

# DIFFUSION ANNEALED LANGEVIN DYNAMICS: A THEORETICAL STUDY.

BY PATRICK CATTIAUX<sup>1,a</sup>, PAULA CORDERO-ENCINAR<sup>2,b</sup> AND ARNAUD GUILLIN<sup>3,c</sup>

<sup>1</sup>*Institut de Mathématiques de Toulouse. CNRS UMR 5219.  
Université de Toulouse, 118 route de Narbonne, F-31062 Toulouse cedex 09.  
<sup>a</sup>[patrick.cattiaux@math.univ-toulouse.fr](mailto:patrick.cattiaux@math.univ-toulouse.fr)*

<sup>2</sup>*Department of Mathematics. Imperial College London. London, UK.  
<sup>b</sup>[paula.cordero-encinar22@imperial.ac.uk](mailto:paula.cordero-encinar22@imperial.ac.uk)*

<sup>3</sup>*Université Clermont Auvergne, CNRS UMR 6620, LMBP, F-63000 Clermont-Ferrand, France.  
<sup>c</sup>[arnaud.guillin@uca.fr](mailto:arnaud.guillin@uca.fr)*

In this work we study the diffusion annealed Langevin dynamics, a score-based diffusion process recently introduced in the theory of generative models and which is an alternative to the classical overdamped Langevin diffusion. Our goal is to provide a rigorous construction and to study the theoretical efficiency of these models for general base distribution as well as target distribution. As a matter of fact these diffusion processes are a particular case of Nelson processes i.e. diffusion processes with a given flow of time marginals.

Providing existence and uniqueness of the solution to the annealed Langevin diffusion leads to proving a Poincaré inequality for the conditional distribution of  $X$  knowing  $X + Z = y$  uniformly in  $y$ , as recently observed by one of us and her coauthors. Part of this work is thus devoted to the study of such Poincaré inequalities. Additionally we show that strengthening the Poincaré inequality into a logarithmic Sobolev inequality improves the efficiency of the model.

**1. Introduction.** The aim of this paper is to give a rigorous presentation of the recently introduced *diffusion annealed Langevin dynamics* [39]. This stochastic process is a score based generative model and provides an alternative to the well known overdamped Langevin process and its reversed in time version commonly used for sampling purpose. In particular, we will fill some gaps in the main arguments used for building the annealed Langevin dynamics discussed in [39, 30, 24]. We will not discuss its practical efficiency nor its numerical counterparts, that is we will not introduce nor discuss the corresponding discrete algorithms, presented in [24] by the second author, and the references therein. However, some quantitative aspects, useful for discretization schemes or important from the statistical point of view, are discussed in details.

Also, for distributions like the gaussian, an important idea introduced in the papers on diffusion annealed Langevin dynamics consists in using a functional inequality (namely the Poincaré inequality) to control some covariance. This inequality is crucial in [24] for proving that the score of the intermediate distributions is Lipschitz continuous, which, as we explain in Section 2, ensures the existence and uniqueness of strong solutions for the annealed Langevin diffusion. As a matter of fact, heavy tailed base distributions are also particularly well suited for the model as will see in an example.

More generally, Section 2 provides a rigorous introduction to annealed Langevin dynamics, and concludes in the main theorem, which establishes conditions for the strong

---

*MSC2020 subject classifications:* Primary 60J60; Secondary 47D07, 60H10, 60J28, 60J35, 39B62.

*Keywords and phrases:* generative diffusion model, diffusions, annealed langevin dynamics, Poincaré inequality, score based model.

and weak existence and uniqueness of solutions, as well as a quantitative estimate of the Kullback–Leibler divergence between the annealed Langevin dynamics and the flow of time marginals targeted by the annealed Langevin diffusion, i.e. a neat control of the bias of the algorithm.

We then show that a Poincaré inequality for the conditional distribution is necessary to guarantee existence and uniqueness of solutions. In doing so, we complete the results in [24], by recalling and applying several known results, concerning the Poincaré constant of perturbed measures, as explained in [16], see Sections 3 and 4. We then prove new results for this Poincaré constant, Section 5.

Finally, in Section 6, we demonstrate that one can improve the bias between the target distribution and the final distribution of the annealed Langevin dynamics by using the stronger logarithmic Sobolev inequality.

## 2. Score based Stochastic dynamics.

**2.1. Generative diffusion model.** A central point in modern Statistics is to learn a probability distribution  $\pi$  from a given set of samples. Recent advances proposed to introduce dynamical systems bridging the target distribution  $\pi$  and a base (or noising) distribution  $\nu$ . Several methods have been used, see e.g. [2], in particular using stochastic differential equations.

We only consider the case where  $\mathbb{R}^d$  is the state space. A popular method is to build an ergodic diffusion process with invariant distribution given by  $\nu$  and initial distribution  $\pi$  (overdamped Langevin which can be implemented using the Langevin-Monte Carlo algorithm). If  $\nu(dx) = e^{-V(x)}dx$  for some smooth enough  $V$  (say  $C^2$ ), the natural associated process is the solution of

$$(2.1) \quad dX_t = \sqrt{2} dB_t - \nabla V(X_t) dt \quad \text{with} \quad \mathcal{L}(X_0) = \pi.$$

where  $B_t$  is a standard Brownian motion. Since  $\nabla V$  is local Lipschitz, existence and uniqueness hold true up to an explosion time. In order to guarantee that the explosion time is infinite some additional assumptions are required like for instance

(H1) there exists some  $\psi$  such that  $\psi(x) \rightarrow +\infty$  as  $|x| \rightarrow +\infty$  and  $\Delta\psi - \nabla V \cdot \nabla\psi$  is bounded above,

$$(H2) \quad \int |\nabla V|^2 d\nu < +\infty.$$

(H1) is immediate using Itô formula, (H2) is more intricate and requires the use of Dirichlet forms and some tools in Potential Theory (see e.g. [13]). In these cases  $\nu$  is a reversible (hence invariant) measure for the dynamics (2.1).

Using the time reversed process, one is able to rebuild the target measure  $\pi$ . Unfortunately, the invariant measure is attained in an infinite time (a possible method to overcome this difficulty is to consider bridges, see [33] for a survey). Hence one is obliged to stop the procedure at a given time  $T$ , to evaluate the bias introduced by this cut-off using an appropriate distance between measures, and then to study the time reversed diffusion.

More precisely, if  $p_t(\cdot)$  denotes the density of  $X_t$  w.r.t. the Lebesgue measure (which exists thanks to ellipticity and regularity), the time reversed process  $X_t^T = X_{T-t}$  would satisfy

$$(2.2) \quad dX_t^T = \sqrt{2} dB_t + \nabla V(X_t^T) dt + 2 \nabla \cdot \ln p_{T-t}(X_t^T) dt$$

with initial distribution  $\mathcal{L}(X_0^T) = p_T(x)dx$ . Such time reversal argument requires to be careful with the assumptions. One possible set of assumptions is that  $\nabla V$  is (global) Lipschitz and

$\nabla \cdot p(\cdot) \in \mathbb{L}_{loc}^1(dt, dx)$ . For this result see e.g. [35] Theorem 2.3. Other sets of assumptions are possible. We will not discuss them here but in another paper.

Unfortunately, there is no explicit method to recover the density  $p$ . (or the gradient of its logarithm  $\nabla \cdot \ln p$ , called the score in the statistical literature), except in the particular case where  $\nu$  is a gaussian measure. Assume for simplicity that  $\nu$  is the standard normal distribution, then  $X_\cdot$  is the classical Ornstein-Uhlenbeck process, starting from  $\pi$ , so that

$$(2.3) \quad X_t = e^{-t} X_0 + \sqrt{1 - e^{-2t}} G$$

where  $G$  is a standard normal distribution independent of  $X_0$ , implying that the law at time  $t$  is given by

$$(2.4) \quad p_t(x) dx = \frac{\pi(x/e^{-t})}{e^{-td/2}} * \frac{\nu(x/(1 - e^{-2t})^{\frac{1}{2}})}{(1 - e^{-2t})^{d/2}}.$$

The remaining problem is that at the stopping time  $T$ ,  $X_T$  is not equal to the gaussian distribution. However it is not far from it. Indeed, if we assume that the Kullback-Leibler divergence (or relative entropy)

$$d_{KL}(\pi, \nu) := \int \ln(d\pi/d\nu) d\pi$$

is finite, then

$$(2.5) \quad d_{KL}(\pi_T, \nu) \leq e^{-2T} d_{KL}(\pi, \nu),$$

since the gaussian distribution  $\nu$  satisfies a logarithmic Sobolev inequality

$$\text{Ent}_\nu(f) := \int f \ln \left( \frac{f}{\int f d\nu} \right) d\nu \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\nu,$$

for any smooth and positive  $f$ .

If the SDE

$$(2.6) \quad dZ_t = \sqrt{2} dB_t + Z_t dt + 2 \nabla \cdot \ln p_{T-t}(Z_t) dt$$

with initial distribution  $\mathcal{L}(Z_0) = \nu$  admits a unique strong solution, one has

$$\begin{aligned} d_{KL}(\pi, \mathcal{L}(Z_T)) &= d_{KL}(\mathcal{L}(X_T^T), \mathcal{L}(Z_T)) \leq d_{KL}(\mathcal{L}(X_0^T), \mathcal{L}(Z_0)) \\ &= d_{KL}(\pi_T, \nu) \leq e^{-2T} d_{KL}(\pi, \nu). \end{aligned}$$

In order to ensure that (2.6) has a unique solution, the most standard way is to get conditions for  $\nabla \cdot \ln p_{T-t}$  to be Lipschitz in  $x$  uniformly in  $t \in [0, T - \varepsilon]$  for any  $\varepsilon > 0$  (we shall say uniformly on  $[0, T^-)$ ), in other words conditions for  $\nabla^2 \ln p_t$  to be bounded, uniformly in time on  $(0^+, T]$ . For a unique weak solution it is enough to have that  $\nabla \cdot \ln p_{T-t}$  is bounded in  $(x, t)$  for  $t \in [0, T - \varepsilon]$  for any  $\varepsilon > 0$ .

**2.2. Annealed Langevin process.** A natural alternative way is to introduce some interpolation between  $\pi$  and  $\nu$ . Several interpolation procedures are possible, see e.g. [1]. From the probabilistic point of view a very natural one is the following analogue of what we have done before in the gaussian case: let  $X$  and  $Z$  be two independent random variables with respective distributions  $\pi$  and  $\nu$ , consider the interpolation flow

$$(2.7) \quad \tilde{X}_t = \sqrt{\lambda_t} X + \sqrt{1 - \lambda_t} Z$$

for an increasing flow  $t \in [0, T] \mapsto \lambda_t$  of parameters in  $[0, 1]$  (called the schedule in the machine learning literature) such that hopefully  $\lambda_T = 1$  and  $\lambda_0 = 0$ . We denote  $\tilde{\pi}$  (resp.  $\tilde{\nu}$ ) the distribution of  $\tilde{X}_T$  (resp.  $\tilde{X}_0$ ).

At time  $t$  the process has a distribution given by

$$p_t(x) dx = \frac{\pi(x/\sqrt{\lambda_t})}{\lambda_t^{d/2}} * \frac{\nu(x/\sqrt{1-\lambda_t})}{(1-\lambda_t)^{d/2}}.$$

The question is then to represent  $X_t$  as the solution of some Stochastic Differential Equation.

For this purpose, the first thing to do is to show that  $p_t$  does satisfy a Fokker Planck equation. For convenience, one introduces a renormalization parameter  $\kappa$  and the time-changed flow

$$(2.8) \quad \hat{p}_t = p_{\kappa t} \quad \text{for } t \in [0, T/\kappa].$$

Inspired by optimal transport, one way consists in finding a vector field  $v_t$  satisfying the transport (continuity) equation

$$(2.9) \quad \partial_t \hat{p}_t + \nabla \cdot (v_t \hat{p}_t) = 0$$

so that, at least formally,

$$(2.10) \quad \partial_t \hat{p}_t = \Delta \hat{p}_t - \nabla \cdot ((\nabla \ln \hat{p}_t + v_t) \hat{p}_t).$$

To this end one can for instance apply Theorem 8.3.1 in [3], i.e

**THEOREM 2.1.** *Assume that for all  $t \in [0, T]$ ,  $\int |x|^2 p_t(x) dx < +\infty$ . Define, when it exists,*

$$|p'_t| = \lim_{h \rightarrow 0} \frac{1}{|h|} W_2(p_{t+h}, p_t)$$

where  $W_2$  denotes the 2-Wasserstein distance. Assume that  $|p'_t| \in \mathbb{L}^1([0, T])$ . Remark that  $|(\hat{p})'_t| = \kappa |p'_{\kappa t}|$  for all  $t \in [0, T/\kappa]$ .

Then there exists a Borel vector field  $v_t$  such that (2.9) is satisfied and

$$\|v_t\|_{\mathbb{L}^2(\hat{p}_t dx)} = \kappa |p'_{\kappa t}| \quad \text{for almost all } t \in [0, T/\kappa].$$

This result is recalled and used in [30, 24]. The next step in these papers is to introduce first the so called *diffusion annealed Langevin dynamics* given by the S.D.E. on  $[0, T/\kappa]$ ,

$$(2.11) \quad dY_t = \sqrt{2} dB_t + \nabla \ln \hat{p}_t(Y_t) dt \quad \text{with} \quad \mathcal{L}(Y_0) = \tilde{\nu},$$

which, in a sense, is mimicking (2.2). If the drift is explicit, the law at time  $t$  is not, once again. To ensure that the law at time  $t$  is given by  $\hat{p}_t$  one has to consider

$$(2.12) \quad dX_t = \sqrt{2} dB_t + (\nabla \ln \hat{p}_t + v_t)(X_t) dt \quad \text{with} \quad \mathcal{L}(X_0) = \tilde{\nu}.$$

This time the drift is not explicit but the law at time  $t$  should be  $\hat{p}_t$ .

The first problem is to show existence and, if necessary uniqueness, of the solution of (2.11) and (2.12). One also has to check that the marginals flow of a (or the) solution of (2.12) is actually  $\hat{p}_t$ . The next problem is to use Girsanov theory in order to control the Kullback-Leibler divergence between the distributions on the path space. To this end, uniqueness for (2.11) is required. These problems are not completely discussed in [30, 24] and some arguments are missing.

We shall thus fill the gap(s) and complete the proofs of the main results in these references.

Existence and strong uniqueness for (2.11) are ensured as soon as  $\nabla^2 \ln \hat{p}_t$  is bounded uniformly in  $t$  for  $t \in [0, (T/\kappa)^-]$ . If  $\nabla \ln \hat{p}_t$  is bounded, one may use Girsanov theorem and get existence and weak uniqueness.

(2.12) is much more delicate. Indeed the only known property on  $v_t$  is its square integrability w.r.t.  $\hat{p}_t dx$ , so that none of the standard existence theorem applies. If in addition to the uniqueness of the solution of (2.11), we assume that

$$\int_0^{(T/\kappa)} \|v_t\|_{\mathbb{L}^2(\hat{p}_t dx)}^2 dt < +\infty,$$

i.e. the finite energy condition for  $v_t$ , one can use the results in [19] and build a solution of (2.12) satisfying nice properties. The main tool for doing this is Theorem A.1 in the Appendix shown in [19] Theorems 4.29, 4.42 and 4.48.

