# Rates in the Central Limit Theorem and diffusion approximation via Stein's Method

Thomas Bonis
DataShape Team
Inria Saclay

June 30, 2016

**Abstract**

We present a way to apply Stein's method in order to bound the Wasserstein distance between a, possibly discrete, measure and another measure assumed to be the invariant measure of a diffusion operator. We apply this construction to obtain convergence rates, in terms of $p$-Wasserstein distance for $p \geq 2$, in the Central Limit Theorem in dimension 1 under precise moment conditions. We also establish a similar result for the Wasserstein distance of order 2 in the multidimensional setting. In a second time, we study the convergence of stationary distributions of Markov chains in the context of diffusion approximation, with applications to density estimation from geometric random graphs and to sampling using the Langevin Monte Carlo algorithm.

## 1 Introduction

Consider a diffusion process with infinitesimal generator $\mathcal{L}_\mu = b.\nabla + < a, Hess >$ admitting an invariant measure $\mu$. Let $\nu$ be the invariant measure of a discrete Markov chain approximating the previous diffusion. In this work, we provide a way to quantify the proximity between $\nu$ and $\mu$ using Stein's method. As $\nu$ is discrete, metrics such as the total variation distance or the relative entropy are not suited to compare $\mu$ and $\nu$ since their respective values would always be 1 and $\infty$. Instead, we focus on deriving bounds in terms of Wasserstein distance.

Let $\nu$ be a measure on $\mathbb{R}^d$ and $X$ be a random variable drawn from $\nu$. A measurable map $\tau_\nu$ is said to be a Stein kernel for $\nu$ if, for every smooth test function $\phi$,

$$\mathbb{E}\left[b(X).\nabla\phi(X) + < \tau_\nu(X), Hess(\phi)(X) >_{HS}\right] = 0, \tag{1}$$

where $< ., . >_{HS}$ is the usual Hilbert-Schmidt scalar product. Under technical assumptions on $\mathcal{L}_\mu$, Ledoux, Nourdin and Peccati [18] have proved that if $\tau_\nu$ and $a$ are close, so are $\nu$ and $\mu$.

In many cases, for instance when $\nu$ is discrete, a Stein kernel does not exist. For sums of Rademacher random variables, Chatterjee [9] proposed to overcome this issue by computing the Stein kernel of a smoothed version of $\nu$ instead. This approach was then generalized to more general discrete random variables in dimension one by Bhattacharjee and Goldstein [6] using the zero-bias distribution. As Goldstein and Reinert [1] have proposed a generalization of the zero-bias distribution to higher dimensions, it may be possible to extend the smoothing approach to measures in $\mathbb{R}^d$. In this work, we propose to bypass this smoothing procedure by extending the approach of [18] to more general operators $\mathcal{L}_\nu$ such that, for a suitable class of functions $\phi$,

$$\mathbb{E}[\mathcal{L}_\nu \phi(X)] = 0,$$

in which case we say that $\nu$ is invariant under $\mathcal{L}_\nu$.

For instance, suppose there exists a coupling $(X, X')$ where both $X$ and $X'$ are drawn from $\nu$. Then, for any renormalization factor $s > 0$ $\nu$ is invariant under the operator $\mathcal{L}_\nu$ defined on integrable functions $\phi$ by

$$\mathcal{L}_\nu \phi(X) = \frac{1}{s}\mathbb{E}\left[\phi(X) - \phi(X')|X\right].$$

Furthermore, in dimension one, if $\phi$ is real analytic, a Taylor expansion gives

$$\mathcal{L}_\nu \phi(X) = \frac{1}{s}\mathbb{E}\left[\sum_{k=1}^{\infty} \frac{(X' - X)^k}{k!}\phi^{(k)} \mid X\right].$$

In Theorem 1, for the Gaussian measure, and in Theorem 4, for more general measures, we show that if $\mathcal{L}_\nu$ is close to $\mathcal{L}_\mu$, then $\nu$ is close to $\mu$ in terms of Wasserstein distance of order 2. In dimension one and when the target measure $\mu$ is the Gaussian measure, we also derive a similar result for the Wasserstein distance of order $p \geq 1$ in Theorem 6. Let us note that such couplings have already been used in a different approach to Stein's method by Röllin [24].

As an application of our results, we provide convergence rates for the Central Limit Theorem in terms of Wasserstein distances. More precisely, if we consider i.i.d random variables $X_1, ..., X_n$ in $\mathbb{R}^d$ with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1 X_1^T] = I_d$ admitting a finite moment of order $2 + m$ for $m \leq 2$, then the Wasserstein distance of order 2 between the measure of $S_n = n^{-1/2}\sum_{i=1}^n X_i$ and the Gaussian measure decreases in $n^{-1/2+(2-m)/4}$. Moreover, if $d = 1$ and $X_1$ admits a finite moment of order $p + m$ for some $0 \leq m \leq 2 \leq p$, then $W_p(\nu_n, \gamma)$ decreases in $n^{-1/2+(2-m)/2p}$.

Finally, we show how our approach can be used to quantify the Wasserstein distance between the invariant measure of a continuous diffusion process and the invariant measure of an approximation of the diffusion through a Markov chain. We then apply this result to random walks on random geometric graphs to perform density estimation and to study the complexity of a Monte-Carlo algorithm for approximate sampling.

## 2 Notations

Let $x \in \mathbb{R}^d$ and $k \in \mathbb{N}$, we denote by $x^{\otimes k} \in (\mathbb{R}^d)^{\otimes k}$ the tensor of order $k$ of $x$,

$$\forall j_1, \ldots, j_k \in \{1, \ldots, d\}, (x^{\otimes k})_{j_1,\ldots,j_k} = \prod_{i=1}^{k} x_{j_i}.$$

For any $x, y \in (\mathbb{R}^d)^{\otimes k}$ and any symmetric positive-definite $d \times d$ matrix $A$, let

$$< x, y >_A = \sum_{l,j \in \{1,\ldots,d\}^k} x_l y_j \prod_{i=1}^{k} A_{j_i,l_i},$$

and, by extension,

$$\|x\|_A^2 = < x, x >_A .$$

For any smooth function $\phi$ and $x \in \mathbb{R}^d$, let $\nabla^k \phi \in (\mathbb{R}^d)^{\otimes k}$ where

$$\forall j_1, \ldots, j_k \in \{1, \ldots, d\}, (\nabla^k \phi(x))_{j_1,\ldots,j_k} = \frac{\partial^k \phi}{\partial x_{j_1} \ldots \partial x_{j_k}}(x).$$

The Wasserstein distance of order $p \geq 1$ between two measures $\mu$ and $\nu$ on $\mathbb{R}^d$ is defined as

$$W_p(\mu, \nu) = \inf_{\pi} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \pi(dx, dy) \right)^{1/p}$$

where $\pi$ has marginals $\mu$ and $\nu$.

## 3 The approach

Let $E$ be a convex domain of $\mathbb{R}^d$ and $\nu$ and $\mu$ be two measures with support $E$. Suppose $\mu$ is invariant under the operator $\mathcal{L}_\mu = b.\nabla + < a, Hess >_{HS}$ where $b$ and $a$ are $C^\infty$ on $E$ and $a$ is symmetric positive-definite on all of $E$. For any measure $d\eta = hd\mu$, the Fisher information of $\eta$ with respect to $\mu$ is given by

$$I_\mu(\eta) = \int_E \frac{\|\nabla h\|_a^2}{h} d\mu.$$

Let $(P_t)_{t \geq 0}$ be the Markov semigroup with infinitesimal generator $\mathcal{L}_\mu$. For any measure $d\eta = hd\mu$, let $d\eta^t = P_t h d\mu$. We first assume that $d\nu = hd\mu$ and $I_\mu(\nu^t) < \infty$ for any $t > 0$.

Since $\mu$ is the invariant measure of $\mathcal{L}_\mu$, under reasonable assumptions, $\nu^t$ converges to $\mu$ as $t$ goes to infinity. We can thus control the distance between $\mu$ and $\nu$ by controlling the distance between between $\nu^t$ and $\nu$ at any time. The latter can be achieved via the following inequality (see [28]),

$$\frac{d^+}{dt} W_2(\nu, \nu_t) \leq I_\mu(\nu_t)^{1/2}, \tag{2}$$

along with a bound on $I_\mu(\nu_t)$. We have

$$I_\mu(\nu) = \int_E \frac{\|\nabla h\|_a^2}{h} d\mu = \int_E < \nabla h, \nabla(\log h) >_a d\mu.$$

If we write $v_t = \log(P_t h)$,

$$I_\mu(\nu_t) = \int_E < \nabla P_t h, \nabla v_t >_a d\mu.$$

Since $\mu$ is the invariant measure of $\mathcal{L}_\mu$, it satisfies the following integration by parts formula: for any smooth compactly supported functions $f$ and $g$,

$$\int_E < \nabla f, \nabla g >_a d\mu = -\int_E f \mathcal{L}_\mu g d\mu.$$

Since $I_\mu(\nu)$ is finite and $h\mathcal{L}_\mu v_t$ can be shown to be integrable by the results of the following sections, we can apply this integration by parts formula to obtain

$$I_\mu(\nu) = \int_E < \nabla P_t h, \nabla v_t >_a d\mu = -\int_E P_t h \mathcal{L}_\mu v_t d\mu.$$

Using the symmetry of $\mu$ with respect to $P_t$ and the commutativity of $P_t$ and $\mathcal{L}_\mu$,

$$I_\mu(\nu_t) = -\int_E h P_t \mathcal{L}_\mu v_t d\mu = -\int_E \mathcal{L}_\mu P_t v_t d\nu.$$

