \leq -1$ and $y \geq 1$ correspondingly will never meet.

In case when the drift vector is not of gradient type, the condition 3.3 is formulated as follows

$$
\exists K \geq 0 \text{ s.t. } (x-y)\cdot(b(x)-b(y)) \leq -\frac{K}{2}|x-y|^2 \quad \forall x, y \in \mathbb{R}^d. \tag{3.11}
$$

The question we are interested in is: is it possible to profit from the random noise added to the ODE 3.2 in such a way that exponential convergence becomes possible even when the potential function is not strongly convex? This question leads to analysing the behaviour of the SDE 3.1.

3.2 Stochastic differential equations

We denote the transition kernels of the stochastic process defined by 2.2 as $p_t(x, A) = \mathbb{P}[X_t \in A | X_0 = x],\ (x, A) \in \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$. With a given initial value $X_0 \sim \mu$ we understand μp_t as a probability measure of X_t at a given time $t \geq 0$:

$$\mu p_t(A) = \mathcal{L}(X_t)(A) = \mathbb{P}[X_t \in A | X_0 \sim \mu] \quad \forall A \in \mathcal{B}(\mathbb{R}^d). \tag{3.12}$$

Since our system is no longer deterministic but random, we will operate with the law of the solution from now on instead of looking at the deterministic position X_t. We first need to define the distance between two probability measures. Let $p \geq 1$ and $\mathcal{P}_p$ be the set of *Borel probability measures* on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\int_{\mathbb{R}^d} |x|^p \mu(dx)$ for $\forall \mu \in \mathcal{P}_p$.

Definition 3.2.1 (Set of couplings)**.** *Let $\mu, \nu \in \mathcal{P}_p$. Then the **set of couplings** of μ and ν is defined as*

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \ \pi(A \times \mathbb{R}^d) = \mu(A),\ \pi(\mathbb{R}^d \times A) = \nu(A) \\ \textit{for } \forall A \in \mathcal{B}(\mathbb{R}^d)\}.$$

In other words, a coupling represents a random vector (X, Y) such that $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$.

Definition 3.2.2 (Wasserstein distance)**.** *Let $p \geq 1$ and $\mu, \nu \in \mathcal{P}_p$. Let $d_f(x, y) = f(|x - y|),\ x, y \in \mathbb{R}^d$, be a distance function. The **Wasserstein distance** of order p is*

$$W_f^p(\mu, \nu) = \inf_{\eta \in \Pi(\mu,\nu)} \Big(\int d_f(x, y)^p \eta(dxdy) \Big)^{\frac{1}{p}} = \inf_{(X,Y): X \sim \mu, Y \sim \nu} \Big(\mathbb{E}[d_f(X, Y)^p] \Big)^{\frac{1}{p}}. \tag{3.13}$$

Proposition 3.2.1. *The Wasserstein distance W_f^p is a metric on $\mathcal{P}_p$:*

1. $W_f^p(\mu, \nu) = 0 \iff \mu = \nu \in \mathcal{P}_p$,
2. $W_f^p(\mu, \nu) = W_f^p(\nu, \mu)$ *for* $\forall \mu, \nu \in \mathcal{P}_p$,
3. $W_f^p(\mu_1, \mu_2) \leq W_f^p(\mu_1, \mu_3) + W_f^p(\mu_3, \mu_2)$ *for* $\forall \mu_1, \mu_2, \mu_3 \in \mathcal{P}_p$.

Next statement clarifies the relation between the convergence in Wasserstein distances and other types of convergence.

Proposition 3.2.2. *A sequence of probability measures $(\mu_n)_{n\geq 1} \in \mathcal{P}_p$ converges to $\mu \in \mathcal{P}_p$ if and only if $(\mu_n)_{n\geq 1} \in \mathcal{P}_p$ converges to μ weakly and $\int_{\mathbb{R}^d} |x|^p \mu_n(dx)$ converges to $\int_{\mathbb{R}^d} |x|^p \mu(dx)$.*

With this being said, we are now able to quantify how far on average from each other two solutions of the SDE are.

3.3 Synchronous coupling

How do we study the long-time behaviour of the SDE? Here a coupling of SDEs comes in help. Let Brownian motions $(B_t)_{t\geq 0}$, $(\check{B}_t)_{t\geq 0}$ and random vectors $\xi_X \sim \mu, \xi_Y \sim \nu$ be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 3.3.1 (Coupling). *A stochastic process $(X_t, Y_t)_{t\geq 0} \in \mathbb{R}^{2d}$ with initial distribution $(X_0, Y_0) \sim \eta$ is called a* ***coupling*** *of two solutions of the SDE 3.1 if $(X_t)_{t\geq 0}$ is a strong solution of 2.2 for the Brownian motion $(B_t)_{t\geq 0}$ and the initial distribution μ and $(Y_t)_{t\geq 0}$ is a strong solution of 2.2 for the Brownian motion $(\check{B}_t)_{t\geq 0}$ and the initial distribution ν and $\eta \in \Pi(\mu, \nu)$.*

In this subsection we will use p-Wasserstein distance with the Euclidean underlying distance function, i.e. $d_f(x, y) = |x - y|$, and we denote it as follows

$$W^p := \inf_{\eta \in \Pi(\mu,\nu)} \Big(\int |x - y|^p \eta(dxdy) \Big)^{\frac{1}{p}}.$$

If $(X_t, Y_t)_{t\geq 0}$ is a coupling of two solutions of 2.2, then $\mathcal{L}(X_s, Y_s) \in \Pi(\mathcal{L}(X_s), \mathcal{L}(Y_s))$ for $\forall s \geq 0$. In order to minimize the distance between the laws of two solutions, we have to couple the corresponding SDEs in such a way that $\mathbb{E}[|X_t - Y_t|]$ is the smallest possible. We will introduce two approaches for coupling and study the long-time behaviour of the system depending on the approach. The simplest one is synchronous coupling.