Uniqueness of the solution is not explicitly addressed in [19]. This discussion is made in Remark 2.12 of [11].

We may thus apply all what precedes in our situation. We gather all this in the following main theorem.

**THEOREM 2.2.** *Let  $p_t$  be defined by (2.2), and  $\hat{p}_t = p_{\kappa t}$  for some  $\kappa \in (0, 1)$ . Assume the following*

- (1) (i) *Either  $\nabla^2 \ln p_t$  is uniformly bounded on  $[0, T^-)$ , meaning that*

$$\|\nabla^2 \ln p_t(\cdot)\|^2 := \sum_i \sup_x \sum_j |\partial_{i,j}^2 \ln p_t(x)|^2 \leq a_t^2,$$

*and  $\sup_{t \leq T-\varepsilon} a_t = a(\varepsilon) < +\infty$  for all  $\varepsilon > 0$ ,*

- (ii) or  $\nabla \ln p_t$  is uniformly bounded on  $[0, T^-)$ , i.e.*

$$\sup_x \sum_i |\partial_i \ln p_t(x)|^2 \leq b_t^2$$

*where  $\sup_{t \leq T-\varepsilon} b_t = b(\varepsilon) < +\infty$  for all  $\varepsilon > 0$ ,*

- (2)  $\int_0^T |p'_t|^2 dt < +\infty$ ,  $|p'|$  *being defined in Theorem 2.1.*

*Then there exists a (strongly in case (i), weakly in case (ii)) unique solution of (2.11), and a (weakly) unique solution of (2.12) up to  $(T/\kappa)^-$ .*

*If  $Q_X$  and  $Q_Y$  denote the distribution on the path space  $C^0([0, (T/\kappa)^-), \mathbb{R}^d)$  of the processes  $X_t$  and  $Y_t$ ,  $Q_X$  is absolutely continuous w.r.t.  $Q_Y$  and*

$$d_{KL}(Q_X, Q_Y) = \frac{1}{4} \int_0^{T/\kappa} \int |v_t|^2 \hat{p}_t dx dt = \frac{\kappa}{4} \int_0^T |p'_t|^2 dt.$$

*In addition, the distribution of  $X_t$  is exactly  $\hat{p}_t dx$  and we have*

$$(2.13) \quad d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq \frac{\kappa}{4} \int_0^T |p'_t|^2 dt,$$

*where  $Y_{T/\kappa}$  denotes any weak limit of  $Y_{(T/\kappa)-\varepsilon}$  as  $\varepsilon \rightarrow 0$ . If  $\lambda_T = 1$ ,  $\tilde{\pi} = \pi$ .*

**PROOF.** With our assumptions it is immediate that  $\hat{p}_t$  is a weak solution of (2.10) i.e. of the weak forward equation in Theorem A.1. Under assumption (1) (i),  $\nabla \ln p_t$  is  $L$ -Lipschitz

continuous with  $L \leq a(\varepsilon)$  for  $t \in [0, T - \varepsilon)$ . So is  $\nabla \ln \hat{p}_t$  for  $t \in [0, (T - \varepsilon)/\kappa)$ . We may thus apply Theorem A.1 (2) with  $P = Q_Y$ . Under assumption (1) (ii),  $\nabla \ln p_t + v_t$  is of finite energy and we may apply Theorem A.1 with  $P$  equal to the Wiener measure with initial distribution  $\tilde{\nu}$ . Since  $P$  is equivalent (in this situation) to  $Q_Y$ ,  $Q_X$  is absolutely continuous w.r.t.  $Q_Y$ . The density is given by the appropriate Girsanov exponential martingale and the conclusion regarding the Kullback-Leibler divergence between the path-space distributions  $Q_X, Q_Y$  follows.

The last inequality then follows immediately since the Kullback-Leibler divergence is non-increasing under measurable transformation and is lower semi continuous w.r.t. the weak convergence topology. Note that, under our assumptions, it is not clear that  $Y_{(T/\kappa)-\varepsilon}$  is converging to some  $Y_{T/\kappa}$ , this has to be discussed on each example.  $\square$

REMARK 2.3. It is important to remark that  $a_t$  in the previous Theorem is exactly what is needed to control the Lipschitz norm of  $\nabla \ln p_t$ . Of course

$$a_t \leq \sup_x \|\nabla^2 \ln p_t(x)\|_{HS}$$

where  $\|\cdot\|_{HS}$  denotes the Hilbert-Schmidt (or Frobenius) norm of a matrix.

Since  $\nabla^2 \ln p_t(x)$  is symmetric, another natural “pseudo-norm” is given by

$$-C_t(x) Id \leq \nabla^2 \ln p_t(x) \leq C_t(x) Id$$

in the sense of quadratic forms.

Recall that  $C_t^2(x) \leq \|\nabla^2 \ln p_t(x)\|_{HS}^2 \leq d C_t^2(x)$  hence in case (i),

$$a_t \leq \sqrt{d} C_t := \sup_x \sqrt{d} C_t(x).$$

$\diamond$

As is common for interpolations of the form (2.7), some degeneracy may occur if  $\lambda_0 = 0$  or/and  $\lambda_T = 1$ . Below, we give a result showing how to overcome possible difficulties in these cases.

COROLLARY 2.4. Define  $m_\pi = \int |x| \pi(dx)$  and  $V_\pi = \int |x|^2 \pi(dx)$ , and similarly  $m_\nu$  and  $V_\nu$ . Recall that

$$\tilde{\pi} = \mathcal{L}(\sqrt{\lambda_T} X + \sqrt{1 - \lambda_T} Z) \quad \text{and} \quad \tilde{\nu} = \mathcal{L}(\sqrt{\lambda_0} X + \sqrt{1 - \lambda_0} Z).$$

Then

$$W_1(\pi, \tilde{\pi}) \leq (1 - \sqrt{\lambda_T}) m_\pi + \sqrt{1 - \lambda_T} m_\nu \quad \text{and} \quad W_2^2(\pi, \tilde{\pi}) \leq 2(1 - \sqrt{\lambda_T})^2 V_\pi + 2(1 - \lambda_T) V_\nu,$$

and similar bounds are satisfied for  $\nu$  and  $\tilde{\nu}$ .

Consequently,

$$d_{BL}(\pi, \mathcal{L}(Y_{T/\kappa})) \leq (1 - \sqrt{\lambda_T}) m_\pi + \sqrt{1 - \lambda_T} m_\nu + \sqrt{\kappa/2} \left( \int_0^T |p'_t|^2 dt \right)^{\frac{1}{2}},$$

where  $d_{BL}$  denotes the bounded Lipschitz distance i.e.

$$d_{BL}(\mu, \theta) = \sup \left\{ \int f d\mu - \int f d\theta ; \|f\|_\infty \leq 1 \text{ and } \|\nabla f\|_\infty \leq 1 \right\}.$$

PROOF. The first part is immediate since  $(X, \tilde{X}_T)$  is a coupling of  $\pi$  and  $\tilde{\pi}$ . Note that if one of  $\pi$  or  $\nu$  is centered one can skip the factor 2 for  $W_2$ . The second part follows from the following facts

1.  $d_{BL} \leq W_1$ , since  $W_1(\mu, \theta) = \sup\{\int f d\mu - \int f d\theta; \|\nabla f\|_\infty \leq 1\}$ ,
2.  $d_{BL} \leq d_{TV}$  where  $d_{TV}(\mu, \theta) = \sup\{\int f d\mu - \int f d\theta; \|f\|_\infty \leq 1\}$  denotes the total variation distance,
3.  $d_{TV} \leq \sqrt{2d_{KL}}$  which is the celebrated Pinsker inequality.

□

REMARK 2.5. As we said the above strategy is described in [30] but with missing assumptions (see Lemma 1 therein) or references (in particular for (2.12)).

Very interesting is that the choice of  $v$  in Theorem 2.1 is the optimal choice in the set of vector fields satisfying the continuity equation with a finite  $\mathbb{L}^2$  norm. It is shown in [3] that this  $v$  belongs to the  $\mathbb{L}^2$  closure of the gradients. Looking at section 4 in [19] one sees that it corresponds to the optimal (minimal entropy) choice of  $Q_X$  on the path space among all probability measures with finite relative entropy w.r.t.  $Q_Y$  and marginals flow  $\hat{p}_t dx$ . ◇

**REMARK 2.6. About the literature.**

Building a diffusion process with a given flow of time marginals is an old problem. It was stated and solved in [10] for drifted Brownian motion in connection with Nelson stochastic mechanics. The proof in [10] is purely analytic and does not discuss the nature of the path measure (here  $Q_X$ ). The construction in [19] is using (delicate) stochastic calculus. Another proof based on large deviations arguments is contained in [20, 21]. The link between Nelson problem and large deviations was pointed out in Föllmer's lecture notes at Saint Flour [28]. Another related problem is the construction of (Schrödinger) bridges where only the initial and the final marginals are given. For this aspect we refer to [28, 33]. The second reference in particular contains the fundamental relationship between these bridges and entropic transport of measures. ◇

REMARK 2.7. If we compare the final bias  $d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa}))$  with the one obtained with the overdamped Langevin process and time reversal at  $T/\kappa$ , one can be disappointed. In the annealed case we get something of size  $\kappa$  while it is  $e^{-T/\kappa}$  for the overdamped case. Of course, if one wants to preserve the fact that the density at time  $t$  is explicit, the latter is limited to a gaussian base distribution, while the former does not assume anything on the base distribution, except some regularity for the score function. In addition one does not have to assume that  $d_{KL}(\pi, \nu)$  is finite in the annealed case.

One can nevertheless ask whether the final bias can be improved if we assume that  $\nu$  satisfies a logarithmic Sobolev inequality, as the gaussian distribution does. We shall come back to this question in Section 6, where our main result will be presented. ◇

Sections 3-5 are devoted to analyzing conditions on  $\pi$  and  $\nu$  under which Theorem 2.2 holds. In particular, we establish how this requirement is connected to controlling the Poincaré constant of the conditional distribution. This analysis leads to the study of Poincaré inequalities for perturbed measures, where we recall relevant results from [16] and present new contributions.

**3. When functional inequalities enter the game, or not.** For

$$(3.1) \quad p_t(x)dx = \frac{\pi(x/\sqrt{\lambda_t})}{\lambda_t^{d/2}} * \frac{\nu(x/\sqrt{1-\lambda_t})}{(1-\lambda_t)^{d/2}},$$

we thus have to find conditions for assumptions (1) and (2) of Theorem 2.2 to be satisfied.

We start with (2).



PROPOSITION 3.1. Define  $\int |x|^2 d\pi = V_\pi$  and  $\int |x|^2 d\nu = V_\nu$ . Let  $t \mapsto \lambda_t$  be increasing, non-negative and  $C^1$  on  $[0, T]$ . Define

$$A_0 = \int_0^T \frac{|\lambda'_t|^2}{\lambda_t} dt \quad ; \quad A_1 = \int_0^T \frac{|\lambda'_t|^2}{1 - \lambda_t} dt.$$

Then

$$\int_0^T |p'_t|^2 dt \leq \frac{1}{2} (V_\pi A_0 + V_\nu A_1).$$

PROOF. If  $X$  and  $Z$  are independent with respective distributions  $\pi$  and  $\nu$ , the pair

$$(\sqrt{\lambda_t}X + \sqrt{1 - \lambda_t}Z; \sqrt{\lambda_{t+h}}X + \sqrt{1 - \lambda_{t+h}}Z)$$

is a coupling of  $p_t$  and  $p_{t+h}$  for  $t \in (0, T)$  and  $h$  small enough. Thus

$$\begin{aligned} W_2^2(p_t, p_{t+h}) &\leq \mathbb{E}[(\sqrt{\lambda_{t+h}} - \sqrt{\lambda_t})X + (\sqrt{1 - \lambda_{t+h}} - \sqrt{1 - \lambda_t})Z]^2 \\ &\leq 2\mathbb{E}[(\sqrt{\lambda_{t+h}} - \sqrt{\lambda_t})^2 |X|^2] + 2\mathbb{E}[(\sqrt{1 - \lambda_{t+h}} - \sqrt{1 - \lambda_t})^2 |Z|^2] \end{aligned}$$

It follows immediately that

$$|p'_t|^2 \leq \frac{|\lambda'_t|^2}{2} \left( \frac{V_\pi}{\lambda_t} + \frac{V_\nu}{(1 - \lambda_t)} \right)$$

hence the result.  $\square$

REMARK 3.2. Notice that the condition on  $\lambda$  is satisfied for  $\lambda_t = 2(t/T)^2 \mathbf{1}_{t \leq T/2} + (1 - 2(1 - t/T)^2) \mathbf{1}_{t > T/2}$  and  $\lambda_t = 1/2(1 + \cos(\pi(1 - (t/T)^\alpha)))$  with  $\alpha > 1/2$ .

It is worth noting that the  $C^1$  assumption on the schedule can be weakened to  $t \mapsto \lambda_t$  being right differentiable.

Moreover, if  $X$  or  $Z$  is centered, one can replace  $1/2$  by  $1/4$  in the final estimate.

Finally, the conditions  $A_0$  and  $A_1$  finite are much weaker than Assumption A6 (or A10) in [24].  $\diamond$

We turn to condition (1). For simplicity we shall assume that both  $\pi$  and  $\nu$  are absolutely continuous w.r.t. the Lebesgue measure so that

$$\pi(dx) = e^{-U(x)} dx \text{ and } \nu(dx) = e^{-W(x)} dx.$$

We thus have

$$p_t(x) = c(t) \int e^{-U_t(y)} e^{-W_t(x-y)} dy = c(t) \int e^{-U_t(x-y)} e^{-W_t(y)} dy,$$

with

$$c(t) = (\lambda_t (1 - \lambda_t))^{-d/2}, \quad U_t(z) = U(z/\lambda_t^{1/2}), \quad W_t(z) = W(z/(1 - \lambda_t)^{1/2}).$$

We shall now make formal calculations. They will be justified later. First introduce for all  $x$  the conditional probability density

$$(3.2) \quad q_t^x(y) = (Z_t^x)^{-1} e^{-(U_t(y) + W_t(x-y))}, \quad Z_t^x = \int e^{-(U_t(y) + W_t(x-y))} dy.$$

Let  $Y_t^x$  denote a random variable with probability density  $q_t^x$ . One has

$$\begin{aligned} \nabla \ln p_t(x) &= - \int \nabla W_t(x-y) q_t^x(y) dy = -(1 - \lambda_t)^{-1/2} \mathbb{E}[\nabla W((x - Y_t^x)/(1 - \lambda_t)^{1/2})] \\ &= - \int \nabla U_t(y) q_t^x(y) dy = -\lambda_t^{-1/2} \mathbb{E}[\nabla U(Y_t^x/\lambda_t^{1/2})] \end{aligned}$$



and

$$\begin{aligned}\nabla^2 \ln p_t(x) &= (1 - \lambda_t)^{-1} \left( -\mathbb{E}[\nabla^2 W((x - Y_t^x)/(1 - \lambda_t)^{1/2})] + \text{Cov}[\nabla W((x - Y_t^x)/(1 - \lambda_t)^{1/2})] \right) \\ &= \lambda_t^{-1} \left( -\mathbb{E}[\nabla^2 U(Y_t^x/\lambda_t^{1/2})] + \text{Cov}[\nabla U(Y_t^x/\lambda_t^{1/2})] \right) \\ &= (\lambda_t (1 - \lambda_t))^{-1/2} \text{Cov}[\nabla W((x - Y_t^x)/(1 - \lambda_t)^{1/2}), \nabla U(Y_t^x/\lambda_t^{1/2})].\end{aligned}$$

REMARK 3.3. One can of course replace  $q_t^x$  by  $\bar{q}_t^x(y) = q_t^x(x - y) = (\bar{Z}_t^x)^{-1} e^{-(U_t(x-y) + W_t(y))}$ ,  $Y_t^x$  by  $\bar{Y}_t^x$  and then exchange  $W$  and  $U$  (as well as  $\lambda_t$  and  $1 - \lambda_t$ ) in the preceding formulas. We shall use this remark at several places.  $\diamond$

It remains to make some assumptions for the previous calculations to be rigorous.

