C. R. Acad. Sci. Paris, Ser. I 349 (2011) 1101–1104

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability Theory

Reflection coupling and Wasserstein contractivity without convexity

Couplage de réflection et contractivité de Wasserstein sans convexité

Andreas Eberle

University of Bonn, Institute for Applied Mathematics, Endenicher Allee 60, 53115 Bonn, Germany

ARTICLE INFO

Article history:
Received 16 July 2011
Accepted 2 September 2011
Available online 1 October 2011

Presented by the Editorial Board

ABSTRACT

We note that even if convexity of the potential U fails locally, overdamped Langevin diffusions in $\mathbb{R}^d$ are contractions w.r.t. the Kantorovich–Rubinstein-Wasserstein distance based on an appropriately chosen concave distance function equivalent to the Euclidean distance. The choice of the distance function is then optimized to obtain a large exponential decay rate. The results yield dimension-independent bounds of optimal order in $R, L \in [0,\infty)$ and $K \in (0,\infty)$ if $(x-y)\cdot(\nabla U(x)-\nabla U(y))$ is bounded from below by $-L|x-y|^2$ for $|x-y|<R$ and by $K|x-y|^2$ for $|x-y|\geqslant R$.

RÉSUMÉ

On considére diffusions de Langevin sur $\mathbb{R}^d$ dans un potentiel U non convex dans un ensemble borné. A l'aide du couplage de réflection, on observe que ces diffusions sont des contractions pour la distance de Kantorovich–Rubinstein–Wasserstein basée sur une distance concave appropriée, équivalente à la distance Euclidienne. Le choix de la distance est optimisé pour obtenir un grand taux de décroissance exponentielle. Les résultats impliquent bornes optimales pour $R, L \in [0,\infty)$ et $K \in (0,\infty)$, indépendamment de la dimension, sous la condition que $(x-y)\cdot(\nabla U(x)-\nabla U(y))$ est borné inférieurement par $-L|x-y|^2$ pour $|x-y|<R$ et par $K|x-y|^2$ pour $|x-y|\geqslant R$.

1. Introduction

Consider a diffusion process $(X_t)_{t\geqslant 0}$ in $\mathbb{R}^d$ defined by a stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t. \tag{1}$$

Here $(B_t)_{t\geqslant 0}$ is a d-dimensional Brownian motion, $\sigma \in \mathbb{R}^{d\times d}$ is a constant $d\times d$ matrix with $\det\sigma > 0$, and $b:\mathbb{R}^d \to \mathbb{R}^d$ is a locally Lipschitz continuous function. We assume that the unique strong solution of (1) is non-explosive, which is essentially a consequence of the assumptions imposed further below. The transition kernels of the diffusion process on $\mathbb{R}^d$ defined by (1) will be denoted by $p_t(x,\mathrm{d}y)$. We are interested in upper bounds for Kantorovich–Rubinstein–Wasserstein distances of the distributions μp_t and νp_t at a given time $t\geqslant 0$ w.r.t. two different initial distributions μ and ν.

E-mail address: eberle@uni-bonn.de.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2011.09.003

Example 1 *(Overdamped Langevin dynamics).* Suppose $\sigma = I_d$ and $b(x) = -\frac{1}{2}\nabla U(x)$ for a function $U \in C^2(\mathbb{R}^d)$ that is strictly convex (i.e. $\nabla^2 U \geqslant K \cdot I_d$ for some $K > 0$) outside a given ball $B \subset \mathbb{R}^d$. Then $Z := \int \exp(-U(x))\,\mathrm{d}x < \infty$, and $\mathrm{d}\mu := Z^{-1}\exp(-U)\,\mathrm{d}x$ is a stationary distribution for the diffusion process (X_t). The results below imply upper bounds for the L^1 Wasserstein distances between the law νp_t of X_t and μ for an arbitrary initial distribution ν and $t \geqslant 0$.