Now, suppose there exists an operator $\mathcal{L}_\nu$ such that,

$$\int_E \mathcal{L}_\nu P_t v_t d\nu = 0,$$

then

$$I_\mu(\nu_t) = \int_E (\mathcal{L}_\nu - \mathcal{L}_\mu) P_t v_t d\nu.$$

In [18], $\mathcal{L}_\nu$ is given by the Stein kernel but it can be defined in many other ways. For example, as mentioned in the introduction, if $(X, X')$ is a couple of random variables drawn from $\nu$ then, taking

$$\mathcal{L}_\nu P_t v_t(x) = \frac{1}{s}\mathbb{E}\left[\phi(X') - \phi(X) | X = x\right],$$

we have

$$\int_{\mathbb{R}^d} \mathcal{L}_\nu P_t v_t d\nu = 0.$$

Now, suppose $P_t v_t$ is real analytic on $E$, we then have

$$\mathcal{L}_\nu P_t v_t(x) = \frac{1}{s}\mathbb{E}\left[\sum_{k=1}^{\infty} < \frac{(X' - X)^{\otimes k}}{k!}, \nabla^k P_t v_t > | X = x\right],$$

In which case,

$$\begin{aligned} I_\mu(\nu_t) =&\mathbb{E}\left[< \mathbb{E}[X' - X \mid X] - b(X), \nabla P_t v_t(X) >\right] \\ &+ \mathbb{E}\left[< \mathbb{E}[\frac{(X'-X)^{\otimes 2}}{2} \mid X] - a(X), \nabla^2 P_t v_t >\right] \\ &+ \mathbb{E}\left[\sum_{k=3}^{\infty} < \mathbb{E}[\frac{(X'-X)^{\otimes k}}{k!} \mid X], \nabla^k P_t v_t(X) >\right]. \end{aligned} \tag{3}$$

The last step of the approach consists in using the regularizing properties of the semigroup $P_t$ in order to bound the last equation by a quantity involving $P_t\|\nabla v_t\|_a^2$. Then, since $\mathbb{E}[P_t\|\nabla v_t\|_a^2(X)]^{1/2} = I_\mu(\nu_t)^{1/2}$ and $I_\mu(\nu_t)$ is finite, we obtain a bound on $I_\mu(\nu_t)^{1/2}$ and conclude. Let us note that, since $a$ is positive-definite on all of $E$, the bounds we derive on $\|\nabla^k P_t v_t\|_a$ imply $P_t v_t$ is real analytic on all of $E$ [15].

In order to deal with discrete measures, let us note that for $\epsilon > 0$, $\nu^\epsilon$ is well-defined. Thus, if it has a finite Fisher information with respect to $\mu$, we can apply the previous approach to any $\nu^\epsilon$ and let $\epsilon$ go to 0 to obtain a bound on $W_2(\nu, \mu)$ even when $\nu$ is discrete.

Our goal in the remainder of this section will thus consist in providing bounds for Equation 3. We start wtih the Gaussian case where such a bound can be directly obtained using the integral representation of $(P_t)_{t\geq 0}$ and integrations by parts. We then turn ourselves to more general measures $\mu$ and derive a bound using Gamma calculus.

## 3.1 Gaussian case

Let $d\mu = d\gamma = (2\pi)^{-d/2} e^{-\frac{|x|^2}{2}} dx$ be the Gaussian measure in $\mathbb{R}^d$. $\gamma$ is the invariant measure of $\mathcal{L}_\gamma = -x.\nabla + \Delta$ and the associated semigroup $(P_t)_{t\geq 0}$ is the Ornstein-Uhlenbeck semigroup. For any test function $\phi$ and any $x \in \mathbb{R}^d$, $P_t\phi$ admits the following representation

$$P_t\phi(x) = \int_{\mathbb{R}^d} \phi(xe^{-t} + \sqrt{1-e^{-2t}}y) d\gamma(y).$$

Using an integration by part, we obtain

$$\begin{aligned} \nabla P_t\phi(x) &= e^{-t} \int_{\mathbb{R}^d} \nabla\phi(xe^{-t} + \sqrt{1-e^{-2t}}y) d\gamma(y) \\ &= \frac{e^{-t}}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}^d} y\phi(xe^{-t} + \sqrt{1-e^{-2t}}y) d\gamma(y). \end{aligned}$$

For any $k > 0$ and any $i \in \{1, \ldots, d\}^k$, let $H_i$ be the multivariate Hermite polynomial of index $i$,

$$H_i = (-1)^k e^{\frac{|x|^2}{2}} \frac{\partial^k}{\partial x_{i_1} \ldots \partial x_{i_k}} e^{-\frac{|x|^2}{2}}.$$

Multiple integrations by parts thus yield

$$(\nabla^k P_t \phi(x))_i = \frac{e^{-kt}}{(1-e^{-2t})^{k/2}} \int_{\mathbb{R}^d} H_i(y) f(xe^{-t} + \sqrt{1-e^{-2t}}y) d\gamma(y).$$

Hermite polynomials form an ortogonal basis of $L_2(\gamma)$ and their respective norm is

$$\forall i \in \{1,\ldots,d\}^k, \|H_i\|_\gamma = \int_{\mathbb{R}^d} H_i^2(y) d\gamma(y) = \prod_{j=1}^d \left(\sum_{l=1}^k \delta_{i_l,j}\right)!.$$

Therefore, letting

$$\begin{aligned} S(t) =& e^{-2t}\mathbb{E}\left[\left\|\mathbb{E}[\frac{X'-X}{s} \mid X] + X\right\|^2\right] \\ &+ \frac{e^{-4t}}{1-e^{-2t}}\mathbb{E}\left[\left\|\mathbb{E}[\frac{(X'-X)^{\otimes 2}}{2s} \mid X] - I_d\right\|^2\right] \\ &+ \sum_{k=3}^{\infty} \frac{e^{-2kt}}{(sk!)^2(1-e^{-2t})^k} \sum_{i\in\{1,\ldots,d\}^k} \|H_i\|_\gamma \mathbb{E}\left[\|\mathbb{E}[(X'-X)_i^{\otimes k} \mid X]\|^2\right], \end{aligned}$$

and applying Cauchy-Scharwtz's inequality to Equation 3 yields

$$I_\gamma(\nu_t) \leq S(t)^{1/2}\mathbb{E}[P_t\|\nabla v_t(X)\|^2]^{1/2} = S(t)^{1/2} I_\gamma(\nu_t)^{1/2}. \tag{4}$$

We have thus bounded $I_\gamma(\nu_t)$. Now, according to Equation 2, integrating our bound on $I_\gamma(\nu_t)$ for $t \in \mathbb{R}^+$ would yield a bound on $W_2(\nu,\gamma)$. However, we encounter integrability issues for the higher order terms for small values of $t$. To circumvent this issue, we use a family of couplings $(X, X'_t)_{t\geq 0}$ interpolating between $X'_0 = X$, ensuring the problematic terms are 0 at $t = 0$, and $X'_\infty = X'$. Finally, for any discrete measure $\nu$, $\nu^t$ has finite Fisher information as long as the second moment of $\nu$ is finite (see Remark 2.1 [19]). We are now ready to state the first result of this work.

**Theorem 1.** *Let $\nu$ be a measure on $\mathbb{R}^d$ with finite second moment and let $X$ and $(X'_t)_{t\geq 0}$ be random variables drawn from $\nu$. For any $s > 0$,*

$$W_2(\nu,\gamma) \leq \int_0^\infty \sqrt{S(t)}dt,$$

*with*

$$\begin{aligned} S(t) =& e^{-2t}\mathbb{E}\left[\left\|\mathbb{E}[\frac{X'_t-X}{s} \mid X] + X\right\|^2\right] \\ &+ \frac{e^{-4t}}{1-e^{-2t}}\mathbb{E}\left[\left\|\mathbb{E}[\frac{(X'_t-X)^{\otimes 2}}{2s} \mid X] - I_d\right\|^2\right] \\ &+ \sum_{k=3}^{\infty} \sum_{i\in\{1,\ldots,d\}^k} \frac{e^{-2kt}\|H_i\|_\gamma}{(sk!)^2(1-e^{-2t})^k} \mathbb{E}\left[\|\mathbb{E}[(X'_t-X)_i^{\otimes k} \mid X]\|^2\right]. \end{aligned}$$

## 3.2 General case

In general, $(P_t)_{t\geq 0}$ does not admit a closed form formula so we cannot rely on a direct approach. Let us first apply Cauchy-Schwartz's inequality to Equation 3 in order to obtain

$$\begin{aligned} I_\mu(\nu_t) \leq & \mathbb{E}\left[\|\mathbb{E}[X' - X \mid X] - b(X)\|^2_{a^{-1}(X)}\right]^{1/2} \mathbb{E}[\|\nabla P_t v_t(X)\|^2_{a(X)}]^{1/2} \\ & + \mathbb{E}\left[\|\mathbb{E}[\frac{(X'-X)^{\otimes 2}}{2} \mid X] - a(X)\|^2_{a^{-1}(x)}\right]^{1/2} \mathbb{E}[\|\nabla^2 P_t v_t(X)\|^2_{a(X)}]^{1/2} \\ & + \sum_{k=3}^{\infty} \mathbb{E}\left[\|\mathbb{E}[\frac{(X'-X)^{\otimes k}}{k!} \mid X]\|^2_{a^{-1}(x)}\right]^{1/2} \mathbb{E}[\|\nabla^k P_t v_t(X)\|^2_{a(X)}]^{1/2}. \end{aligned} \tag{5}$$

Our objective is to bound $\|\nabla^k P_t v_t\|^2_a$ by a quantity involving $P_t\|\nabla v_t\|^2_a$ using the framework of $\Gamma$-calculus described in [3]. This approach relies on the study of the iterated gradients $\Gamma_i$ defined recursively for any smooth test functions $f, g$ by

$$\begin{aligned} &\Gamma_0(f,g) = fg; \\ &\Gamma_{i+1}(f,g) = \frac{1}{2}\left[\mathcal{L}_\mu(\Gamma_i(f,g)) - \Gamma_i(\mathcal{L}_\mu f, g) - \Gamma_i(f, \mathcal{L}_\mu g)\right]. \end{aligned}$$

The triple $(E, \mu, \Gamma_1)$ is called a Markov triple, a structure extensively studied in [3]. In particular if there exists $\rho \in \mathbb{R}$ such that

$$\Gamma_2 \geq \rho \Gamma_1,$$

the Markov triple is said to satisfy a *curvature-dimension inequality* or $CD(\rho, \infty)$ condition under which $(P_t)_{t\geq 0}$ has many interesting properties. For instance, it is known that, under a $CD(\rho,\infty)$ condition, $(P_t)_{t\geq 0}$ satisfies the following gradient bound (see e.g. Theorem 3.2.3 [3])

$$\|\nabla P_t f\|^2_a \leq e^{-2\rho t} P_t(\|\nabla f\|^2_a),$$

*Remark* 2. Under a curvature-dimension inequality, if $\nu = h d\mu$, then according to Theorem 5.5.2 [3],

$$I_\mu(\nu^t) \leq \frac{2\rho}{1 - e^{-2\rho t}}\left(P_t(h \log h) - P_t h \log(P_t h)\right).$$

Thus, under a curvature-dimension assumption, if $\nu^\epsilon$ has finite entropy with respect to $\mu$ for any $\epsilon > 0$ then $I_\mu(\nu^t)$ is finite for any $t > 0$.