ASSUMPTION 3.4.  $e^{-U}$  and  $e^{-W}$  are bounded,  $V_\pi$  and  $V_\nu$  are finite.

Under these assumptions, if  $|\nabla W|$  (resp.  $|\nabla U|$ ) is bounded, one can differentiate under the integral sign and get the first (resp. second) expression for  $\nabla \ln p_t$ .

If  $|\nabla^2 W| \leq C_W \text{Id}$ ,  $|\nabla W|$  has at most linear growth, so that thanks to the Assumption 3.4 it is integrable w.r.t.  $q_t^x(y) dy$ . We deduce the first expressions for both  $\nabla \ln p_t$  and  $\nabla^2 \ln p_t$ . If  $|\nabla^2 U| \leq C_U \text{Id}$  one obtains the second and the third expressions.

A first result is thus immediate

THEOREM 3.5. Assume that  $e^{-U}$  and  $e^{-W}$  are bounded, that  $V_\pi$  and  $V_\nu$  are finite, and that  $t \mapsto \lambda_t$  is increasing. We will introduce some assumptions

(HbW)  $\sup_x |\nabla W(x)| \leq M_W < +\infty$ ,

(H2bW)  $|\nabla^2 W| \leq C_W \text{Id}$ , meaning that  $-C_W \text{Id} \leq \nabla^2 W(x) \leq C_W \text{Id}$  in the sense of quadratic forms, for all  $x$ ,

(HbU)  $\sup_x |\nabla U(x)| \leq M_U < +\infty$ ,

(H2bU)  $|\nabla^2 U| \leq C_U \text{Id}$ , meaning that  $-C_U \text{Id} \leq \nabla^2 U(x) \leq C_U \text{Id}$  in the sense of quadratic forms, for all  $x$ .

(1) Assume that (HbW) is satisfied. Then

$$|\nabla \ln p_t| \leq (1 - \lambda_t)^{-1/2} M_W$$

so that (1) (ii) in Theorem 2.2 is satisfied, so that one may apply Theorem 2.2 up to time  $(T/\kappa)^-$ . If in addition (HbU) is also satisfied, one has for all  $t$

$$\begin{aligned}|\nabla \ln p_t| &\leq \min((1 - \lambda_t)^{-1/2} M_W ; \lambda_t^{-1/2} M_U) \\ &\leq \max(\lambda_{1/2}^{-1/2}, (1 - \lambda_{1/2})^{-1/2}) \max(M_W, M_U)\end{aligned}$$

and one may apply Theorem 2.2 up to (and including) time  $T/\kappa$ .

(2) Assume that (HbW) and (H2bW) are satisfied. Then for all  $t \in [0, T]$ ,

$$-(1 - \lambda_t)^{-1} C_W \text{Id} \leq \nabla^2 \ln p_t \leq (1 - \lambda_t)^{-1} (C_W + M_W^2) \text{Id}.$$

If in addition (HbU) and (H2bU) are also satisfied, then for all  $t \in [0, T]$ ,

$$-c \max(C_U, C_W) Id \leq \nabla^2 \ln p_t \leq c \max(C_W + M_W^2, C_U + M_U^2) Id,$$

where  $c = \max(\lambda_{1/2}^{-1}, (1 - \lambda_{1/2})^{-1})$ . (1) (i) in Theorem 2.2 is satisfied and the conclusion of the Theorem is in force.

As a by product one can apply Corollary 2.4.

REMARK 3.6. As we previously mentioned, (2) is interesting in order to prove strong existence for (2.11), and also furnishes important useful controls for discretization schemes. It is worth to note that one can take  $\lambda_0 = 0$  as soon as (HbW) and (H2bW) are satisfied, and also  $\lambda_T = 1$  if (HbU) and (H2bU) are satisfied.

Also notice that the lower bounds in (2) are still true without assuming (HbW) or (HbU).

Another important point is to get some explicit upper bound for the Lipschitz constant  $L_t$  of  $\nabla \ln p_t$ . What precedes shows that

$$L_t \leq \frac{\sqrt{d}}{1 - \lambda_t} (C_W + M_W^2) \quad \text{or} \quad L_t \leq \frac{\sqrt{d}}{\min(\lambda_{1/2}, (1 - \lambda_{1/2}))} \max(C_W + M_W^2, C_U + M_U^2).$$

◇

EXAMPLE 3.7. In section 4 of [24] the authors consider the case where  $\nu$  is given by a multivariate Student distribution, while  $\pi$  is a compactly supported perturbation of another multivariate Student distribution.

Recall that the multivariate Student distribution  $t(0, \sigma^2 Id, \alpha)$  is given by a density

$$q(\alpha, \sigma, d, y) = z^{-1} \left( 1 + \frac{|y|^2}{\alpha \sigma^2} \right)^{-(\alpha+d)/2}$$

so that one has to assume that  $\alpha > 2$  for its variance to be finite (the mean being 0 in this case), i.e.  $V_W < +\infty$ . One also uses the terminology generalized Cauchy distribution.

If  $\nu$  is a  $t(0, \sigma^2 Id, \alpha)$ , then

$$W(y) = \ln z + \frac{\alpha + d}{2} \ln \left( 1 + \frac{|y|^2}{\alpha \sigma^2} \right) \quad \text{so that} \quad \nabla W(y) = (\alpha + d) \frac{y}{\alpha \sigma^2 + |y|^2}.$$

In particular  $|\nabla W|$  is bounded by  $M_W = (\alpha + d)/(2\sigma\sqrt{\alpha})$ .

Since

$$\partial_{i,j}^2 W(y) = \frac{(\alpha + d)}{\alpha \sigma^2 + |y|^2} \left( \delta_{i,j} - \frac{2y_i y_j}{\alpha \sigma^2 + |y|^2} \right)$$

it is a simple exercise to check that

$$\frac{(\alpha + d)}{\alpha \sigma^2} Id \geq \nabla^2 W \geq -\frac{(\alpha + d)}{2\alpha \sigma^2} Id \quad \text{so that} \quad C_W \leq \frac{(\alpha + d)}{\alpha \sigma^2}.$$

One can similarly consider Subbotin (exponential power) distributions

$$\nu(dx) = z^{-1} e^{-(1+|x|^2)^{\alpha/2}} dx$$

for  $0 < \alpha < 1$ . Explicit calculations are left to the reader.

◇

Assume that (H2bW) is satisfied. If we do no more assume that  $\nabla W$  is bounded, as already remarked, the lower bound in (2) of Theorem 3.5 is still true. For the upper bound we shall follow the idea in [24] and use a functional inequality. Indeed on one hand, we have to bound  $\nabla^2 W$  and possibly  $\nabla^2 U$  from below, and on the other hand we have to bound  $\text{Cov}[\nabla W((x - Y_t^x)/(1 - \lambda_t)^{1/2})]$  and possibly  $\text{Cov}[\nabla U(Y_t^x/\lambda_t^{1/2})]$  from above. This leads to

LEMMA 3.8. Assume that there exists some constant  $C_P(t)$  such that for all  $x$  the distribution  $q_t^x(y)dy$  satisfies a Poincaré inequality with constant less than  $C_P(t)$ , i.e. for all nice function  $g$ ,

$$\text{Var}(g(Y_t^x)) \leq C_P(t) \mathbb{E}(|\nabla g|^2(Y_t^x)).$$

- (1) Assume that  $|\nabla^2 W| \leq C_W \text{Id}$ , meaning that  $-C_W \text{Id} \leq \nabla^2 W \leq C_W \text{Id}$  in the sense of quadratic forms, so that  $\|\nabla^2 W\|^2 \leq dC_W^2$ . Then

$$|\nabla^2 \ln p_t| \leq \frac{C_W}{1 - \lambda_t} \left( 1 + \frac{dC_W C_P(t)}{1 - \lambda_t} \right) \text{Id}.$$

- (2) Assume in addition that  $|\nabla^2 U| \leq C_U \text{Id}$ . Then

$$|\nabla^2 \ln p_t| \leq c \max(C_W, C_U) (1 + cd \max(C_W, C_U) C_P(t)) \text{Id},$$

where  $c$  is as in Theorem 3.5.

Following remark 3.3 we may replace  $q_t^x(y)dy$  by  $\bar{q}_t^x(y)dy$  that satisfies a Poincaré inequality with the same constant  $C_P(t)$  for all  $x$ .

REMARK 3.9. Recall that the Lipschitz constant  $L_t$  of  $\nabla \ln p_t$  is less than  $\sqrt{d}$  times the constant obtained in the bound for  $|\nabla^2 \ln p_t|$ . In particular even if  $C_P(t)$  is dimension free, the dimension dependence becomes of order at least  $d^{3/2}$ , except in some special cases discussed in Remark 3.11.  $\diamond$

PROOF. It is enough to show that the right hand side is an upper bound for  $\nabla^2 \ln p_t$ , since it is larger than the (opposite of) the lower bound we already know. The proof is immediate using the straightforward extension of Poincaré inequality to multivalued functions yielding (see Lemma B.8 in [24])

$$\begin{aligned} \text{Cov}[\nabla W(x - Y_t^x/(1 - \lambda_t)^{1/2})] &\leq \sup_{|\xi|=1} \text{Var}(\langle \nabla W(x - Y_t^x/(1 - \lambda_t)^{1/2}), \xi \rangle) \text{Id} \\ &\leq C_P(t) (1 - \lambda_t)^{-1} \sup_{|\xi|=1} \mathbb{E} \left[ \sum_i \left| \sum_j \partial_{ij}^2 W(x - Y_t^x/(1 - \lambda_t)^{1/2}) \xi_j \right|^2 \right] \text{Id} \\ &\leq C_P(t) (1 - \lambda_t)^{-1} \mathbb{E}[\|\nabla^2 W(x - Y_t^x/(1 - \lambda_t)^{1/2})\|_{HS}^2] \text{Id}. \end{aligned}$$

□

Everything in what precedes is in a sense satisfactory in order to apply the results in the previous section, except, at a first glance, the existence and overall the control w.r.t.  $t$  of  $C_P(t)$ . In particular, one can take  $\lambda_0 = 0$  in (1) and also  $\lambda_T = 1$  in (2), provided one can bound  $C_P(t)$  in a neighborhood of 0 or  $T$ . In the next section we investigate these problems.

REMARK 3.10. One may ask whether it is really necessary to use a heavy tool like the Poincaré inequality in order to get an upper bound for  $\text{Cov}[\nabla W((x - Y_t^x)/(1 - \lambda_t)^{1/2})]$ . Indeed if we assume that  $|\nabla_{i,j}^2 W| \leq C_W$  for all  $(i, j)$ ,

$$|\nabla W(y)|^2 \leq dC_W^2 |y|^2 + c$$

for some constant  $c$ , so that

$$\text{Cov}[\nabla W((x - Y_t^x)/(1 - \lambda_t)^{1/2})] \leq \frac{dC_W^2 \mathbb{E}[|x - Y_t^x|^2]}{1 - \lambda_t} + c.$$

Unfortunately the right hand side a priori depends on  $x$ . Of course we may exchange the role of  $W$  and  $U$ , but similarly, we will have to control  $\mathbb{E}(|Y_t^x|^2)$  which also depends on  $x$ .  $\diamond$

REMARK 3.11. The prototype of  $\nu$  entering the framework of the previous lemmata is the gaussian measure  $\mathcal{N}(0, \sigma^2 Id)$ , centered for the sake of simplicity.  $\nu$  is thus strictly log-concave (meaning that  $W$  is strictly convex) i.e.

$$C_W Id \geq \nabla^2 W \geq D_W Id$$

for some  $D_W > 0$ . In this situation the estimates in Lemma 3.8 improve as

$$(3.3) \quad \nabla^2 \ln p_t \leq \left( -\frac{D_W}{1-\lambda_t} + \frac{d C_W^2 C_P(t)}{(1-\lambda_t)^2} \right) Id.$$

In the gaussian case one has the stronger

$$(3.4) \quad \nabla^2 \ln p_t \leq \frac{1}{\sigma^2(1-\lambda_t)} \left( -1 + \frac{C_P(t)}{\sigma^2(1-\lambda_t)} \right) Id.$$

Indeed  $\text{Var}(\langle \xi, \nabla W(Z) \rangle) = 1/\sigma^4 \text{Var}(\langle \xi, Z \rangle)$  for any unit vector  $\xi$ , so that one can directly use the second inequality in the proof of Lemma 3.8 (without using the final Cauchy-Schwarz inequality). In particular, the dimension dependence disappears in the conclusion of the lemma.

More generally if  $\nu$  is a strictly log-concave product measure with  $W(x) = \sum W_i(x_i)$  and  $W_i'' \geq (1/\sigma_i^2) > 0$ , one has

$$(3.5) \quad \nabla^2 \ln p_t \leq \left( -\frac{1}{\max_i \sigma_i^2(1-\lambda_t)} + \frac{C_P(t)}{\min_i \sigma_i^4(1-\lambda_t)^2} \right) Id.$$

◇

**4. About the Poincaré constant  $C_P(t)$ .** The literature on the Poincaré inequality is almost impossible to master. We shall only try here to understand how to find sufficient conditions for  $C_P(t)$  to exist and how it depends, in these situations, on  $t$ .

4.1. *General properties of the Poincaré inequality.* First recall some basic facts. The Poincaré inequality is written as

$$\text{Var}(g(Z)) \leq C_P(Z) \mathbb{E}(|\nabla g|^2(Z))$$

i.e.

$$\int g^2(x) \mu_Z(dx) \leq C_P(\mu_Z) \int |\nabla g|^2(x) \mu_Z(dx) + \left( \int g(x) \mu_Z(dx) \right)^2$$

where  $\mu_Z$  denotes the distribution of the random variable  $Z$ ,  $C_P(Z) = C_P(\mu_Z)$  denoting in general the optimal constant in the previous inequality.