A *coupling by reflection* of two solutions of (1) with initial distributions μ and ν is a diffusion process (X_t, Y_t) with values in $\mathbb{R}^{2d}$ defined by $(X_0, Y_0) \sim \eta$ where η is a coupling of μ and ν,

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t \quad \text{for } t \geqslant 0,\\ \mathrm{d}Y_t &= b(Y_t)\,\mathrm{d}t + \sigma\left(I - 2e_t e_t^\top\right)\mathrm{d}B_t \quad \text{for } t < T, \qquad Y_t = X_t \quad \text{for } t \geqslant T. \end{aligned} \tag{2}$$

Here $e_t e_t^\top$ is the orthogonal projection onto the unit vector $e_t := \sigma^{-1}(X_t - Y_t)/|\sigma^{-1}(X_t - Y_t)|$, and $T = \inf\{t \geqslant 0\colon\ X_t = Y_t\}$ is the coupling time, i.e., the first hitting time of the diagonal $\Delta = \{(x, y) \in \mathbb{R}^{2d}\colon\ x = y\}$, cf. [5,1]. The reflection coupling can be realized as a diffusion process in $\mathbb{R}^{2d}$, and the marginal processes $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ are solutions of (1) w.r.t. the Brownian motions B_t and $\tilde{B}_t = \int_0^t (I_d - 2I_{\{s<T\}} e_s e_s^\top)\,\mathrm{d}B_s$. The difference vector $Z_t := X_t - Y_t$ solves the s.d.e.

$$\mathrm{d}Z_t = \big(b(X_t) - b(Y_t)\big)\,\mathrm{d}t + 2|\sigma^{-1}Z_t|^{-1} Z_t\,\mathrm{d}W_t \quad \text{for } t < T, \qquad Z_t = 0 \quad \text{for } t \geqslant T, \tag{3}$$

w.r.t. the *one-dimensional* Brownian motion $W_t = \int_0^t e_s^\top\,\mathrm{d}B_s$.

Lindvall and Rogers [5] introduced coupling by reflection in order to derive upper bounds for the total variation distance of the distributions of X_t and Y_t at a given time $t \geqslant 0$. Here we are instead considering the Kantorovich–Rubinstein (L^1-Wasserstein) distances

$$W_f(\mu, \nu) = \inf_\eta \int d_f(x, y)\eta(\mathrm{d}x\,\mathrm{d}y), \qquad d_f(x, y) = f\big(\|x - y\|\big) \quad \big(x, y \in \mathbb{R}^d\big), \tag{4}$$

of probability measures μ, ν on $\mathbb{R}^d$, where the infimum is over all couplings η of μ and ν, $f : [0, \infty) \to [0, \infty)$ is an appropriately chosen concave increasing function with $f(0) = 0$, and $\|z\| = \sqrt{z \cdot Gz}$ with $G \in \mathbb{R}^{d\times d}$ symmetric and strictly positive definite. Typical choices for the norm are the Euclidean norm $\|z\| = |z|$ and the intrinsic metric $\|z\| = |\sigma^{-1}z|$ corresponding to $G = I_d$ and $G = (\sigma\sigma^\top)^{-1}$ respectively.

2. Results

Similarly to Lindvall and Rogers [5], we define for $r \in (0, \infty)$:

$$\kappa(r) = \inf\left\{-2\frac{|\sigma^{-1}(x - y)|^2}{\|x - y\|^2}\,\frac{(x - y)\cdot G(b(x) - b(y))}{\|x - y\|^2}\colon\ x, y \in \mathbb{R}^d \text{ with } \|x - y\| = r\right\}.$$

Note that the factor $|\sigma^{-1}(x - y)|^2/\|x - y\|^2$ equals 1 if $\|\cdot\|$ is the intrinsic metric. In Example 1 with $G = I_d$, we have $\kappa(r) = \inf\{\int_0^1 \partial^2_{(x-y)/|x-y|} U((1 - t)x + ty)\,\mathrm{d}t\colon\ x, y \in \mathbb{R}^d \text{ s.t.} |x - y| = r\}$. We assume from now on that $\liminf_{r\to\infty} \kappa(r) > 0$, and we define constants $R_0, R_1 \in [0, \infty)$ with $R_0 \leqslant R_1$ by

$$R_0 = \inf\big\{R \geqslant 0\colon\ \kappa(r) \geqslant 0\ \forall r \geqslant R\big\}, \qquad R_1 = \inf\big\{R \geqslant R_0\colon\ \kappa(r)R(R - R_0) \geqslant 8\ \forall r \geqslant R\big\}.$$