In the proof of Theorem 4.1 [18], the authors show that, under a $CD(\rho,\infty)$ inequality for $\rho > 0$ and assuming there exists $\kappa, \sigma > 0$ such that $\Gamma_3 \geq \kappa \Gamma_2$ and $\Gamma_2 \geq \sigma \|\nabla^2 f\|_a$,

$$\|\nabla^2 P_t f\|^2_a \leq \frac{\kappa}{\sigma(e^{\kappa t} - 1)} P_t \|\nabla f\|^2_a. \tag{6}$$

We could use a similar approach and suppose that for any $k > 1$ there exists some $\kappa_k$ and $\sigma_k$ such that $\Gamma_{k+1} \geq \kappa_k \Gamma_k$ and $\Gamma_k \geq \sigma \|\nabla^k f\|_a$ in order to bound $\|\nabla^k P_t f\|_a$. However, such assumptions would be quite restrictive in practice. Instead, we derive bounds relying on a simple a $CD(\rho, \infty)$ condition.

**Proposition 3.** *Suppose that $\mathcal{L}$ satisfies a $CD(\rho,\infty)$ condition for $\rho \in \mathbb{R}$. Then, for any $k \in \mathbb{N}^\star$, $t > 0$ and any smooth compactly supported function $\phi$,*

$$\|\nabla^k P_t \phi\|_a \leq f_k(t)\sqrt{P_t \|\nabla \phi\|_a^2},$$

*where*

$$f_k(t) = \begin{cases} e^{-\rho t \max(1,k/2)} \left(\frac{(2\rho d)}{e^{2\rho t/(k-1)}-1}\right)^{(k-1)/2} & \text{if } \rho \neq 0 \\ t^{(1-k)/2} \text{ if } \rho = 0. \end{cases}$$

Unfortunately, our bound is not dimension-independent as one could expect from Equation 6, we believe this dependency to be an artifact of the proof. Nevertheless, injecting these bounds in Equation 5 gives us a bound on $I_\mu(\nu^t)^{1/2}$ leading to the following result.

**Theorem 4.** *Let $\nu$ be a measure on $\mathbb{R}^d$. Assume the entropy of $\nu^\epsilon$ with respect to $\mu$ is finite for any $\epsilon > 0$ and let $X$ and $(X'_t)_{t\geq 0}$ be random variables drawn from $\nu$. If $\mathcal{L}_\mu$ satisfies a $CD(\rho,\infty)$ condition for $\rho \in \mathbb{R}$. Then, for any $s > 0$, $T > 0$,*

$$\begin{aligned} W_2(\nu, \nu^T) \leq & \int_0^T e^{-\rho t} \mathbb{E}\left[\left\| \mathbb{E}\left[\frac{X'_t - X}{s} \mid X\right] - b(X)\right\|_{a^{-1}}^2\right] dt \\ & + \int_0^T f_2(t) \mathbb{E}\left[\left\| \mathbb{E}\left[\frac{(X'_t - X)^{\otimes 2}}{2s} \mid X\right] - a(X)\right\|_{a^{-1}}^2\right] dt \\ & + \sum_{k=3}^{\infty} \int_0^T \frac{f_k(t)}{(sk!)^2} \mathbb{E}\left[\left\| \mathbb{E}[(X'_t - X)^{\otimes k} \mid X]\right\|_{a^{-1}}^2\right] dt, \end{aligned}$$

*where the functions $(f_k)_{k\geq 1}$ are defined in Proposition 3.*

If $\rho > 0$, we can let $T$ go to infinity and bound $W_2(\nu, \mu)$. On the other hand, if $\rho \leq 0$, it is still possible to bound $W_2(\nu, \mu)$ as long as $\nu^t$ converges exponentially fast to $\nu$.

**Lemma 5.** *Suppose there exists $\kappa$ such that for any measure $\eta$ and any $t > 0$, we have*

$$W_2(\eta^t, \mu) \leq e^{-\kappa t} W_2(\eta, \nu).$$

*Then, for any $T > 0$,*

$$W_2(\nu, \mu) \leq \frac{W_2(\nu, \nu^T)}{1 - e^{-\kappa T}}.$$

*Proof.* Indeed, we have

$$\begin{aligned} W_2(\nu, \mu) &\leq W_2(\nu, \nu_t) + W_2(\nu_t, \mu) \\ &\leq W_2(\nu, \nu_t) + e^{-\kappa t} W_2(\nu, \mu). \end{aligned}$$

□

Such an exponential convergence to $\mu$ can be verified under weaker conditions than a $CD(\rho, \infty)$ inequality for $\rho > 0$. For example, if $a$ is the identity and $b$ is the gradient of some potential $V$ then this assumption is satisfied whenever $V$ is strongly convex outside a bounded set $C$ with bounded first and second order derivatives on $C$ [13] which is equivalent to satisfying a $CD(\rho_1, \infty)$ condition and having $\Gamma_2 \geq \rho_2 \Gamma_1$ with $\rho_2 > 0$ outside of $C$. An extension of this result for more general $a$ and for manifolds is proposed in [31].

# 4 Gaussian measure in dimension one

The Stein kernel can also be used to bound the Wasserstein distance of order $p \geq 1$ between a measure $\nu$ and the Gaussian measure $\gamma$ under a stronger definition. Let $X$ be a random variable drawn from $\nu$, we say that $\tau_\nu$ is a strong Stein kernel for $\nu$ if

$$\mathbb{E}\left[-X\phi(X) + \tau_\nu(X)\nabla\phi(X)\right] = 0$$

for every compactly supported smooth function $\phi$.

In dimension one, if $\tau_\nu$ satisfies Equation 1 then it satisfies the previous condition, hence we can expect our coupling approach to work for replacing the Stein kernel. Let $\mu = \gamma_1$ be the one-dimensional Gaussian measure. For $k \in \mathbb{N}$, we denote by $H_k$ the $k$-th Hermite polynomial,

$$H_k = (-1)^k e^{\frac{|x|^2}{2}} \frac{d^k e^{-\frac{|x|^2}{2}}}{dx^k}.$$

First, a modification of the proof of Lemma 2 from [21] yields the general estimate

$$\frac{d^+}{dt} W_p(\nu, \nu_t) \leq \left( \int_{\mathbb{R}} |v_t'|^p d\nu_t \right)^{1/p}. \tag{7}$$

Let us provide a version fo $v_t'$. Let $(X, X')$ be random variables drawn from $\nu$ and $Z$ be a Gaussian random variable. For $t > 0$, let $F_t = e^{-t}X + \sqrt{1 - e^{-2t}}Z$ and consider the function $\rho_t$ defined for any $x \in \mathbb{R}$ as follows

$$\begin{aligned}\rho_t(x) = \mathbb{E}\left[e^{-t}(\frac{X' - X}{s} + X) + \frac{e^{-2t}}{\sqrt{1 - e^{-2t}}}(\frac{(X' - X)^2}{2s} - 1)H_1(Z) | F_t = x\right] \\ + \mathbb{E}\left[\sum_{k=3}^{i-1} \frac{e^{-kt}}{s\sqrt{1 - e^{-2t}}^{k-1}} \frac{(X' - X)^k}{k!} H_{k-1}(Z) | F_t = x\right].\end{aligned}$$

For any compactly supported smooth function $\phi : \mathbb{R} \to \mathbb{R}$, we obtain, after successive integrations by parts with respect to $Z$,

$$\begin{aligned}\mathbb{E}[\rho_t(F_t)\phi(F_t)] = \mathbb{E}\left[e^{-t}X\phi(F_t) - e^{-2t}\phi'(F_t)\right] \\ + \mathbb{E}\left[\sum_{k=1}^{\infty} e^{-kt} \frac{(X' - X)^k}{sk!} \phi^{(k-1)}(F_t)\right].\end{aligned}$$

Let $\Phi$ be a primitive function of $\phi$, by the results of Section 3.1, $\mathbb{E}[\Phi(F_t) \mid X = x] = P_t\Phi(x)$ is real analytic. Hence, since $X'$ and $X$ have the same measure we have

$$\mathbb{E}\left[\sum_{k=1}^{\infty} e^{-kt}\frac{(X'-X)^k}{sk!}\phi^{(k-1)}(F_t)\right] = \frac{1}{s}\mathbb{E}[\Phi(e^{-t}X' + \sqrt{1-e^{-2t}}Y) - \Phi(e^{-t}X + \sqrt{1-e^{-2t}}Y)] = 0.$$

Therefore,

$$\begin{aligned}\mathbb{E}[(\rho_t(F_t) - F_t)\phi(F_t)] &= \mathbb{E}[(-F_t + e^{-t}X)\phi(F_t) - e^{-2t}\phi'(F_t)] \\ &= \mathbb{E}[-(1-e^{-2t})\nabla\phi(F_t) - e^{-2t}\phi'(F_t)] \\ &= -\mathbb{E}[\nabla\phi(F_t)].\end{aligned}$$

Therefore, $\rho_t$ satisfies the characterization of $v_t'$ presented in Equation 2.28 [19]: it is thus a version of $v_t'$. We are thus able to bound

$$\left(\int_{\mathbb{R}} |v_t'|^p d\nu_t\right)^{1/p} = \mathbb{E}[|\rho_t(F_t)|^p]^{1/p}$$

using the $L_p(\gamma_1)$-norm of the Hermite polynomials $\|H_k\|_{p,\gamma_1}^p = \int_{\mathbb{R}} |H_k|^p d\gamma_1$. Injecting this bound in Equation 7, we are able to bound $W_p(\nu, \gamma_1)$.

**Theorem 6.** *Let $\nu$ be a measure on $\mathbb{R}$ and let $X$ and $(X_t)_{t\geq 0}$ be random variables drawn from $\nu$. We have, for any $p \geq 1$, $s > 0$,*

$$\begin{aligned}W_p(\nu,\gamma_1) \leq& \int_0^{\infty} e^{-t}\mathbb{E}[\left|\mathbb{E}[\frac{X_t'-X}{s} \mid X] + X\right|^p]^{1/p} dt \\ &+ \int_0^{\infty} \frac{e^{-2t}\|H_1\|_{p,\gamma_1}}{\sqrt{1-e^{-2t}}}\mathbb{E}[\left|\mathbb{E}[\frac{(X_t'-X)^2}{2s} \mid X] - 1\right|^p]^{1/p} dt \\ &+ \sum_{k=3}^{\infty}\int_0^{\infty} \frac{e^{-kt}\|H_{k-1}\|_{p,\gamma_1}}{s\sqrt{1-e^{-2t}}^{k-1}k!}\mathbb{E}[\left|\mathbb{E}[(X_t'-X)^k \mid X]\right|^p]^{1/p} dt.\end{aligned}$$

*Remark* 7. [17] gives the asymptotic of the $p$-norm of Hermite polynomials with respect to the Gaussian measure, more precisely there exist constants $C(p)$ such that

$$\|H_k\|_p \leq \begin{cases} C(p)\sqrt{k!}k^{-1/4}(1+O(k^{-1})) \text{ if } 0 < p < 2 \\ C(p)\sqrt{k!}(p-1)^{k/2}(1+O(k^{-1})) \text{ if } p > 2 \end{cases}$$

# 5 Applications

## 5.1 Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. random variables and let $\nu_n$ be the measure of $S_n = n^{-1/2}\sum_{i=1}^n X_i$. According to the Central Limit Theorem, $S_n$ should converge to the Gaussian measure $\gamma$. Let $X'_1, \dots X'_n$ be independent copies of $X_1, \dots, X_n$ and let $I$ be a uniform random variable on $\{1, \dots, n\}$. For any $t > 0$, we pose

$$S'_{n,t} = S_n + n^{-1/2}(X'_I - X_I)1_{\|X'_I\|, \|X_I\| \leq \sqrt{nt}}.$$

By construction, for any $t > 0$, $S'_{n,t}$ is drawn from $\nu_n$. Let $p \geq 2, 0 \leq m \leq 2$ and suppose $\mathbb{E}[\|X_1\|^{p+m}] < \infty$. Then, after some technical computations presented in Subsection 6.2, taking $l = \min(4 - (p+m), 2)$, $\alpha = -1/2 + (2-m)/2p$ and $Y_t = S'_{n,t} - S_n$, there exists a constant $C$ depending on the law of $X_1$, $d$ and $p$ such that,

- $\mathbb{E}[\|\mathbb{E}[Y_t \mid S_n]\|^p]^{1/p} \leq Cn^{\alpha}(t^{-m/2p} + t^{-l/4})$;
- $\mathbb{E}\left[\left\|\mathbb{E}[\frac{Y_t^{\otimes 2}}{2} \mid S_n] - I_d\right\|^p\right]^{1/p} \leq C((nt)^{-l/2} + n^{\alpha}t^{1/2}(t^{-m/2p} + t^{-l/4}))$;
- $\mathbb{E}[\left\|\mathbb{E}[Y_t^{\otimes 3} \mid S_n]\right\|^p]^{1/p} \leq Cn^{\alpha}t(t^{-m/2p} + t^{-l/4})$;
- $\forall k \geq 4, \mathbb{E}[\left\|\mathbb{E}[Y_t^{\otimes k} \mid S_n]\right\|^p]^{1/p} \leq Ct^{k/2-2}(n^{-l/2}t^{1-l/2} + n^{\alpha}t^{3/2}(t^{-m/2p} + t^{-l/4}))$.

We can now apply both Theorem 1 and Theorem 6 with a timestep $s = \frac{1}{n}$ to obtain the following result.

**Theorem 8.** *Let $X_1, \dots, X_n$ be i.i.d. random variables in $\mathbb{R}^d$ with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1 X_1^T] = I_d$. If $\mathbb{E}[\|X_1\|^{2+m}] < \infty$ for some $m \in [0, 2]$, then*

$$W_2(\nu_n, \gamma) = O(n^{-1/2+(2-m)/4}).$$

*Moreover, if $d = 1$ then for any $p \geq 2$, if $\mathbb{E}[|X_1|^{p+m}] < \infty$ for some $m \in [0, 2]$,*

$$W_p(\nu_n, \gamma) = O(n^{-1/2+(2-m)/2p}).$$

The one-dimensional result completes a result obtained by Rio [22] who considered the case $1 \leq p \leq 2, m = 2$ and generalizes a result obtained by Sakhanenko [25] treating the case $p > 2, m = 0$. Bobkov [7] also recovered the case $p = 2, m = 2$ using an entropic approach and recently proved the case $m = 2$ for any $p > 2$ [8]. To our knowledge, the multidimensional result is new although the entropic approach from Bobkov [7] might be generalized to the multidimensional setting at the expense of stronger assumptions on the moments of the variables.

## 5.2 Diffusion approximation

Let $\mu$ be the invariant measure of the diffusion process with infinitesimal generator $\mathcal{L}_\mu = b.\nabla + < a, Hess >_{HS}$. Consider a discretization of this diffusion process by a Markov chain $M$ with transition kernel $K$ and invariant measure $\pi$ and let $s$ be the timestep of this discretization. Let $X$ be a random variable drawn from $\pi$ and let $\xi$ be a random jump from $X$. Then for any $t > 0, T > 0$,

$$X_t = X + 1_{t \geq T}\xi$$

and $X$ follow the same law, and we can apply Theorem 4.

**Corollary 9.** *Under the assumptions Theorem 4, we have, for any $T_1 > T_2 > 0$,*

$$
\begin{aligned}
W_2(\pi_s, (\pi_s)_{T_1}) \leq & \int_0^{T_2} e^{-\rho t} \mathbb{E}[\|b(X)\|_{a^{-1}}^2]^{1/2} + f_2(t) dt \\
& + \int_{T_2}^{T_1} e^{-\rho t} \mathbb{E}[\left\| \mathbb{E}[\frac{\xi}{s} \mid X] - b(x) \right\|_{a^{-1}}^2 ]^{1/2} dt \\
& + \int_{T_2}^{T_1} f_2(t) \mathbb{E}[\left\| \mathbb{E}[\frac{\xi^{\otimes 2}}{2s} \mid X] - a(x) \right\|_{a^{-1}}^2 ]^{1/2} dt \\
& + \int_{T_2}^{T_1} \sum_{k=3}^{\infty} f_k(t) \mathbb{E}[\left\| \mathbb{E}[\frac{\xi^{\otimes k}}{k! s} \mid X] \right\|_{a^{-1}}^2 ]^{1/2} dt,
\end{aligned}
$$

*where the functions $(f_k)_{k \geq 1}$ are defined in Proposition 3.*

The quantities involved in this Corollary also appear in standard results proving the weak convergence of a familiy $(M_s)_{s>0}$ of Markov chains with state space $S_s$ to the diffusion process with infinitesimal generator $\mathcal{L}_\mu$. An example of such result can be found in [26] and states that such a convergence occurs if

$$
\begin{aligned}
& \lim_{s \to 0} \sup_{x \in S_s} \mathbb{E}\left[\frac{\xi}{s} \mid X = x\right] = b; \\
& \lim_{s \to 0} \sup_{x \in S_s} \mathbb{E}\left[\frac{\xi^{\otimes 2}}{2s} \mid X = x\right] = a; \\
& \forall r > 0, \lim_{s \to 0} \sup_{x \in S_s} P\left(\left\|\frac{\xi}{s}\right\| > r\right) = 0; \\
& \lim_{s \to 0} M_s^0 = X_0.
\end{aligned}
$$

### 5.2.1 Density approximation on $k$-nearest neighbor graphs

Let $X_1, \ldots, X_n$ br i.i.d. random variables on $\mathbb{R}^d$ drawn from a measure $\mu$ with a smooth density $f$. Let $\mathcal{X} = (X_1, \ldots X_n)$ and let $r_{\mathcal{X}}$ be a function from $\mathbb{R}^d$ to $\mathbb{R}^+$. Let $G$ be a graph with vertices $\mathcal{X}$ and edges $\{(x,y) \in \mathcal{X}^2 \mid \|x-y\|^2 \leq r_{\mathcal{X}}(x)\}$, such a graph is called a *random geometric graph.* These graphs are widely spread in machine learning applications as input for various algorithms such as spectral clustering [29], semi-supervised label propagations [5], dimensionality reduction [4] and many more. Of particular interest is the $k$-nearest neighbor graph, for $k \in \mathbb{N}$, corresponding to

$$
r_{\mathcal{X}}(x) = \inf \left\{ s \in \mathbb{R}^+ \middle| \sum_{i=1}^n 1_{\|X_i - x\| \leq s} \geq k \right\},
$$

which is one of the most frequently used random geometric graphs because of its sparsity. One may wonder whether the sole graph structure preserve all the information from the original data. One way to answer this question is to consider the possibility to recover the density $f$ from the graph structure when $n$ gets to infinity and $r_{\mathcal{X}}$ goes to zero. For the $k$-nearest-neighbor graph, Luxburg and Alamgir [30] have shown this is possible, in dimension one and under conditions on the growth of $k$, and conjecture their estimator to be consistent in higher dimension. For general random geometric graphs, another approach consists in using the invariant measure of the random walk on the graph $\pi$ to estimate $f$. Indeed, if $r_{\mathcal{X}}$ is constant over $T$ and large enough, the random walk on the graph is reversible and its invariant measure corresponds to a ball density estimator. However, when the graph is directed, the random walk is not reversible and the invariant measure of the random walk has no closed-form expression. Instead, if $r_{\mathcal{X}}$, after a proper rescaling, converges uniformly to $\tilde{r}$, Ting, Huang and Jordan [27] have shown the random walk is a discrete approximation of a diffusion process with infinitesimal generator

$$\mathcal{L} = \tilde{r}^2 \nabla \log f . \nabla + \frac{\tilde{r}^2}{2} \Delta.$$

The invariant measure $\tilde{\mu}$ of this diffusion process has a density proportional to $f^2 \tilde{r}^2$. In this framework, Hashmioto, Sun and Jaakkola [14] proved a weak convergence of $\pi$ to $\tilde{\mu}$. Using our results, we can obtain quantative convergence rates for this approach.