The following properties are easily shown:

$$(4.1) \quad \text{for any } x \in \mathbb{R}^d, \quad C_P(x + Z) = C_P(Z),$$

$$(4.2) \quad \text{for any } \lambda \in \mathbb{R}, \quad C_P(\lambda Z) = \lambda^2 C_P(Z),$$

if  $Z_1$  and  $Z_2$  are independent and  $\lambda \in [0, 1]$

$$(4.3) \quad C_P(Z_1, Z_2) = \max(C_P(Z_1), C_P(Z_2)),$$

$$C_P(\sqrt{\lambda} Z_1 + \sqrt{1-\lambda} Z_2) \leq \lambda C_P(Z_1) + (1-\lambda) C_P(Z_2).$$

For the latter one can see e.g. [7] Proposition 1. One also easily sees that

$$(4.4) \quad \text{If } \mu_n \text{ weakly converges to } \mu, \quad C_P(\mu) \leq \liminf C_P(\mu_n).$$

4.2. *Convexity and Poincaré inequality.* As emphasized by the Bakry–Émery curvature-dimension criterion, (strictly) log-concave measures play a special role in the realm of functional inequalities, and in particular Poincaré inequality.

LEMMA 4.1. *Assume that  $H = H_1 + H_2$  where for all  $x$ ,  $\nabla^2 H_1(x) \geq C_1 \text{Id}$  for some  $C_1 > 0$ , and  $H_2$  is bounded. Then*

$$C_P(e^{-H(x)} dx) \leq \frac{e^{\text{Osc}(H_2)}}{C_1},$$

where  $\text{Osc}(h) = \sup h - \inf h$  denotes the oscillation of  $h$ .

The previous Lemma is a simple consequence of Bakry–Émery criterion and Holley–Stroock perturbation argument (see [6]). The following is known.

LEMMA 4.2. *If  $H$  is  $C^2$  and satisfies  $\nabla^2 H(x) \geq C \text{Id}$  for some  $C > 0$  and all  $|x| \geq R$ , then one can build  $H_1$  and  $H_2$  as in Lemma 4.1. Precisely one can obtain*

$$\nabla^2 H_1(x) \geq (C/2) \text{Id},$$

and

$$\text{Osc}(H_2) \leq 16 R^2 L_H,$$

where  $L_H$  denotes the Lipschitz constant of  $\nabla H$ . Recall that  $L_H \leq \|\nabla^2 H\|$  where  $\|\nabla^2 H\|^2 = \sup_x \sum_i \left| \sum_j \partial_{i,j}^2 H(x) \right|^2$ .

There are several proofs of this type of result. A very simple one is sketched in Annex B of [32] (with some gaps), we may also mention [29] where  $H_2$  is not only bounded but also compactly supported. The version stated in the previous Lemma is shown in the supplementary information of [34] (see Lemma 1 therein).

Replacing  $H$  by  $H + \varepsilon|x|^2$  and letting  $\varepsilon$  go to 0, it is immediately seen that the previous Lemma extends to the case  $C = 0$ . We can thus extend the Lemma to the case  $C \leq 0$  as follows.

LEMMA 4.3. *Assume that  $H$  is  $C^2$  and satisfies  $\nabla^2 H(x) \geq C \text{Id}$  for some  $C \in \mathbb{R}$  and all  $|x| \geq R$ . Then one can build  $H_1$  and  $H_2$  such that*

$$\nabla^2 H_1 \geq 0$$

and

$$\text{Osc}(H_2) \leq 16 R^2 (L_H + |C|)$$

where  $L_H$  denotes the Lipschitz constant of  $\nabla H$ , such that for all  $y$ ,

$$H(y) = \frac{C}{2}|y|^2 + H_1(y) + H_2(y).$$

PROOF. It is enough to apply the previous Lemma replacing  $H(y)$  by  $\tilde{H}(y) = H(y) - (C/2)|y|^2$ . The only thing to remark is that  $L_{\tilde{H}} \leq L_H + |C|$ .  $\square$

We may thus state a first result

**THEOREM 4.4.** *Let  $p_t(x)dx$  be given by (3.1) and  $q_t^x(y)$  (resp.  $\bar{q}_t^x(y)$ ) be given by (3.2) (resp. remark 3.3). Assume that  $t \mapsto \lambda_t$  is increasing. Let  $R \geq 0$ .*

*Assume that for all  $|x| \geq R$ ,  $\nabla^2 W(x) \geq D_W^R Id$  and  $\nabla^2 U(x) \geq D_U^R Id$  for some  $D_W^R$  and  $D_U^R$  in  $\mathbb{R}$ . Define*

$$c^R(t) := \frac{D_W^R}{1 - \lambda_t} + \frac{D_U^R}{\lambda_t},$$

*and  $t_W^R = \inf\{t; c^R(t) > 0\}$  and  $t_U^R = \sup\{t; c^R(t) > 0\}$ .*

*Then, for all  $t$  such that  $t_U^R > t > t_W^R$ , both  $q_t^x(y)dy$  and  $\bar{q}_t^x(y)dy$  satisfy a Poincaré inequality with*

$$C_P(q_t^x(y)dy) \leq \frac{1}{c^R(t)} e^{16 R^2 (L_W + |D_W^R| + L_U + |D_U^R|)},$$

*where  $L_W$  and  $L_U$  denote the corresponding Lipschitz constants.*

**PROOF.** According to the previous Lemma, for all  $z$ , we may decompose

$$W(z) = \frac{D_W^R}{2} |z|^2 + W_1(z) + W_2(z)$$

where  $W_1$  is convex and  $\text{Osc}(W_2) \leq 16 R^2 (L_W + |D_W^R|)$ . We have a similar decomposition for  $U$ . We may use these decompositions with  $z = x - y$  for  $W$  and  $z = y$  for  $U$ .

It follows, for any  $y$  and any  $x$ ,

$$W_t(x - y) + U_t(y) = (1 - \lambda_t)^{-1} D_W^R \frac{|x - y|^2}{2} + \lambda_t^{-1} D_U^R \frac{|y|^2}{2} + \theta_t^x(y) + \eta_t^x(y)$$

where for all  $x$ ,  $\nabla^2 \theta_t^x(\cdot) \geq 0$  and  $\text{Osc}(\eta_t^x) \leq 16 R^2 (L_W + |D_W^R| + L_U + |D_U^R|)$ . It thus holds

$$W_t(x - y) + U_t(y) = A_t^x(y) + \eta_t^x(y)$$

where for all  $x$ ,  $\nabla^2 A_t^x \geq c^R(t) Id$ .

The result for  $q_t^x(y)dy$  follows from Lemma 4.1. The case of  $\bar{q}_t^x(y)dy$  is similar.  $\square$

**REMARK 4.5.** The decomposition with an explicit quadratic term is crucial for the proof. It yields a loss on the perturbation constant but allows a simple result.

If both  $D_W^R$  and  $D_U^R$  are non-negative, one of them being positive, one has  $t_W^R = 0$  and  $t_U^R = T$ . If  $D_W^R > 0$  but  $D_U^R \leq 0$ ,  $t_U^R = T$  but  $t_W^R = 0$  only if  $\lambda_0 \geq |D_U^R|/(D_W^R + |D_U^R|)$ . Note that  $W_1(\tilde{\nu}, \nu) \leq \sqrt{\lambda_0} m_\pi + (1 - \sqrt{1 - \lambda_0}) m_\nu$ , so that one can control the initial bias.  $\diamond$

**COROLLARY 4.6.** *Assume that  $e^{-W}$  and  $e^{-U}$  are bounded, that  $V_\pi$  and  $V_\nu$  (see Proposition 3.1) are finite and that  $t \mapsto \lambda_t$  is increasing.*

*Assume in addition that  $D_U Id \leq \nabla^2 U \leq C_U Id$  and that  $D_W^R Id \leq \nabla^2 W(x) \leq C_W^R Id$  for some  $R \geq 0$ , all  $|x| \geq R$  and some  $C_W^R, D_W^R > 0$  (which imply that  $|\nabla^2 W| \leq C'_W Id$ ).*

*Finally assume that  $\lambda_0(D_W^R - D_U) + D_U > 0$ .*

*Then (2.13), Theorem 2.2, applies with  $\tilde{\pi} = \pi$ , the right hand side being given by Proposition 3.1.*

**REMARK 4.7.** In the previous Remark and Corollary we pushed forward the properties of  $W$  since it is chosen by the user, while the data have to fit  $U$ . One can of course exchange the roles of  $W$  and  $U$  (exchanging  $\lambda_t$  and  $1 - \lambda_t$  too) in the previous statements.  $\diamond$

This result contains the main result in [24]. If one wants  $\lambda_0 = 0$ , one has to assume convexity at infinity of  $U$ , as in [24] Assumption A.4. If  $\lambda_0 > 0$ , from a practical point of view one has first to estimate  $p_0(x)$ . Since one may choose  $W$ , one can choose a large  $D_W^R$  so that  $\lambda_0$  is small, possibly equal to the step size of some discretization scheme. The counterpart is that the bounds for  $\nabla^2 \ln p_t$  deteriorate.

**4.3. Perturbation of Poincaré inequality.** The previous Theorem is based on convexity results for  $W + U$ . We will now adopt a perturbative point of view: starting from  $\nu_t^x(y)dy := e^{-W((x-y)/\sqrt{1-\lambda_t})}dy$ , look at  $q_t^x$  as a perturbation of  $\nu_t^x$ . This approach is studied in [16]. We shall give below the most tractable existing results we know in this direction. We will come back to the perturbation point of view in the next section.

**THEOREM 4.8.** *Let  $p_t(x)dx$  be given by (3.1) and  $q_t^x(y)$  be given by (3.2). Assume that  $t \mapsto \lambda_t$  is increasing.*

- (1) *Assume that  $\nabla^2 W \geq D_W Id$  with  $D_W > 0$  and that  $|\nabla U| \leq M_U$ . Then*

$$C_P(q_t^x(y)dy) \leq \frac{2(1-\lambda_t)}{D_W} e^{4\sqrt{2d/\pi} \frac{M_U^2(1-\lambda_t)}{\lambda_t D_W}}.$$

(Due to Miclo, see lemma 2.1 in [8] recalled in Theorem 1.3 (1) of [16] with a typo.)

- (2) *Under the same assumptions,*

$$C_P(q_t^x(y)dy) \leq 2 \left( \frac{M_U(1-\lambda_t)}{\sqrt{\lambda_t} D_W} + \sqrt{\frac{2(1-\lambda_t)}{D_W}} \right)^2 e^{\frac{M_U^2(1-\lambda_t)}{2\lambda_t D_W}}.$$

(Example 3 section 7.1 in [14], recalled in Theorem 1.3 (2) of [16].)

- (3) *If  $\nabla^2 W(x) \geq D_W^R Id$  for some  $D_W^R > 0$  and all  $|x| \geq R$ , and  $|\nabla U| \leq M_U$ ,*

$$C_P(q_t^x(y)dy) \leq \frac{4(1-\lambda_t)}{D_W^R} e^{16\sqrt{2d/\pi} \frac{M_U^2(1-\lambda_t)}{\lambda_t D_W^R}} e^{16R^2 L_W},$$

and

$$C_P(q_t^x(y)dy) \leq 8 \left( \frac{M_U(1-\lambda_t)}{\sqrt{\lambda_t} D_W^R} + \sqrt{\frac{1-\lambda_t}{D_W^R}} \right)^2 e^{\frac{M_U^2(1-\lambda_t)}{\lambda_t D_W^R}} e^{16R^2 L_W}.$$

- (4) *Assume that for some  $\varepsilon > 0$ ,*

$$s := \frac{(1+\varepsilon)(1-\lambda_t)}{4\lambda_t} C_P(\nu) M_U^2 < 1,$$

then

$$C_P(q_t^x(y)dy) \leq \frac{1+(1/\varepsilon)}{1-s} (1-\lambda_t) C_P(\nu).$$

(See Theorem 2.1 in [16] and recall that  $\nu(dy) = e^{-W(y)}dy$ .)

**COROLLARY 4.9.** *Under the basic assumptions of Corollary 4.6, the assumptions in any of the items of the previous Theorem and the additional assumptions  $|\nabla^2 W| \leq C_W Id$  (if not already implied) and  $|\nabla^2 U| \leq C_U Id$ , (2.13), Theorem 2.2, applies with  $\tilde{\pi} = \pi$ , the right hand side being given by Proposition 3.1.*



REMARK 4.10. Again, in the previous Theorem and Corollary we pushed forward the properties of  $W$  since it is chosen by the user, while the data have to fit  $U$ .  $\diamond$

To get (3) from (1) and (2) in the Theorem, simply use Lemma 4.2 and Holley-Stroock argument, for (4) use (4.1) and (4.2) saying that  $C_P(\nu_t^x(y)dy) = (1 - \lambda_t)C_P(e^{-W(y)}dy)$ .

Even if the value of the Poincaré constant can be desperately big if  $\lambda_0$  is close to 0, the first three items of the previous Theorem complete Theorem 4.4 (except if  $U$  is strictly convex at infinity too) for Lipschitz  $U$ . In addition there is no restriction on  $t$ .

(4) does not require any convexity assumption but it is limited to a small range of  $t$ 's. Indeed the condition  $s < 1$  amounts to

$$(4.5) \quad \lambda_t > \frac{(1 + \varepsilon)C_P(\nu)M_U^2}{4 + (1 + \varepsilon)C_P(\nu)M_U^2} := \lambda_{min}.$$

REMARK 4.11. The previous Theorem completes the range of potential examples, including for instance the case of a gaussian  $\nu$  with covariance  $\sigma^2 Id$  and an heavy tailed  $\pi$ . Recall that  $C_P(\nu) = \sigma^2$ .

In this case, in order to get an upper bound for  $\nabla^2 \ln p_t$  one has to find an upper bound for  $\text{Var}(\langle \xi, (x - Y_t^x) \rangle) = \text{Var}(\langle \xi, Y_t^x \rangle)$  for any unit vector  $\xi$ . If  $\pi$  is compactly supported in the (euclidean) ball  $B(0, R)$ , the latter is bounded by  $\lambda_t R^2$ . Notice that in this case we can drop the regularity assumption for  $\pi$ . We thus have

PROPOSITION 4.12. *Let  $\nu$  be a  $\mathcal{N}(0, \sigma^2 Id)$  and  $\pi$  be compactly supported in the euclidean ball  $B(0, R)$ . Then for all  $t$ ,*

$$-\frac{1}{\sigma^2(1 - \lambda_t)} Id \leq \nabla^2 \ln p_t \leq -\frac{\sigma^2(1 - \lambda_t) - \lambda_t R^2}{\sigma^4(1 - \lambda_t)^2} Id.$$

This result is not new and is (up to the presentation) contained in subsection 2.1 of [8]. The previous proposition can be extended to the case where  $\pi$  is given by the convolution of a gaussian with a compactly supported distribution.

PROPOSITION 4.13. *Let  $\nu$  be a  $\mathcal{N}(0, \sigma^2 Id)$  and let  $\pi$  be the convolution of  $\mathcal{N}(0, \tau^2 Id)$  with a distribution supported in the euclidean ball  $B(0, R)$ . Then for all  $t$ ,*

$$-\frac{1}{\alpha_t^2} Id \leq \nabla^2 \ln p_t \leq -\frac{\alpha_t^2 - \lambda_t R^2}{\alpha_t^4} Id,$$

where  $\alpha_t^2 = \sigma^2(1 - \lambda_t) + \tau^2 \lambda_t$ .