We consider the particular distance function $d_f(x, y) = f(\|x - y\|)$ given by

$$f(r) = \int_0^r \varphi(s)g(s)\,\mathrm{d}s, \qquad \varphi(r) = \exp\left(-\frac{1}{4}\int_0^r s\kappa(s)^-\,\mathrm{d}s\right), \qquad g(r) = 1 - \frac{1}{2}\int_0^{r\wedge R_1} \frac{\Phi(s)}{\varphi(s)}\,\mathrm{d}s \Bigg/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)}\,\mathrm{d}s, \tag{5}$$

where $\Phi(r) = \int_0^r \varphi(s)\,\mathrm{d}s$. Note that Φ and f are concave, because φ and g are decreasing. Moreover, $\Phi(r)/2 \leqslant f(r) \leqslant \Phi(r)$ for any $r \geqslant 0$. Hence d_f and d_Φ as well as W_f and W_Φ differ at most by a factor 2. The choice of f is obtained by trying to maximize the decay rate of W_f, cf. the proof below.

Theorem 1. *Let $\alpha := \sup\{|\sigma^{-1}z|^2\colon\ z \in \mathbb{R}^d \text{ with } \|z\| = 1\}$, and define $c \in (0, \infty)$ by*

$$\frac{1}{c} = \alpha\int_0^{R_1} \Phi(s)\varphi(s)^{-1}\,\mathrm{d}s = \alpha\int_0^{R_1}\int_0^s \exp\left(\frac{1}{4}\int_t^s u\kappa(u)^-\,\mathrm{d}u\right)\mathrm{d}t\,\mathrm{d}s. \tag{6}$$

Then for d_f given by (4) and (5), the function $t \mapsto e^{ct}\mathbb{E}[d_f(X_t, Y_t)]$ is decreasing on $[0, \infty)$.

The theorem yields exponential contractivity at rate $c > 0$ for the transition kernels p_t of (1) w.r.t. the Kantorovich–Rubinstein–Wasserstein distance W_f. Moreover, it implies upper bounds for the standard KRW distance $W = W_{\mathrm{id}}$ w.r.t. the distance function $\mathrm{d}(x, y) = \|x - y\|$:

Corollary 2.1. *For any $t \geqslant 0$ and any probability measures μ, ν on $\mathbb{R}^d$,*

$$W_f(\mu p_t, \nu p_t) \leqslant e^{-ct} W_f(\mu, \nu), \quad \text{and} \quad W(\mu p_t, \nu p_t) \leqslant 2\varphi(R_0)^{-1} e^{-ct} W(\mu, \nu). \tag{7}$$

The second estimate follows from the first, because $\varphi(R_0)\|x - y\|/2 \leqslant d_f(x, y) \leqslant \|x - y\|$ for any $x, y \in \mathbb{R}^d$. For the Wasserstein mixing times, the corollary yields the upper bound

$$\tau_W(\varepsilon) := \inf\bigl\{t \geqslant 0\colon\ W(\mu p_t, \nu p_t) \leqslant \varepsilon W(\mu, \nu)\ \forall \mu, \nu\bigr\} \leqslant c^{-1} \log\bigl(2/\bigl(\varepsilon \varphi(R_0)\bigr)\bigr) \quad \text{for any } \varepsilon > 0.$$

Proof of Theorem 1. Let $r_t = \|Z_t\| = \|X_t - Y_t\|$. By (3) and Itô's formula,

$$\mathrm{d}f(r_t) = 2\bigl|\sigma^{-1} Z_t\bigr|^{-1} r_t f'(r_t)\,\mathrm{d}W_t + r_t^{-1} Z_t \cdot G\bigl(b(X_t) - b(Y_t)\bigr) f'(r_t)\,\mathrm{d}t + 2\bigl|\sigma^{-1} Z_t\bigr|^{-2} r_t^2 f''(r_t)\,\mathrm{d}t \tag{8}$$

a.s. for $t < T$. The drift is bounded from above by $\beta_t := 2|\sigma^{-1} Z_t|^{-2} r_t^2 (f''(r_t) - r_t \kappa(r_t) f'(r_t)/4)$. We show that by our choice of f, this expression is smaller than $-cf(r_t)$. Indeed, for $r < R_1$,

$$f''(r) = -\frac{1}{4} r\kappa(r)^- \varphi(r) g(r) - \frac{1}{2}\Phi(r) \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)}\,\mathrm{d}s \leqslant \frac{1}{4} r\kappa(r) f'(r) - \frac{1}{2} f(r) \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)}\,ds. \tag{9}$$

