An initiation to coupling methods for partial differential equations

Pierre Le Bris

These notes are quite informal, and not meant to be published

Contents

These notes are intended for an audience with a background in partial differential equations (PDE), and only a basic familiarity with probability theory. In particular, no prior knowledge of nor expertise in stochastic calculus are assumed from the reader. The goal is to present, in a concise and simple manner, the main results on the long time behavior of some PDEs. They are not meant to replace a genuine course on the topic, of which there are many. They are only giving a somewhat less traditional perspective on PDEs.

These notes are articulated as follows. We collect and recall in Section 1 some elementary notions of stochastic calculus, in particular regarding Brownian motion and Itô's formula, in order to explain how some PDE can be linked to a stochastic differential equation (SDE), the definition of which will be given later. In Section 2, we then describe how coupling methods work, and explain how the Wasserstein distance, closely linked to the theory of Optimal Transport, can be understood as a coupling problem. While the first two sections may seem unrelated to one another, and can thus be read independently, we explain in Section 3 how a probabilistic point of view using the underlying SDE and a coupling method may be be useful in studying some PDEs.

Essentially, we only wish to convey a basic understanding of how a PDE can be seen as the evolution of the probability distribution of a random process, and give a flavor of how things work. We will refer to more complete courses for more details.

1 On the link between SDE and PDE

We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, that is, a triple composed of the set Ω of all possible outcomes, $\mathcal{F}$, a σ-algebra, the set of events (i.e set of "things" we can give the probability of), and $\mathbb{P}$ a measure. For a random variable X, we write $X \sim \mu$ if the probability density of X is μ. We denote by $\mathbb{E}(\cdot)$ the expectation

of a random variable with respect to $\mathbb{P}$, and $\mathbb{E}(\cdot|A)$ the conditional expectation of a random variable knowing A.

Throughout these notes, we will be interested in *stochastic processes*, intuitively random variables evolving with time. Given two sets T and E (respectively representing the set of time and the set in which the process lives), a random variable $X : T \times \Omega \mapsto E$, and a point $\omega \in \Omega$, we will consider $\{X(t, \omega) : t \in T\}$. For any point $\omega \in \Omega$, the mapping $X(\cdot, \omega) : T \mapsto \mathbb{R}^d$ is called a *sample path*. For simplicity, we will often omit the variable ω.

One the most basic example of a stochastic process is the well-known (symmetrical) random walk in dimension 1. In this case, let $T = \mathbb{N}$, $E = \mathbb{Z}$, and $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of independent identically distributed (i.i.d) random variables such that $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = \frac{1}{2}$. The random walk is thus given by $X_n = X_0 + \sum_{i=1}^{n} \xi_i$. At each (discrete) time step, as if you were flipping a coin, the process either goes left (-1) or right (+1) with probability 1/2 on the integer line.

1.1 A quick recap on Brownian motion

The specific stochastic process on which these notes rely is the Brownian motion. The notion was introduced by the Scottish botanist Robert Brown in 1828 to describe the movement of minute particles ejected by pollen grain suspended in water. Since then it has been widely used to model various irregular movements in physics, biology, finance and other fields.

The purpose of this section is to introduce the Brownian motion, and give some of its main properties. We start by the case of dimension 1 and with a quick reminder on the Gaussian distribution, which is crucial to define the process.

Definition 1.1 (Gaussian distribution)**.** *We say a real-valued random variable X follows a Gaussian distribution of expectation μ and standard deviation σ if its probability density function is*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We denote $X \sim \mathcal{N}(\mu, \sigma^2)$. In particular, for any measurable function g, we have

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(x) f(x) dx.$$

We now give the main definition of this section, omitting for simplicity the considerations on measurability.

Definition 1.2 (Brownian motion)**.** *Let $B = \{B_t, t \in \mathbb{R}^+\}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. B is a standard Brownian motion if*

(i) *$B_0 = 0$ and the sample paths $B_\cdot(\omega)$ are continuous for a.e $\omega \in \Omega$,*

(ii) *B has independent increments : $B_t - B_s$ is independent of $\sigma\left(B_u, u \leq s\right)$ for all $t > s \geq 0$,*

(iii) *the distribution of $B_t - B_s$ is $\mathcal{N}(0, t - s)$ for all $t > s \geq 0$.*

The notation $\sigma\left(B_u, u \leq s\right)$ refers to the σ-algebra generated by the random variables B_u for $0 \leq u \leq s$. Point (ii) thus describes the fact that the increment $B_t - B_s$ is independent of the path of B before time s.

Notice how this definition is close to the one of the random walk in $\mathbb{Z}$: when for the latter we choose independent increments $(\xi_i)_{i\in\mathbb{N}}$, (ii) implies similar independent increments but in a continuous case. And when for the random walk those increments are chosen to be either +1 or -1 with probability 1/2, in the non-discrete setting of the Brownian motion we choose Gaussian increments (which is also a symmetrical law). The Brownian motion could thus be understood as a sort of symmetrical random walk in $\mathbb{R}$. Keeping in mind those similarities, although it is not rigorous, might help intuitively understand some properties of the Brownian motion.

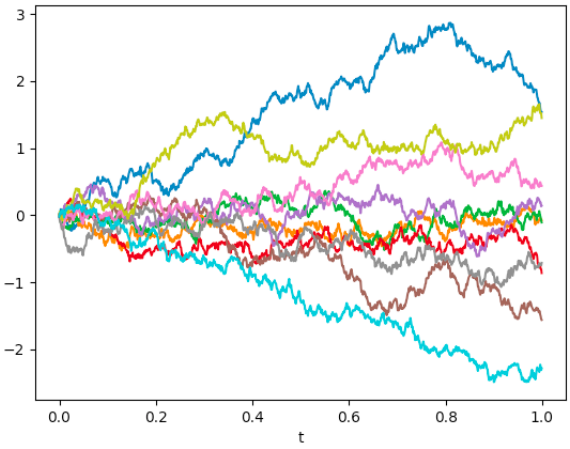

Figure 1: Approximation of 10 sample paths of Brownian motion.

Remark 1.1. *We give a quick explanation on the numerical simulation of the Brownian motion (see for instance Figure 1). We construct the Brownian motion as if we were constructing a random walk at discrete times* $(t_i^n)_{i\in\{1,..n\}}$ *through the Euler-Maruyama method* $B_0 = 0$ *and*

$$B_{t_i^n} - B_{t_{i-1}^n} = G_i^n,$$

where $G_i^n \sim \mathcal{N}(0, t_i^n - t_{i-1}^n)$ *is a simulated Gaussian variable. Thanks to results on the variance, we have* $\mathcal{N}(0, t_i^n - t_{i-1}^n) \overset{law}{=} \sqrt{t_i^n - t_{i-1}^n}\mathcal{N}(0,1)$*, i.e simulating a random variable according to* $\mathcal{N}(0, t_i^n - t_{i-1}^n)$ *amounts to simulating a random variable according to* $\mathcal{N}(0,1)$ *and then multiplying it by* $\sqrt{t_i^n - t_{i-1}^n}$*. This implies (once again in a very formal way) that "*$dB_t \simeq \sqrt{dt}$*".*

The fact that a Brownian motion exists is non trivial, and for the sake of conciseness we do not wish to go down that rabbit hole. We admit its existence and refer to Jean-François Le Gall [LG18] for a rigorous presentation on the matter. Let us however list some basic, yet important for what follows, properties that are directly implied by the definition.

Lemma 1.1. *Let* B *and* $\tilde{B}$ *be two independent Brownian motions,* $t_0 > 0$*,* $c > 0$ *and let* $a, b \in \mathbb{R}$ *such that* $a^2 + b^2 = 1$*. Then the following processes are also Brownian motions*

- $\{-B_t, t \geq 0\}$ *(symmetry),*
- $\{c^{-1/2}B_{ct}, t \geq 0\}$ *(scaling),*
- $\{B_{t_0+t} - B_{t_0}, t \geq 0\}$ *(time translation),*
- $\{B_{t_0-t} - B_{t_0}, 0 \leq t \leq t_0\}$ *(time reversal),*
- $\{aB_t + b\tilde{B}_t, t \geq 0\}$ *(rotation).*

All these properties can be easily deduced from the definition of the Brownian motion. The symmetry, time translation and time reversal properties are also true for the random walk, and the scaling property can be understood using Remark 1.1.

We have been comparing the Brownian motion to the random walk in dimension 1, admittedly because the comparison may help explaining various results. We however wish to quickly show that this comparison is not that nonsensical. Denoting S_n a random walk, we consider a sequence of rescaled process

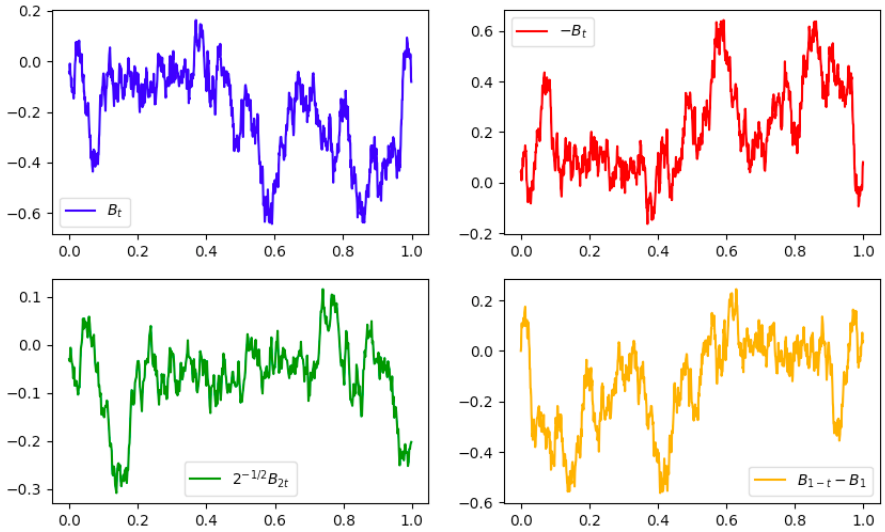

Figure 2: Illustration of Lemma 1.1. **Upper left corner :** Approximation of a sample path of a Brownian motion B_t. **Upper right corner :** Approximation of the path of $-B_t$ (*symmetry*). **Lower left corner :** Approximation of the path of $\frac{B_{2t}}{\sqrt{2}}$ (*scaling*). **Lower right corner :** Approximation of the path of $B_{1-t} - B_1$ (*time reversal*)

$W_t^{(k)} = \frac{S_{\lfloor kt \rfloor}}{\sqrt{k}}$ (notice once again the scaling time / square root of time). A result known as Donsker's theorem ensures that the process $W^{(k)}$ converges in law to a Brownian motion (see Figure 3). It can be seen as an extension of the central limit theorem. Donsker's theorem holds no importance for the following of these notes, but it remains an interesting and possibly enlightening result.

Maybe the most interesting property of the Brownian motion in the context of this document is the following. By definition, for all $t \geq 0$, the distribution of the random variable B_t is a $\mathcal{N}(0, t)$

$$p(t, x)dx := \mathbb{P}(B_t \in [x, x + dx]) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx.$$

An important observation is that this density function satisfies the heat equation, as

$$\frac{\partial}{\partial t} p = \frac{1}{2} \frac{\partial^2}{\partial x^2} p$$

Thus, we have a link between the Brownian motion and the heat equation : the former is a random process whose probability density is the solution of the latter. Now imagine a cloud of particles all starting from 0 and only under the influence of thermal agitation. From a PDE point of view, the shape of the cloud satisfies the heat equation. From a probabilistic point of view, you consider each particle individually to be a Brownian motion. Hence how we circle back to the initial motivation behind the Brownian motion to describe the movement of pollen suspended in water, as for the first time we notice two points of view : the particle and the density.

This duality between the PDE and the particle is, we believe, an idea someone with a background in PDE might already be familiar with. Considering a function b, the transport equation

$$\partial_t p + \text{div}(b\, p) = 0,$$

is linked through the method of characteristic to the ordinary differential equation

$$x_t = x + \int_0^t b\, ds,$$

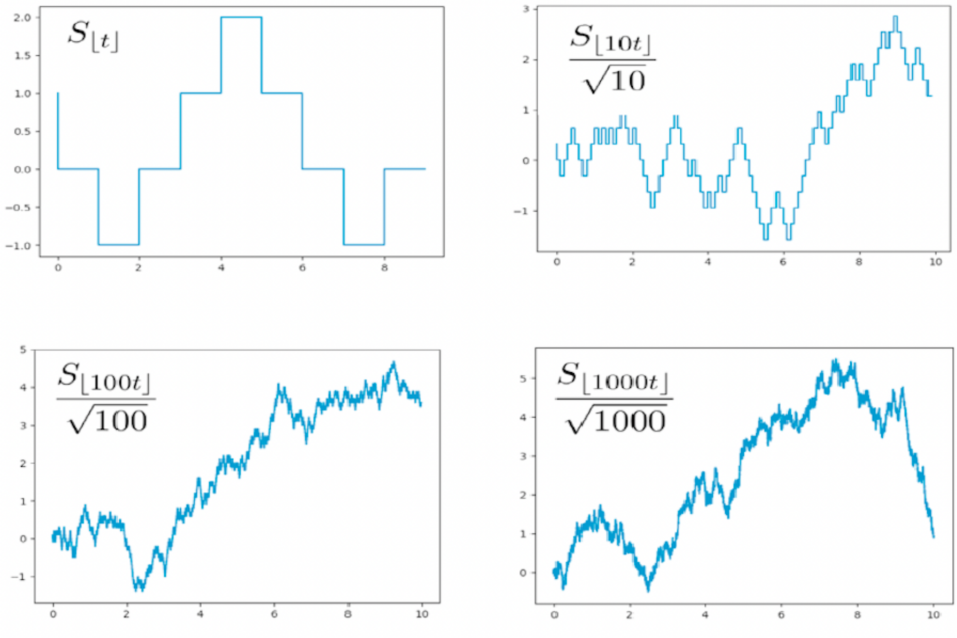

Figure 3: Donsker's theorem : Brownian motion as a limit of symmetrical random walk

which we will rewrite

$$\begin{cases} dx_t = & b\, dt \\ x_0 = & x. \end{cases}$$

The proof of this link will de done latter, as it will be the basis for an interesting result on SDE. Our goal now is to prove that the transport-diffusion equation

$$\partial_t p + \text{div}(b\, p) = \frac{1}{2}\Delta p,$$

is linked to what we will refer to as an SDE

$$X_t = x + \int_0^t b\, ds + B_t,$$

which we will rewrite

$$\begin{cases} dX_t = & b\, dt + dB_t \\ X_0 = & x. \end{cases}$$

This is the subject of this first part.

We conclude this section by extending the definition of Brownian motion to dimensions higher than 1 for the sake of completeness.

Definition 1.3 (Brownian motion in dimension d)**.** *Let $B = \{B_t, t \in \mathbb{R}^+\}$ be a $\mathbb{R}^d$-valued stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{F}$ a filtration. B is a $\mathbb{F}$-standard Brownian motion if the components $B^i, i = 1, ..., d$ are independent $\mathbb{F}$-standard Brownian motions.*

Everything we have said so far holds true in dimensions higher than one, and in particular the density of a Brownian motion in dimension d also satisfies the heat equation in dimension d.

1.2 Itô differential calculus

In this section, we give a brief explanation of the stochastic integral with respect to the Brownian motion, and we mention an important tool in stochastic calculus : Itô's formula. Once again, we will not give any proof and refer to [LG18, Ber21] for more details.

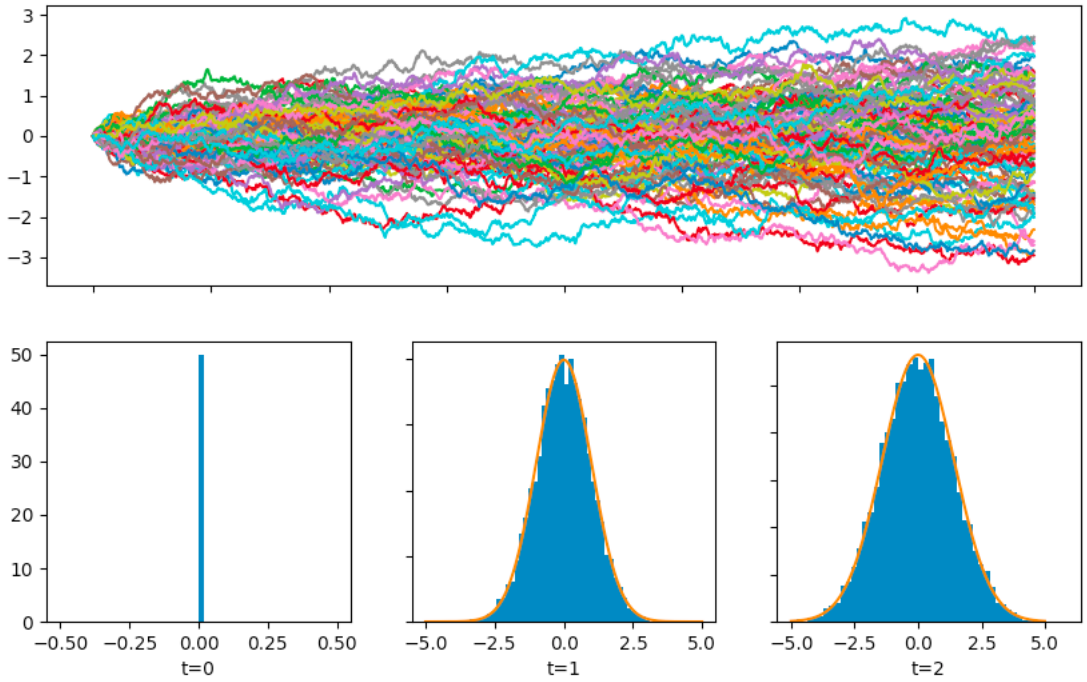

Figure 4: Simulation of the law of Brownian motion. **On top :** 100 sample paths of Brownian motion between $t = 0$ and $t = 2$. **Lower left corner :** Empirical law of the Brownian motion for $t = 0$, using 10000 sample paths. **Lower middle :** Empirical law of the Brownian motion for $t = 1$, using 10000 sample paths, and theoretical law in yellow. **Lower right corner :** Empirical law of the Brownian motion for $t = 2$, using 10000 sample paths, and theoretical law in yellow.

Integration with respect to the Brownian motion. The construction of the stochastic integral might be a bit cumbersome. It is sufficient to understand, in the context of this document, that for the stochastic processes σ_s considered here the quantity $\int_0^t \sigma_s dB_s$ exists (as an L^2 limit of $\sum_{i=1}^n \sigma_{t_i}(B_{t_{i+1}} - B_{t_i})$ for a partition $(t_i)_{i\geq 1}$) and that

$$\mathbb{E}\left(\int_0^t \sigma_s dB_s\right) = 0.$$

Itô's formula. The main result of this section is Itô's formula, a crucial tool in stochastic calculus. Let B be a Brownian motion in dimension n and X be a general Itô process with values in $\mathbb{R}^d$

$$X_t := x + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s, \tag{1.1}$$

where $x \in \mathbb{R}^d$, μ and σ are adapted processes with values respectively in $\mathbb{R}^d$ and $\mathbb{R}^{d\times n}$, and satisfying

$$\int_0^T |\mu_s|^2 ds + \int_0^T |\sigma_s|^2 ds < \infty \text{ a.s.}$$

The process (1.1) will often be denoted

$$\begin{cases} dX_t = \mu_t dt + \sigma_t dB_t, \\ X_0 = x \end{cases} \tag{1.2}$$

with the very natural interpretation of : "small variation of X" can be written as the sum of a "small variation of t" and a "small variation of Brownian motion". Itô's formula then gives us a mean to calculate the dynamic of a function of such a process.

Theorem 1.1 (Itô's formula in dimension 1)**.** *Let X be a Itô process in dimension 1 defined by* (1.2)*, and $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be a function in $\mathcal{C}^{1,2}([0,T],\mathbb{R})$. Then almost surely we have*

$$df(t,X_t) = \partial_x f(t,X_t)\cdot dX_t + \left(\partial_t f(t,X_t) + \frac{1}{2}\sigma_t^2 \partial^2_{xx} f(t,X_t)\right) dt \tag{1.3}$$

for every $0 \leq t \leq T$.

When looking at the form (1.4) takes, the fact that the quantity $\partial_x f(t,X_t)\cdot dX_t + \partial_t f(t,X_t)dt$ appears in $df(t,X_t)$ feels natural for anyone who has studied differential calculus at some point. Notice however the term $\frac{1}{2}\sigma_t^2\partial^2_{xx} f(t,X_t)dt$: it is what may be new and possibly off-putting. We do not wish to write down the full proof, but still want to give an intuition on why this term appears. Let us **very** informally write, for B a Brownian motion in dimension 1

$$f(t,B_t) - f(0,B_0) \simeq \sum_i \left(f(t_i,B_{t_i}) - f(t_{i-1},B_{t_i})\right) + \sum_i \left(f(t_{i-1},B_{t_i}) - f(t_{i-1},B_{t_{i-1}})\right).$$

Then, using a Taylor expansion on the first part we have $f(t_i,B_{t_i}) - f(t_{i-1},B_{t_i}) \simeq \partial_t f(t_i',B_{t_i})(t_i - t_{i-1})$. Notice how this is a "deterministic" term, as it is only parametrized by a single random variable. However, when we consider the Taylor expansion of the second term, we get

$$f(t_{i-1},B_{t_i}) - f(t_{i-1},B_{t_{i-1}}) \simeq \partial_x f(t_{i-1},\xi_i)(B_{t_i} - B_{t_{i-1}}).$$

Recall from the definition of the Brownian motion $B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0,dt) = \sqrt{dt}\mathcal{N}(0,1)$. We hence need to go one term further in the Taylor expansion

$$f(t_{i-1},B_{t_i}) - f(t_{i-1},B_{t_{i-1}}) \simeq \partial_x f(t_{i-1},B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) + \frac{1}{2}\partial^2_{xx} f(t_{i-1},\xi_i)(B_{t_i} - B_{t_{i-1}})^2.$$

The first part gives us $\partial_x f(t_{i-1},B_{t_{i-1}})dB_t$, and we may use for the second part the fact that $(B_{t_i} - B_{t_{i-1}})^2 \simeq t_i - t_{i-1}$.

This theorem can be extended in dimensions higher than one with the same ideas.

Theorem 1.2 (Itô's formula in any dimension)**.** *Let X be a Itô process defined by* (1.2)*, and $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ be a function in $\mathcal{C}^{1,2}([0,T],\mathbb{R}^d)$. Then almost surely we have*

$$df(t,X_t) = \nabla f(t,X_t)\cdot dX_t + \left(\partial_t f(t,X_t) + \frac{1}{2}Tr[\sigma_t^*(H_X f)\sigma_t]\right) dt \tag{1.4}$$

for every $0 \leq t \leq T$, where $Tr(\cdot)$ is the trace operator, σ_t^* is the transpose matrix of σ_t and $H_X f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{1\leq i,j\leq d}$ is the Hessian matrix of f.

1.3 Some links

Heat equation. To understand the link between SDE and PDE, let us begin with a toy model as we come back to the Brownian motion in dimension 1. In the next few lines, the calculations are very formal and only serve to give a better understanding of the ideas that follow behind. Let $X_t = \sqrt{2}B_t$.

A direct application of Itô's formula gives us, for $f : \mathbb{R} \to \mathbb{R}$ a sufficiently regular test function (for instance $f \in \mathcal{C}^2_c(\mathbb{R})$).

$$\begin{aligned} f(X_t) &= \int_0^t f'(X_s)dX_s + \int_0^t \frac{1}{2}(\sqrt{2})^2 f''(X_s)ds, \\ \mathbb{E}f(X_t) &= \mathbb{E}\left(\int_0^t \sqrt{2}f'(X_s)dB_s\right) + \mathbb{E}\left(\int_0^t f''(X_s)\right)ds \\ &= \int_0^t \mathbb{E}f''(X_s)ds, \end{aligned}$$

where the last equality is a consequence of the fact that the expectation of the stochastic integral is 0. Hence why, denoting $p(t,x)$ the probability density of X_t, we have

$$\begin{aligned}\frac{d}{dt}\mathbb{E}f(X_t) =& \mathbb{E}f''(X_t),\\ \frac{d}{dt}\int_{\mathbb{R}} f(x)p(t,x)dx &= \int_{\mathbb{R}} f''(x)p(t,x)dx,\\ \int_{\mathbb{R}} f(x)\partial_t p(t,x)dx &= \int_{\mathbb{R}} f(x)\partial_x^2 p(t,x)dx,\end{aligned}$$

and thus p satisfies the heat equation

$$\partial_t p(t,x) = \partial_x^2 p(t,x).$$

Fokker-Planck equation (or Kolmogorov forward). In a slightly more general setting, let $b : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^{d\times n}$ be two measurable and locally bounded functions. We consider, for a n-dimensional Brownian motion B_t, the stochastic differential equation

$$\begin{cases} dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \ \ t\in[0,T],\\ X_0 = x \end{cases} \tag{1.5}$$

for some $T \in \mathbb{R}$. A *strong solution* of (1.5) is a process X such that

$$\int_0^T \left(|b(t,X_t)| + |\sigma(t,X_t)|^2\right) dt < \infty, \text{ a.s,}$$

and

$$X_t = x + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dB_s, \ \ t\in[0,T].$$

We write that X is a solution of $E_x(\sigma,b)$. Let, for $u \in \mathcal{C}_c^2(\mathbb{R}^+\times\mathbb{R}^d)$, $t\in\mathbb{R}^+$ and $x\in\mathbb{R}^d$,

$$\mathcal{L}u(t,x) := \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^*)_{i,j}(t,x)\frac{\partial^2 u}{\partial x_i \partial x_j}(t,x) + \sum_{i=1}^d b_i(t,x)\frac{\partial u}{\partial x_i}(t,x),$$

where σ^* is the transpose matrix of σ.

Theorem 1.3. *If X is a solution of $E_x(\sigma,b)$, and if we denote $p_t := p(t,\cdot) \in \mathcal{P}(\mathbb{R}^d)$ the law of X_t, then for all $\phi \in \mathcal{C}_c^2(\mathbb{R}^d)$, and every $t \geq 0$*

$$\int_{\mathbb{R}^d} \phi(x)p_t(dx) = \int_{\mathbb{R}^d}\phi(x)p_0(dx) + \int_0^t\int_{\mathbb{R}^d}\mathcal{L}\phi(x)p_s(dx)ds.$$

$(p_t)_{t\geq 0}$ is then said to be a weak solution of

$$\partial_t p = \mathcal{L}^* p, \tag{1.6}$$

with

$$\mathcal{L}^* p(t,x) := \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i\partial x_j}\left((\sigma\sigma^*)_{i,j}p\right)(t,x) - \sum_{i=1}^d \frac{\partial}{\partial x_i}(b_i(t,x)p)(t,x).$$

Proof. Direct use of Itô's formula and integration by parts. □

Results on the uniqueness of solutions for (1.6) give us the converse theorem.