PROOF. The result follows from the observation that  $X_t = \sqrt{\lambda_t}X + \sqrt{1 - \lambda_t}\sigma Z \stackrel{d}{=} \sqrt{\lambda_t}U + \sqrt{(1 - \lambda_t)\sigma^2 + \lambda_t\tau^2}Z'$  where  $X \sim \pi$ ,  $Z, Z' \sim \mathcal{N}(0, I)$ ,  $U$  is compactly supported, and by applying Proposition 4.12.  $\square$

$\diamond$

REMARK 4.14. An important point in Statistics is the dimension dependence of all constants. In the strictly convex case estimates are dimension free, while in the convex at infinity case  $\|\nabla W\|$  grows linearly with  $d$ . Notice however that  $V_\pi$  and  $V_\nu$  also grow linearly.

In case (4) if we choose  $\nu$  as a product measure, the Poincaré constant is dimension free. If it is log-concave (but not strictly log-concave) it is known that the Poincaré constant at most grows as  $\ln d$ , see [31]. Log-concave distributions are now playing an important role in Statistics (see e.g. [38, 23]).  $\diamond$

**5. New perturbation results for the Poincaré constant.** The perturbative results we used for Theorem 4.8 are based on various techniques for proving a Poincaré inequality: (1) relates a Lipschitz perturbation to a bounded one, (2) uses the equivalence between Poincaré inequality and exponential  $\mathbb{L}^2$  convergence to equilibrium for the semi-group, the latter being obtained via the reflection coupling method introduced in [26, 27], (4) is using a direct rough approach.

Another method was introduced in [4, 5]: the use of (Foster)-Lyapunov functions. Foster-Lyapunov functions are a central tool in the MCMC community, and their use for exponential stabilization of diffusion processes was pointed out in [25]. Recall two main results for the Poincaré inequality.

**THEOREM 5.1.** *Let  $\mu(dx) = e^{-V(x)}dx$ , for some smooth  $V$ , be a probability measure. Denote  $L_V := \Delta - \nabla V \cdot \nabla$ .*

*Assume that there exists a  $C^2$  function  $F$  such that  $F \geq 1$ , some  $R > 0$ , constants  $b \geq 0$  and  $\theta > 0$  such that*

$$L_V F(x) \leq -\theta F(x) + b \mathbf{1}_{|x| \leq R},$$

*such a  $F$  is called a Lyapunov function (for  $L_V$ ).*

*Then  $\mu$  satisfies a Poincaré inequality with constant*

$$C_P(\mu) \leq \frac{1}{\theta} (1 + C_P(\mu_R))$$

*where  $\mu_R(dx) = \frac{e^{-V(x)} \mathbf{1}_{|x| \leq R}}{\mu(B(0, R))} dx$  is the (normalized) restriction of  $\mu$  to the euclidean ball  $B(0, R)$ .*

*If in addition  $\frac{\partial F}{\partial n} \leq 0$  on  $S_R := \{|x| = R\}$ , where  $n$  denotes the inward normal to  $S_R$ , then*

$$C_P(\mu) \leq \frac{1}{\theta} + C_P(\mu_R).$$

*For  $d \geq 2$ ,*

$$C_P(\mu_R) \leq \frac{d+2}{d(d-1)} R^2 e^{O_{SC_{B(0, R)}} V},$$

*while for  $d = 1$  the pre-factor has to be replaced by  $4/\pi^2$ .*

The first part is Theorem 1.4 in [4] while the second one is shown in [18] Remark 3.3. The control of  $C_P(\mu_R)$  is a consequence of Holley-Stroock perturbation result and the known upper bound (asymptotically sharp as  $d \rightarrow +\infty$ ) for the Poincaré constant of the uniform measure on  $B(0, R)$  (see [9] Corollary 4.1 or [17] Example 5.5). The case  $d = 1$  is well known.

Actually, as shown in Theorem 2.3 of [18], there is an equivalence between the existence of a Lyapunov function and the finiteness of the Poincaré constant. One can also control  $\theta$  in the definition of a Lyapunov function by the Poincaré constant, see Proposition 3.1 in [18] and Theorem 2.1 in [15].

As a direct consequence, if  $V = W + U$  is such that  $F$  is a  $L_W$  Lyapunov function, it is also a  $L_V$  Lyapunov function provided

$$(5.1) \quad (-\nabla U \cdot \nabla F)(x) \leq \theta' F(x) \quad \text{for } |x| > R' \text{ for some } R' > R \text{ and } \theta' < \theta.$$

Of course in the framework of the previous sections, one has to consider  $V(y) = U(x - y) + W(y)$  (recall that the Poincaré constant is unchanged by translation) and get a result independent of  $x$ . Clearly the only tractable situation is the one where  $U$  is Lipschitz continuous with  $|\nabla U| \leq M_U$  and such that

$$M_U |\nabla F(x)| \leq \theta' F(x) \quad \text{for } |x| \text{ large, with } \theta' < \theta.$$

We shall focus on a class of examples, namely potentials  $W$  satisfying some “almost convex” property

$$(5.2) \quad \text{there exist } \alpha > 0, \beta \geq 1 \text{ and } R \geq 0 \text{ such that for } |x| \geq R, \langle x, \nabla W(x) \rangle \geq \alpha |x|^\beta.$$

Prototypical examples are Subbotin distributions, i.e.  $W(x) = |x|^\beta$  for  $\beta \geq 1$ , so that  $\alpha = \beta$  and  $R = 0$ .

REMARK 5.2. More generally, as shown in [4] Lemma 2.2, if  $W$  is convex ( $\nabla^2 W \geq 0$ ), it satisfies (5.2) for  $\beta = 1$ .

We will revisit the mentioned proof in a quantitative perspective, assuming for simplicity that  $W(0) = \min_y W(y)$  which is not a restriction. First let  $M = \sup_{|y| \leq 1} |\nabla W(y)|$  so that  $W(y) - W(0) \leq M$  for  $|y| \leq 1$ . Denote  $A_M = \{W(y) - W(0) \leq M\}$ , so that  $B(0, 1) \subseteq A_M$ . Since  $e^{-W}$  is a probability density, it holds  $\text{vol}(A_M) \leq e^{|W(0)|+M}$  where  $\text{vol}$  denotes Lebesgue volume. For  $x \in A_M$ , since  $A_M$  is convex, the cone with basis  $B_{d-1}(0, 1)$  (the unit ball in dimension  $d-1$ ) and vertex  $x$  is a subset of  $A_M$ . The volume of this cone is thus less than the one of  $A_M$ , i.e.

$$\frac{\text{vol}(B_{d-1}(0, 1)) |x|}{d} \leq e^{|W(0)|+M} \quad \text{i.e. } |x| \leq \frac{d \Gamma((d+1)/2) e^{|W(0)|+M}}{\pi^{(d-1)/2}} := R_W.$$

If  $|u| = R_W + 1$ ,  $u$  does not belong to  $A_M$ , so that using that  $t \mapsto \frac{1}{t} (W(tu) - W(0))$  is non-decreasing since  $W$  is convex we obtain for  $|x| \geq R_W + 1$  that  $W(x) - W(0) \geq \frac{M}{1+R_W} |x|$ . Since  $\langle x, \nabla W(x) \rangle \geq W(x) - W(0)$  thanks to convexity again, we thus have obtained that (5.2) is satisfied with  $\beta = 1$  and  $\alpha = M/(1 + R_W)$ .

Notice that if the previous proof is elementary, the dimension dependence  $\alpha \sim (d/c)^{-d}$  is a disaster for large  $d$ . Using more sophisticated results one can greatly improve upon the dimension dependence, the price to pay being that quantities are not explicit.

First of all, according to [31], there exists a universal constant  $C_{Klar}$  such that if  $\nu = e^{-W(x)} dx$  is a log-concave probability distribution,

$$C_P(\nu) \leq C_{Klar} \sigma^2(\nu) (1 + \ln d)$$

where  $\sigma^2(\nu)$  is the largest eigenvalue of the covariance matrix of  $\nu$ . It is not easy (but probably not impossible) to trace  $C_{Klar}$  in the various papers needed to get the previous result. It is conjectured (Kannan-Lovasz-Simonovits, for short KLS conjecture) that the bound is actually dimension free i.e.

$$C_P(\nu) \leq C_{KLS} \sigma^2(\nu)$$

for some universal constant. Klartag’s result is the most recent step in the direction of this conjecture. In some specific situations, KLS conjecture is true: of course if  $\nu$  is a product measure using the tensorization property of the Poincaré constant or if  $\nu$  is radial (or spherically symmetric and  $d \geq 2$ ) where  $C_{KLS} \leq 2$  according to Theorem 1.2 in [9] (actually one can choose  $d/(d-1)$  in dimension  $d \geq 2$ ). It is also true for some families of explicit measures (like the uniform measures on  $l^p$  balls for instance).

It thus immediately follows from Theorem 2.1 in [15] that, for any  $R > 0$ , there exists a  $L_W$  Lyapunov function  $F_R$  with

$$\theta_R = \nu(B(0, R)) \min \left( \frac{1}{8}, \frac{1}{4C_{Klar}\sigma^2(\nu)(1 + \ln d)} \right).$$

This time however  $F_R$  is not explicit at all, since it is obtained via Lax-Milgram theorem.  $\diamond$

If  $W$  satisfies (5.2),  $F(x) = e^{\gamma|x|}$  is a  $L_W$  Lyapunov function. Indeed

$$\begin{aligned} L_W F &= \gamma \left( \frac{d-1}{|x|} + \gamma - \left\langle \frac{x}{|x|}, \nabla W \right\rangle \right) F(x) \\ &\leq \gamma \left( \frac{d-1}{|x|} + \gamma - \alpha |x|^{\beta-1} \right) F(x), \end{aligned}$$

for  $|x| \geq R$ . Choosing  $R' \geq R$  large enough and  $\gamma$  small enough for

$$\theta = \gamma \left( \alpha(R')^{\beta-1} - \gamma - \frac{d-1}{R'} \right) > 0$$

we thus have built a Lyapunov function for  $|x| \geq R'$ . Notice that  $\partial F / \partial n$  is non-positive on any sphere  $S_A$ .

With all these points in mind we may state several perturbation results

**THEOREM 5.3.** *Let  $\mu(dx) = e^{-V(x)}dx = e^{-(U+W)(x)}dx$ , for smooth  $U$  and  $W$  be a probability measure. Assume that*

- (1)  *$W$  satisfies the “quasi-convex” condition (5.2) for some  $\alpha_W > 0$ ,  $\beta_W \geq 1$  and  $R \geq 0$ ,*
- (2)  *$U$  satisfies  $|\langle x, \nabla U(x) \rangle| \leq \kappa_U |x|^{\beta_U}$  for some  $\kappa_U \geq 0$ ,  $\beta_U \leq \beta_W$  and  $|x| \geq R$ .*

*Then, if for some  $R' \geq R$  one can find  $\gamma > 0$  such that*

$$c(W, U) := \alpha_W (R')^{\beta_W-1} - \frac{d-1}{R'} - \gamma - \kappa_U (R')^{\beta_U-1} > 0,$$

*$\mu$  satisfies a Poincaré inequality with*

$$C_P(\mu) \leq \frac{1}{\gamma c(W, U)} + C_P(\mu_{R'}),$$

*where a bound for  $C_P(\mu_{R'})$  is given in Theorem 5.1.*

*If  $\beta_U < \beta_W$  such a  $\gamma$  always exists, while if  $\beta_U = \beta_W$  it exists provided  $\alpha_W > \kappa_U$ .*

Of course we can add an optimization step by increasing  $R'$ , hence decreasing  $1/\gamma c$  but increasing  $C_P(\mu_{R'})$ . This is left to the courageous reader.

Remark that, for  $c(W, U)$  to be positive we need  $R' > ((d-1)/\alpha_W)^{1/\beta_W}$ , so that in Theorem 5.1 the bound for  $C_P(\mu_{R'})$  is of order  $d^{\frac{2-\beta_W}{\beta_W}}$  times the exponential of the Oscillation of  $V$ . Note that the larger  $\beta$  is, the larger the Oscillation of  $W$  is. If  $W$  is convex we can (in general) improve this dependence.

**THEOREM 5.4.** *In addition to the assumptions of Theorem 5.3 assume that  $W$  is convex. Denote  $\nu(dx) = e^{-W(x)}dx$ , assumed to be a probability measure as well, and for simplicity a centered probability measure. Then*

$$C_P(\mu_{R'}) \leq \frac{C_{Klar} \sigma^2(\nu)(1 + \ln d)}{\nu(B(0, R'))} e^{O_{sc_{B(0, R')}} U}.$$

*If in addition  $W$  is radially symmetric we may replace  $C_{Klar}(1 + \ln d)$  by 2.*

PROOF. Since  $\nu$  is log-concave and  $B(0, R')$  is convex,  $\nu_{R'}$  is still log-concave. As we recalled in Remark 5.2,  $C_P(\nu_{R'}) \leq C_{Klar} \sigma^2(\nu_{R'}) (1 + \ln d)$  and we may replace  $C_{Klar} (1 + \ln d)$  by 2 in the radially symmetric case which is preserved by restricting  $\nu$  to a ball. Now, if  $\xi$  is a unit vector

$$\text{Cov}_{\nu_{R'}}(\langle \xi, X \rangle) \leq \frac{\mathbb{E}_{\nu}(\langle \xi, X \rangle^2)}{\nu(B(0, R'))} = \frac{\text{Cov}_{\nu}(\langle \xi, X \rangle^2)}{\nu(B(0, R'))} \leq \frac{\sigma^2(\nu)}{\nu(B(0, R'))}.$$

It remains to apply Holley-Stroock bounded perturbation to get the result.  $\square$

In a sense the previous results complete the perturbation result by Miclo we recalled and used in (1) of Theorem 4.8, for log-concave but non strictly log-concave  $\nu$ 's. It turns out however that in the uniformly convex case the method also furnishes new estimates.