For $r \geqslant R_1$, we have $f'(r) = \varphi(r)/2 = \varphi(R_0)/2$ and $\kappa(r) R_1 (R_1 - R_0) \geqslant 8$ by definition of R_1, whence

$$\begin{aligned} f''(r) - \frac{1}{4} r\kappa(r) f'(r) &\leqslant -\frac{1}{8} r\kappa(r)\varphi(R_0) \leqslant -\frac{\varphi(R_0)}{R_1 - R_0} \cdot \frac{r}{R_1} \leqslant -\frac{\varphi(R_0)}{R_1 - R_0} \cdot \frac{\Phi(r)}{\Phi(R_1)} \\ &\leqslant -\frac{1}{2}\Phi(r) \Big/ \int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}\,\mathrm{d}s \leqslant -\frac{1}{2} f(r) \Big/ \int_0^{R_1} \Phi(s)\varphi(s)^{-1}\,\mathrm{d}s. \end{aligned} \tag{10}$$

Here we have used that for $r \geqslant R_0$, we have $\varphi(r) = \varphi(R_0)$, $\Phi(r) = \Phi(R_0) + (r - R_0)\varphi(R_0)$, and hence

$$\begin{aligned} \int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}\,\mathrm{d}s &= \int_{R_0}^{R_1} \bigl(\Phi(R_0) + (s - R_0)\varphi(R_0)\bigr)\varphi(R_0)^{-1}\,\mathrm{d}s = \frac{\Phi(R_0)}{\varphi(R_0)}(R_1 - R_0) + \frac{1}{2}(R_1 - R_0)^2 \\ &\geqslant (R_1 - R_0)\bigl(\Phi(R_0) + (R_1 - R_0)\varphi(R_0)\bigr)\varphi(R_0)^{-1}/2 \geqslant (R_1 - R_0)\Phi(R_1)\varphi(R_0)^{-1}/2. \end{aligned}$$

By (9) and (10), we conclude that $\beta_t \leqslant -cf(r_t)$. Optional stopping in (8) at $T_k = \inf\{t \geqslant 0\colon\ r_t \notin (k^{-1}, k)\}$ now implies $\mathbb{E}[f(r_t); t < T_k] \leqslant -c \int_0^t \mathbb{E}[f(r_s); s < T_k]\,\mathrm{d}s$ for any $k \in \mathbb{N}$ and $t \geqslant 0$. The assertion follows for $k \to \infty$ since $r_t = 0$ for $t \geqslant T$, and $T = \sup T_k$ by non-explosiveness. □

A first application. To illustrate that the bounds derived above are fairly sharp, let us suppose that $\kappa(r) \geqslant -L$ for $r \leqslant R$ and $\kappa(r) \geqslant K$ for $r > R$ with constants $R, L \in [0, \infty)$ and $K \in (0, \infty)$. Then, since $\varphi(r) = \varphi(R_0)$ and $\Phi(r) = \Phi(R_0) + (r - R_0)\varphi(R_0)$ for $r \geqslant R_0$,

$$\alpha^{-1} c^{-1} = \int_0^{R_1} \Phi(s)\varphi(s)^{-1}\,\mathrm{d}s = \int_0^{R_0} \Phi(s)\varphi(s)^{-1}\,\mathrm{d}s + (R_1 - R_0)\Phi(R_0)\varphi(R_0)^{-1} + (R_1 - R_0)^2/2. \tag{11}$$

The lower bounds on the function κ imply the upper bounds $R_0 \leqslant R$, $R_1 - R_0 \leqslant \min(8/(KR_0), \sqrt{8/K})$, $\Phi(r)\varphi(r)^{-1} \leqslant \int_0^r \exp(L(r^2 - t^2)/8)\,\mathrm{d}t \leqslant \min(\sqrt{2\pi/L}, r)\exp(Lr^2/8)$ for $r \leqslant R_0$, and

$$\int_0^{R_0} \Phi(r)\varphi(r)^{-1}\,\mathrm{d}r \leqslant \begin{cases} 4L^{-1}(\exp(LR_0^2/8) - 1) \leqslant (e - 1)R_0^2/2 & \text{if } LR_0^2/8 \leqslant 1, \\ 8\sqrt{2\pi} L^{-3/2} R_0^{-1} \exp(LR_0^2/8) & \text{if } LR_0^2/8 \geqslant 1. \end{cases}$$