The idea of *synchronous coupling* is to couple such solutions of SDEs to which the same Brownian motion is applied ($B_t = \check{B}_t$). To simplify, we consider the equations with a drift of a gradient type:

$$dX_t = -\frac{1}{2}\nabla U(X_t)dt + dB_t, \quad X_0 = \xi_X, \tag{3.14}$$

$$dY_t = -\frac{1}{2}\nabla U(Y_t)dt + dB_t, \quad Y_0 = \xi_Y, \tag{3.15}$$

where $(B_t)_{t\geq 0}, \xi_X$ and ξ_Y are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Equation of type 3.14 is known as a *overdamped Langevin diffusion*.

Proposition 3.3.1. *Let $(X_t, Y_t)_{t\geq 0}$ be a synchronous coupling of 3.14, 3.15. If $\forall x, y \in \mathbb{R}^d$ and $K > 0$ the inequality*

$$(x - y) \cdot (\nabla U(x) - \nabla U(y)) \geq K|x - y|^2$$

holds, then $\forall \xi_X \sim \mu, \xi_Y \sim \nu$ initial values, where ξ_X, ξ_Y are random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$,

$$|X_t - Y_t| \leq e^{-\frac{Kt}{2}}|\xi_X - \xi_Y| \quad \mathbb{P}\text{-almost surely} \quad \forall t \geq 0, \tag{3.16}$$

and

$$W^p(\mu p_t, \nu p_t) \leq e^{-\frac{Kt}{2}} W^p(\mu, \nu) \quad \forall p \geq 1. \tag{3.17}$$

Proof. The approach for proving 3.16 is basically the same as in 3.1.1. We first rewrite 3.14 and 3.15 in the following form

$$dX_t - dY_t = -\frac{1}{2}(\nabla U(X_t) - \nabla U(Y_t))dt,$$
$$\implies \frac{d}{dt}(X_t - Y_t) = -\frac{1}{2}(\nabla U(X_t) - \nabla U(Y_t)) \quad \mathbb{P}\text{-a.s.}$$

Choosing function $v(t) = |X_t - Y_t|^2$ the rest of the proof for 3.16 is analogous to 3.1.1.

Now, let us prove 3.17. Taking expectation in 3.16 we get

$$W^p(\mu p_t, \nu p_t) = \inf_{(X,Y):X_t\sim\mu, Y_t\sim\nu} \Big(\mathbb{E}[|X_t - Y_t|^p]\Big)^{\frac{1}{p}} \leq (\mathbb{E}[|X_t - Y_t|^p])^{\frac{1}{p}}$$
$$\leq e^{-\frac{Kt}{2}}(\mathbb{E}[|\xi_X - \xi_Y|^p])^{\frac{1}{p}} = e^{-\frac{Kt}{2}}\Big(\int |x-y|^p \eta(dxdy)\Big)^{\frac{1}{p}},$$

where $\eta \in \Pi(\mu, \nu)$. To obtain the exponential contraction we take the infimum over all $\eta \in \Pi(\mu, \nu)$ and we are done.

□

In analogy with the deterministic system 3.2 we have a corollary about the existence of an invariant measure for the 3.1.

Corollary 3.3.1. *There exists a unique invariant measure μ_∞ such that*

$$W^p(\mu p_t, \mu_\infty) \leq e^{-\frac{Kt}{2}} W^p(\mu, \mu_\infty) \quad \forall p \geq 1, t \geq 0, K > 0. \tag{3.18}$$

and μ_∞ is given by

$$\mu_\infty(dx) = e^{-U(x)}dx. \tag{3.19}$$

Therefore, we have shown that the dynamical system, the coefficients of which are perturbed with Brownian motion, is still stable if the potential function is strongly convex. But can we profit from the randomness that is added to a deterministic system in such a way that it is possible to get an exponentially fast convergence to equilibrium even when the potential function does not satisfy the necessary condition? For this purpose, let us introduce another coupling.

3.4 Coupling by reflection

The method developed in this section was originally described in [2]. The idea of a coupling by reflection is to couple two strong solutions of SDE relative to different Brownian motions that are constructed in such a way that the increments perpendicular to the difference vector of solutions are the same and the increments parallel to the difference vector are reflected. That is to say, we consider the following equations with the given initial distributions

$$dX_t = b(X_t)dt + \sigma dB_t \quad for\ t \geq 0, \quad X_0 \sim \mu, \tag{3.20}$$
$$dY_t = b(Y_t)dt + \sigma d\check{B}_t \quad for\ t \geq 0, \quad Y_0 \sim \nu, \tag{3.21}$$

where

$$d\check{B}_t = \begin{cases} (I - 2e_t e_t^T) dB_t, & t < T \\ dB_t, & t \geq T. \end{cases} \tag{3.22}$$

Here $e_t = \frac{\sigma^{-1}(X_t - Y_t)}{|\sigma^{-1}(X_t - Y_t)|}$ and $T = \inf_{s \geq 0}\{X_s = Y_s\}$, i.e. T is a first hitting time when the two solutions meet. It is easy to see that after time T the synchronous coupling is applied.