THEOREM 5.5. *Assume that  $\nabla^2 W \geq D_W \text{Id}$  for some  $D_W > 0$  and that  $|\nabla U| \leq M_U$ . Then*

$$C_P(\mu) \leq \frac{1}{D_W} \left( 1 + e^{2M_U R / \sqrt{D_W}} \right)$$

with

$$R = \frac{1}{2} \left( (2 + (M_U / \sqrt{D_W})) + \sqrt{(2 + (M_U / \sqrt{D_W}))^2 + 4(d-1)} \right).$$

PROOF. Assume first that  $D_W = 1$ . We may apply Theorem 5.3 with  $R = 0$ ,  $\beta_W = 2$ ,  $\alpha_W = D_W$ ,  $\beta_U = 1$  and  $\kappa_U = M_U$ . Choose arbitrarily  $\gamma = 1 \leq c(W, U)$ . It yields

$$R' \geq \frac{1}{2} \left( (2 + M_U) + \sqrt{(2 + M_U)^2 + 4(d-1)} \right).$$

This time  $\nu_{R'}$  is strictly log-concave so that  $C_P(\nu_{R'}) \leq 1$ . Hence

$$C_P(\mu) \leq 1 + e^{\text{Osc}_{B(0, R')} U} \leq 1 + e^{2M_U R'}.$$

It remains to use the dilation property of the Poincaré constant using the change of variable  $x \mapsto x / \sqrt{D_W}$  and replace  $M_U$  by  $M_U / \sqrt{D_W}$ .  $\square$

REMARK 5.6. This result is better than the one by Miclo

$$C_P(\mu) \leq \frac{2}{D_W} e^{4\sqrt{2d/\pi} M_U^2 / D_W}$$

for large values of  $M_U / \sqrt{D_W}$  or large dimension  $d$ . It is however worse than the one in [14] we recalled

$$C_P(\mu) \leq 2 \left( \frac{M_U}{D_W} + \sqrt{\frac{2}{D_W}} \right)^2 e^{M_U^2 / 2D_W},$$

when the dimension is much larger than  $M_U / \sqrt{D_W}$ . In general they are not comparable. Notice that the previous result is the only one where the dimension  $d$  does not appear in the controls.  $\diamond$

Coming back to the situation of the preceding section, we may state

COROLLARY 5.7. *Assume that  $|\nabla U| \leq M_U$ . Then Theorem 5.3, Theorem 5.4 and Theorem 5.5 apply to  $C_P(\bar{q}_t^x(y)dy)$  with  $\beta_U = 1$ , replacing  $\alpha_W (= D_W \text{ if } \beta_W = 2)$ ,  $R$ ,  $\kappa_U = M_U$ ,  $\sigma^2(\nu)$  respectively by  $\alpha_W / (1 - \lambda_t)^{\beta_W/2}$ ,  $R\sqrt{1 - \lambda_t}$ ,  $M_U / \sqrt{\lambda_t}$  and  $(1 - \lambda_t) \sigma^2(\nu)$ .*

Indeed  $|\langle y, \nabla_y U(x - y) \rangle| \leq M_U |y|$  for all  $x$ , so that all the previous results apply, uniformly in  $x$  to  $(\bar{Z}_t^x)^{-1} e^{-(W_t(y) + U_t(x - y))} dy$ .

All our results now almost cover the full spectrum of typical noise distributions, heavy tailed distributions whose Hamiltonians have bounded gradients, strictly log-concave distributions as the gaussian one and now intermediate distributions. In particular, we may apply one of our results to any Subbotin distribution with any power  $0 < \alpha \leq 2$ .

**6. Using logarithmic Sobolev inequalities.** A measure  $\mu$  satisfies a logarithmic Sobolev inequality if the following holds for all  $f$  nice enough,

$$\text{Ent}_\mu(f^2) := \int f^2 \ln \left( \frac{f^2}{\int f^2 d\mu} \right) d\mu \leq 2C_{LS}(\mu) \int |\nabla f|^2 d\mu = 2C_{LS}(\mu) \mathbb{E}[|\nabla f|^2],$$

where  $C_{LS}(\mu) > 0$  denotes the optimal constant in the previous inequality. We shall use the equivalent formulation

$$\int \rho \ln(\rho) d\mu \leq (C_{LS}(\mu)/2) \int \frac{|\nabla \rho|^2}{\rho} d\mu$$

if  $\rho$  is a nice density of Probability. One recognizes the Fisher information of  $\rho$  in the right hand side. If the random variable  $Z$  satisfies  $\mathcal{L}(Z) = \mu$ , we will denote  $C_{LS}(Z)$  instead of  $C_{LS}(\mu)$  for convenience.

Similar to the Poincaré inequality the following properties hold:

$$(6.1) \quad \text{for any } x \in \mathbb{R}^d, \quad C_{LS}(x + Z) = C_{LS}(Z),$$

$$(6.2) \quad \text{for any } \lambda \in \mathbb{R}, \quad C_{LS}(\lambda Z) = \lambda^2 C_{LS}(Z),$$

if  $Z_1$  and  $Z_2$  are independent and  $\lambda \in [0, 1]$

$$(6.3) \quad C_{LS}(Z_1, Z_2) = \max(C_{LS}(Z_1), C_{LS}(Z_2)),$$

$$C_{LS}(\sqrt{\lambda} Z_1 + \sqrt{1 - \lambda} Z_2) \leq \lambda C_{LS}(Z_1) + (1 - \lambda) C_{LS}(Z_2).$$

For the latter one can see e.g. [14] Proposition 18. One also easily sees that

$$(6.4) \quad \text{If } \mu_n \text{ weakly converges to } \mu, C_{LS}(\mu) \leq \liminf C_{LS}(\mu_n).$$

We shall try to use here a log-Sobolev inequality in the spirit of the semi-group proof of exponential decay of relative entropy (Kullback-Leibler distance). Our aim is to improve upon the bound

$$d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq \frac{\kappa}{4} \int_0^T |p'_t|^2 dt$$

given in (2.13), Theorem 2.2. Indeed this bound is far from being sharp, as it is a consequence of the decay of the Kullback-Leibler distance by any measurable push forward. The following theorem presents the main result of this section.

**THEOREM 6.1.** *Let  $\hat{p}_t$  denote the density at time  $t$  of the law of  $X_t$  defined by (2.12), which is given by a time change of the convolution product (2.2). Let  $\hat{\mu}_t$  the density at time  $t$  of the law of  $Y_t$  defined by (2.11), i.e. the solution of the annealed Langevin dynamics driven by the score  $\nabla \ln \hat{p}_t$ . Consider  $\lambda_t$  such that  $\lambda_0 = 0$ . If  $\nu$  is strictly log-concave and  $\nabla \ln p_t$  is  $L(t)$  Lipschitz with  $L(t) \leq c$  for  $t \in [0, T]$ , then*

$$(6.5) \quad d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq \frac{\kappa}{2} \int_0^T |p'_s|^2 e^{-\frac{1}{\kappa} \int_s^T C_{LS}^{-1}(\hat{\mu}_{u/\kappa}) du} ds.$$

Note that the integral term is expected to be small since  $\kappa$  is small.

To understand what we intend to do, we will first give a formal derivation, assuming that there are no technical problems. We will then show how to overcome all the (unfortunately existing) technical problems.

**FORMAL PROOF.** Expanding the time derivative of the Kullback–Leibler distance, we have

$$\begin{aligned} \partial_t d_{KL}(\hat{p}_t, \hat{\mu}_t) &= \partial_t \int \ln \frac{\hat{p}_t}{\hat{\mu}_t} d\hat{p}_t = \int \ln \frac{\hat{p}_t}{\hat{\mu}_t} \partial_t \hat{p}_t - \int \frac{\hat{p}_t}{\hat{\mu}_t} \partial_t \hat{\mu}_t \\ &= - \int \ln \frac{\hat{p}_t}{\hat{\mu}_t} \nabla \cdot (v_t \hat{p}_t) dx + \int \frac{\hat{p}_t}{\hat{\mu}_t} \nabla \cdot \left( \hat{\mu}_t \nabla \left( \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right) \right) dx \\ &= \int \left\langle v_t, \nabla \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right\rangle \hat{p}_t dx - \int \left\| \nabla \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right\|^2 \hat{p}_t dx \\ &\leq - \left(1 - \frac{\lambda}{2}\right) \int \left\| \nabla \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right\|^2 \hat{p}_t dx + \frac{1}{2\lambda} \int \|v_t\|^2 \hat{p}_t dx \\ &= - \left(1 - \frac{\lambda}{2}\right) \int \left\| \nabla \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right\|^2 \hat{p}_t dx + \frac{\kappa^2}{2\lambda} |p'_{\kappa t}|^2 \end{aligned}$$

for all  $\lambda > 0$ . Recall that we have selected the Borel vector field  $v_t$  such that it minimizes the  $\mathbb{L}^2(\hat{p}_t)$ -norm, i.e.  $\|v_t\|_{\mathbb{L}^2(\hat{p}_t dx)} = \kappa |p'_{\kappa t}|$ . Notice that for  $\lambda = 2$  we recover (2.13), Theorem 2.11.

Assume that  $\mu_t$  satisfies a logarithmic Sobolev inequality (this will be proved later). It follows that

$$- \int \left\| \nabla \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right\|^2 \hat{p}_t dx = - \int \left\| \nabla \ln \frac{\hat{p}_t}{\hat{\mu}_t} \right\|^2 \frac{\hat{p}_t}{\hat{\mu}_t} \hat{\mu}_t dx \leq - \frac{2}{C_{LS}(\hat{\mu}_t)} d_{KL}(\hat{p}_t, \hat{\mu}_t).$$

Substituting this with  $\lambda = 1$  we obtain

$$\partial_t d_{KL}(\hat{p}_t, \hat{\mu}_t) \leq - \frac{1}{C_{LS}(\hat{\mu}_t)} d_{KL}(\hat{p}_t, \hat{\mu}_t) + \frac{\kappa^2}{2} |p'_{\kappa t}|^2.$$

Applying Gronwall's inequality, we have

$$d_{KL}(\hat{p}_t, \hat{\mu}_t) \leq \frac{\kappa^2}{2} \int_0^t |p'_{\kappa s}|^2 e^{-\int_s^t C_{LS}^{-1}(\hat{\mu}_u) du} ds.$$

Therefore, for  $t = T/\kappa$

$$\begin{aligned} d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) &\leq \frac{\kappa^2}{2} \int_0^{T/\kappa} |p'_{\kappa s}|^2 e^{-\int_s^{T/\kappa} C_{LS}^{-1}(\hat{\mu}_u) du} ds \\ &= \frac{\kappa}{2} \int_0^T |p'_s|^2 e^{-\frac{1}{\kappa} \int_s^T C_{LS}^{-1}(\hat{\mu}_{u/\kappa}) du} ds, \end{aligned}$$



where we have used that  $\mathcal{L}(X_0) = \mathcal{L}(Y_0)$ .  $\square$

The proof of Theorem 6.1 consists of two main components: establishing that  $\mu_t$  satisfies a log-Sobolev inequality and justifying the formal computation of the time derivative of the relative entropy.

The following result regarding functional inequalities for drifted Brownian motion will play a key role in showing that  $\mu_t$  satisfies a log-Sobolev inequality.

LEMMA 6.2. *Let  $Z_t = Z_0 + B_t + \int_0^t b(s, Z_s) ds$  where  $b(t, \cdot)$  is  $L(t)$ -Lipschitz for  $t \in [0, T]$ . Then*

$$C_{LS}(Z_t) \leq e^{\int_0^t L(u) du} C_{LS}(Z_0) + \int_0^t e^{\int_0^s L(u) du} ds.$$

Alternatively, if  $b(t, \cdot)$  satisfies  $\nabla b(t, \cdot) \leq -K(t)Id$ , with  $K(t) > 0$  for all  $t \in [0, T]$ . Then

$$C_{LS}(Z_t) \leq e^{-\int_0^t K(u) du} C_{LS}(Z_0) + \int_0^t e^{-\int_0^s K(u) du} ds.$$

This result is exactly Theorem 5 in [14] since

$\langle b(t, x) - b(t, y), x - y \rangle \geq -L(t) |x - y|^2$  or  $\langle b(t, x) - b(t, y), x - y \rangle \leq -K(t) |x - y|^2$  for all  $t$ , which is denoted (H.C.-L(t)) or (H.C.K(t)) in [14]. The proof, given in detail for the homogeneous case in Proposition 3 of [14], does not require any additional regularity for  $b$ . The only difference is our normalization of the log-Sobolev constant ( $2C_{LS}$  here is  $C_{LS}$  there).

Before providing the proof of Theorem 6.1, we present conditions under which the bound in Theorem 6.1 improves upon (2.13).

PROPOSITION 6.3. *Let  $\kappa \in (0, 1/2)$ ,  $\alpha \in (0, 1/2]$  such that  $\kappa^\alpha < 1/2$ ,  $\nu$  be a  $\mathcal{N}(0, \sigma^2 I)$  and  $\pi$  be compactly supported in the euclidean ball  $B(0, R)$ . Choose*

$$\lambda(t) = 2(t/T)^2 \mathbf{1}_{t < T/2} + (1 - \kappa^\alpha - 2(1 - 2\kappa^\alpha)(1 - t/T)^2) \mathbf{1}_{t \geq T/2}.$$

and  $\sigma^2 = \frac{R^2}{\kappa^\alpha}$ . Then, there exists a constant  $C := C(R, T)$  such that

$$d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq C d \kappa^{2(1-\alpha)}.$$

PROOF. With our choice

$$|p'_t|^2 \leq \frac{|\lambda'_t|^2}{2} \left( \frac{V_\pi}{\lambda_t} + \frac{V_\nu}{1 - \lambda_t} \right) \leq \frac{8}{T^2} (R^2 + \sigma^2 d) \leq 16 \frac{R^2}{T^2 \kappa^\alpha} d.$$

According to Proposition 4.12, we have that for all  $t \in [0, T]$

$$\nabla^2 \ln p_t \leq -\frac{\kappa^\alpha(1 - \lambda_T) - \kappa^{2\alpha} \lambda_T}{R^2(1 - \lambda_T)^2} Id = -\frac{\kappa^\alpha}{R^2} Id.$$

Therefore,  $p_t$  is strictly log-concave and applying Lemma 6.2 (Theorem 5 in [14]), it follows that

$$C_{LS}(Y_t) \leq C_{LS}(\nu) e^{-T\kappa^\alpha/R^2} + \frac{4R^2(1 - e^{-T\kappa^\alpha/R^2})}{\kappa^\alpha} \leq \sigma^2 + \frac{4R^2}{\kappa^\alpha} = \frac{5R^2}{\kappa^\alpha}.$$

Substituting all this into (6.5), we obtain

$$d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq C(R, T) d \kappa^{2(1-\alpha)}.$$

$\square$

Furthermore, if  $\pi$  can be expressed as the convolution of a gaussian with a compactly supported distribution, an alternative result can be established.

**PROPOSITION 6.4.** *Let  $\kappa \in (0, 1/2)$ ,  $\nu$  be a  $\mathcal{N}(0, \sigma^2 I)$  and  $\pi$  be the convolution of  $\mathcal{N}(0, \tau^2 I)$  with a distribution supported in the euclidean ball  $B(0, R)$ , with  $\tau^2 \geq R^2$ . Choose*

$$\lambda(t) = 2(t/T)^2 \mathbf{1}_{t \leq T/2} + (1 - 2(1 - t/T)^2) \mathbf{1}_{t > T/2}.$$

*Then, there exists a constant  $C := C(\tau^2, \sigma^2, R, T)$  such that*

$$d_{KL}(\pi, \mathcal{L}(Y_{T/\kappa})) \leq C d\kappa^2.$$

**PROOF.** With our choice

$$|p'_t|^2 \leq \frac{|\lambda'_t|^2}{2} \left( \frac{V_\pi}{\lambda_t} + \frac{V_\nu}{1 - \lambda_t} \right) \leq \frac{4}{T^2} (R^2 + (\tau^2 + \sigma^2) d).$$

On the other hand, according to Proposition 4.13, we have that for all  $t \in [0, T]$

$$\nabla^2 \ln p_t \leq - \frac{\min(\sigma^2, \tau^2 - R^2)}{\max(\sigma^2, \tau^2)} Id = -K Id,$$

where  $K \geq 0$ . Therefore,  $p_t$  is strictly log-concave and applying Lemma 6.2 (Theorem 5 in [14]), it follows that

$$C_{LS}(Y_t) \leq C_{LS}(\nu) e^{-KT} + \frac{4(1 - e^{-KT})}{K} \leq \min \left( \sigma^2 + 4T, \sigma^2 + \frac{4}{K} \right).$$

For  $K = 0$ , the second term in the first expression is  $T$  instead of  $(1 - e^{-KT})/K$ , however, the final bound still holds. Plugging this into (6.5), we obtain

$$d_{KL}(\pi, \mathcal{L}(Y_{T/\kappa})) \leq \frac{2}{T^2} (R^2 + (\tau^2 + \sigma^2) d) \min \left( \sigma^2 + 4T, \sigma^2 + \frac{4}{K} \right) \kappa^2.$$

Note that we can take  $\tilde{\pi} = \pi$  since  $\lambda_T = 1$ . □

**REMARK 6.5.** As shown in the proof below, thanks to Lemma 6.2,  $\mathcal{L}(Y_{T/\kappa})$  satisfies a logarithmic Sobolev inequality. Therefore, we can use the implication between the log Sobolev inequality and the Talagrand  $T_2$  inequality, Theorem 22.17 in [41] to derive a bound in Wasserstein-2 distance

$$W_2^2(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq 2C_{LS}(Y_{T/\kappa}) d_{KL}(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})).$$

In particular, under the assumptions of Propositions 6.3 and 6.4, we have

$$W_2^2(\tilde{\pi}, \mathcal{L}(Y_{T/\kappa})) \leq C(R, T) d\kappa^{2(1-3/2\alpha)} \quad \text{and} \quad W_2^2(\pi, \mathcal{L}(Y_{T/\kappa})) \leq C(\tau^2, \sigma^2, R, T) d\kappa^2,$$

respectively. ◇

**PROOF OF THEOREM 6.1.** First, we note that, under the assumptions of the theorem,  $\mathcal{L}(\sqrt{\lambda_0} X + \sqrt{1 - \lambda_0} Z) = \mathcal{L}(Z)$  satisfies a log-Sobolev inequality and  $L(t)$  is uniformly bounded on  $[0, T]$ . Therefore, Lemma 6.2 provides control over the log-Sobolev constant for  $\mu_t$  (uniformly in  $t$ ).