Combining these estimates, we obtain by (11),

$$\alpha^{-1}c^{-1} \leqslant \begin{cases} (e-1)R^2/2 + e\sqrt{8/K}R + 4/K \leqslant (3e/2)\max(R^2, 8/K) & \text{if } LR_0^2/8 \leqslant 1, \\ 8\sqrt{2\pi}R^{-1}L^{-1/2}(L^{-1}+K^{-1})\exp(LR^2/8) + 32R^{-2}K^{-2} & \text{if } LR_0^2/8 \geqslant 1. \end{cases}$$

In the first case, c is at least of order $\min(R^{-2}, K)$. Even if $L = 0$ (convex case), this order can not be improved as one-dimensional Langevin diffusions with potential $U(x) = Kx^2/2$, or, respectively, with vanishing drift on $(-R/2, R/2)$ demonstrate. In the second case ($LR_0^2 \geqslant 8$), if $K \geqslant \text{const.} \cdot L$ then the upper bound for c^{-1} is of order $R^{-1}L^{-3/2}\exp(LR^2/8)$. This order in R and L is again optimal:

Example 2 *(Double-well with $U''(x) = -L$ for $|x| \leqslant R/2$).* Consider a Langevin diffusion in $\mathbb{R}^1$ with a symmetric potential $U \in C^2(\mathbb{R})$ satisfying $U(x) = -Lx^2/2$ for $x \in [-R/2, R/2]$, $U'' \geqslant -L$, and $\liminf_{|x|\to\infty} U''(x) > 0$. If $\|\cdot\|$ is the Euclidean norm then $\kappa(r) = -L$ for $r \in (0, R]$. On the other hand,

$$\lim_{t\to\infty} t^{-1}\log P_{R/2}[T_0 > t] = -\lambda_1(0,\infty) \geqslant -(2e-2)^{-1}(eL)^{3/2}R\exp\bigl(-LR^2/8\bigr) \quad \text{for } LR^2 \geqslant 4, \tag{12}$$

where T_0 denotes the first hitting time of 0 for the process starting at $R/2$, and $\lambda_1(0,\infty)$ is the lowest Dirichlet eigenvalue of the generator on $(0,\infty)$, cf. [3]. The bound for λ_1 follows by inserting the function $g(x) = \min(\sqrt{L}x, 1)$ into the variational characterization of the Dirichlet eigenvalue. By (12), the L^1 Wasserstein distance $W(\delta_{-R/2}p_t, \delta_{R/2}p_t)$ decays at most with a rate of order $L^{3/2}R\exp(-LR^2/8)$.

Remark. The idea to study Wasserstein contractivity w.r.t. concave distance functions goes back to Chen and Wang [2], where it is implicitly contained in the proofs. Indeed, in [2] and [6], Chen and Wang apply very similar methods to estimate spectral gaps of diffusion generators on $\mathbb{R}^d$ and on manifolds. Related arguments have also been applied in [4] to quantify exponential ergodicity in infinite dimensional situations. The techniques presented have natural extensions to non-constant diffusion coefficients and diffusions on manifolds, Euler discretizations of s.d.e., and high and infinite dimensional diffusions (dimension-independent bounds) that will be studied in detail in forthcoming work.

References

[1] M.F. Chen, S.F. Li, Coupling methods for multidimensional diffusion processes, Ann. Probab. 17 (1989) 151–177.
[2] M.F. Chen, F.-Y. Wang, Estimation of spectral gap for elliptic operators, Trans. Amer. Math. Soc. 349 (1997) 1239–1267.
[3] M. Freidlin, Functional Integration and Partial Differential Equations, Princeton University Press, Princeton, 1985.
[4] M. Hairer, J.C. Mattingly, Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations, Ann. Probab. 36 (2008) 2050–2091.
[5] T. Lindvall, L.C.G. Rogers, Coupling of multidimensional diffusions by reflection, Ann. Probab. 14 (1986) 860–872.
[6] F.-Y. Wang, Functional Inequalities, Markov Processes and Spectral Theory, Science Press, Beijing, 2004.