Time Averages of Markov Processes and Applications to Two-Timescale Problems

Bob Pepin *

Mathematics Research Unit, FSTC, University of Luxembourg
Maison du Nombre, 4364 Esch-sur-Alzette, Grand-Duchy of Luxembourg

February 8, 2018

Abstract

We show a decomposition into the sum of a martingale and a deterministic quantity for time averages of the solutions to non-autonomous SDEs and for discrete-time Markov processes. In the SDE case the martingale has an explicit representation in terms of the gradient of the associated semigroup or transition operator. We show how the results can be used to obtain quenched Gaussian concentration inequalities for time averages and to provide deeper insights into Averaging principles for two-timescale processes.

1 Introduction

For a Markov process $(X_t)_t$ with $t \in [0,T]$ or $t = 0, 1, \ldots, T$ let

$$S_T f = \int_0^T f(t, X_t)dt$$

in the continuous-time case or

$$S_T f = \sum_{t=0}^{T-1} f(t, X_t)$$

in discrete time.

In the first part of this work, we will show a decomposition of the form

$$S_T f = \mathbb{E}S_T f + M_T^{T,f}$$

where $M^{T,f}$ is a martingale depending on T and f for which we will give an explicit representation in terms of the transition operator or semigroup associated to X.

*bob.pepin@uni.lu or bobpepin@gmail.com

We then proceed to illustrate how the previous results can be used to obtain Gaussian concentration inequalities for S_T when X is the solution to an Itô SDE.

The last part of the work showcases a number of results on two-timescale processes that follow from our martingale representation.

2 Martingale Representation

Consider the following SDE with time-dependent coefficients on $\mathbb{R}^n$:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, X_0 = x$$

where B is a standard Brownian motion on $\mathbb{R}^n$ with filtration $(\mathcal{F}_t)_{t\geq 0}$ and $b(t,x), \sigma(t,x)$ are continuous in t and locally Lipschitz continuous in x. We assume that X_t does not explode in finite time.

Denote C_c^∞ the set of smooth compactly supported space-time functions on $\mathbb{R}_+ \times \mathbb{R}^n$.

Let $P_{s,t}$ be the evolution operator associated to X,

$$P_{s,t}f(x) = \mathbb{E}\left[f(t, X_t)|X_s = x\right], \quad f \in C_c^\infty.$$

For $T > 0$ fixed consider the martingale

$$M_t = \mathbb{E}^{\mathcal{F}_t} \int_0^T f(s, X_s)ds.$$

and observe that since X is adapted and by the Markov property

$$M_t = \int_0^t f(s, X_s)ds + \mathbb{E}^{\mathcal{F}_t} \int_t^T f(s, X_s)ds = \int_0^t f(s, X_s)ds + R_t^T f(X_t)$$

with

$$R_t^T f(x) = \int_t^T P_{t,s} f(x)ds.$$

By applying the Itô formula to $R_t^T f$ we can identify the martingale M. This is the content of the following short theorem.

Theorem 2.1. *For $T > 0$ fixed, $t \in [0, T]$ and $f \in C_c^\infty$*

$$\int_0^t f(s, X_s)ds + R_t^T f(X_t) = \mathbb{E} \int_0^T f(s, X_s)ds + M_t^{T,f}$$

with

$$M_t^{T,f} = \int_0^t \nabla R_s^T f(X_s) \cdot \sigma(s, X_s)dB_s.$$

Proof. From the Kolmogorov backward equation $\partial_t P_{t,s} f = -L_t P_{t,s} f$ and since $P_{t,t} f = f$ we have

$$\partial_t R_t^T f(x) = -f(t,x) - \int_t^T L_t P_{t,s} f(x) ds = -f(t,x) - L_t R_t^T f(x).$$

By Itô's formula

$$\begin{aligned} R_t^T f(X_t) &= R_0^T f(X_0) + \int_0^t \partial_s R_s^T f(X_s) ds + \int_0^t L_s R_s^T f(X_s) ds + \int_0^t \nabla R_s^T f(X_s) \cdot \sigma(s, X_s) dB_s \\ &= \mathbb{E} \int_0^T f(t, X_t) dt - \int_0^t f(s, X_s) ds + \int_0^t \nabla R_s^T f(X_s) \cdot \sigma(s, X_s) dB_s \end{aligned}$$

and we are done.

□

Remark 2.2 (Poisson Equation). In the time-homogeneous case $P_{t,s} = P_{s-t}$ and when the limit below is finite then it is independent of t and we have

$$R^\infty f := \lim_{T\to\infty} R_t^T f = \lim_{T\to\infty} \int_t^T P_{s-t} f ds = \lim_{T\to\infty} \int_0^{T-t} P_s f ds = \int_0^\infty P_s f ds.$$

This is the resolvent formula for the solution to the Poisson equation $-Lg = f$ with $g = R^\infty f$.

By taking $t = T$ in Theorem 2.1 we can identify the martingale part in the martingale representation theorem for $\int_0^T f(t, X_t) dt$.

Corollary 2.3. *For $T > 0$ fixed, $f \in C_c^\infty$*

$$\int_0^T f(t, X_t) dt - \mathbb{E} \int_0^T f(t, X_t) dt = \int_0^T \nabla \int_t^T P_{t,s} f(X_t) ds \cdot \sigma(t, X_t) dB_t.$$

By applying the Itô formula to $P_{t,T} f(X_t)$ we obtain for $T > 0$ fixed

$$dP_{t,T} f(X_t) = \nabla P_{t,T} f(X_t) \cdot \sigma(t, X_t) dB_t \tag{2.1}$$

and by integrating from 0 to T

$$f(T, X_T) = \mathbb{E}\left[f(T, X_T)\right] + \int_0^T \nabla P_{t,T} f(X_t) \cdot \sigma(t, X_t) dB_t.$$

This was observed at least as far back as [EK89] and is commonly used in the derivation of probabilistic formulas for $\nabla P_{s,t}$.

Combining the formula (2.1) with Theorem 2.1 we obtain the following expression for $S_t - \mathbb{E} S_t$ in terms of $\nabla P_{s,t} f$.