Next, it remains to justify the formal calculation we have made for the time derivative of the relative entropy. To this end we will first collect some properties of  $\hat{p}_t$  and  $\hat{\mu}_t$ .

First  $\hat{p}_t$  is the law of  $X_t = \sqrt{\lambda_t} X + \sqrt{1 - \lambda_t} Z$  so that

$$V_{\hat{p}_t} \leq 2(\lambda_t V_\pi + (1 - \lambda_t) V_\nu) \leq 2(V_\pi + V_\nu).$$

Next  $t \mapsto \nabla \ln(\hat{p}_t)$  is continuous, simply using Lebesgue's continuity theorem for integrals. It follows that  $\nabla \ln(\hat{p}_t)(0)$  is bounded w.r.t  $t \in [0, T]$ . Since  $\nabla \ln(\hat{p}_t)$  is  $c$  global Lipschitz, and thanks to what precedes, it is at most of linear growth, uniformly in  $t$ , so that

$$\sup_t \int |\nabla \ln(\hat{p}_t)|^2 \hat{p}_t dx < +\infty.$$

Also notice that, since  $e^{-W}$  (or  $e^{-U}$ ) is bounded,  $\hat{p}_t \in \mathbb{L}^1(dx) \cup \mathbb{L}^\infty(dx)$  hence belongs to all the  $\mathbb{L}^q$  by interpolation.

We turn to the properties of  $\hat{\mu}_t$ . Using that  $Y$  is a drifted Brownian motion with a  $c$ -Lipschitz drift,  $Y_t$  has bounded (w.r.t.  $t \in [0, T]$ ) moments of order  $k$  as soon as  $Y_0$  has a finite moment of order  $k$ . Since  $\nu$  is strictly log-concave,  $Y_0$  has finite moments of any order. In particular

$$\sup_t \mathbb{E}^{Q_Y} [|\nabla \ln \hat{p}_t(\omega_t)|^2] = \sup_t \int |\nabla \ln \hat{p}_t|^2 \hat{\mu}_t dx < +\infty.$$

Let  $B^\nu$  be the law of a Brownian motion with initial distribution  $\nu$  (and variance 2), denoted by  $B_t^\nu$ . The previous inequality shows that

$$d_{KL}(Q_Y, B^\nu) < +\infty.$$

Denote by  $R$  the time reversal operator at time  $T$ . Since the Kullback-Leibler distance decays using measurable push-forward,  $d_{KL}(Q_Y \circ R, B^\nu \circ R) \leq d_{KL}(Q_Y, B^\nu) < +\infty$ .

Let  $\gamma_t$  be the density of the law  $\mathcal{L}(Z + \sqrt{2t}G)$  where  $G$  is a standard gaussian measure.  $\gamma_t$  is thus the density of  $B_t^\nu$ . All what is done in section 2 and section 3 is unchanged if we replace  $X_0$  by  $G$ ,  $\sqrt{\lambda_t}$  by  $\sqrt{2t}$  and  $\sqrt{1 - \lambda_t}$  by 1. We may thus again apply Theorem 4.4 and Lemma 3.8 to deduce that  $\nabla \ln(\gamma_t)$  is  $c'$  global Lipschitz uniformly in  $t \in [0, T]$ . In particular, as for  $\nabla \ln(\hat{p}_t)$ , it is at most of linear growth uniformly in  $t$ .

Standard results on time reversal thus say that  $B^\nu \circ R$  is the law of a Brownian motion with drift  $-2\nabla \ln(\gamma_{T-t})$  and initial measure  $\gamma_T dx$ .  $Q_Y \circ R$  is thus the law of a Brownian motion with drift  $r_{T-t} = -2\nabla \ln(\gamma_{T-t}) + h_{T-t}$  according to the Girsanov transform theory, and

$$\int_0^T \int \mathbb{E}^{Q_Y} [|h_{T-t} \circ R|^2] dt < +\infty$$

thanks to the finite entropy condition (implying in particular  $d_{KL}(\mathcal{L}(Y_T), \mathcal{L}(B_T^\nu)) < +\infty$ ).

Using the results in [12] (see Theorem 4.9 and formula (4.12) therein, also see the unpublished preliminary version [22] Corollary 3.15 and Theorem 4.8) one deduces the so called duality equation

$$\nabla \ln(\hat{\mu}_t) = \nabla \ln(\hat{p}_t) + \bar{h}_t$$

for  $t > 0$ , where  $\bar{h}_t(x) = \mathbb{E}^{Q_Y} [h_{T-t} \circ R | X_t = x]$ . In particular, using Cauchy-Schwarz inequality for the conditional expectation,

$$\int_0^T \int |\bar{h}_t|^2 \hat{\mu}_t(dx) dt < +\infty.$$

Since  $\nabla \ln(\hat{p}_t)$  is of linear growth, it is square integrable w.r.t.  $\hat{\mu}_t$  uniformly in  $t$ . All this implies that

$$\int_0^T \int |\nabla \ln(\hat{\mu}_t)|^2 \hat{\mu}_t dx dt < +\infty.$$

Unfortunately this is not enough to ensure that  $\nabla \ln(\hat{p}_t/\hat{\mu}_t) \in \mathbb{L}^1([0, T], \mathbb{L}^2(\hat{p}_t dx))$ .

We will thus have to use a (space) smooth cutoff function  $\psi$  satisfying  $\mathbf{1}_{B(0,r)} \leq \psi \leq \mathbf{1}_{B(0,R)}$  for some well chosen  $0 < r < R < +\infty$ . Recall that  $Y$  is a drifted Brownian motion with a drift  $b(t, x)$  which is  $c$ -Lipschitz uniformly in  $t$  and at most of linear growth, uniformly in  $t$ . One can thus use (a simplified and less precise form of) Theorem 3.1 in [37] saying that the transition density of the process satisfies for all  $T \geq t \geq t_0 > 0$ ,

$$\begin{aligned} \hat{\mu}_t(x, y) &\geq C(t_0, T) \exp - \left( c_1(t_0, T) |y - x|^2 e^{2\alpha t} + c_2(t_0, T) \sum_j |x_j - y_j| \right) \\ &\geq C(t_0, T) \exp - \left( 2c_1(t_0, T) |xe^{\alpha t}|^2 + c_2(t_0, T) \sum_j |x_j| \right) \\ &\quad \exp - \left( 2c_1(t_0, T) |y|^2 + c_2(t_0, T) \sum_j |y_j| \right), \end{aligned}$$

for some  $\alpha > 0$ , where all the constants depend on  $C_W, D_W, C_U$  and  $\sup_t |\nabla \ln(\hat{p}_t)(0)|$ .

Actually this Theorem is only stated for a time homogeneous drift. A careful reading of the proof shows that it extends to a time inhomogeneous drift, provided it is of at most linear growth, uniformly in  $t$ . Indeed the proof is based on a comparison result with two other, time homogeneous, diffusion processes (Theorem 2.7) first shown in [36]. This Theorem immediately extends to time inhomogeneous drifts of at most linear growth uniformly in  $t$  as conditions 1 and 2 in the Theorem are preserved. It is worth highlighting that the preliminary Corollary 2.6 in [37] is written for a time dependent perturbation, and furnishes the proof of Theorem 2.7.

We immediately deduce a rough lower bound

$$(6.6) \quad \hat{\mu}_t(x) \geq C'(t_0, T) e^{-c'_1(t_0, T)|x|^2 - c'_2(t_0, T)|x|}.$$

**REMARK 6.6.** Estimates for general heat kernels have a long history, starting with Nash and overall Aronson in 1967, for generators in divergence form with very weak regularity. For non divergence generators a lot of work has been done for regular coefficients. For only Lipschitz, but unbounded, drifts the literature is not as rich. In addition to [36, 37], we may mention the more recent [40] covering more general situations (including non constant diffusion coefficient). We also refer to the Bibliography of the latter.  $\diamond$

Now consider  $t \geq t_0 > 0$ . Recall that  $\psi$  is smooth and satisfies  $\mathbf{1}_{B(0,r)} \leq \psi \leq \mathbf{1}_{B(0,R)}$  for some well chosen  $0 < r < R < +\infty$ .

First

$$\begin{aligned} \int \psi \ln \frac{\hat{p}_t}{\hat{\mu}_t} \hat{p}_t dx - \int \psi \ln \frac{\hat{p}_s}{\hat{\mu}_s} \hat{p}_s dx &= \int_s^t \int \psi \left( 1 + \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right) \partial_u \hat{p}_u du - \int_s^t \int \psi \frac{\hat{p}_u}{\hat{\mu}_u} \partial_u \hat{\mu}_u du \\ &= - \int_s^t \int \psi \left( 1 + \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right) \nabla \cdot (v_u \hat{p}_u) dx du \\ &\quad + \int_s^t \int \psi \frac{\hat{p}_u}{\hat{\mu}_u} \nabla \cdot \left( \hat{\mu}_u \nabla \left( \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right) \right) dx du \\ &= \int_s^t \int \psi \left\langle v_u, \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right\rangle \hat{p}_u dx du \end{aligned}$$

$$\begin{aligned}
& - \int_s^t \int \psi \left\| \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right\|^2 \hat{p}_u dx du \\
& + \int_s^t \int \langle \nabla \psi, v_u \rangle \left( 1 + \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right) \hat{p}_u dx du \\
& - \int_s^t \int \langle \nabla \psi, \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \rangle \hat{p}_u dx du \\
& = A_1 - A_2 + A_3 - A_4.
\end{aligned}$$

Everything is justified since

- (1)  $v_u$  and  $\nabla \ln \hat{p}_u$  belong to  $\mathbb{L}^2(\hat{p}_u dx)$ ,
- (2)  $\sqrt{\psi} \nabla \ln \hat{\mu}_u$  belongs to  $\mathbb{L}^2(\hat{p}_u dx)$ , since  $\hat{\mu}_u$  is bounded from below by a positive constant on the support of  $\psi$ , hence  $\psi \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u}$  belongs to  $\mathbb{L}^2(\hat{p}_u dx)$
- (3)  $\ln \hat{p}_u$  belongs to  $\mathbb{L}^2(\hat{p}_u dx)$  since  $\hat{p}_u$  belongs to all the  $\mathbb{L}^q(dx)$ ,  $\nabla \psi \ln \hat{\mu}_u$  is bounded, hence  $\nabla \psi (1 + \ln \frac{\hat{p}_u}{\hat{\mu}_u})$  belongs to  $\mathbb{L}^2(\hat{p}_u dx)$ ,
- (4)  $\langle \nabla \psi, \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \rangle$  belongs to  $\mathbb{L}^2(\hat{p}_u dx)$  hence to  $\mathbb{L}^1(\hat{p}_u dx)$ ,

and all the  $\mathbb{L}^2$  norms in the previous items are integrable in time.

We now study each term  $A_j$ . First for  $\lambda > 0$ ,

$$A_1 \leq \frac{\lambda}{2} \int_s^t \int \psi \left\| \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right\|^2 \hat{p}_u dx du + \frac{1}{2\lambda} \int_s^t \int \|v_u\|^2 \hat{p}_u dx du.$$

Next we have

$$A_3 \leq \frac{\|\nabla \psi\|_\infty}{2} \left( \int_s^t \int \mathbf{1}_{|x| \geq r} \|v_u\|^2 \hat{p}_u dx du + \int_s^t \int \mathbf{1}_{|x| \geq r} \left( 1 + \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right)^2 \hat{p}_u dx du \right),$$

and

$$\begin{aligned}
\frac{1}{3} \int_s^t \int \mathbf{1}_{|x| \geq r} \left( 1 + \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right)^2 \hat{p}_u dx du & \leq \int_s^t \int \mathbf{1}_{|x| \geq r} \hat{p}_u dx du \\
& + \int_s^t \int \mathbf{1}_{|x| \geq r} (\ln^2 \hat{p}_u) \hat{p}_u dx du \\
& + \int_s^t \int \mathbf{1}_{|x| \geq r} (\ln^2(1/\hat{\mu}_u)) \hat{p}_u dx du,
\end{aligned}$$

and finally thanks to (6.6),

$$\int_s^t \int \mathbf{1}_{|x| \geq r} (\ln^2(1/\hat{\mu}_u)) \hat{p}_u dx du \leq C \int_s^t \int \mathbf{1}_{|x| \geq r} (1 + |x|^2)^2 \hat{p}_u dx du.$$

For  $A_4$  we perform one more integration by parts. First we introduce a new cutoff  $\eta$  similar to  $\psi$  replacing  $(r, R)$  by  $(r/2, 2R)$ . This is done for the boundary term in the integration by parts to vanish.