To avoid boundary effects, we assume $\mu$ is a measure on the flat torus $T = (\mathbb{R}/\mathbb{Z})^d$ and we focus on the case of the $k$-nearest neighbor graph. In this case, the approximation timestep is

$$s = \left(\frac{k}{n}\right)^{2/d} \frac{\int_{\|x\| \leq 1} x_1^2 dx}{\left(\int_{\|x\| \leq 1} 1 dx\right)^{1+2/d}}$$

and $\tilde{r} = f(x)^{-1/d}$. While $T$ is not a domain of $\mathbb{R}^d$, the arguments used in Theorem 4 still hold, let us check its assumptions. As $T$ is compact and $f$ is smooth and strictly positive, $f^{-2/d} \nabla \log f$ and $f^{-2/d}$ are smooth, hence, a $CD(\rho, \infty)$ condition is verified for some $\rho \in \mathbb{R}$. Moreover, for any $\epsilon > 0$, $\pi^\epsilon$ is a measure with strictly positive smooth density, it has thus finite Fisher information with respect to $\tilde{\mu}$. Finally, the assumption of Lemma 5 is verified thanks to Corollary 2.2 [31].

Let $X$ be drawn from $\pi$ and $\xi$ be a random jump on one of the $k$-nearest neighbor of $X$, there exists a constant $C$ such that with probability $\frac{C}{n}$,

- $\|\mathbb{E}[\frac{\xi}{s} \mid X] - f^{-2/d} \nabla f\| \leq C(\frac{\sqrt{\log n} n^{1/d}}{k^{1/2+1/d}} + (\frac{k}{n})^{2/d})$;
- $\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2s} \mid X] - f^{-2/d} I_d\| \leq C(\sqrt{\frac{log n}{k}} + (\frac{k}{n})^{2/d})$;
- $\|\mathbb{E}[\frac{\xi^{\otimes 3}}{s} \mid X]\| \leq C(\frac{\sqrt{\log n} k^{1/d}}{n^{1/d} k^{1/2}} + (\frac{k}{n})^{2/d})$;
- $\forall j > 3, \|\mathbb{E}[\frac{\xi^{\otimes j}}{s} \mid X]\| \leq C^j (\frac{k}{n})^{(j-2)/d}$.

Plugging these bounds in Corollary 9 and using Lemma 5, we are able to quantify the distance between $\pi$ and $\tilde{\mu}$.

**Proposition 10.** *There exists $C > 0$ such that, with probability $\frac{C}{n}$,*

$$W_2(\pi, \tilde{\mu}) \le C\left(\frac{\sqrt{logn}n^{1/d}}{k^{1/2+1/d}} + \left(\frac{k}{n}\right)^{1/d}\right).$$

#### 5.2.2 Analysis of lower order schemes for the Langevin Monte Carlo algorithm

Quite often in Bayesian statistics, one is interested in sampling points from a probability measure $d\mu = e^{-u}dx$ on $\mathbb{R}^d$. Many Monte-Carlo algorithms have been proposed and analyzed to solve this task, we want to show how our result can be used to study the convergence rate of a simple Monte-Carlo algorithm.

The measure $\mu$ is the stationary measure of the diffusion process $(Y_t)_{t\geq 0}$ solution of the following stochastic differential equation

$$dY_t = -\nabla u(Y_t)dt + \sqrt{2}dW_t,$$

where $W_t$ is a $d$-dimensional Brownian motion. This diffusion process has the following infinitesimal generator

$$\mathcal{L}_\mu = -\nabla u.\nabla + \Delta.$$

Since, under some assumptions on $\mu$, the measure of $Y_t$ converges to $\mu$ as $t$ goes to infinity, one may want to sample points from $\mu$ by approximating $Y_t$. Using the Euler-Maruyama approximation with timestep $s$, we discretize $Y_t$ using a Markov chain $M$ with $M^0 = 0$ and transitions given by

$$M^{n+1} = M^n - s\nabla u(M^n) + \sqrt{2s}\mathcal{N}_n,$$

where $\mathcal{N}_1, \dots, \mathcal{N}_n$ is a sequence of independent normal random variables with mean 0 and covariance matrix $I_d$. If the timestep is small enough, the invariant measure of $M^n$, which we call $\pi$, should be close to $\mu$. Hence, for $n$ large enough, the measure of $M^n$ should be close to its invariant measure and thus be close to $\mu$. Approximate sampling for $\mu$ using this approach is known as the Langevin Monte-Carlo (LMC) algorithm [23].

One may then wonder how many iterations of the algorithm are required to achieve a given accuracy. Answering this question is linked to the choice of $s$ as this parameter must satisfy a trade-off: large values lead to a poor approximation of $\mu$ by $\pi$, but the smaller $s$ is, the larger the number of iterations required for the measure of $M^n$ to be close to $\pi$. Recently, Dalalyan [10] proved that whenever $\mu$ is a strictly log-concave measure (i.e. satisfying a $CD(\rho, \infty)$ condition for $\rho > 0$), the LMC algorithm can reach an $\epsilon$ accuracy in total variation distance in $O(\epsilon^{-2}(d^3 + d\log(1/\epsilon))$ steps. For the Wasserstein distance, this complexity was later improved to $O(\epsilon^{-1}\sqrt{d}\log(1/\epsilon))$ by Durmus and Moulines [11]. A second order discretization, called the Ozaki discretization, was also considered by Dalalyan [10]. Under this scheme, the number of iterations required to achieve an $\epsilon$ accuracy in total variation distance is smaller than $O(\epsilon^{-1}\dim(d+\log(1/\epsilon))^{3/2})$. Here, we propose to do the opposite by considering an example of a smaller order scheme with non-normal increments. Let $(B_n)_{n\geq 0}$ be independent multivariate Rademacher random variables, we consider the following scheme

$$M^{n+1} = M^n - s\nabla u(M^n) + \sqrt{2s}B_n. \tag{8}$$

Let us assess the performance of the LMC algorithm using such a scheme.

Let $\mu$ be log-concave a log-concave measure, i.e. $d\mu = e^{-u}d\lambda$ and there exists $\rho > 0$ such that

$$\forall x \in \mathbb{R}^d, q\nabla u(x) - \nabla u(y), (x-y) >\geq \rho\|x-y\|_2^2.$$

Taking $\Gamma_1(f,g) =< \nabla f, \nabla g >$, this is equivalent to saying the Markov Triple $(\mathbb{R}^d, \mu, \Gamma_1)$ satisfies a $CD(\rho,\infty)$ condition for $\rho > 0$. Moreover, as shown in Subsection 6.4, $\pi$ has finite second moment which implies, by Theorem 5.1 [2], that $\pi^\epsilon$ has finite entropy with respect to $\mu$ for $\epsilon > 0$. Together with Remark 2 this implies $\pi^\epsilon$ has finite Fisher information with respect to $\mu$ for any $\epsilon > 0$. Let $X$ be a random variable drawn from $\pi$ and $\xi$ be an increment from state $X$. If $\mu$ is log-concave and $\nabla u$ is Lipschitz continous, then there exists $C > 0$ such that

- $\mathbb{E}[\frac{\xi}{s} - \nabla u(X) \mid X] = 0$;
- $\mathbb{E}[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2s} - I_d \mid X]\|^2]^{1/2} \leq C(sd)^{1/2}$;
- $\mathbb{E}[\|\mathbb{E}[\frac{\xi^{\otimes 3}}{s} \mid X]\|^2]^{1/2} \leq C^3 sd$;
- $\mathbb{E}[\|\mathbb{E}[\frac{\xi^{\otimes k}}{s} \mid X]\|^2]^{1/2} \leq C^k s^{k/2-1} d^{(k-1)/2}$.

Computations for the previous inequalities can be found in Subsection 6.4. Applying Corollary 9 with $T = sd^2$ allows us to bound $W_2(\pi,\mu)$.

**Proposition 11.** *Suppose $\mu$ is log concave. Then, if $\|\nabla u\| \leq L$,*

$$W_2(\pi,\mu) \leq O(d^2 s^{-1/2}).$$

Combining this result with the exponential convergence of $M^n$ to $\pi$, which can be obtained using the coarse Ricci curvature framework introduced by Ollivier [20], only $O(\epsilon^{-2}d^4\log(1/\epsilon))$ iterations are required to achieve an $\epsilon$ accuracy in Wasserstein distance between the measure sampled by the LMC algorithm and $\mu$. We believe our result to be suboptimal due to the dependency on the dimension of the function $f_k$ defined in Proposition 3, we conjecture the correct complexity to be $O(\epsilon^{-2}d^2\log(1/\epsilon))$.

# 6 Proofs

## 6.1 Proof of Proposition 3

By Theorem 3.2.4 [3], under a $CD(\rho,\infty)$, we have for any compactly supported smooth function $\phi$,

$$\|\nabla P_t\phi\|_a \leq e^{-2\rho t}P_t\|\nabla\phi\|_a. \tag{9}$$

In order to prove the proposition, we need to replace the integration by parts used in the Gaussian case.

**Lemma 12.** *Suppose $\mathcal{L}$ satisfies a $CD(\rho,\infty)$ condition, then for all compactly supported smooth function $\phi$, and any $t>0$,*

$$\|\nabla P_t\phi\|_a^2 \leq \frac{2\rho}{e^{2\rho t}-1}P_t|\phi|^2.$$

*Proof.* Let $t>0$, for any $0<s<t$ let

$$\Lambda(s) = P_s(\Gamma_0(P_{t-s}\phi)),$$

the first two derivatives of this function are

$$\begin{aligned}
\Lambda'(s) &= 2P_s(\Gamma_1(P_{t-s})\phi);\\
\Lambda''(s) &= 4P_s(\Gamma_2(P_{t-s})\phi).
\end{aligned}$$

By our assumption, $\Lambda''(s) \geq 2\rho\Lambda'(s)$. Hence, by Gronwall's Lemma, $\Lambda'(s) \geq e^{2\rho s}\Lambda'(0)$. Now, we have

$$\begin{aligned}
\Gamma_1(P_t\phi) &\leq \frac{2\rho}{e^{2\rho t}-1}\int_0^t e^{2\rho s}\Gamma_1(P_t\phi)ds\\
&\leq \frac{2\rho}{e^{2\rho t}-1}\int_0^t \Lambda'(0)ds\\
&\leq \frac{2\rho}{e^{2\rho t}-1}\int_0^t \Lambda'(s)ds\\
&\leq \frac{2\rho}{e^{2\rho t}-1}(P_t(\Gamma_0(\phi)) - \Gamma_0(P_t\phi))\\
&\leq \frac{2\rho P_t(\Gamma_0(\phi))}{e^{2\rho t}-1}.
\end{aligned}$$

□

Let $(e_1,\ldots,e_d)$ be an orthonormal basis of $\mathbb{R}^d$ with respect to the $a$-scalar product $< .,. >_a$.

**Lemma 13.** *For any smooth function $\phi$ and any $k>0$, we have*

$$\|\nabla^k\phi\|_a = \sup_{\alpha\in\mathbb{R}^d,\|\alpha\|=1}\sum_{i=1}^d \alpha_i\|\nabla^{k-1} < \nabla\phi, e_i >_a\|.$$

*Proof.* By duality of the $a$-norm, we have

$$\begin{aligned}
\|\nabla^k \phi\|_a &= \sup_{h\in(\mathbb{R}^d)^{\otimes k}, \|h\|_a=1} < \nabla^k \phi, h >_a \\
&= \sup_{\alpha\in\mathbb{R}^d, \|\alpha\|=1} \sum_{i=1}^d \sup_{h\in(\mathbb{R}^d)^{\otimes k-1}, \|h\|_a=1} < \nabla^k \phi, \alpha_i e_i \otimes h >_a \\
&= \sup_{\alpha\in\mathbb{R}^d, \|\alpha\|=1} \sum_{i=1}^d \sup_{h\in(\mathbb{R}^d)^{\otimes k-1}, \|h\|_a=1} \alpha_i < \nabla^{k-1} < \nabla\phi, e_i >_a, h >_a \\
&= \sup_{\alpha\in\mathbb{R}^d, \|\alpha\|=1} \sum_{i=1}^d \alpha_i \|\nabla^{k-1} < \nabla\phi, e_i >_a \|_a.
\end{aligned}$$

□

Let us prove Proposition 3 by induction. Take $x \in \mathbb{R}^d$ and let $\phi$ be a compactly supported smooth function $\phi$. The inequality holds for $k = 1$ by Equation 9, suppose it is true for some $k \in \mathbb{N}$. By Lemma 13 we only need to bound

$$\|\nabla^k < \nabla P_t\phi, e_i >_a \|_a = \lim_{\epsilon\to 0} \|\nabla^k \left(P_t\phi(x+\epsilon a e_i) - P_t\phi(x)\right)\|_a$$

for any $e_i$. Let $\pi_\epsilon$ be a coupling between the measures of two diffusion processes with infinitesimal generator $\mathcal{L}_\mu$ started respectively at $x$ and $x + \epsilon a e_1$ at time $t$, then

$$P_t\phi(x+\epsilon a e_1) - P_t\phi(x) = P_t\left(\frac{\phi\circ(id+\pi_\epsilon)-\phi}{\epsilon}\right),$$

Applying the induction hypothesis,

$$\begin{aligned}
&\|\nabla^k P_t\left(\frac{\phi\circ(id+\pi_\epsilon)-\phi}{\epsilon}\right)\|_a^2 \leq \\
&\qquad e^{-\rho t\max(2,k)\frac{k-1}{k}}\left(\frac{2\rho d}{e^{2\rho dt/k}-1}\right)^{k-1} P_{t\frac{k-1}{k}}\|\nabla P_{t/k}\left(\frac{\phi\circ(id+\pi_\epsilon)-\phi}{\epsilon}\right)\|_a^2.
\end{aligned}$$

By applying Lemma 12,

$$\begin{aligned}
&\|\nabla^k P_t\left(\frac{\phi\circ(id+\pi_\epsilon)-\phi}{\epsilon}\right)\|_a^2 \leq \\
&\qquad\qquad e^{-\rho t(k-1)}d^{k-1}\left(\frac{2\rho}{e^{2\rho dt/k}-1}\right)^{k} P_t\left|\frac{\phi\circ(id+\pi_\epsilon)-\phi}{\epsilon}\right|^2.
\end{aligned}$$

By Theorem 2.2 [16], Equation 9 implies that we can take $\pi_\epsilon$ such that, for any $y \in \mathbb{R}^d$, $\|\pi_\epsilon(y)\|_{a^{-1}} \leq \epsilon e^{-\rho t} + o(\epsilon)$, therefore

$$\lim_{\epsilon\to 0}\left|\frac{\phi\circ(id+\pi_\epsilon)-\phi}{\epsilon}\right| = \lim_{\epsilon\to 0}\left|\frac{< \nabla\phi, a^{-1}\pi_\epsilon >_a + o(\|\pi_\epsilon\|)}{\epsilon}\right| \leq e^{-\rho t}\|\nabla\phi\|_a.$$

Since a similar result holds for any $e_i$, we have, using Lemma 13,

$$\|\nabla^{k+1} P_t \phi\|_a^2 \leq$$
$$e^{-\rho t(k+1)} d^{k-1} \left(\frac{2\rho}{e^{2\rho d t/k} - 1}\right)^k \sup_{\alpha \in \mathbb{R}^d, \|\alpha\|=1} \left(\sum_{i=1}^d \alpha_i \sqrt{P_t \|\nabla \phi\|^2}\right)^2.$$

Finally, since the supremum is obtained for $\alpha_1 = \cdots = \alpha_d = \frac{1}{\sqrt{d}}$, the proof is complete.

## 6.2 Proof of Theorem 8

For any $t > 0$, let

$$\begin{aligned}
&\epsilon_{\nu_n,1}(S_n, t) = \mathbb{E}[S'_{n,t} - S_n \mid S_n] + S_n;\\
&\epsilon_{\nu_n,2}(S_n, t) = \mathbb{E}[\frac{(S'_{n,t} - S_n)^{\otimes 2}}{2} \mid S_n] - I_d;\\
&\epsilon_{\nu_n,3}(X, t) = \mathbb{E}[(S'_{n,t} - S_n)^{\otimes 3} \mid X];\\
&\forall k \geq 4, \epsilon_{\nu_n,k}(S_n, t) = \mathbb{E}[(S'_{n,t} - S_n)^{\otimes k} \mid S_n].
\end{aligned}$$

Now, let $X$ and $X'$ be two independent random variables drawn from the measure of $X_1$ and $t > 0$. Let $Y_t = (X' - X)1_{\|X\|,\|X'\| \leq \sqrt{nt}}$, we pose

$$\begin{aligned}
&\epsilon_{\nu,1}(X, t) = \mathbb{E}[Y_t \mid X] + X;\\
&\epsilon_{\nu,2}(X, t) = \mathbb{E}[\frac{Y_t^{\otimes 2}}{2} \mid X] - I_d;\\
&\epsilon_{\nu,3}(X, t) = \mathbb{E}[Y_t^{\otimes 3} \mid X];\\
&\forall k \geq 4, \epsilon_{\nu,k}(X, t) = \mathbb{E}[\|Y_t\|^k].
\end{aligned}$$

Since the $X_i$ are independent, by the multidimensional version of Rosenthal's inequality given in Lemma 1 [12], there exists a constant $C_p$ such that

$$\mathbb{E}[\|\epsilon_{\nu_n,k}(S_n, t)\|^p]^{1/p} \leq$$
$$C_p n^{-k/2} \left(n\|\mathbb{E}[\epsilon_{\nu,k}(X, t)]\| + (n\mathbb{E}[\|\epsilon_{\nu,k}(X, t)\|^2])^{1/2} + (n\mathbb{E}[\|\epsilon_{\nu,k}(X, t)\|^p])^{1/p}\right).$$

Let us bound $\|\mathbb{E}[\epsilon_{\nu,k}(X, t)]\|$, $\mathbb{E}[\|\epsilon_{\nu,k}(X, t)\|^2]$ and $\mathbb{E}[\|\epsilon_{\nu,k}(X, t)\|^p]$ using the moments of $\nu$ of orders smaller than $p + m$. Let us pose $l = \min(4 - (p + m), 2)$. Starting with $\epsilon_{\nu,1}$, we have $\mathbb{E}[\epsilon_{\nu,1}(X, t)] = 0$. Moreover, since $\mathbb{E}[X] = 0$, we have, for $t > n^{-1/2}$,

$$
\begin{aligned}
&\mathbb{E}[\|\epsilon_{\nu,1}(X,t)\|^p] \\
&\quad = \mathbb{E}\left[1_{\|X\|\geq\sqrt{nt}}\|X\|^p + 1_{\|X\|\leq\sqrt{nt}}\|\mathbb{E}[X'1_{\|X\|\leq\sqrt{nt}}] + P(\|X\|\leq\sqrt{nt})X\|^p\right] \\
&\quad = \mathbb{E}\left[1_{\|X\|\geq\sqrt{nt}}\|X\|^p + 1_{\|X\|\leq\sqrt{nt}}\|\mathbb{E}[-X'1_{\|X\|\geq\sqrt{nt}}] + P(\|X\|\leq\sqrt{nt})X\|^p\right] \\
&\quad \leq \mathbb{E}[1_{\|X\|\geq\sqrt{nt}}\|X\|^p] + 2^{p-1}(\mathbb{E}[1_{\|X\|\geq\sqrt{nt}}\|X'\|^p] + P(\|X\|\leq\sqrt{nt})^p\mathbb{E}[\|X\|^p]) \\
&\quad \leq (2^{p-1}+1)\mathbb{E}[1_{\|X\|\geq\sqrt{nt}}\|X\|^p] + 2^{p-1}P(\|X\|\leq\sqrt{nt})^p\mathbb{E}[\|X\|^p] \\
&\quad \leq (2^{p-1}+1)(\sqrt{nt})^{-m}\mathbb{E}[\|X\|^{p+m}] + 2^{p-1}(\sqrt{nt})^{-2p}\mathbb{E}[\|X\|^2]^{1/2}\mathbb{E}[\|X\|^p] \\
&\quad \leq 2^p(\sqrt{nt})^{-m}\mathbb{E}[\|X\|^2]^{1/2}\mathbb{E}[\|X\|^{p+m}],
\end{aligned}
$$

and

$$\mathbb{E}[\epsilon_{\nu,1}(X,t)^2] \leq 8(\sqrt{nt})^{-l}\mathbb{E}[\|X\|^2]^{1/2}\mathbb{E}[\|X\|^{2+l}].$$

Thus, there exists $C_{1,p,\nu}$ such that

$$\mathbb{E}[\|\epsilon_{\nu_n,1}(S_n,t)\|^p]^{1/p} \leq C_{1,p,\nu,d}(t^{-m/2p}+t^{-l/4})n^{-1/2+(2-m)/2p}.$$

We now deal with $\epsilon_{\nu,2}$. Since $\mathbb{E}[X]=0$ and $\mathbb{E}[X^{\otimes 2}]=I_d$, for $t\geq n^{-1/2}$,

$$
\begin{aligned}
\|\mathbb{E}[\epsilon_{\nu,2}(X,t)]\| &= \|\mathbb{E}[1_{\|X\|,\|X'\|\leq\sqrt{nt}}\frac{(X'-X)^{\otimes 2}}{2} - I_d]\| \\
&= \|\mathbb{E}[\frac{1_{\|X\|\leq\sqrt{nt}}X^{\otimes 2}}{2} - \frac{I_d}{2}] + \mathbb{E}[\frac{1_{\|X'\|\leq\sqrt{nt}}X'^{\otimes 2}}{2} - \frac{I_d}{2}] \\
&\qquad + \mathbb{E}[1_{\|X\|\leq\sqrt{nt}}X]\otimes\mathbb{E}[1_{\|X'\|\leq\sqrt{nt}}X']\| \\
&= \|\mathbb{E}[1_{\|X\|\geq\sqrt{nt}}X^{\otimes 2}] + \mathbb{E}[X1_{\|X\|\leq\sqrt{nt}}]^{\otimes 2}\| \\
&\leq 2\mathbb{E}[1_{\|X\|\geq\sqrt{nt}}\|X\|^2] \\
&\leq 2(\sqrt{nt})^{-l}\mathbb{E}[\|X\|^{2+l}].
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\mathbb{E}[\|\epsilon_{\nu,2}\|^p] &\leq d^{p/2}\mathbb{P}(\|X\|\geq\sqrt{nt}) + \mathbb{E}[\|1_{\|X\|,\|X'\|\leq\sqrt{nt}}\frac{(X'-X)^{\otimes 2}}{2} - I_d\|^p] \\
&\leq d^{p/2}\mathbb{P}(\|X\|\geq\sqrt{nt}) + \frac{1}{2}\mathbb{E}[(1_{\|X\|,\|X'\|\leq\sqrt{nt}}\|X'-X\|^{2p}] + 2^{p-1}d^{p/2} \\
&\leq d^{p/2}(\mathbb{E}[\|X\|^2]^{1/2} + 2^{p-1}) + \frac{1}{2}\mathbb{E}[1_{\|X\|,\|X'\|\leq\sqrt{nt}}\|X'-X\|^{2p}] \\
&\leq d^{p/2}(\mathbb{E}[\|X\|^2]^{1/2} + 2^{p-1}) + 2^{2p-1}\mathbb{E}[1_{\|X\|\leq\sqrt{nt}}\|X\|^{2p}] \\
&\leq d^{p/2}(\mathbb{E}[\|X\|^2]^{1/2} + 2^{p-1}) + 2^{2p-1}(\sqrt{nt})^{p-m}\mathbb{E}[\|X\|^{p+m}] \\
&\leq 2^{2p}d^{p/2}(\sqrt{nt})^{p-m}\mathbb{E}[\|X\|^2]^{1/2}\mathbb{E}[\|X\|^{p+m}].
\end{aligned}
$$

and,

$$\mathbb{E}[\|\epsilon_{\nu,2}(X,t)\|^2] \leq 16d(\sqrt{nt})^{2-l}\mathbb{E}[\|X\|^2]^{1/2}\mathbb{E}[\|X_1\|^{2+m}].$$

putting everything together, there exists $C_{2,p,\nu}$ such that,

$$\mathbb{E}[\|\epsilon_{\nu_n,2}(S_n,t)\|^p]^{1/p} \leq C_{2,p,\nu,d}((nt)^{-l/2} + n^{-1/2+(2-m)/2p}(t^{1-m/2p} + t^{1-l/4})).$$

Similarly, there exists $C_{p,\nu,d} > 0$ such that

$$\begin{aligned}
&\|\mathbb{E}[\epsilon_{\nu,3}(X,t)]\| = 0;\\
&\mathbb{E}[\|\epsilon_{\nu,3}(X,t)\|^p] \leq C_{p,\nu,d}(\sqrt{nt})^{2p-m};\\
&\mathbb{E}[\|\epsilon_{\nu,3}(X,t)\|^2] \leq C_{p,\nu,d}(\sqrt{nt})^{4-l};\\
&\|\mathbb{E}[\epsilon_{\nu,4}(X,t)]\| \leq C_{p,\nu,d}(\sqrt{nt})^{2-l};\\
&\mathbb{E}[\|\epsilon_{\nu,4}(X,t)\|^p] \leq C_{p,\nu,d}(\sqrt{nt})^{3p-m};\\
&\mathbb{E}[\|\epsilon_{\nu,4}(X,t)\|^2] \leq C_{p,\nu,d}(\sqrt{nt})^{6-l}.
\end{aligned}$$

Hence, there exist $C_{3,p,\nu}$ and $C_{4,p,\nu}$ such that

- $\mathbb{E}[\|\epsilon_{\nu_n,3}(S_n,t)\|^p]^{1/p} \leq C_{3,p,\nu,d}n^{-1/2+(2-m)/2p}(t^{1-m/2p} + t^{1-l/4});$
- $\mathbb{E}[\|\epsilon_{\nu_n,2}(S_n,t)\|^p]^{1/p} \leq C_{4,p,\nu,d}(n^{-l/2}t^{1-l/2} + n^{-1/2+(2-m)/2p}(t^{3/2-m/2p} + t^{3/2-l/4}))$ .

Finally, since $\|S'_{n,t} - S_n\|$ is bounded by $\sqrt{t}$ we have, for any $k > 4$, $\|\epsilon_{\nu_n,k}\| \leq \sqrt{t}^{k-4}\|\epsilon_{\nu_n,4}\|$ which concludes the proof.

## 6.3 Proof of Proposition 10

Let $x \in \mathbb{R}^d$. In the remainder of this proof $C$ denotes a generic constant depending only on $d$ and $f$. For any $r > 0$, we denote by $\mathcal{B}(x,r)$ the ball centered in $x$ with radius $r$. Let $P_r = \int_{\mathcal{B}(x,r)} \mu(dx)$ and, for $k > 0$, we pose $V_k = \int_{\mathcal{B}(0,1)} x_1^k dx$. Let $N_r$ be the number of points in $\mathcal{B}(x,r)$, for any $0 < \epsilon < 1$, Chernoff's bound yields

$$P(|N_r - nP_r| \geq n\epsilon P_r) \leq 2e^{-\frac{\epsilon^2 nP_r}{3}}. \tag{10}$$

Take $r_M = \left(\frac{2k}{nV_0 \min f}\right)^{1/d}$, we have $P_{r_M} \geq \frac{2k}{n} + C\left(\frac{k}{n}\right)^{1+2/d}$. Hence, for $\frac{k}{n}$ sufficiently small, with probability greater than $1 - \frac{1}{n^2}$, $N_{r_M} \geq k$. Thus the $k$-th nearest neighbor of $x$ is at most at distance $r_M$ of $x$. Applying a union-bound, this is true for any $x \in \mathcal{X} = \{X_1, \ldots, X_n\}$ with probability $1 - \frac{1}{n}$. Hence, for any $k \geq 4$,

$$\|\xi^{\otimes k}\| = \|\xi\|^k \leq r_M^k.$$

Let us now prove the first inequality, we have

$$\begin{aligned}
\mathbb{E}[(X_i - x)1_{X_i \in \mathcal{B}(x,r)}] \\
&= \int_{\mathcal{B}(x,r)} (y-x)\mu(dy) \\
&= \int_{\mathcal{B}(x,r)} (y-x) f(y) dy \\
&= \int_{\mathcal{B}(x,r)} (y-x) f(x) + (y-x)^{\otimes 2} \nabla f(x) + (y-x)^{\otimes 3} \nabla^2 f(x) + Cr^4 dy \\
&= V_2 r^{d+2} \nabla f(x) + Cr^{d+4}.
\end{aligned}$$

Therefore, if we let $b_1 = \sum_{X_i \in \mathcal{B}(x,r)} X_i - x$ and apply Bernstein's inequality,

$$P\left(\left|b_1 - nV_2 r^{d+2} \nabla f(x)\right| \geq C\left(r\sqrt{nP_r \log n} + nr^{d+4}\right)\right) \leq \frac{2}{n^2}.$$

For $r = \left(\frac{k}{nV_0 f}\right)^{1/d}$, we have $|P_r - \frac{k}{n}| \leq C\left(\frac{k}{n}\right)^{1+2/d}$. Hence , by Equation 10, $|N_r - k| \leq C(\sqrt{k \log n} + \frac{k^{1+2/d}}{n^{2/d}})$ holds with probability $1 - \frac{1}{n^2}$. Hence, letting $b_2 = \sum_{X_i \in \mathcal{B}(x,\tilde{r})} X_i - x$, we have

$$|b_1 - b_2| \leq Cr_M\left(\sqrt{k \log n} + \frac{k^{1+2/d}}{n^{2/d}}\right).$$

Putting everything together, we have, with probability $1 - \frac{C}{n^2}$

$$\begin{aligned}
\| \frac{1}{ks} \sum_{X_i \in \mathcal{B}(x,\tilde{r}} (X_i - x) - f^{-2/d} \nabla \log f \| \\
&= \| \frac{b_2}{ks} - f^{-2/d} \nabla \log f \| \\
&\leq \left\| \frac{b_1}{ks} - f^{-2/d} \nabla \log f \right\| + C\left(\frac{\sqrt{\log n} n^{2/d}}{k^{1/2+2/d}} + \frac{1}{k}\right) \\
&\leq C\left(\frac{\sqrt{\log n} n^{1/d}}{k^{1/2+1/d}} + \left(\frac{k}{n}\right)^{2/d}\right).
\end{aligned}$$

Using a union bound, we obtain that

$$\|\mathbb{E}[\xi \mid X] - f(X)^{-2/d} \nabla \log f(X)\| \leq C\left(\frac{\sqrt{\log n} n^{1/d}}{k^{1/2+1/d}} + \left(\frac{k}{n}\right)^{2/d}\right),$$

with probability $1 - \frac{C}{n^2}$. Similar computations give the convergence of the two remaining terms.

### 6.4 Proof of Proposition 11

Let $B$ be a multivariate Rademacher random variable. By construction,

- $\mathbb{E}[\frac{\xi}{s} - \nabla u(X) \mid X] = 0$;
- $\mathbb{E}[\|\mathbb{E}[\frac{\xi^{\otimes 2}}{2s} - I_d \mid X]\|^2]^{1/2} = s\mathbb{E}[\|\nabla u(X)\|^4]^{1/2}$;

Since $\nabla u(0)$ is assumed to be 0 and Lipschitz continuous, we have $\|\nabla u(X)\| \leq L_1\|X\|$. Let us bound $\mathbb{E}[\|X\|^2]$. Since $X$ and $X+\xi$ have the same law,

$$\begin{aligned}\mathbb{E}[\|X\|^2] &= \mathbb{E}[\|X+\xi\|^2] \\ &= \mathbb{E}[\|X + s\nabla u(X)\|^2] + 2ds \\ &\leq \mathbb{E}[\|X\|^2 + 2sX.\nabla u(X) + s^2\|\nabla(u)(X)\|^2] + 2ds.\end{aligned}$$

And, since $\mu$ is log-concave,

$$\mathbb{E}[\|X\|^2] \leq (1 - 2s\rho + L_1^2 s^2)\mathbb{E}[\|X\|^2] + 2ds,$$

therefore

$$\mathbb{E}[\|X\|^2] \leq \frac{2ds}{2s\rho - L_1^2 s^2} \leq \frac{d}{\rho} + O(s).$$

Now, since $\mu$ is strongly log-concave and by construction of our increments, $\|X\| < \frac{\sqrt{2}}{\rho\sqrt{sd}}$. Therefore $\mathbb{E}[\|\nabla u(X)\|^4]^{1/2} \leq \left(\frac{d}{s}\right)^{1/2}$. For $k \geq 3$,

$$\begin{aligned}&\mathbb{E}[|\mathbb{E}[(\xi)^{\otimes k}] \mid X|^2]^{1/2} \\ &\quad = \sum_{j=0}^{k} \binom{k}{j} 2^{j/2} s^{k-j/2}\mathbb{E}[\|\nabla u(X)\|^{2(k-j)}]^{1/2}\|\mathbb{E}[B^{\otimes j}]\| \\ &\quad = \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{j} 2^j L_1^{k-j} s^{k-2j} d^{j-1/2}\mathbb{E}[\|X\|^{2(k-2j)}]^{1/2}.\end{aligned}$$

Finally, for $j < \frac{k}{2}$

$$\mathbb{E}[\|X\|^{2(k-2j)}]^{1/2} \leq s^{k-2j-1}\mathbb{E}[\|X\|^2]^{1/2} \leq \frac{s^{k-2j-1}d^{1/2}}{\rho} + O(s),$$

and the proof is complete.

## 7 Acknowledgements

The author would like to thank Jérôme Dedecker, Michel Ledoux and Yvik Swan for fruitful discussions as well as Frédéric Chazal for his help during the redaction of the paper. This work was supported by the French Délégation Générale de l'Armement (DGA) and by ANR project TopData ANR-13-BS01-0008.

## References

[1] Zero biasing in one and higher dimensions, and applications. *Proceedings of the conference in honor of Charles Stein*, 2005.

[2] Luigi Ambrosio, Giuseppe Savar, and Lorenzo Zambotti. Existence and stability for fokkerplanck equations with log-concave reference measure. *Probability Theory and Related Fields*, 145(3-4):517–564, 2009.

[3] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion operators*. Grundlehren der mathematischen Wissenschaften, Vol. 348. Springer, Jan 2014.

[4] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, June 2003.

[5] Mikhail Belkin and Partha Niyogi. Semi-supervised learning on riemannian manifolds. *Machine Learning*, 56(1):209–239, 2004.

[6] Chinmoy Bhattacharjee and Larry Goldstein. On strong embeddings by steins method. *Electron. J. Probab.*, 21:30 pp., 2016.

[7] Sergey G. Bobkov. Entropic approach to e. rios central limit theorem for w2 transport distance. *Statistics and Probability Letters*, 83(7):1644–1648, 2013.

[8] Sergey G. Bobkov. Berry-esseen bounds and edgeworth expansions in the central limit theorem for transport distances. *preprint*, 2016.

[9] Sourav Chatterjee. A new approach to strong embeddings. *Probability Theory and Related Fields*, 152(1):231–264, 2010.

[10] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from a smooth and log-concave density. submitted 1412.7392, arXiv, December 2014.

[11] A. Durmus and E. Moulines. Sampling from strongly log-concave distributions with the Unadjusted Langevin Algorithm. *ArXiv e-prints*, May 2016.

[12] F. Götze and A. Yu. Zaitsev. Rates of approximation in the multidimensional invariance principle for sums of i.i.d. random vectors with finite moments. *Journal of Mathematical Sciences*, 167(4):495–500, 2010.

[13] Arnaud Guillin, Ivan Gentil, and Franois Bolley. Convergence to equilibrium in wasserstein distance for fokker-planck equations. *Journal of Functional Analysis*, 263(8):2430–2457, 2012.

[14] Tatsunori B. Hashimoto, Yi Sun, and Tommi S. Jaakkola. Metric recovery from directed unweighted graphs. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2015, San Diego, California, USA, May 9-12, 2015*, 2015.

[15] Hikosaburo Komatsu. A characterization of real analytic functions. *Proc. Japan Acad.*, 36(3):90–93, 1960.

[16] Kazumasa Kuwada. Duality on gradient estimates and wasserstein controls. *Journal of Functional Analysis*, 258(11):3758 – 3774, 2010.

[17] Lars Larsson-Cohn. L p-norms of hermite polynomials and an extremal problem on wiener chaos. *Arkiv fr Matematik*, 40(1):133–144, 2002.

[18] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati. Stein's method, logarithmic Sobolev and transport inequalities. 52 pages. To appear in GAFA, October 2014.

[19] Ivan Nourdin, Giovanni Peccati, and Yvik Swan. Entropy and the fourth moment phenomenon. *Journal of Functional Analysis*, 266(5):3170 – 3207, 2014.

[20] Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810 – 864, 2009.

[21] F. Otto and C. Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173(2):361 – 400, 2000.

[22] Emmanuel Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. H. Poincar Probab. Statist.*, 45(3):802–817, 08 2009.

[23] Gareth O. Roberts and Richard L. Tweedie. Exponential Convergence of Langevin Distributions and Their Discrete Approximations. *Bernoulli*, 2(4):341–363, 1996.

[24] Adrian Röllin. A note on the exchangeability condition in steins method. *Statistics and Probability Letters*, 78(13):1800 – 1806, 2008.

[25] A. I. Sakhanenko. Estimates in the invariance principle. *Proc. Inst. Math. Novosibirsk*, 5:2744, 1985.

[26] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*. Die Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin, Heidelberg, New York, 1979.

[27] Daniel Ting, Ling Huang, and Michael I. Jordan. An analysis of the convergence of graph laplacians. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel*, pages 1079–1086, 2010.

[28] Cédric Villani. *Optimal transport : old and new.* Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.

[29] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[30] Ulrike von Luxburg and Morteza Alamgir. Density estimation from unweighted k-nearest neighbor graphs: a roadmap. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 225–233, 2013.

[31] F.-Y. Wang. Exponential Contraction in Wasserstein Distances for Diffusion Semigroups with Negative Curvature. *ArXiv e-prints*, March 2016.