A parabolic equation and method of characteristics. Let us consider a backward approach on the solution of a PDE. Suppose b and σ independent of time. We are interested in the following PDE

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = \mathcal{L}u(t,x), & t>0, x\in\mathbb{R}^d \\ u(0,x) = f(x), & x\in\mathbb{R}^d, \end{cases} \tag{1.7}$$

where $f:\mathbb{R}^d\to\mathbb{R}^d$ is given. Let us start by considering the case $\sigma=0$. The PDE (1.7) can then be written

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = b(x)\cdot\nabla_x u(t,x), & t>0, x\in\mathbb{R}^d \\ u(0,x) = f(x), & x\in\mathbb{R}^d. \end{cases} \tag{1.8}$$

To solve (1.8), we use the well-known method of characteristics. Let $y\in\mathbb{R}^d$, and assume there is a solution to the ordinary differential equation

$$\begin{cases} x'(t) = b(x(t)) & t>0 \\ x(0) = y. \end{cases}$$

For $\phi\in\mathcal{C}_c^1(\mathbb{R}^+\times\mathbb{R}^d)$

$$\begin{aligned} \frac{d}{dt}\phi(t,x(t)) =& \partial_t\phi(t,x(t)) + \sum_{i=1}^d \partial_{x_k}\phi(t,x(t))\frac{dx_k}{dt} \\ =& \partial_t\phi(t,x(t)) + \sum_{i=1}^d b_k(x(t))\partial_{x_k}\phi(t,x(t)) \\ =& \left(\partial_t\phi + b\cdot\nabla_x\phi\right)(t,x(t)). \end{aligned}$$

Let $t_0>0$ and apply the formula above to $\phi(t,x)=u(t_0-t,x)$, for $0\le t\le t_0$,

$$\frac{d}{dt}u(t_0-t,x(t)) = \left(-\partial_t u + b\cdot\nabla_x u\right)(t_0-t,x(t)) = 0.$$

This way

$$u(t_0-t,x(t)) = u(t_0,x(0)),$$

and by taking $t_0=t$

$$u(t,y) = f(x(t)).$$

We thus get the expression of the solution of this transport PDE. This shows that, using the same notation we have used so far, (1.8) is linked to the process defined by

$$dX_t = b(X_t)dt \;\; t\in[0,T],$$

i.e $(X_t)_{t\ge0}$ solution of $E_y(0,b)$. Now, we use the same idea in a diffusive case $\sigma\neq0$ and, by considering the related SDE, we obtain a similar result "in average".

Theorem 1.4. *If $u\in\mathcal{C}_c^2(\mathbb{R}^+\times\mathbb{R}^d)$ satisfies* (1.7)*, then for $x\in\mathbb{R}^d$ and $(X_t)_{t\ge0}$ a solution of $E_x(\sigma,b)$, we have*

$$u(t,x) = \mathbb{E}f(X_t), \;\; t\ge0.$$

Proof. The proof is very similar to the method of characteristic, but we need to use Itô's differential calculus. If $\phi \in \mathcal{C}_c^2(\mathbb{R}^+ \times \mathbb{R}^d)$, thanks to Itô's formula

$$\phi(t, X_t) = \phi(0, x) + \int_0^t \left[\frac{\partial \phi}{\partial t}(s, X_s) + \mathcal{L}\phi(s, X_s)\right] ds + \sum_{i=1}^{d}\sum_{j=1}^{n} \int_0^t \frac{\partial \phi}{\partial x_i}(X_s)\sigma_{i,j}(s, X_s) dB_s^j.$$

Let $t_0 > 0$. We apply the formula above for $\phi(t, x) = u(t_0 - t, x)$, for $0 \leq t \leq t_0$,

$$u(t_0 - t, X_t) = u(t_0, x) + \int_0^t \left[-\frac{\partial u}{\partial t}(t_0 - s, X_s) + \mathcal{L}u(t_0 - s, X_s)\right] ds + M_t = u(t_0, x) + M_t,$$

where M_t is a stochastic integral with expectation 0. Considering the expectation of the expression above

$$\mathbb{E}u(t_0 - t, X_t) = u(t_0, x).$$

We conclude by taking $t_0 = t$ □

Remark 1.2. *A direct consequence of the theorem above is the maximum principle :* $||u||_\infty \leq ||f||_\infty$.

Dirichlet or Neumann boundary conditions. We end this section with two examples.

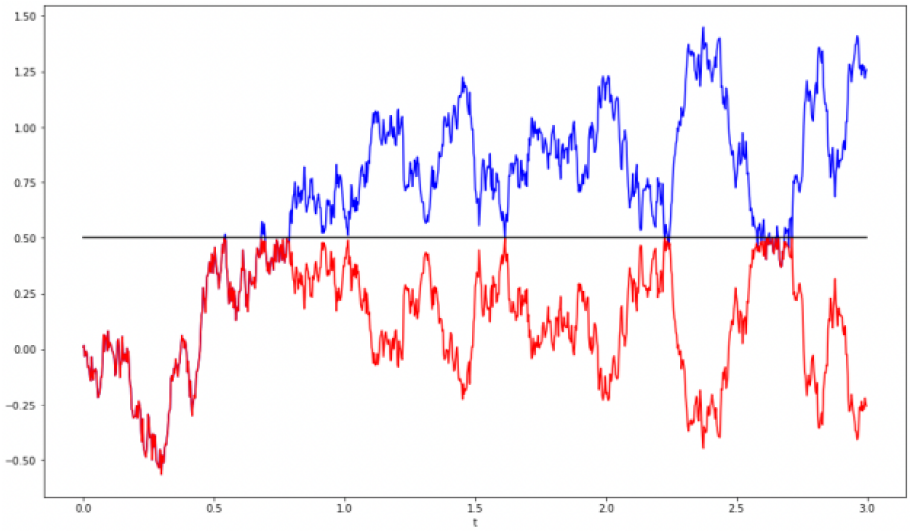

Figure 5: In blue, simulation of a Brownian motion. In red, simulation of the same Brownian motion reflected at level $H = 0.5$.

Let B_t be a Brownian motion and denote B_t^r its reflection at level H (see Figure 5). For any $x \leq H$ we can write

$$\begin{aligned}\mathbb{P}(B_t^r \leq x) =& \mathbb{P}(B_t \leq x) + \mathbb{P}(B_t \geq 2H - x) \\ =& \Phi\left(\frac{x}{\sqrt{t}}\right) + \Phi\left(\frac{x - 2H}{\sqrt{t}}\right),\end{aligned}$$

where $\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is the cumulative distribution function of the standard normal law. Taking the derivative with respect to x above, we get the probability density of the reflected Brownian motion

$$p^r(t, x) = \frac{1}{\sqrt{2\pi t}}\left(e^{-\frac{x^2}{2t}} + e^{-\frac{(2H-x)^2}{2t}}\right) \quad \text{for } x \leq H,$$

which solves

$$\begin{cases} \partial_t p^r(t,x) = \frac{1}{2}\Delta p^r(t,x) \text{ for } x \leq H, \\ \nabla p^r(t,H) = 0, \end{cases}$$

that is, the heat equation with Neumann boundary conditions.

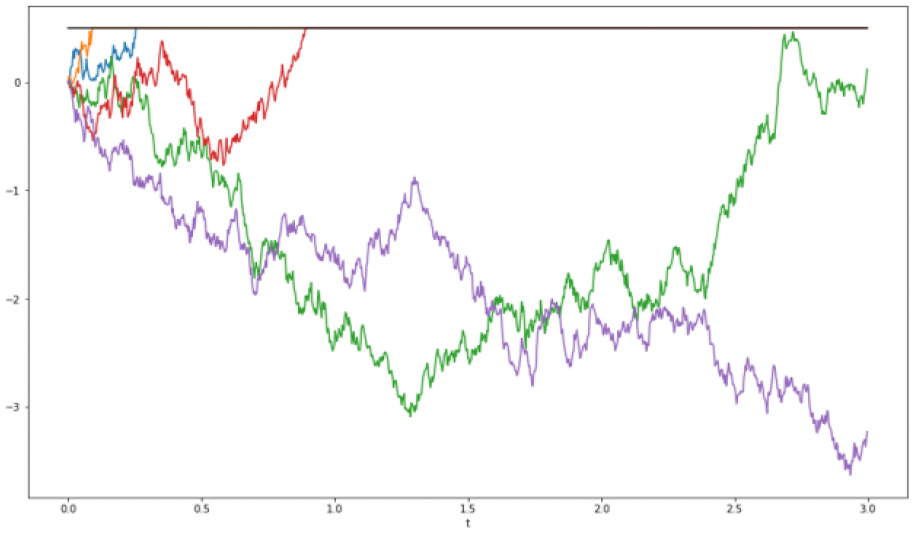

Figure 6: Simulation of Brownian motions killed at level $H = 0.5$

Next, let B_t be a Brownian motion and consider B_t^k a Brownian motion killed upon reaching level H, i.e, denoting $\tau = \inf\{t \geq 0 : B_t \geq H\}$, consider $B_t^k = B_t$ for $t \leq \tau$ and $B_t^k = \theta$ a cemetery state for $t \geq \tau$ (see Figure 6). We have for any $x \leq H$

$$\mathbb{P}(B_t^k \leq x) = \mathbb{P}(B_t \leq x, t \leq \tau).$$

Thanks to the continuity of Brownian motion paths, we can write, for $y \geq H$

$$\begin{aligned} \mathbb{P}(B_t \geq y) =& \mathbb{P}(B_t \geq y, t \geq \tau) \\ =& \mathbb{P}(2H - B_t \leq 2H - y, t \geq \tau). \end{aligned}$$

One can easily see that the process defined by $B_t^* = B_t$ if $t \leq \tau$ and $B_t^* = 2H - B_t$ if $t \geq \tau$ is also a Brownian motion. Thus

$$\mathbb{P}(B_t \geq y) = \mathbb{P}(B_t^* \leq 2H - y, t \geq \tau) = \mathbb{P}(B_t \leq 2H - y, t \geq \tau).$$

Taking $y = 2H - x$, this yields

$$\mathbb{P}(B_t^k \leq x) = \mathbb{P}(B_t \leq x) - \mathbb{P}(B_t \leq x, t \geq \tau) = \mathbb{P}(B_t \leq x) - \mathbb{P}(B_t \geq 2H - x).$$

Derivating with respect to x, we get the following density for the killed Brownian motion

$$p^k(t,x) = \frac{1}{\sqrt{2\pi t}} \left(e^{-\frac{x^2}{2t}} - e^{-\frac{(2H-x)^2}{2t}} \right) \text{ for } x \leq H,$$

which solves

$$\begin{cases} \partial_t p^k(t,x) = \frac{1}{2}\Delta p^k(t,x) \text{ for } x \leq H, \\ p^k(t,H) = 0, \end{cases}$$

that is, the heat equation with Dirichlet boundary conditions.

2 Coupling and Wasserstein distance

So far, we studied the equivalence between the points of view PDE and SDE. We now see how there may be a possibility of using probabilistic tools to study PDEs. In this section, we consider one of those tools, namely coupling methods, and describe how it works.

2.1 Optimal transport as a coupling

Coupling. Let A and B be two sets. For p a probability density on $A \times B$, we call μ and ν its *marginal distributions* on A and B respectively if

$$\mu(x) = \int_B p(x, dy) \quad \text{and} \quad \nu(y) = \int_A p(dx, y).$$

For two random variables (or distributions) X and Y, there are in general several choices for a random vector Z whose marginal distributions correspond to X and Y respectively. Making sure that X and Y are thus related in a particularly interesting way is what we refer to as a *coupling method*.

First example. Imagine two people, named A and B, that don't know each other, each asks you to toss a coin in order for them to take a decision. The first solution would be for you to do two consecutive tosses, give the first result to A and the second to B. But you could also choose to be lazier, and only throw the coin once and give the same result to both. Or give the result to A and the opposite result to B. Neither would notice, as from the perspective of A for instance, the probability of heads or tails would still be one half. These three possibilities (two throws, same result or opposite results) each represents a coupling of the two marginal distributions $\mu = \nu = \frac{1}{2}\delta_H + \frac{1}{2}\delta_T$. Choosing which suits you best can then provide you with some interesting results. This is the topic of the next section.

On optimal transport. Before diving further into coupling arguments, let us explain how the theory of optimal transport is related to what we are discussing here. Let μ and ν be two probability measures on a space Ω. Our goal is to study the most effective way to transport mass from μ to ν. We denote $U(\mu, \nu)$ the set of admissible transport maps

$$U(\mu, \nu) := \{\mathbb{P} \text{ probability on } \Omega \times \Omega \text{ s.t. } \mathbb{P}(E \times \Omega) = \mu(E), \mathbb{P}(\Omega \times E) = \nu(E) \text{ for all } E \in \mathcal{F}\}.$$

We call $\mathbb{P} \in U(\mu, \nu)$ a *transport map* or a *coupling* of μ and ν without distinction. Intuitively, $\mathbb{P}(dx, dy)$ is the amount of mass transported from point x to point y, and thus the marginals of $\mathbb{P}$ are μ and ν. We now assume that moving this quantity has a cost, that depends on x and y according to a certain function $c(x, y)$. Finding the best way (the cheapest) to transport the mass amounts to the minimization problem

$$\mathcal{W}_c(\mu, \nu) = \inf \left\{ \int_{\Omega^2} c(x, y)\mathbb{P}(dx, dy), \text{ s.t. } \mathbb{P} \in U(\mu, \nu) \right\}. \tag{2.1}$$

This way, optimal transport theory gives a natural metric between measures. How much would it cost to go from μ to ν at best ? We consider on $\mathcal{P}_p(\mathbb{R}^d)$, set of probability measures on $\mathbb{R}^d$ with finite p moments (i.e such that $\mathbb{E}(|\cdot|^p) < \infty$), the *Wasserstein distance*

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\mathbb{P} \in U(\mu,\nu)} \int |y - x|^p \mathbb{P}(dx, dy) \right)^{1/p}$$

where $|x|^p = \sum_{k=1}^d |x^k|^p$ is the L^p norm on $\mathbb{R}^d$. Important results on optimal transport can be found in *Optimal Transport : Old and New* of C. Villani [Vil08] or in *Optimal Transport for Applied Mathematicians* of F. Santambrogio [San15].

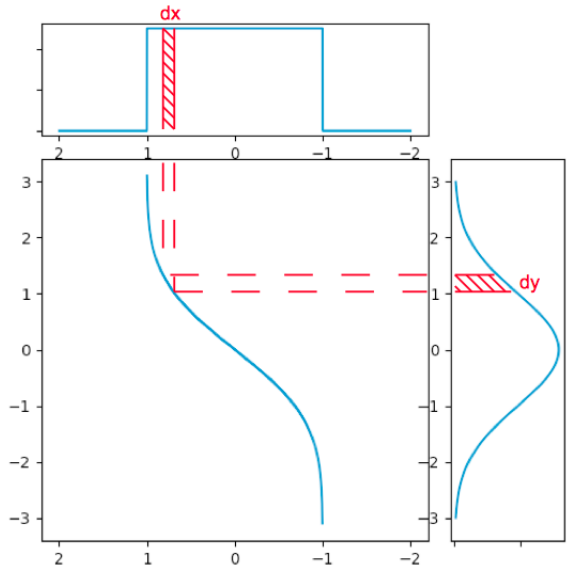

Figure 7: Example of a transport map (center) between the unifom distribution on [-1,1] (up) and the Gaussian law $\mathcal{N}(0,1)$ (right). Mass on dx is transported to dy.

We may write the Wasserstein distance between two probability measures μ and ν for an underlying distance c using probabilistic notations

$$\mathcal{W}_c(\mu,\nu) = \inf_{(X,Y)\sim\Gamma\in U(\mu,\nu)} \mathbb{E}\left(c(X,Y)\right).$$

It is the minimum over all coupling Γ of μ and ν of $\mathbb{E}\left(c(X,Y)\right)$, where (X,Y) is distributed according to Γ. Any coupling then yields an upper bound on the Wasserstein distance.

2.2 A toy model

Let us show, on a toy model, how a coupling argument works. The following example is adapted from the course "Mixing times of Markov chains" by Justin Salez.

We consider the space $\mathbb{Z}/n\mathbb{Z}$, for a given $n \in \mathbb{N}$, and we consider the (lazy) random walk on this space. This process is defined, given a sequence of i.i.d variables $(\xi_i)_{i\geq 1}$ such that $\mathbb{P}(\xi = 0) = \frac{1}{2}$ and $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = \frac{1}{4}$, by

$$X_n = \left(X_0 + \sum_{i=1}^{n} \xi_i\right) \mod n.$$

We wish to prove the following intuitive result : *in the long run, it doesn't matter where the walk started.* We start with the invariant distribution. Denote $(X_t)_{t\geq 0}$ a random walk on $\mathbb{Z}/n\mathbb{Z}$ (see Figure 8). From the law of total probability we get

$$\begin{aligned}\mathbb{P}(X_t = k) =& \mathbb{P}(X_t = k|X_{t-1} = k+1)\mathbb{P}(X_{t-1} = k+1) + \mathbb{P}(X_t = k|X_{t-1} = k)\mathbb{P}(X_{t-1} = k)\\ &+ \mathbb{P}(X_t = k|X_{t-1} = k-1)\mathbb{P}(X_{t-1} = k-1)\\ =& \frac{1}{4}\mathbb{P}(X_{t-1} = k+1) + \frac{1}{2}\mathbb{P}(X_{t-1} = k) + \frac{1}{4}\mathbb{P}(X_{t-1} = k-1).\end{aligned}$$

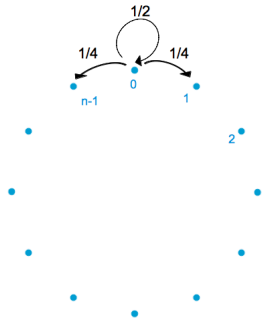

Figure 8: Illustration of the (lazy) random walk on the cycle $\mathbb{Z}/n\mathbb{Z}$.

Then $p := (p_k)_{k=1,..,n}$ is an invariant measure if and only if

$$\begin{aligned} p_k &= \frac{1}{4}p_{k+1} + \frac{1}{2}p_k + \frac{1}{4}p_{k-1} \\ \frac{1}{4}(p_k - p_{k-1}) &= \frac{1}{4}(p_{k+1} - p_k) \end{aligned}$$

The increments $(p_k - p_{k-1})_{k=1,..,n}$ are therefore constant. Since

$$0 = p_0 - p_0 = p_n - p_0 = \sum_{k=1}^{n} p_k - p_{k-1} = n(p_1 - p_0),$$

we get that the invariant measure is the uniform distribution. Conversely we prove that the uniform distribution is invariant. We thus have existence and uniqueness of the invariant measure for this process. Results on Markov chains then tell us that, given any initial distribution, the process will tend to this distribution. The remaining question is : how quickly will it converge ?

Figure 9: Evolution of the probability density of the (lazy) random walk on the cycle $\mathbb{Z}/n\mathbb{Z}$ starting in 0.

We consider on the cycle $\mathbb{Z}/n\mathbb{Z}$ the norm $||x||_c = \min(|x|, n - |x|)$. Let μ and ν be two probability measures on $\mathbb{Z}/n\mathbb{Z}$, we denote

$$\mathcal{W}_c(\mu, \nu) = \inf_{(X,Y)\sim\Gamma\in U(\mu,\nu)} \mathbb{E}\left(||X - Y||_c\right), \tag{2.2}$$

the Wasserstein distance associated to the norm $||\cdot||_c$. We may now prove the following result.

Theorem 2.1. *Let $x, y \in \mathbb{Z}/n\mathbb{Z}$. Let μ_t (resp. ν_t) be the distribution of the (lazy) random walk X_t (resp. Y_t) on $\mathbb{Z}/n\mathbb{Z}$ starting in $X_0 \sim \mu_0$ (resp. $Y_0 \sim \nu_0$) We have*

$$\mathcal{W}_c(\mu_t, \nu_t) \leq \frac{n^2}{4t}\mathcal{W}_c(\mu_0, \nu_0)$$

Proof. Recall the definition of the Wasserstein distance (2.2). Instead of considering the minimum over all coupling of μ_t and ν_t, we will construct a specific one that will, hopefully, converge. To this end, we may consider any vector (X_t, Y_t) as long as it has the correct marginals. To prove the result, we are thus looking for a coupling of X_t and Y_t that would encourage the two walks to meet, i.e that would decrease $\mathbb{E}\left(||X - Y||_c\right)$, which would give us a better bound on $\mathcal{W}_c(\mu_t, \nu_t)$. Let Γ be any coupling of (μ_0, ν_0), and $(X_0, Y_0) \sim \Gamma$.

- **First coupling : synchronous.** We write $X_t := X_0 + \sum_{i=1}^t \xi_i$. The first idea, as a coupling of μ_t and ν_t, could be to give the same increments to both walks, and consider $Y_t := Y_0 + \sum_{i=1}^t \xi_i$. However, one can easily see that for all t, the difference $X_t - Y_t$ would be constant, and then

$$\forall t, \mathbb{E}\left(||X_t - Y_t||_c\right) = \mathbb{E}\left(||X_0 - Y_0||_c\right).$$

  We cannot hope to use this coupling.

- **Second coupling : mirror.** To help both walks meet, it is natural to consider opposite increments, i.e $X_t := X_0 + \sum_{i=1}^t \xi_i$ and $Y_t := Y_0 - \sum_{i=1}^t \xi_i$. Here, another problem arises, as when n is even, the difference $X_t - Y_t$ keeps its parity, such that

$$\forall t, \mathbb{E}\left(||X_t - Y_t||_c\right) > 1,$$

  when $X_0 - Y_0$ is odd.

- **Third coupling : alternate.** The solution to this issue of parity is to have one process jump when the other doesn't until they meet, and then use a synchronous coupling to have them stick together. Let

$$T := \min\{t \in \mathbb{N}, X_t = Y_t\}.$$

  Using this coupling

$$W_t := X_t - Y_t = \begin{cases} X_0 - Y_0 + \sum_{i=1}^t \tilde{\xi}_i, & t < T \\ 0, & t \geq T \end{cases}$$

  where $(\tilde{\xi}_i)_{i\in\mathbb{N}}$ are i.i.d random variables satisfying $\mathbb{P}(\tilde{\xi}_1 = -1) = \mathbb{P}(\tilde{\xi}_1 = 1) = \frac{1}{2}$. In other words, W_t is the usual symmetric random walk on $\mathbb{Z}/n\mathbb{Z}$. We can see T as the stopping time W_t either reaches 0 or n starting from $X_0 - Y_0$, and we have $\{X_t \neq Y_t\} = \{T > t\}$. Using Markov's inequality, we have the very coarse upper bound on the expectation of the distance

$$\mathbb{E}\left(||X_t - Y_t||_c\right) \leq \frac{n}{2}\mathbb{P}(X_t \neq Y_t) = \frac{n}{2}\mathbb{P}(T > t) \leq \frac{n}{2}\frac{\mathbb{E}(T)}{t}.$$

  $\mathbb{E}(T)$ can then be calculated, using usual results on martingales, to show that

$$\mathbb{E}(T) = \mathbb{E}\left((n - |X_0 - Y_0|)|X_0 - Y_0|\right) \leq \frac{n}{2}\mathbb{E}\left(||X_0 - Y_0||_c\right).$$

  We now have

$$\mathbb{E}\left(||X_t - Y_t||_c\right) \leq \frac{n^2}{4t}\mathbb{E}\left(||X_0 - Y_0||_c\right).$$

  On one hand, by definition of the Wasserstein distance $\mathcal{W}_c(\mu_t, \nu_t) \leq \mathbb{E}\left(||X_t - Y_t||_c\right)$, and on the other hand, the inequality above being true for all initial coupling Γ, we obtain

$$\mathcal{W}_c(\mu_t, \nu_t) \leq \frac{n^2}{4t}\mathcal{W}_c(\mu_0, \nu_0).$$

□

Remark 2.1. *One can notice that most of the inequalities above are far from being optimal. This theorem is indeed not made to be used as a result, only to be shown as an example of coupling.*

3 Coupling methods for the study of the long time behavior of PDE

Let us tackle the main point of these notes. We have seen, in Section 1, how one can interpret some PDEs as the evolution of the probability distribution of a random process. Then, in Section 2, we studied how coupling methods may be used to prove convergence results for random processes. We now combine those two ideas to prove convergence results for PDEs, and use various coupling to do so.

3.1 Synchronous coupling

Given a probability distribution μ_0, consider the PDE

$$\begin{cases} \partial_t u = -\nabla \cdot (b(x)u) + \Delta u, \ \ t \in [0,T] \\ u(0,x) = \mu_0(x), \end{cases} \tag{3.1}$$

with the following assumption on b

Assumption 1. *b is Lipschitz continuous and satisfies*

$$\exists c > 0, \ \forall x, y \in \mathbb{R}, \ (b(x) - b(y)) \cdot (x - y) \leq -c|x-y|^2$$

As you recall from Theorem 1.3, if you consider the following SDE

$$\begin{cases} dX_t = b(X_t)dt + \sqrt{2}dB_t \\ X_0 \sim \mu_0 \end{cases} \tag{3.2}$$

where B_t is a Brownian motion, then the law of X satisfies (3.1). Under the Lipschitz continuous assumption on b, we have weak (and even strong) existence and uniqueness of the solution X of the SDE, that is, given any two realizations B and $\tilde{B}$ of Brownian motion, the resulting processes X and $\tilde{X}$ have the same law. The solution μ of (3.1) is therefore **the** law shared by all solutions of the SDE. We may now show the following theorem.

Theorem 3.1. *Consider two solutions μ^1 and μ^2 of* (3.1) *with initial conditions μ_0^1 and μ_0^2 respectively, then*

$$\mathcal{W}_2(\mu_t^1, \mu_t^2) \leq e^{-ct}\mathcal{W}_2(\mu_0^1, \mu_0^2)$$

Proof. Consider any coupling Γ of μ_0^1 and μ_0^2 and let X^1 and X^2 be two processes starting respectively from $(X_0^1, X_0^2) \sim \Gamma$. Suppose X^1 and X^2 satisfy

$$\begin{cases} dX_t^1 = b(X_t^1)dt + \sqrt{2}dB_t \\ dX_t^2 = b(X_t^2)dt + \sqrt{2}dB_t. \end{cases} \tag{3.3}$$

In other words, we couple X^1 and X^2 by using the same Brownian motion. The processes X_t^i have for law μ_t^i, $i = 1, 2$, thanks to the weak uniqueness of the solutions of the SDE. We now calculate

$$\begin{aligned} d(X_t^1 - X_t^2) =& (b(X_t^1) - b(X_t^2))dt \\ d|X_t^1 - X_t^2|^2 =& 2(b(X_t^1) - b(X_t^2)) \cdot (X_t^1 - X_t^2)dt. \end{aligned}$$

Under Assumption 1, we have

$$\frac{d}{dt}\mathbb{E}\left(|X_t^1 - X_t^2|^2\right) \leq -2c\mathbb{E}\left(|X_t^1 - X_t^2|^2\right),$$

and using Gronwall's lemma, we get

$$\mathcal{W}_2(\mu_t^1, \mu_t^2)^2 \leq \mathbb{E}\left(|X_t^1 - X_t^2|^2\right) \leq e^{-2ct}\mathbb{E}\left(|X_0^1 - X_0^2|^2\right).$$

This being true for all initial coupling Γ, we get

$$\mathcal{W}_2(\mu_t^1, \mu_t^2)^2 \leq e^{-2ct}\mathcal{W}_2(\mu_0^1, \mu_0^2)^2,$$

hence the result. □

We thus have shown that, provided the dynamic is naturally contracting, which is translated in Assumption 1, a synchronous coupling has the difference satisfy an ODE that will also contract.

Remark 3.1. *With the same proof, you show that, given the same initial condition and the same Brownian motion, two processes satisfying* (3.2) *are indistinguishable, i.e there is strong uniqueness of the solution of the SDE.*

Remark 3.2. *In these notes, we keep going back and forth between the PDE and the SDE. An interesting possibility would be to consider the Fokker-Planck equation associated to* (3.3)*. Applying Itô's lemma to calculate the dynamic of $f(X_t^1, X_t^2)$ for a test function f, one may see that the law Γ_t of the couple (X_t^1, X_t^2) satisfies*

$$\partial_t \Gamma_t(x,y) = -div_x(b(x)\Gamma_t(x,y)) - div_y(b(y)\Gamma_t(x,y)) + \Delta_x\Gamma_t(x,y) + \Delta_y\Gamma_t(x,y) + 2\nabla_x \cdot \nabla_y \Gamma_t(x,y), \tag{3.4}$$

with any initial condition Γ_0 coupling of μ_0^1 and μ_0^2. We can easily check that, by integrating the above equation with respect to y, $\mu_t = \int \Gamma_t(\cdot, y)dy$ satisfies (3.1) *and by integrating with respect to x, $\nu_t = \int \Gamma_t(x,\cdot)dx$ also satisfies* (3.1)*. The solution Γ_t of* (3.4) *is therefore a coupling of μ_t^1 and μ_t^2. If one does not wish to use any result on stochastic calculus, and use a purely PDE approach, one can directly start from* (3.4)*, and prove the same result.*

$$\begin{aligned}
\frac{d}{dt}\int\int |x-y|^2\Gamma_t(x,y)dxdy =& -\int\int |x-y|^2 div_x(b(x)\Gamma_t(x,y)) - \int\int |x-y|^2 div_y(b(y)\Gamma_t(x,y)) \\
&+ \int\int |x-y|^2\Delta_x\Gamma_t(x,y) + \int\int |x-y|^2\Delta_y\Gamma_t(x,y) \\
&+ 2\int\int |x-y|^2\nabla_x\cdot\nabla_y\Gamma_t(x,y) \\
=& 2\int\int (x-y)\cdot b(x)\Gamma_t(x,y) - 2\int\int (x-y)\cdot b(y)\Gamma_t(x,y) + 2d + 2d - 4d \\
=& 2\int\int (x-y)\cdot (b(x) - b(y)\Gamma_t(x,y) \\
\leq& -2c\int\int |x-y|^2\Gamma_t(x,y)dxdy,
\end{aligned}$$

and thus the same result using Gronwall's lemma. We refer to [FP19] for a more complete description of coupling method using only such PDE approaches. By doing so, although it is indeed mathematically correct, you lose the probabilistic interpretation (where does (3.4) *come from ?), which is in itself both interesting and useful, as shown in the next section.*

3.2 Mirror coupling

Let us consider again (3.1), but with a new assumption on b

Assumption 2. *We make the following assumptions on b*

- *There is $L > 0$ such that $\forall x, y \in \mathbb{R}$, $|b(x) - b(y)| \leq L|x-y|$.*
- *There are $\lambda > 0$ and $A \geq 0$ such that $\forall x \in \mathbb{R}$, $x \cdot b(x) \leq A - \frac{\lambda}{2}|x|^2$*

The first assumption ensures that b is Lipschitz continuous, and the second ensures that b tends to bring back particles if they venture at infinity.

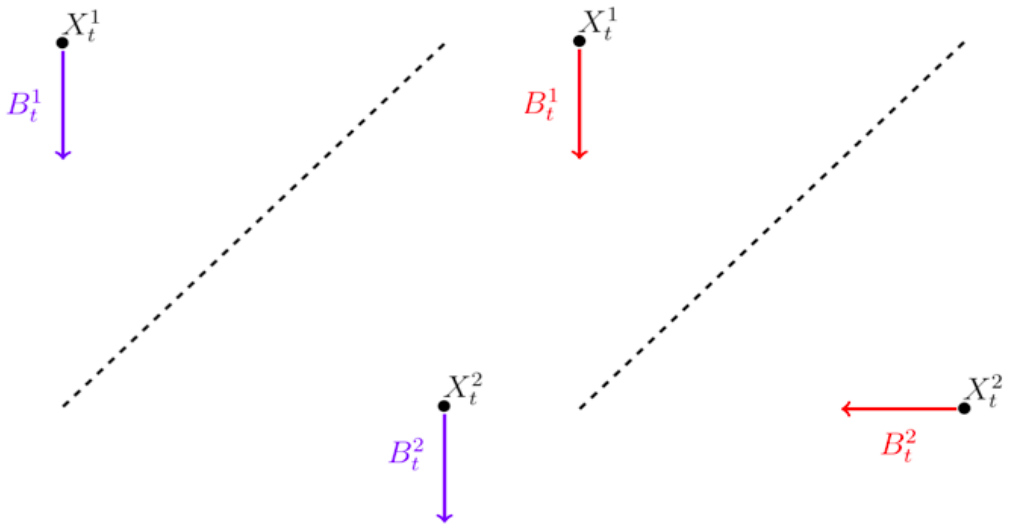

Figure 10: **Left :** Synchronous coupling. **Right :** Mirror, or reflection, coupling.

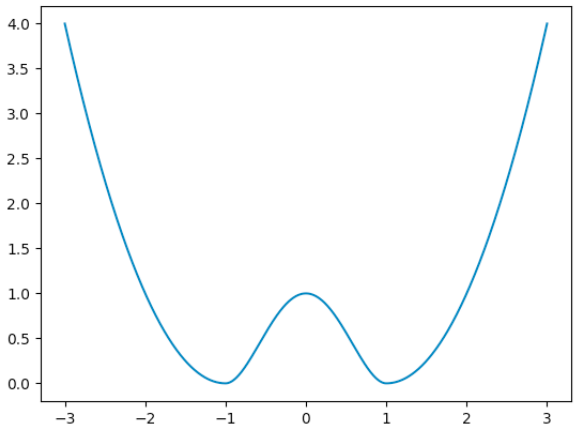

Figure 11: Double well potential U given in Remark 3.1. The force $b = -\nabla U$ satisfies Assumption 2.

Example 3.1. *We consider $U \in \mathcal{C}^1$ given by*

$$U(x) = \begin{cases} \left(x^2-1\right)^2 & \text{if } |x| \leq 1, \\ \left(|x|-1\right)^2 & \text{otherwise.} \end{cases}$$

This defines a double-well potential, for which $b = -\nabla U$ satisfies Assumption 2.

We begin by a small lemma that translates the fact that any particle that ventures at infinity tends to come back. Let, for $u \in \mathcal{C}^2(\mathbb{R}^d)$

$$\mathcal{L}u(x) = \nabla u(x) \cdot b(x) + \Delta u(x)$$

Lemma 3.1. *Let $H(x) = \frac{|x|^2}{2}$, we have*

$$\mathcal{L}H(x) \leq d + A - \lambda H(x),$$

which in particular implies

$$\frac{d}{dt}\mathbb{E}H(X_t) \leq d + A - \lambda \mathbb{E}H(X_t),$$

Proof. We have

$$\mathcal{L}H(x) = x \cdot b(x) + d \leq d + A - \lambda H(x).$$

And using Itô's formula, we have

$$dH(X_t) = \mathcal{L}H(X_t)dt + \sqrt{2}x \cdot dB_t,$$

hence the result. □

We will also use the fact that for all $\epsilon > 0$, there is a constant $R > 0$ such that, for all $x, v \in \mathbb{R}^d$ such that $|x - v|^2 \geq R$

$$|x|^2 + |v|^2 \geq 8\epsilon \frac{A + d}{\lambda}$$

We wish to prove the following result

Theorem 3.2. *There is an explicit $c > 0$ such that, for any two solutions μ^1 and μ^2 of* (3.1) *with initial conditions μ_0^1 and μ_0^2 respectively, there is an explicit constant C_0 depending only on μ_0^1 and μ_0^2 such that for all $t \geq 0$*

$$\mathcal{W}_2(\mu_t^1, \mu_t^2) \leq C_0 e^{-ct},$$

What will happen ? Before constructing our coupling argument, we need to understand how the process will act. We identify two main behaviors. Either the particle ventures at infinity, in which case the assumption ensuring that the potential is sufficiently convex outside a ball will bring back the particle, or the particle moves within that compact ball. When we construct our coupling, we thus see that a synchronous coupling would be sufficient outside the ball, as the dynamic will be naturally contracting, but within, we need another form of coupling to bring the coupled particle closer together. Being inspired by what we did in Section 2.2, we choose mirrored Brownian motions. We call it *mirror* or *reflection* coupling. See Figure 10 for a visualization of the mirror coupling, or below in (3.5) for a mathematical definition. Doing so will maximise the variance of the noise in the direction needed. But there is at this stage no reason for the noise to actually bring the processes closer to one another, as we expect the noise to both contract and expand their distance symmetrically. This is why we apply a concave function f to the quadratic distance $|x-y|^2$: the contraction is produced by the fact that a random decrease in $|x - y|^2$ has more effect on $f(|x - y|^2)$ than a random increase of the same amount.

Creating a distance. Knowing that behavior, we construct a new semimetrics, i.e a quantity we consider the dynamic of. Given two processes X^1 and X^2, we denote

$$r_t = |X_t^1 - X_t^2|^2,$$

and we consider a concave, non-negative, increasing function f, such that $f(0) = 0$ and f is constant for $r \geq R$, where R is the radius of the ball we consider. To tackle the processes at infinity, we then use the function H introduced in Lemma 3.1 and construct

$$G_t = 1 + \epsilon H(X_t^1) + \epsilon H(X_t^2).$$

We thus consider the semimetrics

$$\rho_t = f(r_t)G_t.$$

It will be easy to prove that $|X_t^1 - X_t^2|^2 \lesssim \rho_t$, or in other words that ρ_t controls the L^2 usual norm up to a universal constant. At this point, f and ϵ are not yet specified. We keep studying the dynamic, and construct these objects as we go.

Coupling. Consider any coupling Γ of μ_0^1 and μ_0^2 and let X^1 and X^2 be two processes starting from $(X_0^1, X_0^2) \sim \Gamma$. Suppose X^1 and X^2 satisfy

$$\begin{cases} dX_t^1 = b(X_t^1)dt + \sqrt{2}dB_t \\ dX_t^2 = b(X_t^2)dt + \sqrt{2}(Id - 2e_t e_t^*)dB_t, \end{cases} \tag{3.5}$$

where B_t is a Brownian motion, and

$$e_t = \begin{cases} \frac{X_t^1 - X_t^2}{|X_t^1 - X_t^2|} \text{ if } X_t^1 \neq X_t^2, \\ 0 \text{ otherwise,} \end{cases}$$

and e_t^* is the transpose of e_t. Recall the last point of Lemma 1.1, which guarantees that any rotation of a Brownian motion is a Brownian motion, and thus $(Id - 2e_t e_t^*)B_t$ is indeed a Brownian motion. This way, the law X_t^2 does satisfy the right PDE and, since the distribution of X_0^2 is μ_0^2, the law of X_t^2 is μ_t^2. Furthermore, the operation $Id - 2e_t e_t^*$ consists in taking a *mirror* coupling (see Figure 10) : it is -1 in the direction of space given by the difference of the processes, and $+1$ in the perpendicular direction.

Remark 3.3. *Similarly as Remark 3.2, we may write the coupling using only PDE tools. One can check that the solution of*

$$\begin{aligned} \partial_t \Gamma_t(x,y) = &- div_x(b(x)\Gamma_t(x,y)) - div_y(b(y)\Gamma_t(x,y)) + \Delta_x \Gamma_t(x,y) + \Delta_y \Gamma_t(x,y) \\ &+ 2\nabla_x \cdot \nabla_y \Gamma_t(x,y) - 2\nabla_y \nabla_x : (ee^* \Gamma_t(x,y)) - 2\nabla_x \nabla_y : (ee^* \Gamma_t(x,y)), \end{aligned}$$

where $e = \frac{x-y}{|x-y|}$ if $x \neq y$, 0 otherwise, is a coupling of μ_t^1 and μ_t^2.

Here, we denote $\nabla_y \nabla_x : A = \sum_{\alpha,\beta} \partial_{y_\alpha} \partial_{x_\beta} A_{\alpha,\beta}$.

We compute

$$dr_t = 2(X_t^1 - X_t^2) \cdot (b(X_t^1) - b(X_t^2))dt + 4\sqrt{2}(X_t^1 - X_t^2) \cdot e_t e_t^* dB_t + 8dt,$$

$$\begin{aligned} df(r_t) = &2f'(r_t)(X_t^1 - X_t^2) \cdot (b(X_t^1) - b(X_t^2))dt + 8f'(r_t)dt + 4\sqrt{2}f'(r_t)(X_t^1 - X_t^2) \cdot e_t e_t^* dB_t \\ &+ 16f''(r_t)dt, \end{aligned}$$

and

$$dG_t = \epsilon \left(\mathcal{L}H(X_t^1) + \mathcal{L}H(X_t^2) \right) dt + \epsilon\sqrt{2} \left(X_t^1 - X_t^2 \right) \cdot e_t e_t^* dB_t + \epsilon\sqrt{2} \left(X_t^1 + X_t^2 \right) \cdot (Id - e_t e_t^*) \, dB_t.$$

Then

$$f(r_t)G_t = f(r_0)G_0 + A_t + M_t$$

where M_t is a stochastic integral and

$$\begin{aligned} dA_t = &K_t dt, \\ \text{with } K_t = &2f'(r_t)G_t(X_t^1 - X_t^2) \cdot (b(X_t^1) - b(X_t^2)) + 8f'(r_t)G_t + 16f''(r_t)G_t \\ &+ \epsilon \left(\mathcal{L}H(X_t^1) + \mathcal{L}H(X_t^2) \right) f(r_t) + 8\epsilon |X_t^1 - X_t^2|^2 f'(r_t). \end{aligned}$$

At this point, when considering the expectation of $f(r_t)G_t$, M_t will disappear and we will be left with A_t. We thus need to control K_t. We consider this quantity in the various regions of space we identified, and thus justify its construction.

First region of space : $r \leq R$**.** In this region, we use the construction of the function f to contract the dynamic.

$$\begin{aligned} K_t \leq & f'(r_t) G_t \left(2L + 8\epsilon\right) r_t + 8 f'(r_t) G_t + 16 f''(r_t) G_t + \epsilon \left(2A + 2d - \frac{\lambda}{2}(|X_t^1|^2 + |X_t^2|^2)\right) f(r_t) \\ & \leq \left((2A + 2d)\epsilon f(r_t) + f'(r_t)\left(2L + 8\epsilon\right) r_t + 8 f'(r_t) + 16 f''(r_t)\right) G_t. \end{aligned}$$

We construct f so as to have

$$f'(r_t)\left(2L + 8\epsilon\right) r_t + 8 f'(r_t) + 16 f''(r_t) \leq (-c - 2A\epsilon - 2d\epsilon) f(r_t).$$

We hence consider, using the ideas of [Ebe16]

$$\begin{aligned} f(r) = & \int_0^{\min(r,R)} \phi(s) g(s) ds \\ \text{with } \phi(r) = & \exp(\frac{1}{16}(-(2L + 8\epsilon) r^2 - 8r)) \\ g(r) = & 1 - \frac{c + 2A\epsilon + 2d\epsilon}{16} \int_0^{\min(r,R)} \frac{\Phi(s)}{\phi(s)} ds \\ \text{and } \Phi(r) = & \int_0^r \phi(s) ds. \end{aligned}$$

To keep f increasing, we add the condition $g(r) \geq \frac{1}{2}$. This way

$$\begin{aligned} f'(r_t)\left(2L + 8\epsilon\right) r_t + 8 f'(r_t) + 16 f''(r_t) = & \left(\phi(r_t)\left(2L + 8\epsilon\right) r_t + 8\phi(r_t) + 16\phi'(r_t)\right) g(r_t) + 16\phi(r_t) g'(r_t) \\ = & 16\phi(r_t) g'(r_t) \\ \leq & -(c + 2A\epsilon + 2d\epsilon)\Phi(r) \leq -(c + 2A\epsilon + 2d\epsilon) f(r). \end{aligned}$$

And thus

$$K_t \leq -c f(r_t) G_t.$$

Second region of space : $r \geq R$**.** In this region, $f'(r) = f''(r) = 0$.

$$K_t = \epsilon \left(2A + 2d - \frac{\lambda}{2}(|X_t^1|^2 + |X_t^2|^2)\right) f(r_t).$$

For $r \geq R$, $|X_t^1|^2 + |X_t^2|^2 \geq 8\epsilon \frac{A+d}{\lambda}$. We assume $c\epsilon \leq \frac{\lambda}{2}$. This way we have

$$c + 2A\epsilon + 2d\epsilon + (c\epsilon - \frac{\lambda}{2})(|X_t^1|^2 + |X_t^2|^2) \leq c\left(1 + 8\epsilon^2 \frac{A+d}{\lambda}\right) - 2\epsilon(A + d),$$

and we choose $c \leq \frac{2\epsilon\lambda(A+d)}{\lambda + 8\epsilon^2(A+d)}$ so that

$$\epsilon \left(2A + 2d - \frac{\lambda}{2}(|X_t^1|^2 + |X_t^2|^2)\right) \leq - cG_t.$$

Hence

$$K_t \leq -c f(r_t) G_t.$$

Contraction. We have, in all region of space

$$K_t \leq -cf(r_t)G_t.$$

Taking the expectation in the dynamic,

$$\frac{d}{dt}\mathbb{E}(f(r_t)G_t) \leq -c\mathbb{E}(f(r_t)G_t).$$

We conclude using Gronwall's lemma, and the fact that there is a constant C such that

$$\rho_t = f(r_t)G_t \geq C|X_t^1 - X_t^2|.$$

Explicit parameters. We have accumulated the following conditions

$$\begin{aligned}
\frac{1}{2} =& 1 - \frac{c + 2A\epsilon + 2d\epsilon}{16}\int_0^R \frac{\Phi(s)}{\phi(s)}ds, \text{ so as to have } g \geq \frac{1}{2},\\
c \leq& \frac{\lambda}{2\epsilon}, \text{ in the second region,}\\
c \leq& \frac{2\epsilon\lambda(A+d)}{\lambda + 8\epsilon^2(A+d)}, \text{ also in the second region.}
\end{aligned}$$

We accept that there are positive and explicit constants R, c and ϵ satisfying those conditions.

Conclusion. Even though we ignored some technical difficulties, our goal was to show how, when given a PDE, we may construct a proof using coupling methods and our understanding of what should happen. Furthermore, every quantity can be chosen explicitly. We thus obtain a quantitative and intuitive proof of contraction. Obviously, there are cases where a purely PDE approach does work, but only wished to give an alternative proof. This type of proof can be extended, with other tools, to more complicated dynamics.

References

[Ber21] Nils Berglund. Long-time dynamics of stochastic differential equations. *arXiv e-prints*, page arXiv:2106.12998, June 2021.

[Ebe16] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Relat. Fields*, 166(3-4):851–886, 2016.

[FP19] Nicolas Fournier and Benoît Perthame. Monge-kantorovich distance for pdes: the coupling method, 2019.

[LG18] Jean-Francois Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer Publishing Company, Incorporated, 2018.

[San15] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 55. Birkäuser, NY, 2015.

[Vil08] Cédric Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.