Definition 3.4.1. *A **coupling by reflection** of two solutions of SDE 2.2 is a stochastic process $(X_t, Y_t) \in \mathbb{R}^{2d}$ with the initial distribution $(X_0, Y_0) \sim \eta$, where X_t and Y_t are strong solutions of 3.20 and 3.21 w.r.t. the Brownian motions B_t and $\check{B}_t$ respectively and $\eta \in \Pi(\mu, \nu)$.*

To prove that this coupling is a stochastic process according to the definition 3.4.1 indeed, we need to show that $\check{B}_t$ is a Brownian motion. To this aim, we introduce the following characteristic of Brownian motion.

Theorem 3.4.1 (Lévy's characterisation of Brownian motion)**.** *Let $\{X_t = (X_t^1, ..., X_t^d)\}$ be a stochastic process defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to filtration $(\mathcal{F}_t)_{t \geq 0}$. If*

1. *$X_0 = 0$ $\mathbb{P}$-almost surely,*
2. *$(X_t)_{t \geq 0}$ is a continuous local martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$,*
3. *The cross-variations are given by $\langle X_t^i, X_t^j \rangle = \delta_{ij} t,\ 1 \leq i, j \leq d$ (i.e. $X_t^i X_t^j - \delta_{ij} t$ is a continuous local martingale),*

then $(X_t)_{t \geq 0}$ is a d-dimensional Brownian motion.

Proof. We have to show that the process $(X_t)_{t \geq 0}$ satisfies the properties which define Brownian motion. This means that we to have to show that $X_t - X_s$ is independent of σ-algebra $\mathcal{F}_s$ generated by X_s and that $X_t - X_s$ has a normal distribution with mean zero and covariance matrix $(t - s)I$ for $0 < s < t$. It is enough to show that, conditionally on $\mathcal{F}_s$, the characteristic function of $X_t - X_s$ is equal to the characteristic function of a random variable with the distribution $\mathcal{N}(0, t - s)$, i.e.

$$\mathbb{E}[e^{i(u, X_t - X_s)} | \mathcal{F}_s] = e^{-\frac{\|u\|^2}{2}(t-s)} \quad \forall u \in \mathbb{R}^d \quad \mathbb{P}\text{-a.s.} \tag{3.23}$$

Let $f(X_v) = e^{i(u, X_v)}$. Then, applying Itô formula and using the fact that $\langle X_t^i, X_t^j \rangle = \delta_{ij} t,\ 1 \leq i, j \leq d$, we obtain

$$\begin{aligned} de^{i(u, X_v)} &= (\nabla(e^{i(u, X_v)}))^T dX_v + \frac{1}{2} \operatorname{Tr}\left(\nabla^2 (e^{i(u, X_v)})\right) dv \\ &= iue^{i(u, X_v)} dX_v - \frac{1}{2} \operatorname{Tr}\left(uu^T e^{i(u, X_v)}\right) dv \end{aligned}$$

or in the integral notation with $s \leq v \leq t$

$$e^{i(u,X_t)} = e^{i(u,X_s)} + iu \int\limits_s^t e^{i(u,X_v)} dX_v - \frac{1}{2}\|u\|^2 \int\limits_s^t e^{i(u,X_v)} dv. \tag{3.24}$$

Since $\int_s^t e^{i(u,X_v)} dX_v$ is a martingale and due to the given cross-variations of the process X_t, it follows that

$$\mathbb{E}\Big[\int\limits_s^t e^{i(u,X_v)} dX_v | \mathcal{F}_s\Big] = 0 \quad \mathbb{P}\text{-a.s.} \tag{3.25}$$

Multiplying 3.24 by $e^{-i(u,X_s)}\mathbb{1}_A$ with $A \in \mathcal{F}_s$ and using 3.25, we obtain

$$\begin{aligned}
\mathbb{E}[e^{i(u,X_t - X_s)}\mathbb{1}_A] &= \mathbb{E}[\mathbb{1}_A] - \frac{1}{2}\|u\|^2 \int\limits_s^t \mathbb{E}[e^{i(u,X_v - X_s)}\mathbb{1}_A] dv \\
&= \mathbb{P}(A) - \frac{1}{2}\|u\|^2 \int\limits_s^t \mathbb{E}[e^{i(u,X_v - X_s)}\mathbb{1}_A] dv.
\end{aligned}$$

This integral equation with respect to the deterministic function $t \mapsto \mathbb{E}[e^{i(u,X_t - X_s)}\mathbb{1}_A]$ admits the solution

$$\mathbb{E}[e^{i(u,X_t - X_s)}\mathbb{1}_A] = \mathbb{P}(A) e^{-\frac{1}{2}\|u\|^2(t-s)} \quad \forall A \in \mathcal{F}_s. \tag{3.26}$$

Since we consider the sets from $\mathcal{F}_s$, the equation 3.23 follows. Also, from 3.26 it follows that the increments $X_t - X_s$ are independent of σ-algebra $\mathcal{F}_s$.

□

With the use of theorem 3.4.1 it suffices to show that $\check{B}_t$ is a martingale and $\check{B}_t^i \check{B}_t^j - \delta_{ij} t$ is a martingale for $1 \leq i, j \leq d$. Since a stochastic integral with respect to Brownian motion is a martingale, then $\check{B}_t$ is a martingale because $\check{B}_t = \int_0^t (I - 2e_s e_s^T) dB_s,\ t < T$.

Next, we note that for $\forall t \geq 0$ $(I - 2e_t e_t^T)$ is an orthogonal matrix. Indeed,

$$\begin{aligned}
(I - 2e_t e_t^T)^T (I - 2e_t e_t^T) &= (I - 2e_t e_t^T)(I - 2e_t e_t^T) \\
&= I - 4e_t e_t^T + 4e_t(e_t^T e_t)e_t^T = I - 4e_t e_t^T + 4e_t e_t^T = I.
\end{aligned}$$

Next, moving to the matrix notation, we are going to show that $(\check{B}_t \cdot v_1)(\check{B}_t \cdot v_2) - v_1 \cdot v_2 t$ is a martingale, where v_1, v_2 are unit vectors with non-zero entry at i-th and j-th index respectively. Let

$$f(\check{B}_t) = (\check{B}_t \cdot v_1)(\check{B}_t \cdot v_2).$$

Then

$$\begin{aligned}
\nabla f(\check{B}_t) &= v_1^T \check{B}_t^T v_2 + v_2^T \check{B}_t^T v_1, \\
\nabla^2 f(\check{B}_t) &= v_1 v_2^T + v_2 v_1^T.
\end{aligned}$$

We note that

$$\begin{aligned}\mathrm{Tr}((I-2e_te_t^T)(v_1v_2^T+v_2v_1^T)(I-2e_te_t^T)^T) &= \mathrm{Tr}((I-2e_te_t^T)Q\Lambda Q^T(I-2e_te_t^T)^T)\\ &= \mathrm{Tr}((I-2e_te_t^T)Q\Lambda((I-2e_te_t^T)Q)^T) = \mathrm{Tr}(\Lambda),\end{aligned}$$

where Q is an orthogonal matrix, whose columns are eigenvalues of $(v_1v_2^T+v_2v_1^T)$, Λ is a diagonal matrix whose entries are the eigenvalues of $(v_1v_2^T+v_2v_1^T)$, and $\mathrm{Tr}(\Lambda)=0$ if $v_1\neq v_2$, and $\mathrm{Tr}(\Lambda)=2$ if $v_1=v_2$. Putting all the pieces together in the Itô formula we get

$$\begin{aligned}d((\check{B}_t\cdot v_1)(\check{B}_t\cdot v_2)) &= (v_1^T\check{B}_t^Tv_2+v_2^T\check{B}_t^Tv_1)^T(I-2e_te_t^T)dB_t\\ &+\frac{1}{2}\mathrm{Tr}((I-2e_te_t^T)^T(v_1v_2^T+v_2v_1^T)(I-2e_te_t^T))dt\\ &=(I-2e_te_t^T)(v_1^T\check{B}_t^Tv_2+v_2^T\check{B}_t^Tv_1)^TdB_t+v_1^Tv_2dt.\end{aligned}$$

Therefore,

$$(\check{B}_t\cdot v_1)(\check{B}_t\cdot v_2)-v_1\cdot v_2t=\int_0^t(I-2e_te_t^T)(v_1^T\check{B}_t^Tv_2+v_2^T\check{B}_t^Tv_1)^TdB_t,\ t<T,$$

and we are done. Thus, $\check{B}_t$ is a Brownian motion and the SDE 3.21 is a well-defined stochastic process.

We note that $Z_t:=X_t-Y_t$ satisfies the following equation

$$dZ_t=(b(X_t)-b(Y_t))dt+2|\sigma^{-1}Z_t|^{-1}Z_tdW_t,\quad t<T, \tag{3.27}$$
$$Z_t=0,\quad t\geq T, \tag{3.28}$$

$dW_t=e_t^TdB_t$.

Equation 3.27 is well-defined since W_t is a one-dimensional Brownian motion according to theorem 3.4.1. Indeed, $W_t=\int_0^te_s^TdB_s,\ t<T$ is a stochastic integral with respect to the Brownian motion B_t. With the use of Ito's formula we have

$$\begin{aligned}f(W_t)&=W_t^2,\\ dW_t^2&=f'(W_t)dW_t+\frac{1}{2}\mathrm{Tr}(e_tf''(W_t)e_t^T)=2W_tdW_t+\mathrm{Tr}(e_te_t^T)dt\\ &=2W_tdW_t+\|e_t\|^2dt=2W_tdW_t+dt,\end{aligned}$$

Thus, $W_t^2-t=2\int_0^tW_tdW_t$ is a stochastic integral with respect to the Brownian motion W_t.

We now introduce a new underlying distance function for the Wasserstein distance W_f^1. We will see that choosing the right distance function allows us to get contractivity results for the SDE even when the potential function U is locally non-convex. As a result, we get an exponential convergence of the solution law to equilibrium.

Define $r_t=g(X_t-Y_t)=g(Z_t)$ with $g(\cdot)=|\cdot|$.

$$\begin{aligned}\nabla g(Z_t)&=Z_tr_t^{-1},\\ \nabla^2g(Z_t)&=r_t^{-1}-Z_tZ_t^Tr_t^{-3}.\end{aligned}$$

Applying the Itô formula to $g(Z_t)$ we obtain:

$$\begin{aligned}
dg(Z_t) &= \nabla g(Z_t)\cdot dZ_t + \frac{1}{2}\operatorname{Tr}((2|\sigma^{-1}Z_t|^{-1}Z_t)^T\nabla^2 g(Z_t)(2|\sigma^{-1}Z_t|^{-1}Z_t))dt \\
&= Z_t r_t^{-1}\cdot((b(X_t)-b(Y_t))dt + 2|\sigma^{-1}Z_t|^{-1}Z_t dW_t) + 2\operatorname{Tr}\Big((|\sigma^{-1}Z_t|^{-1}Z_t)^T(r_t^{-1} \\
&- Z_t Z_t^T r_t^{-3})(|\sigma^{-1}Z_t|^{-1}Z_t)\Big)dt = Z_t r_t^{-1}\cdot(b(X_t)-b(Y_t))dt \\
&+ 2|\sigma^{-1}Z_t|^{-1}r_t^{-1}Z_t^T Z_t dW_t + 2\operatorname{Tr}\Big(|\sigma^{-1}Z_t|^{-2}r_t^{-1}Z_t^T Z_t - |\sigma^{-1}Z_t|^{-2}r_t^{-3}Z_t^T Z_t Z_t^T Z_t\Big)dt \\
&= Z_t r_t^{-1}\cdot(b(X_t)-b(Y_t))dt + 2\operatorname{Tr}\Big(|\sigma^{-1}Z_t|^{-2}r_t - |\sigma^{-1}Z_t|^{-2}r_t\Big) \\
&= Z_t r_t^{-1}\cdot(b(X_t)-b(Y_t))dt + 2|\sigma^{-1}Z_t|^{-1}r_t dW_t.
\end{aligned}$$

We apply the same approach for $f(r_t)$:

$$\begin{aligned}
df(r_t) &= f'(r_t)dr_t + \frac{1}{2}\operatorname{Tr}(4f''(r_t))dt = f'(r_t)(Z_t r_t^{-1}\cdot(b(X_t)-b(Y_t))dt + 2|\sigma^{-1}Z_t|^{-1}r_t dW_t) \\
&+ 2f''(r_t)dt = 2|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t + (r_t^{-1}Z_t\cdot(b(X_t)-b(Y_t))f'(r_t) + 2f''(r_t))dt.
\end{aligned} \tag{3.29}$$

For $r > 0$ we define the continuous function on $(0,\infty)$

$$\kappa(r) = \inf\{-2\frac{|\sigma^{-1}(x-y)|^2}{|x-y|^2}\frac{(x-y)\cdot(b(x)-b(y))}{|x-y|^2} \text{ for } x,y\in\mathbb{R}^d \text{ such that } |x-y|=r\}$$

and assume that b satisfies

$$\liminf_{r\to\infty}\kappa(r) > 0, \quad \int_0^1 r\kappa(r)^- dr < \infty.$$

I.e., $\kappa(r)$ is the largest constant such that the inequality

$$(x-y)\cdot(b(x)-b(y)) \le -\frac{1}{2}\frac{\kappa(r)|x-y|^4}{|\sigma^{-1}(x-y)|^2}$$

holds for any two points $x,y\in\mathbb{R}^d$ the distance between which is equal to r. We note that when $\sigma\equiv I$, $\kappa(r)$ is defined by the inequality $(x-y)\cdot(b(x)-b(y)) \le -\frac{1}{2}\kappa(r)|x-y|^2$, which is equal to K in the notions of 3.3.1. In case when $0<\kappa(r) =: \kappa$ for $\forall r>0$ and $b(\cdot) = \frac{1}{2}\nabla U(\cdot)$ we have that the function $U(\cdot)$ is κ-convex, and the synchronous coupling can be used for getting exponential contractivity. Nonetheless, the case when $\kappa(r)\ge -L, L\in\mathbb{R}$, for all r and $\kappa(r)\ge K, K\in\mathbb{R}$, for large r, is of our interest, and it exactly describes the case with locally non-convex potential functions. We define the following constants

$$R_0 = \inf\{R\ge 0 : \kappa(r)\ge 0\ \forall r\ge R\}, \tag{3.30}$$

$$R_1 = \inf\{R\ge R_0 : \kappa(r)R(R-R_0)\ge 8\ \forall r\ge R\}, \tag{3.31}$$

Next, we define the underlying distance function $f(|x-y|)$ as follows

$$f(r) = \int_0^r \varphi(s) g(s) ds, \tag{3.32}$$

where

$$\varphi(r) = \exp\Big(-\frac{1}{4}\int_0^r s\kappa(s)^- ds\Big),$$

$$\Phi(r) = \int_0^r \varphi(s) ds,$$

$$g(r) = 1 - \frac{1}{2}\int_0^{r\wedge R_1} \frac{\Phi(s)}{\varphi(s)} ds \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} ds.$$

Remark 3.4.1. *It is important to mention some properties of the given functions:*

1. *φ is decreasing, $\varphi(0) = 1$, and $\varphi(r) = \varphi(R_0)$ for $\forall r \geq R_0$,*
2. *g is decreasing, $g(0) = 1$, and $g(r) = \frac{1}{2}$ for $\forall r \geq R_1$,*
3. *$f(0) = 0$, $f'(0) = 1$, $f''(r) \leq 0$, and $\frac{\Phi(r)}{2} \leq f(r) \leq \Phi(r)$ for $\forall r \geq 0$.*

The particular choice of the distance function allows us in a way to "overcome" the concavity of a potential function. Some details for the verification will be mentioned later, but first, we provide some important results we are obtaining with the current method. It will be shown in the next theorem that the given underlying distance function f together with the reflection coupling allows us to get shrinking of the distance between two solutions as $t \to \infty$.

Theorem 3.4.2 (Exponential contractivity for coupling by reflection)**.** *Let us define*

$$\alpha := \sup\{|\sigma^{-1} z|^2 : z \in \mathbb{R}^d \text{ such that } \|z\| = 1\} \tag{3.33}$$

and

$$\frac{1}{c} = \alpha \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds = \alpha \int_0^{R_1}\int_0^s \exp\Big(\frac{1}{4}\int_t^s u\kappa(u)^- du\Big) dt ds, \quad c \in (0, \infty). \tag{3.34}$$

Then the function $t \longmapsto e^{ct}\mathbb{E}[d_f(X_t, Y_t)]$ is decreasing on $[0, \infty)$ for a given distance function d_f.

Proof. We start from bounding the term $f''(r) - \frac{1}{4} r\kappa(r) f'(r)$. The results we get will help us to manipulate with 3.29 and thus verify the statement.

Let $r < R_1$. Then by direct computation of $\varphi'(r), g'(r)$ and property 3 we have

$$\begin{aligned} f''(r) &= \varphi'(r) g(r) + \varphi(r) g'(r) \\ &= -\frac{1}{4} r\kappa^-(r)\varphi(r) g(r) - \frac{1}{2}\frac{\Phi(r)}{\int_0^{R_1}\Phi(s)\varphi(s)^{-1} ds} \\ &\leq \frac{1}{4} r\kappa(r) f'(r) - \frac{1}{2}\frac{f(r)}{\int_0^{R_1}\Phi(s)\varphi(s)^{-1} ds}. \end{aligned}$$

With this bound we are able to estimate $df(r)$ for $r < R_1$. Now let $r > R_1$. Then we have $f'(r) = \frac{\varphi(R_0)}{2}$ by properties 3.4.1 and $\kappa(r) \geq \frac{8}{R(R-R_0)}$ by 3.31:

$$f''(r) - \frac{1}{4}r\kappa(r)f'(r) = -\frac{1}{8}r\kappa(r)\varphi(R_0) \leq -\frac{\varphi(R_0)}{R_1 - R_0}\frac{r}{R_1}. \tag{3.35}$$

We are going to show that for we get the same bound as for $r < R_1$. We note that $r > R_1$ $\varphi(r) \equiv \text{Const}$ for $\forall r \geq R_0$, therefore, $\Phi(r) = \Phi(R_0) + (r - R_0)\varphi(R_0)$, and

$$\frac{r}{R_1} = \frac{\frac{\Phi(r)}{\varphi(R_0)} - \frac{\Phi(R_0)}{\varphi(R_0)} + R_0}{\frac{\Phi(R_1)}{\varphi(R_0)} - \frac{\Phi(R_0)}{\varphi(R_0)} + R_0} = \frac{\Phi(r) - \Phi(R_0) + R_0\varphi(R_0)}{\Phi(R_0) - \Phi(R_0) + R_0\varphi(R_0)} \geq \frac{\Phi(r)}{\Phi(R_0)}$$

Also,

$$\begin{aligned}
\int\limits_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}ds &= \int\limits_{R_0}^{R_1} (\Phi(R_0) + (r - R_0)\varphi(R_0))\varphi(s)^{-1}ds \\
&= \Phi(R_0)\varphi(R_0)^{-1}(R_1 - R_0) + \frac{(R_1 - R_0)^2}{2} \\
&\geq \frac{(R_1 - R_0)(\Phi(R_0) + (R_1 - R_0)\varphi(R_0))\varphi(R_0)^{-1}}{2} \\
&= \frac{(R_1 - R_0)\Phi(R_1)\varphi(R_0)^{-1}}{2},
\end{aligned}$$

$$\implies \frac{(R_1 - R_0)\Phi(R_1)}{\varphi(R_0)} \leq 2\int\limits_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}ds.$$

And so using these bounds in 3.35 we get

$$\begin{aligned}
f''(r) - \frac{1}{4}r\kappa(r)f'(r) &\leq -\frac{\varphi(R_0)}{R_1 - R_0}\frac{\Phi(r)}{\Phi(R_1)} \leq -\frac{1}{2}\frac{\Phi(r)}{\int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}ds} &(3.36)\\
&\leq -\frac{1}{2}\frac{f(r)}{\int_0^{R_1} \Phi(s)\varphi(s)^{-1}ds}. &(3.37)
\end{aligned}$$

Thus, from 3.35 and 3.36 we have

$$\begin{aligned}
\forall r > 0 \quad & 2|\sigma^{-1}Z_t|^{-2}r_t^2 \cdot (f''(r) - \frac{1}{4}r\kappa(r)f'(r)) \\
&\leq 2|\sigma^{-1}Z_t|^{-2}r_t^2 \cdot \Big(-\frac{1}{2}\frac{f(r)}{\int_0^{R_1} \Phi(s)\varphi(s)^{-1}ds}\Big) = -\frac{|\sigma^{-1}Z_t|^{-2}r_t^2 f(r)}{\int_0^{R_1} \Phi(s)\varphi(s)^{-1}ds},
\end{aligned}$$

and so by definition of f

$$2|\sigma^{-1}Z_t|^{-2}r_t^2 \cdot (f''(r) - \frac{1}{4}r\kappa(r)f'(r)) \leq -cf(r),$$

where c is defined by 3.34. Finally, we bound 3.29:

$$
\begin{aligned}
df(r_t) &= 2|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t \\
&\quad + (2|\sigma^{-1}Z_t|^{-2}r_t^2 f''(r_t) + r_t^{-1}Z_t\cdot(b(X_t)-b(Y_t))f'(r_t))dt \\
&\le 2|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t + (2|\sigma^{-1}Z_t|^{-2}r_t^2 f''(r_t) - \frac{1}{2}|\sigma^{-1}Z_t|^{-2}\kappa(r_t)r_t^3 f'(r_t))dt \\
&\le 2|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t + 2|\sigma^{-1}Z_t|^{-2}r_t^2\cdot(f''(r_t) - \frac{1}{4}r_t\kappa(r_t)f'(r_t))dt \\
&\le 2|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t - cf(r_t).
\end{aligned}
$$

We rewrite this inequality in the integral notation

$$\forall t > 0 \quad f(r_t) \le 2\int_0^t |\sigma^{-1}Z_s|^{-1}r_s f'(r_s)dW_s - c\int_0^t f(r_s)ds.$$

Taking expectation in this inequality and knowing that expectation of a martingale is equal to zero we get

$$\mathbb{E}[f(r_t)] \le -c\int_0^t \mathbb{E}[f(r_s)]ds,$$

or

$$\frac{d}{dt}\mathbb{E}[f(r_t)] \le -c\mathbb{E}[f(r_t)], \quad t \ge 0.$$

Then by the Grönwall's inequality we have

$$e^{ct}\mathbb{E}[f(r_t)] \le \mathbb{E}[f(r_0)], \quad t \ge 0,$$

and

$$\frac{d}{dt}e^{ct}\mathbb{E}[f(r_t)] \le 0,$$

therefore, $t \longmapsto e^{ct}\mathbb{E}[d_f(X_t, Y_t)]$ is decreasing on $[0, \infty)$.

□

The theorem yields an immediate consequence with the exponential contractivity at rate $c > 0$ for the transition kernels p_t of the SDE 3.20 w.r.t. the Wasserstein distance W_f.

Corollary 3.4.1. *For* $\forall t \ge 0$

$$W_f^1(\mu p_t, \nu p_t) \le e^{-ct}W_f^1(\mu, \nu), \tag{3.38}$$

$$W^1(\mu p_t, \nu p_t) \le 2\varphi(R_0)^{-1}e^{-ct}W^1(\mu, \nu). \tag{3.39}$$

Proof. Let (X_t, Y_t) be a coupling by reflection of two solutions of 3.20 with initial distribution $(X_0, Y_0) \sim \eta \in \Pi(\mu, \nu)$. By theorem 3.4.2 we have

$$W_f^1(\mu p_t, \nu p_t) \le \mathbb{E}[d_f(X_t, Y_t)] \le e^{-ct}\mathbb{E}[d_f(X_0, Y_0)] = e^{-ct}\int d_f(x, y)\eta(dx, dy), \quad t \ge 0. \tag{3.40}$$

By optimising $\int d_f(x,y)\eta(dx,dy)$ over $\eta \in \Pi(\mu,\nu)$ we get

$$W_f^1(\mu p_t, \nu p_t) \leq e^{-ct} W_f^1(\mu,\nu), \quad t \geq 0. \tag{3.41}$$

Inequality 3.39 follows from the fact that $\frac{\varphi(R_0)}{2} \leq f' \leq 1$ and hence $\frac{\varphi(R_0)|x-y|}{2} \leq d_f(x,y) \leq |x-y|$.

□

Corollary 3.4.2. *There exists a unique stationary distribution μ_∞ of $(p_t)_{t\geq 0}$ such that $\int |y|\mu_\infty(dy) < \infty$ and*

$$\mathrm{Var}_{\mu_\infty}(g) \leq \frac{1}{2c}\|g\|^2_{Lip(f)} \quad \text{for any Lipschitz continuous } g : \mathbb{R}^d \to \mathbb{R}, \tag{3.42}$$

where

$$\|g\|_{Lip(f)} = \sup\left\{\frac{|g(x)-g(y)|}{d_f(x,y)} : x,y \in \mathbb{R}^d \ s.t.\ x \neq y\right\} \tag{3.43}$$

is a Lipschitz semi-norm w.r.t. d_f. Moreover, for any initial distribution ν we have an exponential contraction for the transitional semigroup $(p_t)_{t\geq 0}$ w.r.t. the Wasserstein distance W_f:

$$W_f^1(\nu p_t, \mu_\infty) \leq e^{-ct} W_f^1(\nu, \mu_\infty) \quad \text{for any } t \geq 0. \tag{3.44}$$

Proof. By corollary 3.4.1 we have that $\nu \longmapsto \nu p_1$ is a contraction map w.r.t. the Wasserstein distance W_f on the complete metric space $\mathcal{P}_1$ defined in 3.2. Hence, it follows from the Banach fixed point theorem that there exists a unique $\mu_0 \in \mathcal{P}_1$ such that $\mu_0 p_1 = \mu_0$. The measure $\mu_\infty = \int_0^1 \mu_0 p_s ds$ satisfies $\mu_\infty p_t = \mu_\infty$ for any $t \in [0,1]$, and hence for any $t \in [0,\infty)$. By corollary 3.4.1

$$W_f^1(\nu p_t, \mu_\infty) = W_f^1(\nu p_t, \mu_\infty p_t) \leq e^{-ct} W(\nu, \mu_\infty), \quad \forall t \geq 0,\ \nu \in \mathcal{P}_1. \tag{3.45}$$

In particular, $p_t(x,\cdot) \to \mu_\infty$ in $\mathcal{P}_1$ as $t \to \infty$ for $\forall x \in \mathbb{R}^d$.

□

These corollaries also imply some important results about quantitative non-asymptotic bounds for the delay of correlations and the bias and variance of ergodic averages. The contractivity w.r.t. W_f can also be used for proving a central limit theorem for the ergodic averages. These results are important in a row of applications.

Underlying distance function f

As it was shown in the proof of the theorem 3.4.2, it is of crucial importance for the distance function to be increasing and concave. The concavity of the distance function (i.e. $f'' < 0$) makes us benefit from the second order derivative in the formula 3.32. The approach for constructing it is the following: we look for a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ with $f(0) = 0,\ f'(0) = 1,\ f'(r_t) > 0,\ f''(r_t) < 0$ such that $e^{ct} f(r_t)$ is a supermartingale for $t < T$ and $c > 0$ with the overall goal to maximize the constant c. Indeed, with the choice of f as in 3.32 we have

$$
\begin{aligned}
d(e^{ct}f(r_t)) &= ce^{ct}f(r_t) + e^{ct}df(r_t) \\
&\le ce^{ct}f(r_t) + 2e^{ct}|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t - ce^{ct}f(r_t) \\
&\le 2e^{ct}|\sigma^{-1}Z_t|^{-1}r_t f'(r_t)dW_t, \quad t < T,
\end{aligned}
$$

hence $e^{ct}f(r_t)$ is a supermartingale for $t < T$. For getting the most optimal constant c we need to find the least concave function among all possible functions in consideration. With the given choice of f the convergence rate c is claimed in [2] to be the best known so far.

4 Summary

In this report, we have step-by-step studied the exponential stability of dynamical systems, starting from the deterministic ones and proceeding with the randomly perturbed ones, which we got by adding random noise to the coefficients of the ordinary differential equation. We have shown that such manipulation leads to a well-defined stochastic differential equation. For studying the exponential ergodicity of an SDE, we described synchronous coupling, which leads to an exponential convergence to a unique equilibrium w.r.t. the Wasserstein distance of the system when the potential function of the system is strongly convex. For getting an exponentially fast convergence to equilibrium w.r.t. Wasserstein distance for a wider class of potential functions, a coupling by reflection was proposed, and a modified underlying distance function was introduced. By means of Itô formula, in particular, it was shown in detail that this coupling is well-defined, and the chosen distance function gives an optimal decay rate for the law of the solution of the SDE.

A natural question that arises with the studying of the described method is if it can somehow be extended to a system of interacting diffusion processes taking values in $\mathbb{R}^{d_i}, d_i \in \mathbb{N}$, i.e. a system of SDEs

$$dX_t^i = b^i(X_t^i) + dB_t^i, \quad i = 1, ...n,$$

where $B^i, i = 1, ...n$, are independent Brownian motions in $\mathbb{R}^{d_i}$, $X = (X^1, ..., X^n)$ with values in $\mathbb{R}^{d_1+...+d_n}$, and $b^i : \mathbb{R}^d \to \mathbb{R}^{d_i}$ are locally Lipschitz continuous functions. It turns out that using a coupling by reflection on $\mathbb{R}^d$ and applying the results based on a new distance function does not lead to the desired exponential contractivity, i.e. with exponential rates independent of n. Instead, a *componentwise coupling by reflection* is proposed by the author in the paper [2]. The idea is to apply the reflection coupling individually for each component (X_t^i, Y_t^i) of the diffusion process (X_t, Y_t) when $X_t^i \neq Y_t^i$, and the synchronous coupling when $X_t^i = Y_t^i$. This approach turns out to be dimension-independent and provides nice convergence results. The described method is of unique interest because it allows one to model and study many applied problems as well as contains many unanswered questions for those who prefer going deep into abstract mathematical theories.

Acknowledgments

I acknowledge funding from the grant SPOT (ANR-20-CE40-0014), without which this project would be impossible. I want to express my endless gratitude to the supervisor of the internship – Professor Giovanni Conforti. Thank you for your guidance and support during these months. Thank you for widening my mathematical knowledge in the direction I like. Last but not least, many thanks to the laboratory CMAP for providing good conditions for the work.

References

[1] Richard Durrett. *Probability: Theory and examples*. Cambridge University Press, 2020.

[2] Andreas Eberle. “Reflection couplings and contraction rates for diffusions”. In: *Probability Theory and Related Fields* 166.3-4 (2015), pp. 851–886. DOI: `10.1007/s00440-015-0673-1`.

[3] Bernt K. Øksendal. *Stochastic differential equations an introduction with applications*. Springer, 1998.

[4] Mörters Peter and Y. Peres. *Brownian motion*. Cambridge University Press, 2012.

[5] David Williams. *Probability with martingales*. Cambridge University Press, 2020.