Corollary 2.4. *For $f \in C_c^\infty$, $T > 0$ fixed and any $t < T$*

$$\int_0^t f(s, X_s) - \mathbb{E}f(s, X_s) ds = M_t^{T,f} - Z_t^{T,f}$$

with

$$\begin{aligned} Z_t^{T,f} &= \int_t^T \int_0^t \nabla P_{r,s} f(r, X_r) \cdot \sigma(r, X_r) \, dB_r \, ds \\ M_t^{T,f} &= \int_0^t \int_r^T \nabla P_{r,s} f(r, X_r) \, ds \cdot \sigma(r, X_r) \, dB_r. \end{aligned}$$

Proof. Let $f_0(t,x) = f(t,x) - \mathbb{E}f(t, X_t) = f(t,x) - P_{0,t}(x_0)$. We have

$$\begin{aligned} R_t^T f_0(X_t) &= \int_t^T P_{t,s} f_0(X_t) ds \\ &= \int_t^T P_{t,s} f(X_t) - P_{0,s} f(X_0) ds \\ &= \int_t^T \int_0^t \nabla P_{r,s} f(r, X_r) \cdot \sigma(r, X_r) dB_r \, ds \end{aligned}$$

where the last equality follows by integrating (2.1) from 0 to t (with $T = s$). Since $R_0^T f_0 = 0$ and $\nabla P_{t,s} f_0 = \nabla P_{t,s} f$ we get from Theorem 2.1 that

$$\int_0^t f_0(s, X_s) ds = M_t^{T,f} - R_t^T f_0(X_t)$$

and the result follows with $Z_t^{T,f} = R_t^T f_0(X_t)$. □

Remark 2.5 (Carré du Champs and Mixing). For differentiable functions f, g let

$$\Gamma_t(f,g)(x) = \tfrac{1}{2} \nabla f(t,x) (\sigma \sigma^\top)(t,x) \nabla g(t,x).$$

Then we have the following expression for the quadratic variation of $M^{T,f}$:

$$\begin{aligned} d\langle M^{T,f} \rangle_t &= \left| \int_t^T \sigma(t, X_t)^\top \nabla P_{t,s} f(X_t) \, ds \right|^2 dt \\ &= \left(4 \int_{t \le s \le r \le T} \Gamma_t(P_{t,s} f, P_{t,r} f)(X_t) \, dr \, ds \right) dt. \end{aligned}$$

Furthermore, since

$$\partial_s P_{r,s}(P_{s,t} f P_{s,t} g) = 2 P_{r,s}(\Gamma_s(P_{s,t} f, P_{s,t} g))$$

and setting $g(t,x) = \int_t^T P_{t,s}f(x)ds$ we have

$$\begin{aligned}
E\langle M^{T,f}\rangle_t &= 2\int_0^T \int_t^T 2P_{0,t}\Gamma_t(P_{t,s}f, P_{t,s}g)ds\,dt \\
&= 2\int_0^T \int_t^T \partial_t P_{0,t}(P_{t,s}fP_{t,s}g)ds\,dt \\
&= 2\int_0^T \partial_t \int_t^T P_{0,t}(P_{t,s}fP_{t,s}g)ds\,dt + 2\int_0^T P_{0,t}(fg)dt \\
&= 2\int_0^T P_{0,t}(fg) - P_{0,t}fP_{0,t}g\,dt \\
&= 2\int_{0\le t\le s\le T} \mathrm{Cov}(f(t,X_t), f(s,X_s))ds\,dt.
\end{aligned}$$

This shows how the expressions we obtain in terms of the gradient of the semigroup relate to mixing properties of X.

Remark 2.6 (Pathwise estimates). We would like to have a similar estimate for

$$\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t f(X_s) - \mathbb{E}f(X_s)ds\right|.$$

Setting

$$f_0(t,x) = f(x) - \mathbb{E}f(X_t) = f(x) - P_{0,t}f(x_0)$$

we have

$$\begin{aligned}
\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t f(X_s) - \mathbb{E}f(X_s)ds\right| &\le \mathbb{E}\sup_{0\le t\le T}|M_t^{T,f_0}| + \mathbb{E}\sup_{0\le t\le T}|R_t^T f_0(X_t)| \\
&\le 2\left(\mathbb{E}\langle M^{T,f_0}\rangle_T\right)^{1/2} + \mathbb{E}\sup_{0\le t\le T}|R_t^T f_0(X_t)|
\end{aligned}$$

and

$$\begin{aligned}
R_t^T f_0(X_t) &= \int_t^T P_{t,s}f(X_t) - P_{0,s}f(x_0)ds \\
&= \int_t^T \int_0^t \nabla P_{r,s}f(X_r)\cdot\sigma(r,X_r)dB_r ds
\end{aligned}$$

where the last equality follows from (for s fixed)

$$dP_{t,s}f(X_t) = \nabla P_{t,s}f(X_t)\cdot\sigma(t,X_t)dB_t.$$

2.1 Discrete time

Consider a discrete-time Markov process $(X_n)_{n=1...N}$ with transition operator

$$P_{m,n}f(x) = \mathbb{E}[f_n(X_n)|X_m = x]$$

and generator

$$L_n f(x) = P_{n,n+1}f(x) - f_n(x).$$

As in the continuous-time setting

$$M_n := f_n(X_n) - f_0(X_0) - \sum_{m=0}^{n-1} L_m f(X_m)$$

is a martingale (by the definition of L) and by direct calculation

$$M_n - M_{n-1} = f_n(X_n) - P_{n-1,n}f(X_{n-1}).$$

Let

$$R_n^N f(x) = \sum_{m=n}^{N-1} P_{n,m}f(x)$$

and observe that

$$L_n R^N f(x) = \sum_{m=n+1}^{N} P_{n,n+1}P_{n+1,m}f(x) - \sum_{m=n}^{N-1} P_{n,m}f(x) = -f_n(x).$$

Note that

$$R_N^N f(x) = 0 \text{ and } R_0^N f(x) = \mathbb{E}\left[\sum_{m=n}^{N-1} f(X_m)\middle| X_0 = x\right].$$

It follows that

$$\sum_{m=0}^{n-1} f_m(X_m) + R_n^N f(X_n) = -\sum_{m=0}^{n-1} L_m R^N f(X_m) + R_n^N f(X_n) = R_0^N f(X_0) + M_n^{N,f}$$

with

$$M_n^{N,f} - M_{n-1}^{N,f} = \sum_{m=n}^{N-1} P_{n,m}f(X_n) - P_{n-1,m}f(X_{n-1}).$$

Analogous to the continuous-time case, we define the carré du champs

$$\begin{aligned}\Gamma_n(f,g) &:= L_n(fg) - g_n L_n f - f_n L_n g \\ &= P_{n,n+1}(fg) - f_n P_{n,n+1}g - g_n P_{n,n+1}f + f_n g_n \\ &= \mathbb{E}\left[(f_{n+1}(X_{n+1}) - f_n(X_n))(g_{n+1}(X_{n+1}) - g_n(X_n))|\mathcal{F}_n\right]\end{aligned}$$

and using the summation by parts formula

$$\begin{aligned}
\langle M^{N,f}\rangle_n - \langle M^{N,f}\rangle_{n-1} &= \mathbb{E}[(M_n^{N,f} - M_{n-1}^{N,f})^2|\mathcal{F}_{n-1}] \\
&= 2 \sum_{n\le k\le m<N-1} \mathbb{E}\left[(P_{n,m}f(X_n) - P_{n-1,m}f(X_{n-1}))(P_{n,k}f(X_n) - P_{n-1,k}f(X_{n-1}))|\mathcal{F}_{n-1}\right] \\
&\quad + \sum_{m=n}^{N-1} \mathbb{E}\left[(P_{n,m}f(X_n) - P_{n-1,m}f(X_{n-1}))^2|\mathcal{F}_{n-1}\right] \\
&= 2\sum_{m=n}^{N-1}\sum_{k=m}^{N-1} \Gamma_{n-1}(P_{n-1,m}f, P_{n-1,k}f)(X_{n-1}) + \sum_{m=n}^{N-1} \Gamma_{n-1}(P_{n-1,m}f)(X_{n-1}).
\end{aligned}$$

3 Concentration inequalities from exponential gradient bounds

In this section we focus on the case where we have uniform exponential decay of $\nabla P_{s,t}$ so that

$$|\sigma(s,x)^\top \nabla P_{s,t}f(x)| \le C_s e^{-\lambda_s(t-s)} \quad (0 \le s \le t \le T) \tag{3.1}$$

for all $x \in \mathbb{R}^n$ and some class of functions f.

We first show that exponential gradient decay implies a concentration inequality.

Proposition 3.1. *For $T > 0$ fixed and all functions f such that* (3.1) *holds we have*

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(t,X_t) - \mathbb{E}f(t,X_t)dt > R\right) \le e^{-\frac{R^2T}{V_T}}, \quad V_T = \frac{1}{T}\int_0^T \left(\frac{C_t}{\lambda_t}\left(1 - e^{-\lambda_t(T-t)}\right)\right)^2 dt$$

Proof. By (3.1)

$$\begin{aligned}
d\langle M^{T,f}\rangle_t &= \left|\int_t^T \sigma(t,X_t)^\top \nabla P_{t,s}f(X_t)ds\right|^2 dt \\
&\le \left(\int_t^T C_t e^{-\lambda_t(s-t)}ds\right)^2 dt = \left(\frac{C_t}{\lambda_t}\left(1 - e^{-\lambda_t(T-t)}\right)\right)^2 dt
\end{aligned}$$

so that $\langle M^{T,f}\rangle_T \le V_T T$.

By Corollary 2.3 and since Novikov's condition holds trivially due to $\langle M^{T,f}\rangle$ being bounded by a deterministic function we get

$$\begin{aligned}
\mathbb{E}\exp\left(a\int_0^T f(t,X_t) - \mathbb{E}f(t,X_t)dt\right) &= \mathbb{E}\exp\left(aM_T^{T,f}\right) \\
&\le \mathbb{E}\left[\exp\left(aM_T^{T,f} - \frac{a^2}{2}\langle M^{T,f}\rangle_T\right)\right]\exp\left(\frac{a^2}{2}\langle M^{T,f}\rangle_T\right) \le \exp\left(\frac{a^2}{2}V_T T\right).
\end{aligned}$$

By Chebyshev's inequality

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(t,X_t)-\mathbb{E}f(t,X_t)dt > R\right) \leq \exp\left(-aRT\right)\exp\left(\frac{a^2}{2}V_T T\right)$$

and the result follows by optimising over a. □

The corresponding lower bound is obtained by replacing f by $-f$.

For the rest of this section, suppose that $\sigma = \mathrm{Id}$ and that we are in the time-homogeneous case so that $P_{s,t} = P_{t-s}$. An important case where bounds of the form (3.1) hold is when there is exponential contractivity in the L^1 Kantorovich (Wasserstein) distance W_1. If for any two probability measures μ, ν on $\mathbb{R}^n$

$$W_1(\mu P_t, \nu P_t) \leq Ce^{-\lambda t} W_1(\mu,\nu). \tag{3.2}$$

then (3.1) holds for all Lipschitz functions f with $C_s = C$, $\lambda_s = \lambda$.

Here the distance W_1 between two probability measures μ and ν on $\mathbb{R}^n$ is defined by

$$W_1(\mu,\nu) = \inf_\pi \int |x-y|\pi(dx\,dy)$$

where the infimum runs over all couplings π of μ. We also have the Kantorovich-Rubinstein duality

$$W_1(\mu,\nu) = \sup_{\|f\|_{\mathrm{Lip}}\leq 1} \int f d\mu - \int f d\nu \tag{3.3}$$

and we use the notation

$$\|f\|_{\mathrm{Lip}} = \sup_{x\neq y} \frac{f(x)-f(y)}{|x-y|}.$$

We can see that (3.2) implies (3.1) from

$$\begin{aligned}|\nabla P_t f|(x) = \lim_{y\to x} \frac{|P_t f(y) - P_t f(x)|}{|y-x|} &\leq \lim_{y\to x} \frac{W_1(\delta_y P_t, \delta_x P_t)}{|y-x|} \\ &\leq \|f\|_{\mathrm{Lip}} Ce^{-\lambda t} \lim_{y\to x} \frac{W_1(\delta_y,\delta_x)}{|y-x|} = \|f\|_{\mathrm{Lip}} Ce^{-\lambda t}\end{aligned}$$

where the first inequality is due to the Kantorovich-Rubinstein duality (3.3) and the second is (3.1).

Bounds of the form (3.2) have been obtained using coupling methods in [Ebe16; EGZ16; Wan16] under the condition that there exist positive constants κ, R_0 such that

$$(x-y)\cdot(b(x)-b(y)) \leq -\kappa|x-y|^2 \text{ when } |x-y| > R_0.$$

Similar techniques lead to the corresponding results for kinetic Langevin diffusions[EGZ17].

Using a different approach, in [CO16] the authors directly show uniform exponential contractivity of the semigroup gradient for bounded continuous functions, focusing on situations beyond hypoellipticity.

Besides gradient bounds, exponential contractivity in W_1 also implies the existence of a stationary measure μ_∞ [Ebe16]. Proposition 3.1 now leads to a simple proof of a deviation inequality that was obtained in a similar setting in [Jou09] via a tensorization argument.

Proposition 3.2. *If* (3.2) *holds then for all Lipschitz functions f and all initial measures μ_0*

$$\mathbb{P}_{\mu_0}\left(\frac{1}{T}\int_0^T f(X_t)dt - \int f d\mu_\infty > R\right) \leq \exp\left(-\left(\frac{\lambda\sqrt{T}\,R}{C\|f\|_{\mathrm{Lip}}(1-e^{-\lambda T})} - \frac{W_1(\mu_0,\mu_\infty)}{\sqrt{T}}\right)^2\right)$$

Proof. We start by applying Proposition 3.1 so that

$$\begin{aligned}
&\mathbb{P}_{\mu_0}\left(\frac{1}{T}\int_0^T f(X_t)dt - \int f d\mu_\infty > R\right) \\
&\quad = \mathbb{P}_{\mu_0}\left(\frac{1}{T}\int_0^T f(X_t) - \mathbb{E}f(X_t)dt > R + \frac{1}{T}\int_0^T \mu_\infty(f) - \mu_0 P_t(f)dt\right) \\
&\quad \leq \exp\left(-\left(R - \left|\frac{1}{T}\int_0^T \mu_\infty(f) - \mu_0 P_t(f)dt\right|\right)^2 \frac{T}{V_T}\right), \quad V_T = \left(\frac{\|f\|_{\mathrm{Lip}} C(1-e^{-\lambda T})}{\lambda}\right)^2.
\end{aligned}$$

By the Kantorovich-Rubinstein duality

$$\begin{aligned}
\left|\frac{1}{T}\int_0^T \mu_\infty(f) - \mu_0 P_t(f)dt\right| &\leq \left|\frac{1}{T}\int_0^T \|\nabla f\|_\infty W_1(\mu_\infty P_t, \mu_0 P_t)dt\right| \\
&\leq \frac{\|\nabla f\|_\infty C}{\lambda}\frac{(1-e^{-\lambda T})}{T} W_1(\mu,\mu_0) = \frac{\sqrt{V_T}}{T} W_1(\mu,\mu_0).
\end{aligned}$$

from which the result follows immediately. ☐

4 Averaging: Two-timescale Ornstein-Uhlenbeck

Consider the following linear multiscale SDE on $\mathbb{R}\times\mathbb{R}$ where the first component is accelerated by a factor $\alpha \geq 0$:

$$\begin{aligned}
dX_t &= -\alpha(X_t - Y_t)dt + \sqrt{\alpha}dB_t^X, \quad X_0 = x_0 \\
dY_t &= -(Y_t - X_t)dt + dB_t^Y, \quad Y_0 = y_0
\end{aligned}$$

with B^X, B^Y independent Brownian motions on $\mathbb{R}$. Denote P_t and L the associated semigroup and infinitesimal generator respectively.

Let $f(x,y) = x - y$ and note that $Lf = -(\alpha+1)f$. We have by the regularity of P_t and the Kolmogorov forward equation

$$\partial_t \partial_x P_t f = \partial_x P_t L f = -(\alpha+1)\partial_x P_t f$$

so that

$$\partial_x P_t f = \partial_x f e^{-(\alpha+1)t} = e^{-(\alpha+1)t}.$$

Repeating the same reasoning for $\partial_y P_t$ and P_t gives

$$\partial_y P_t f = -e^{-(\alpha+1)t} \text{ and } P_t f(x,y) = (x-y)e^{-(\alpha+1)t}.$$

From Corollary 2.3

$$\int_0^T X_t - Y_t \, dt = R_0^T f(x_0, y_0) + M_T^{T,f}$$

with

$$\begin{aligned}
R_t^T f(x,y) &= \int_t^T P_{s-t} f(x,y) ds = (x-y)\frac{1 - e^{-(\alpha+1)(T-t)}}{\alpha+1}, \\
M_T^{T,f} &= \int_0^T \int_t^T \partial_x P_{s-t} f(X_t, Y_t) ds \, \sqrt{\alpha} dB_t^X + \int_0^T \int_t^T \partial_y P_{s-t} f(X_t, Y_t) ds \, dB_t^Y \\
&= \int_0^T \frac{1 - e^{-(\alpha+1)(T-t)}}{\alpha+1} (\sqrt{\alpha} dB_t^X - dB_t^Y).
\end{aligned}$$

This shows that for each T fixed

$$Y_T - (B_T^Y + y_0) = \int_0^T X_t - Y_t dt$$

is a Gaussian random variable with mean

$$R_0^T = (x_0 - y_0)\frac{1 - e^{-(\alpha+1)T}}{\alpha+1}$$

and variance

$$\langle M^{T,f} \rangle_T = \frac{1}{(\alpha+1)} \int_0^T \left(1 - e^{-(\alpha+1)(T-t)}\right)^2 dt.$$

5 Averaging: Exact gradients in the linear case

Consider

$$\begin{aligned}
dX_t &= -\alpha(X_t - Y_t)dt + \sqrt{\alpha} dB_t^X, \quad X_0 = x_0 \\
dY_t &= -(Y_t - X_t)dt - \beta Y_t + dB_t^Y, \quad Y_0 = y_0
\end{aligned}$$

Denote $Z_t((x,y)) = (X_t(x), Y_t(x))$ the solution for $X_0 = x, Y_0 = y$ and let $V_t(z,v) = Z_t(z+v) - Z_t(z)$. Then

$$dV_t = -AV_t\,dt \text{ with } A = \begin{pmatrix} \alpha & -\alpha \\ -1 & (1+\beta) \end{pmatrix}.$$

The solution to the linear ODE for V_t is

$$V_t(z,v) = e^{-At}v$$

Since V_t does not depend on z we drop it from the notation. Now for any continuously differentiable function f on $\mathbb{R}^2$ and $v \in \mathbb{R}^2, z \in \mathbb{R}^2$ we obtain the following expression for the gradient of $P_t f(z)$ in the direction v:

$$\begin{aligned}\nabla_v P_t f(z) &= \lim_{\varepsilon\to 0} \frac{P_t f(z+\varepsilon v) - P_t f(z)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{\mathbb{E} f(Z_t(z+\varepsilon v)) - f(Z_t(z))}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{\mathbb{E}\nabla f(Z_t(z)) \cdot V_t(\varepsilon v) + o(|V_t(\varepsilon v)|)}{\varepsilon} \\ &= \mathbb{E}\nabla f(Z_t(z)) \cdot e^{-At}v.\end{aligned}$$

Since $\nabla_v P_t f = \nabla P_t f \cdot v$ we can identify $\nabla P_t f(z) = \mathbb{E}^z (e^{-At})^\top \nabla f(Z_t)$.

The eigenvalues of A are $(\lambda_0, \alpha\lambda_1)$ with

$$\begin{aligned}\lambda_0 &= \frac{1}{2}\left(\alpha+\beta+1 - \sqrt{(\alpha+\beta+1)^2 - 4\alpha\beta}\right), \\ \lambda_1 &= \frac{1}{2\alpha}\left(\alpha+\beta+1 + \sqrt{(\alpha+\beta+1)^2 - 4\alpha\beta}\right).\end{aligned}$$

By observing that

$$(\alpha+\beta+1)^2 - 4\alpha\beta = (\alpha - (1+\beta))^2 + 4\alpha = (\beta - (\alpha+1))^2 + 4\beta$$

we see that asymptotically as $\alpha \to \infty$

$$\begin{aligned}\lambda_0 &= \beta + O\left(\frac{1}{\alpha}\right) \\ \lambda_1 &= 1 + \frac{1}{\alpha} + O\left(\frac{1}{\alpha^2}\right).\end{aligned}$$

We can compute the following explicit expression for e^{-At}

$$\begin{aligned}e^{-At} &= c_0(t)\,\mathrm{Id} - \frac{c_1(t)}{\alpha}A \\ &= \begin{pmatrix} \frac{c_2(t)}{\alpha} & c_1(t) \\ \frac{c_1(t)}{\alpha} & c_0(t) - \frac{1+\beta}{\alpha}c_1(t) \end{pmatrix}\end{aligned}$$

with

$$
\begin{aligned}
c_0(t) &= \frac{\alpha\lambda_1 e^{-\lambda_0 t} - \lambda_0 e^{-\alpha\lambda_1 t}}{\alpha\lambda_1 - \lambda_0} = \frac{(1+\alpha)e^{-\lambda_0 t} - \beta e^{-\alpha\lambda_1 t}}{\alpha\lambda_1 - \lambda_0} + O\left(\frac{1}{\alpha^2}\right), \\
c_1(t) &= \frac{\alpha}{\alpha\lambda_1 - \lambda_0}\left(e^{-\lambda_0 t} - e^{-\alpha\lambda_1 t}\right), \\
c_2(t) &= \alpha(c_0(t) - c_1(t)) = \frac{\alpha}{\alpha\lambda_1 - \lambda_0}\left(e^{-\lambda_0 t} - (\beta - \alpha)e^{-\alpha\lambda_1 t}\right) + O\left(\frac{1}{\alpha}\right).
\end{aligned}
$$

Note that $\lambda_0, \lambda_1, c_0, c_1$ and c_2 are all of order $O(1)$ as $\alpha \to \infty$.

We obtain

$$
\begin{aligned}
\sigma^\top \nabla P_t f(z) &= \mathbb{E}\left[\begin{pmatrix} \frac{c_2(t)}{\sqrt{\alpha}} & \frac{c_1(t)}{\sqrt{\alpha}} \\ c_1(t) & c_0(t) - \frac{1+\beta}{\alpha}c_1(t) \end{pmatrix} \nabla f(Z_t)\right] \\
&= \frac{\alpha}{1+\alpha}\left(G_0 e^{-\lambda_0 t} + G_1 \alpha e^{-\alpha\lambda_1 t}\right) P_t \nabla f(z)
\end{aligned}
$$

with

$$
\begin{aligned}
G_0 &= \begin{pmatrix} \frac{1}{\sqrt{\alpha}} & \frac{1}{\sqrt{\alpha}} \\ 1 & 1 \end{pmatrix} + O\left(\frac{1}{\alpha}\right) \\
G_1 &= \begin{pmatrix} \frac{1}{\sqrt{\alpha}} - \frac{\lambda_0}{\alpha\sqrt{\alpha}} & -\frac{1}{\alpha\sqrt{\alpha}} \\ -\frac{1}{\alpha} & -\frac{1+\lambda_0+\beta}{\alpha^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\alpha}} & 0 \\ 0 & 0 \end{pmatrix} + O\left(\frac{1}{\alpha}\right)
\end{aligned}
$$

The expression for G_0 shows that $|\sigma^\top \nabla P_t f(z)|$ can be of order $1/\sqrt{\alpha}$ only for functions $f_\alpha(z)$ such that $\mathbb{E}^z[\partial_x f_\alpha(Z_t) + \partial_y f_\alpha(Z_t)] = O(1/\sqrt{\alpha})$.

Furthermore, for any function $f \in C_c^\infty$ we have

$$
\mathrm{Cov}\left(f(Z_t), B_t^X\right) = O\left(\frac{1}{\sqrt{\alpha}}\right)
$$

and

$$
\mathrm{Cov}\left(\int_0^t f(s, Z_s)ds, B_t^X\right) = O\left(\frac{1}{\sqrt{\alpha}}\right).
$$

Indeed, since $dP_{s,t}f(Z_s) = f(Z_s) \cdot \sigma dB_s$ we have

$$
\begin{aligned}
f(Z_t) - \mathbb{E}f(Z_t) &= \int_0^t \nabla P_{s,t} f(Z_s) \cdot \sigma dB_s \\
&= \int_0^t \nabla_x P_{s,t} f(Z_s)\sqrt{\alpha}dB_s^X + \int_0^t \nabla_y P_{s,t} f(Z_s) dB_s^Y
\end{aligned}
$$

we have

$$
\begin{aligned}
\mathrm{Cov}\left(f(Z_t), B_t^X\right) &= \mathbb{E}\left[(f(Z_t) - \mathbb{E}f(Z_t))B_t^X\right] \\
&= \mathbb{E}\left[\int_0^t \nabla_x P_{s,t} f(Z_s)\sqrt{\alpha}ds\right] \\
&= \left(\frac{1}{\sqrt{\alpha}} + O\left(\frac{1}{\alpha}\right)\right)\frac{\alpha}{1+\alpha}\int_0^t e^{-\lambda_0 s} P_{s,t}(\nabla_x f + \nabla_y f)(Z_s)ds.
\end{aligned}
$$

The result for $\int_0^t f(s, Z_s)ds$ follows by the same arguments from the martingale representation for $\int_0^t f(s, Z_s)ds - \mathbb{E}\int_0^t f(s, Z_s)ds$.

6 Averaging: Conditioning on the slow component

Consider the following linear multiscale SDE on $\mathbb{R} \times \mathbb{R}$ accelerated by a factor α:

$$
\begin{aligned}
dX_t &= -\alpha\kappa_X(X_t - Y_t)dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad X_0 = 0 \\
dY_t &= -\kappa_Y(Y_t - X_t)dt + \sigma_Y dB_t^Y, \quad Y_0 = 0
\end{aligned}
$$

where B^X, B^Y are independent Brownian motions and $\alpha, \kappa_X, \kappa_Y, \sigma_X, \sigma_Y$ are strictly positive constants and we are interested in the solution on a fixed inverval $[0, T]$.

We define the corresponding averaged process to be the solution to

$$
d\bar{X}_t = -\alpha\kappa_X(\bar{X}_t - \bar{Y}_t)dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad \bar{X}_0 = 0 \tag{6.1a}
$$

$$
d\bar{Y}_t = \mathbb{E}\left[-\kappa_Y(\bar{Y}_t - \bar{X}_t)\middle|\mathcal{F}_t^{\bar{Y}}\right]dt + \sigma_Y dB_t^Y, \quad \bar{Y}_0 = 0 \tag{6.1b}
$$

where $\mathcal{F}_t^{\bar{Y}}$ is the σ-algebra generated by $(\bar{Y}_s)_{s\le t}$.

The conditional measure $\mathbb{P}(\cdot|\mathcal{F}_T^{\bar{Y}})$ has a regular conditional probability density $u \mapsto \mathbb{P}(\cdot|\bar{Y} = u)$, $u \in C([0,T], \mathbb{R})$. Now observe that B^X remains unchanged under $\mathbb{P}(\cdot|\bar{Y} = u)$ since $\bar{Y}$ and B^X are independent. This means that for all $u \in C([0,T], \mathbb{R})$ and $f \in C_c^\infty(\mathbb{R})$, $\mathbb{P}(\cdot|\bar{Y} = u)$ solves the same martingale problem as the measure associated to

$$
dX_t^u = -\alpha\kappa_X(X_t^u - u(t))dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad X_0^u = 0. \tag{6.2}
$$

It follows that the conditional expectation given $\mathcal{F}_T^{\bar{Y}}$ of any functional involving $\bar{X}$ equals the usual expectation of the same functional with $\bar{X}$ replaced by X^u evaluated at $u = Y$.

For example, since

$$
\mathbb{E}X_t^u = \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X(t-s)}\, u(s)\, ds
$$

the drift coefficient of $\bar{Y}$ is

$$\begin{aligned}\mathbb{E}\left[-\kappa_Y(\bar{Y}_t - \bar{X}_t)\middle|\mathcal{F}_t^{\bar{Y}}\right] &= -\kappa_Y(\bar{Y}_t - \mathbb{E}[\bar{X}_t|\mathcal{F}_T^{\bar{Y}}]) = -\kappa_Y(\bar{Y}_t - \mathbb{E}X_t^u|_{u=\bar{Y}}) \\ &= -\kappa_Y\left(\bar{Y}_t - \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X(t-s)}\,\bar{Y}_s\,ds\right)\end{aligned}$$

so that $\bar{Y}$ solves the SDE

$$dZ_t = -\alpha\kappa_X(Z_t - \bar{Y}_t)dt \tag{6.3a}$$

$$d\bar{Y}_t = -\kappa_Y(\bar{Y}_t - Z_t)dt + \sigma_Y dB_t^Y. \tag{6.3b}$$

The key step in our estimate for $Y_t - \bar{Y}_t$ is the application of the results from the first section to

$$\int_0^T h(t)(X_t^u - \mathbb{E}X_t^u)dt$$

for a certain function $h(t)$.

We begin with a gradient estimate for the evolution operator $P_{s,t}^u$ associated to X^u.

Lemma 6.1. *Let* $\mathrm{id}(x) = x$ *be the identity function and* $h(t) \in C([0,T],\mathbb{R})$*. We have for all* $x \in \mathbb{R}$

$$\partial_x P_{s,t}^u(h\,\mathrm{id})(x) = h(t)e^{-\alpha\kappa_X(t-s)}.$$

Proof. Denote $X_t^{s,x}$ the solution to (6.2) with $X_s^u = x$. Then

$$d(X_t^{s,x+\varepsilon} - X_t^{s,x}) = -\alpha\kappa_X(X_t^{s,x+\varepsilon} - X_t^{s,x})dt$$

so that

$$X_t^{s,x+\varepsilon} - X_t^{s,x} = \varepsilon e^{-\kappa_X\alpha(t-s)}$$

and

$$\partial_x P_{s,t}(h\,\mathrm{id})(x) = \lim_{\varepsilon\to 0}\varepsilon^{-1}\mathbb{E}\left[h(t)X_t^{s,x+\varepsilon} - h(t)X_t^{s,x}\right] = h(t)e^{-\kappa_X\alpha(t-s)}.$$

□

Theorem 6.2.

$$\begin{aligned}\mathbb{E}|Y_T - \bar{Y}_T|^2 &= \frac{\alpha\kappa_Y^2\sigma_X^2}{(\alpha\kappa_X + \kappa_Y)^2}\int_0^T \left(1 - e^{-\alpha\kappa_X(T-t)}\left(2 - e^{-\kappa_Y(T-t)}\right)\right)^2 dt \\ &\leq \frac{T}{\alpha}\frac{\kappa_Y^2\sigma_X^2}{\kappa_X^2}\end{aligned} \tag{6.4}$$

and

$$\begin{aligned}\mathbb{E}|\bar{Y}_T - \sigma_Y B_T^Y|^2 &= \frac{\kappa_Y^2\sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)^2}\int_0^T \left(1 - e^{-(\alpha\kappa_X + \kappa_Y)t}\right)^2 dt \\ &\leq \frac{T}{\alpha^2}\frac{\kappa_Y^2\sigma_Y^2}{\kappa_X^2}.\end{aligned} \tag{6.5}$$

Proof of Theorem 6.2. We now proceed to show the equality (6.4). We decompose

$$Y_T - \bar{Y}_T = \int_0^T \kappa_Y (X_t - Y_t) dt - \int_0^T \kappa_Y (\mathbb{E}[\bar{X}_t | \bar{Y}] - \bar{Y}_t) dt$$
$$= -\kappa_Y \int_0^T (\mathbb{E}[\bar{X}_t | \bar{Y}] - \bar{X}_t) dt - \kappa_Y \int_0^T (Y_t - \bar{Y}_t) - (X_t - \bar{X}_t) dt. \tag{6.6}$$

Using linearity, we now proceed to rewrite this as

$$Y_T - \bar{Y}_T = -\kappa_Y \int_0^T h(T-t)(\mathbb{E}[\bar{X}_t | \bar{Y}] - \bar{X}_t) dt$$

for some function h.

Since

$$d(X_t - \bar{X}_t) = -\alpha\kappa_X (X_t - \bar{X}_t) dt + \alpha\kappa_X (Y_t - \bar{Y}_t) dt$$

we have

$$X_t - \bar{X}_t = \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X (t-s)} (Y_t - \bar{Y}_t) ds.$$

With the notation

$$f(t) = Y_t - \bar{Y}_t, \quad g(t) = \bar{X}_t - \mathbb{E}[\bar{X}_t | \bar{Y}]$$

equation (6.6) reads as

$$\frac{1}{\kappa_Y} f'(t) + f(t) - \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X (t-s)} f(s) ds = g(t).$$

Using capital letters for the Laplace transform, this writes as

$$\frac{s}{\kappa_Y} F(s) + F(s) - \frac{\alpha\kappa_X}{s + \alpha\kappa_X} F(s) = G(s)$$

or, after rearranging,

$$F(s) = \kappa_Y \frac{s + \alpha\kappa_X}{s(s + \alpha\kappa_X + \kappa_Y)} G(s) = \kappa_Y H(s) G(s).$$

Inverting the Laplace transform, we find that

$$h(t) = \frac{\alpha\kappa_X}{\alpha\kappa_X + \kappa_Y} + \frac{\kappa_Y}{\alpha\kappa_X + \kappa_Y} e^{-(\alpha\kappa_X + \kappa_Y)t}$$

so that

$$Y_T - \bar{Y}_T = \kappa_Y \int_0^T h(T-s) \left(\bar{X}_s - \mathbb{E}[\bar{X}_s | \bar{Y}]\right) ds.$$

By the properties of conditional expectation and Corollary 2.3 we have for any integrable function Φ that

$$\mathbb{E}\Phi(Y_T - \bar{Y}_T) = \mathbb{E}[\mathbb{E}\Phi(Y_T - \bar{Y}_T)|\mathcal{F}_T^{\bar{Y}}] = \mathbb{E}[\mathbb{E}(\Phi(Y_T - \bar{Y}_T)|u = \bar{Y})] = \mathbb{E}[(\mathbb{E}\Phi(M_T^u))|_{u=\bar{Y}}]$$

with

$$\begin{aligned}
M_T^u &= \kappa_Y \int_0^T \int_t^T \partial_x P_{t,s}^u (h(T-\cdot)\,\mathrm{id})(X_t)\, ds\, \sqrt{\alpha}\sigma_X dB_t \\
&= \kappa_Y \sqrt{\alpha}\sigma_X \int_0^T \int_t^T h(T-s) e^{-\alpha\kappa_X (s-t)}\, ds\, dB_t \\
&= \frac{\kappa_Y \sqrt{\alpha}\sigma_X}{\alpha\kappa_X + \kappa_Y} \int_0^T \int_t^T \alpha\kappa_X e^{-\alpha\kappa_X(s-t)} ds + \int_t^T \kappa_Y e^{-\kappa_Y(T-s)} e^{-\alpha\kappa_X(T-s)} e^{-\alpha\kappa_X(s-t)}\, ds\, dB_t \\
&= \frac{\sqrt{\alpha}\kappa_Y\sigma_X}{\alpha\kappa_X + \kappa_Y} \int_0^T 1 - e^{-(\alpha\kappa_X + \kappa_Y)(T-t)}\, dB_t.
\end{aligned}$$

Since M_t^u is independent of u we can let $M_t = M_t^u$ for an arbitrary u so that

$$\mathbb{E}\Phi(Y_T - \bar{Y}_T) = \mathbb{E}\Phi(M_T).$$

Now we can compute

$$\mathbb{E}\left|Y_T - \bar{Y}_T\right|^2 = \mathbb{E}\langle M\rangle_T = \frac{\alpha\kappa_Y^2\sigma_X^2}{(\alpha\kappa_X + \kappa_Y)^2} \int_0^T \left(1 - e^{-(\alpha\kappa + \kappa_Y)(T-t)}\right)^2 dt.$$

We now turn to the computation of $\mathbb{E}|\bar{Y}_t - \sigma_Y B_t^Y|^2$.

From equation (6.3) we have

$$d(\bar{Y}_t - Z_t) = -(\alpha\kappa_X + \kappa_Y)(\bar{Y}_t - Z_t)dt + \sigma_Y B_t^Y$$

so that

$$\bar{Y}_t - Z_t = \sigma_Y \int_0^t e^{-(\alpha\kappa_X + \kappa_Y)(t-s)} dB_s^Y. \tag{6.7}$$

is an Ornstein-Uhlenbeck process. This means that

$$\mathbb{E}(\bar{Y}_t - Z_t)(\bar{Y}_s - Z_s) = \frac{\sigma_Y^2 e^{-(\alpha\kappa_X + \kappa_Y)t}}{\alpha\kappa_X + \kappa_Y} \sinh((\alpha\kappa_X + \kappa_Y)s), \quad s \le t.$$

so that

$$
\begin{aligned}
\mathbb{E}|\bar{Y}_t - \sigma_Y B_t^Y|^2 &= \kappa_Y^2 \left| \int_0^t \bar{Y}_s - Z_s ds \right|^2 \\
&= 2\kappa_Y^2 \int_0^t \int_0^s \mathbb{E}(\bar{Y}_s - Z_s)(\bar{Y}_r - Z_r) dr ds \\
&= \frac{2\kappa_Y^2 \sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)} \int_0^t e^{-(\alpha\kappa_X+\kappa_Y)s} \int_0^s \sinh((\alpha\kappa_X + \kappa_Y)r) dr ds \\
&= \frac{2\kappa_Y^2 \sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)^2} \int_0^t e^{-(\alpha\kappa_X+\kappa_Y)s} \left(\cosh((\alpha\kappa_X + \kappa_Y)s) - 1\right) ds \\
&= \frac{\kappa_Y^2 \sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)^2} \left(\int_0^t 1 + e^{-2(\alpha\kappa_X+\kappa_Y)s} - 2e^{-(\alpha\kappa_X+\kappa_Y)s} ds \right)
\end{aligned}
$$

□

7 Approximation by Averaged Measures

In the previous section, the computation for $\mathbb{E}|\bar{Y}_t - \sigma_Y B_t^Y|^2$ relied on the fact that we had an explicit expression for $\mathbb{E}[\bar{X}_t - \bar{Y}_t|Y]$. Here we will see a method that can be used to obtain similar estimates in more general situations.

Consider a diffusion process (X_t, Y_t) on $\mathbb{R}^n \times \mathbb{R}^m$

$$
\begin{aligned}
dX_t &= b_X(X_t, Y_t)dt + \sigma_X(X_t, Y_t)dB_t^X \\
dY_t &= b_Y(Y_t)dt + \sigma_Y(Y_t)dB_t^Y
\end{aligned}
$$

where B^X and B^Y are standard independent Brownian motions. Denote L the generator of (X, Y) and $\mathcal{F}^Y$ the filtration of B^Y.

Let

$$
Q_t f = \mathbb{E}^{\mathcal{F}_t^Y} f(X_t, Y_t)
$$

so that, by the Itô formula and since Y is adapted to $\mathcal{F}^Y$ and B^X and B^Y are independent, we have

$$
\begin{aligned}
Q_t f = \mathbb{E}^{\mathcal{F}_t^Y} \Bigg[& f(X_0, Y_0) + \int_0^t Lf(X_s, Y_s)ds + \int_0^t \nabla_x f(X_s, Y_s) \cdot \sigma_X(X_s, Y_s) dB_s^X \\
& + \int_0^t \nabla_y f(X_s, Y_s) \cdot \sigma_Y(Y_s) dB_s^Y \Bigg] \\
= \mathbb{E}^{\mathcal{F}_0^Y}&[f(X_0, Y_0)] + \int_0^t \mathbb{E}^{\mathcal{F}_s^Y} Lf(X_s, Y_s) ds + \int_0^t (\mathbb{E}^{\mathcal{F}_s^Y} \nabla_y f(X_s, Y_s)) \cdot \sigma_Y(Y_s) dB_s^Y.
\end{aligned}
$$

In other words,

$$
dQ_t f = Q_t Lf dt + (Q_t \nabla_y f) \cdot \sigma_Y(Y_t) dB_t^Y.
$$

Example 7.1 (Averaged Ornstein-Uhlenbeck). Consider again the process $(\bar{X}, \bar{Y})$ from the previous section. In this case, $f(x,y) = x - y$ is an eigenfunction of $-L$ with eigenvalue $\alpha\kappa_X + \kappa_Y$ and we have $\partial_y f = -1$. Therefore

$$dQ_t f = -(\alpha\kappa_X + \kappa_Y) Q_t f dt - \sigma_Y dB_t^Y$$

so that we retrieve the result from (6.7)

$$\mathbb{E}[\bar{X}_t - \bar{Y}_t | \bar{Y}] = Q_t f = -\sigma_Y \int_0^t e^{-(\alpha\kappa_X + \kappa_Y)(t-s)} dB_s^Y.$$

References

[CO16] D. Crisan and M. Ottobre. “Pointwise gradient bounds for degenerate semigroups (of UFG type)”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 472.2195 (Nov. 2016), p. 20160442.

[Ebe16] Andreas Eberle. “Reflection couplings and contraction rates for diffusions”. In: *Probability Theory and Related Fields* 166.3 (2016), pp. 851–886.

[EGZ16] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. “Quantitative Harris type theorems for diffusions and McKean-Vlasov processes”. In: *arXiv preprint arXiv:1606.06012* (2016).

[EGZ17] A. Eberle, A. Guillin, and R. Zimmer. “Couplings and quantitative contraction rates for Langevin dynamics”. In: *ArXiv e-prints* (Mar. 2017).

[EK89] Robert J. Elliott and Michael Kohlmann. “Integration by Parts, Homogeneous Chaos Expansions and Smooth Densities”. In: *The Annals of Probability* 17.1 (1989), pp. 194–207.

[Jou09] Aldéric Joulin. “A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature”. In: *Bernoulli* 15.2 (May 2009), pp. 532–549.

[Wan16] Feng-Yu Wang. “Exponential Contraction in Wasserstein Distances for Diffusion Semigroups with Negative Curvature”. In: *arXiv preprint arXiv:1603.05749* (2016).