We thus have

$$\begin{aligned}
A_4 & = \int_s^t \int \eta \langle \nabla \psi, \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \rangle \hat{p}_u dx du = \\
& = - \int_s^t \int (\eta \hat{p}_u \Delta \psi + \langle \nabla \psi, \nabla (\eta \hat{p}_u) \rangle) \ln \frac{\hat{p}_u}{\hat{\mu}_u} dx du
\end{aligned}$$

$$\begin{aligned}
&\leq \|\Delta\psi\|_\infty \int_s^t \int \mathbf{1}_{|x|\geq r} \left| \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right| \hat{p}_u dx du + \\
&\quad + \|\nabla\psi\|_\infty \int_s^t \int \mathbf{1}_{|x|\geq r} \left| \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right| ( \|\nabla\eta\|_\infty \hat{p}_u + \eta \|\nabla\hat{p}_u\| ) dx du = D_1 + D_2.
\end{aligned}$$

Gathering all this we have obtained, after simple manipulations,

$$\begin{aligned}
&(1 - \frac{\lambda}{2}) \int_s^t \int \psi \left\| \nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right\|^2 \hat{p}_u dx du \leq \\
&\leq - \int \psi \ln \frac{\hat{p}_t}{\hat{\mu}_t} \hat{p}_t dx + \int \psi \ln \frac{\hat{p}_s}{\hat{\mu}_s} \hat{p}_s dx + \int_s^t \int \left( \frac{1}{2\lambda} + \frac{\|\nabla\psi\|_\infty}{2} \mathbf{1}_{|x|\geq r} \right) \|v_u\|^2 \hat{p}_u dx du \\
&\quad + \frac{3\|\nabla\psi\|_\infty}{2} \int_s^t \int \mathbf{1}_{|x|\geq r} (1 + \ln^2 \hat{p}_u + C(1 + |x|^2)^2) \hat{p}_u dx du \\
&\quad + \int_s^t \int \mathbf{1}_{|x|\geq r} (\|\Delta\psi\|_\infty + \|\nabla\psi\|_\infty \|\nabla\eta\|_\infty) \left| \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right| \hat{p}_u dx du \\
&\quad + \|\nabla\psi\|_\infty \int_s^t \int \mathbf{1}_{|x|\geq r} \left| \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right| \|\nabla\hat{p}_u\| dx du.
\end{aligned}$$

The goal now is to let  $r$  go to infinity, i.e.  $\psi$  go to 1. Of course this can be done assuming that  $\nabla\psi$ ,  $\Delta\psi$  and  $\nabla\eta$  are uniformly bounded independently of  $r$ . Look at the right hand side. The first three terms in the sum converge to some limit, the fourth and the fifth one's go to 0 thanks to the integrability properties we recalled (and the fact that  $\int \rho |\ln \rho| dx \leq c + \int \rho \ln \rho dx$ ).

For the last term we remark that

$$\left| \ln \frac{\hat{p}_u}{\hat{\mu}_u} \right| \|\nabla\hat{p}_u\| \leq |\ln \hat{p}_u| \frac{\|\nabla\hat{p}_u\|}{\hat{p}_u} \hat{p}_u + C(1 + |x|^2) \frac{\|\nabla\hat{p}_u\|}{\hat{p}_u} \hat{p}_u.$$

Recalling that  $\ln \hat{p}_u$ ,  $\frac{\nabla\hat{p}_u}{\hat{p}_u}$  and  $|x|^2$  all belong to  $\mathbb{L}^2(\hat{p}_u)$ , uniformly in time, we deduce thanks to Cauchy Schwarz inequality that the previous function is in  $\mathbb{L}^1(\hat{p}_u)$ , so that the last term goes to 0 as  $r \rightarrow \infty$ , uniformly in time.

We have finally obtained that  $\nabla \ln \frac{\hat{p}_u}{\hat{\mu}_u} \in \mathbb{L}^2(\hat{p}_u)$ , and in the limit  $r \rightarrow \infty$  a justification for the final (formal) bound for the time derivative of the Kullback-Leibler distance.

The proof is however not complete since all these derivations are made for  $t \geq t_0 > 0$ . It only means that we have first to integrate between  $t_0$  and  $T$ , pass to the limit and then check that the remaining bounds do not depend on  $t_0$  and conclude. Actually this final step is not necessary, since it only modifies the constants.  $\square$

## APPENDIX

We recall the result from [19], which is used to establish the existence and uniqueness of weak solutions for (2.12) in Section 2.

**THEOREM A.1.** *Assume that  $P$  is a unique weak solution of the martingale problem associated to the generator  $L_P(t) = \frac{1}{2}a(t, \cdot)\partial^2 + b(t, \cdot)\cdot\nabla$ , where  $a = \sigma\cdot\sigma^*$  with initial measure  $\theta_0$ . Let  $\rho$  be an admissible flow in the following sense*

- (i)  $\rho$  satisfies the  $(B, C_b^\infty)$  weak forward equation i.e. for all  $f \in C_b^{1,\infty}(\mathbb{R}^+ \times \mathbb{R}^d)$  and all  $0 \leq s < t \leq T$ ,

$$\int f(t, x) d\rho_t - \int f(s, x) d\rho_s = \int_s^t \int (\partial_u f + L_P(u)f + \langle B(u, x), \nabla_x f \rangle) d\rho_u du.$$

(ii)  $B$  is of finite  $\rho$ -energy, i.e. for all  $T \geq t > 0$ ,

$$\int_0^t \int \langle B, aB \rangle(s, x) \rho_s(dx) ds < +\infty.$$

(iii)  $H(\rho_0|\theta_0) := \int \ln(\frac{d\rho_0}{d\theta_0}) d\rho_0 < +\infty$ .

Assume in addition that one of the following conditions is satisfied

- (1)  $\sigma$  and  $b$  are  $C^{1,2,\alpha}$  i.e. twice differentiable in  $x$ , one time differentiable in  $t$  with derivatives that are  $\alpha$  locally Hölder continuous,
- (2)  $\sigma$  and  $b$  are locally Hölder continuous, and  $a$  is uniformly elliptic i.e.  $\inf_{0 \leq t \leq T} a(t, \cdot) \geq \kappa \text{Id}$  in the sense of quadratic forms for some  $\kappa > 0$ ,
- (3) If  $q_t$  denotes the marginal distribution  $P \circ X_t^{-1}$  of  $P$  at time  $t$ ,  $\rho_t$  is absolutely continuous w.r.t.  $q_t$  and  $d\rho_t/dq_t$  is locally bounded, i.e. bounded on  $[0, T] \times K$  for any compact set  $K \subset \mathbb{R}^d$ .

Then  $Q$  defined by

$$\frac{dQ}{dP}|_{\mathcal{F}_T} = \frac{d\rho_0}{d\theta_0} \exp \left( \int_0^T \langle B(s, \omega_s), dM_s \rangle - \frac{1}{2} \int_0^T \langle B, aB \rangle(s, \omega_s) ds \right)$$

where  $M_\cdot$  is the (local) martingale part of  $X_\cdot$  under  $P$ , is a probability measure satisfying

$$H(Q|P) = H(\rho_0|\theta_0) + \frac{1}{2} \int_0^T \int \langle B, aB \rangle(s, x) \rho_s(dx) ds < +\infty,$$

whose time marginals at time  $t$  are exactly  $\rho_t$ .

In addition,  $Q$  is the unique weak solution (up to time  $T$ ) of the stochastic differential equation associated to  $L_P(s) + aB(s, x) \cdot \nabla$  with initial distribution  $\rho_0$ .

**Acknowledgments.** A. Guillin is supported by the ANR-23-CE-40003, Conviviality and has benefited from a government grant managed by the Agence Nationale de la Recherche under the France 2030 investment plan ANR-23-EXMA-0001. P. Cordero-Encinar is supported by EPSRC through the Modern Statistics and Statistical Machine Learning (StatML) CDT programme, grant no. EP/S023151/1.

## REFERENCES

- [1] ALBERGO, M. S., BOFFI, N. M. and VANDEN-EIJNDEN, E. (2024). A Unifying Framework for Flows and Diffusions. Available on ArXiv: 2303.08797.v3 [cs.LG].
- [2] ALBERGO, M. S. and VANDEN-EIJNDEN, E. (2024). Learning to sample better. *J. Stat. Mech. Theory Exp.* **2024** Paper No. 104014, 30. [MR4832809](#)
- [3] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient flows in metric spaces and in the space of probability measures*, second ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser Verlag, Basel. [MR2401600](#)
- [4] BAKRY, D., BARTHE, F., CATTIAUX, P. and GUILLIN, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures. *Elec. Comm. in Prob.* **13** 60–66.
- [5] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759.
- [6] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov diffusion operators. Grundlehren der mathematischen Wissenschaften* **348**. Springer, Berlin.
- [7] BALL, K., BARTHE, F. and NAOR, A. (2003). Entropy jumps in the presence of a spectral gap. *Duke Math. J.* **119** 41–63. <https://doi.org/10.1215/S0012-7094-03-11912-2> [MR1991646](#)



- [8] BARDET, J. B., GOZLAN, N., MALRIEU, F. and ZITT, P. A. (2018). Functional inequalities for gaussian convolutions of compactly supported measures: explicit bounds and dimension dependence. *Bernoulli* **24** 333–353.
- [9] BONNEFONT, M., JOULIN, A. and MA, Y. (2016). Spectral gap for spherically symmetric log-concave probability measures, and beyond. *J. Funct. Anal.* **270** 2456–2482. <https://doi.org/10.1016/j.jfa.2016.02.007> MR3464047
- [10] CARLEN, E. A. (1984). Conservative diffusions. *Comm. Math. Phys.* **94** 293–315. MR763381
- [11] CATTIAUX, P. (2024). Entropy on the path space and application to singular diffusions and mean-field models. Available on ArXiv: 2404.09552.v2 [math.AP].
- [12] CATTIAUX, P., CONFORTI, G., GENTIL, I. and LÉONARD, C. (2023). Time reversal of diffusion processes under a finite entropy condition. *Ann. Inst. Henri Poincaré Probab. Stat.* **59** 1844–1881. <https://doi.org/10.1214/22-aihp1320> MR4663509
- [13] CATTIAUX, P. and FRADON, M. (1996). Entropy, reversible diffusion processes, and Markov uniqueness. *J. Funct. Anal.* **138** 243–272. <https://doi.org/10.1006/jfan.1996.0064> MR1391638
- [14] CATTIAUX, P. and GUILLIN, A. (2014). Semi log-concave Markov diffusions. In *Séminaire de Probabilités XLVI. Lecture Notes in Math.* **2123** 231–292. Springer, Cham. [https://doi.org/10.1007/978-3-319-11970-0\\_9](https://doi.org/10.1007/978-3-319-11970-0_9) MR3330820
- [15] CATTIAUX, P. and GUILLIN, A. (2017). Hitting times, functional inequalities, Lyapunov conditions and uniform ergodicity. *J. Funct. Anal.* **272** 2361–2391. <https://doi.org/10.1016/j.jfa.2016.10.003> MR3603301
- [16] CATTIAUX, P. and GUILLIN, A. (2022). Functional inequalities for perturbed measures with applications to log-concave measures and to some Bayesian problems. *Bernoulli* **28** 2294–2321. <https://doi.org/10.3150/21-bej1419> MR4474544
- [17] CATTIAUX, P., GUILLIN, A. and WU, L. (2022). Poincaré and logarithmic Sobolev inequalities for nearly radial measures. *Acta Math. Sin. (Engl. Ser.)* **38** 1377–1398. <https://doi.org/10.1007/s10114-022-0501-3> MR4473859
- [18] CATTIAUX, P., GUILLIN, A. and ZITT, P. A. (2013). Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré. Prob. Stat.* **49** 95–118.
- [19] CATTIAUX, P. and LÉONARD, C. (1994). Minimization of the Kullback information of diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **30** 83–132. MR1262893
- [20] CATTIAUX, P. and LÉONARD, C. (1995). Large deviations and Nelson processes. *Forum Math.* **7** 95–115. <https://doi.org/10.1515/form.1995.7.95> MR1307957
- [21] CATTIAUX, P. and LÉONARD, C. (1996). Minimization of the Kullback information for some Markov processes. In *Séminaire de Probabilités, XXX. Lecture Notes in Math.* **1626** 288–311. Springer, Berlin. <https://doi.org/10.1007/BFb0094655> MR1459490
- [22] CATTIAUX, P. and PETIT, F. (2001). Singular diffusions, time reversal and applications to Fokker-Planck equations. Not to be published. Available on <https://perso.math.univ-toulouse.fr/cattiaux/publications-2/>.
- [23] CHEWI, S. (2024). Log-Concave Sampling. Unfinished draft. Available at URL <https://chewisinho.github.io>.
- [24] CORDERO-ENCINAR, P., AKYILDIZ, O. D. and DUNCAN, A. B. (2024). Non-asymptotic Analysis of Diffusion Annealed Langevin Monte Carlo for Generative Modelling. Available on ArXiv: 2502.09306.v1 [stat.ML].
- [25] DOWN, D., MEYN, S. P. and TWEEDIE, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** 1671–1691. MR1379163
- [26] EBERLE, A. (2011). Reflection coupling and Wasserstein contractivity without convexity. *C. R. Acad. Sci. Paris, Ser. I* **349** 1101–1104.
- [27] EBERLE, A. (2016). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** 851–886. <https://doi.org/10.1007/s00440-015-0673-1> MR3568041
- [28] FÖLLMER, H. (1988). Random fields and diffusion processes. In *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87. Lecture Notes in Math.* **1362** 101–203. Springer, Berlin. <https://doi.org/10.1007/BFb0086180> MR983373
- [29] GHOMI, M. (2002). The problem of optimal smoothing for convex functions. *Proc. Amer. Math. Soc.* **130** 2255–2259. <https://doi.org/10.1090/S0002-9939-02-06743-6> MR1896406
- [30] GUO, W., TAO, M. and CHEN, Y. (2025). Provable Benefit of Annealed Langevin Monte Carlo for Non-log-concave Sampling. To appear in ICLR.
- [31] KLARTAG, B. (2023). Logarithmic bounds for isoperimetry and slices of convex sets. *Ars Inven. Anal. Paper No. 4*, 17. MR4603941

- [32] KURTZMANN, A. (2007). Comportement asymptotique de diffusions renforcées sur  $\mathbb{R}^d$ , Thèse de Doctorat, Université de Neuchâtel, Switzerland.
- [33] LÉONARD, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. <https://doi.org/10.3934/dcds.2014.34.1533> MR3121631
- [34] MA, Y.-A., CHEN, Y., JIN, C., FLAMMARION, N. and JORDAN, M. I. (2019). Sampling can be faster than optimization. *Proc. Natl. Acad. Sci. USA* **116** 20881–20885. <https://doi.org/10.1073/pnas.1820003116> MR4025861
- [35] MILLET, A., NUALART, D. and SANZ, M. (1989). Integration by parts and time reversal for diffusion processes. *Ann. Probab.* **17** 208–238. MR972782
- [36] QIAN, Z., RUSSO, F. and ZHENG, W. (2003). Comparison theorem and estimates for transition probability densities of diffusion processes. *Probab. Theory Related Fields* **127** 388–406. <https://doi.org/10.1007/s00440-003-0291-1> MR2018922
- [37] QIAN, Z. and ZHENG, W. (2004). A representation formula for transition probability densities of diffusions and applications. *Stochastic Process. Appl.* **111** 57–76. <https://doi.org/10.1016/j.spa.2003.12.004> MR2049569
- [38] SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: a review. *Statistics Surveys*. **8** 45–114.
- [39] SONG, Y., SOHL-DICKSTEIN, J., KINGMA, D., KUMAR, A., ERMON, S. and POOLE, B. (2021). Score-based generative modeling through stochastic differential equations. In *ICLR, outstanding paper session I*.
- [40] TAGUCHI, D. and TANAKA, A. (2020). Probability density function of SDEs with unbounded and path-dependent drift coefficient. *Stochastic Process. Appl.* **130** 5243–5289. <https://doi.org/10.1016/j.spa.2020.03.006> MR4127329
- [41] VILLANI, C. (2008). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin.