

## Transportation inequalities for Markov kernels and their applications\*

Fabrice Baudoin<sup>†</sup>      Nathaniel Eldredge<sup>‡</sup>

### Abstract

We study the relationship between functional inequalities for a Markov kernel on a metric space  $X$  and inequalities of transportation distances on the space of probability measures  $\mathcal{P}(X)$ . Extending results of Luise and Savaré on Hellinger–Kantorovich contraction inequalities for the particular case of the heat semigroup on an  $RCD(K, \infty)$  metric space, we show that more generally, such contraction inequalities are equivalent to reverse Poincaré inequalities. We also adapt the “dynamic dual” formulation of the Hellinger–Kantorovich distance to define a new family of divergences on  $\mathcal{P}(X)$  which generalize the Rényi divergence, and we show that contraction inequalities for these divergences are equivalent to the reverse logarithmic Sobolev and Wang Harnack inequalities. We discuss applications including results on the convergence of Markov processes to equilibrium, and on quasi-invariance of heat kernel measures in finite and infinite-dimensional groups.

**Keywords:** Markov kernels; optimal transport; functional inequalities; reverse Poincaré inequality; reverse logarithmic Sobolev inequality; Kantorovich–Wasserstein distance; Hellinger distance; Kuwada duality.

**MSC2020 subject classifications:** Primary 47D07, Secondary 49Q22; 28A33; 58J65.

Submitted to EJP on June 26, 2020, final version accepted on March 12, 2021.

Supersedes arXiv:2004.02050.

## 1 Introduction

The goal of this paper is to build upon recent results of G. Luise and G. Savaré [28] on contraction properties of the flow of a heat semigroup in spaces of measures. There, the authors study a “dynamic dual” formulation of various distances between probability measures on a metric measure space, including the Kantorovich–Wasserstein

---

\*Fabrice Baudoin has been supported by National Science Foundation grant DMS-1901315. Nathaniel Eldredge has been supported by a grant from the Simons Foundation (#355659, N.E.).

<sup>†</sup>Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA. E-mail: [fabrice.baudoin@uconn.edu](mailto:fabrice.baudoin@uconn.edu)

<sup>‡</sup>School of Mathematical Sciences, University of Northern Colorado, Greeley, CO 80369, USA. E-mail: [nelldredge@unco.edu](mailto:nelldredge@unco.edu)

and Hellinger distances as well as a family of Hellinger–Kantorovich distances  $\mathbb{H}_\alpha$  introduced in [27]. They focus on the setting of  $\text{RCD}(K, \infty)$  spaces, in which the canonical heat semigroup  $P_t$  generated by the Cheeger energy satisfies a Bakry–Émery curvature condition; these spaces are, very roughly speaking, more general analogues of Riemannian manifolds with Ricci curvature bounded from below. Under this assumption, they obtain contraction inequalities of the form

$$\text{He}_2(\mu_0 P_t, \mu_1 P_t) \leq \mathbb{H}_{\alpha(t)}(\mu_0, \mu_1) \tag{1.1}$$

where  $\mu P_t$  denotes the dual action of the heat semigroup  $P_t$  on the probability measure  $\mu$ ,  $\text{He}$  and  $\mathbb{H}$  are the Hellinger and Hellinger–Kantorovich distances respectively, and  $\alpha(t)$  depends on  $K$ . The proof is based on the fact that  $\text{RCD}(K, \infty)$  spaces satisfy a reverse Poincaré inequality of the form

$$|\nabla P_t f|^2 \leq \frac{K}{e^{2Kt} - 1} (P_t(f^2) - (P_t f)^2). \tag{1.2}$$

Indeed, the inequality (1.2), with its specific form of the time-dependent constant  $\frac{K}{e^{2Kt} - 1}$ , is one of many functional inequalities that are equivalent to the Bakry–Émery curvature condition; see for instance [2, Proposition 3.3].

The first goal of the present paper is to further study the relationship between reverse Poincaré inequalities and Hellinger–Kantorovich contraction inequalities. Our first main result is Theorem 3.7, in which we show that the implication between the two holds in a much more general setting than  $\text{RCD}(K, \infty)$  spaces. We suppose only that we have a Markov operator  $P$  acting on a sufficiently nice metric space  $X$ , and we show that if  $P$  satisfies a reverse Poincaré inequality of the form

$$|\nabla P f|^2 \leq C(P(f^2) - (P f)^2), \quad f \in \text{Lip}_b(X) \tag{1.3}$$

then we obtain a Hellinger–Kantorovich contraction of the form

$$\text{He}_2(\mu_0 P, \mu_1 P) \leq \mathbb{H}_{4/C}(\mu_0, \mu_1) \tag{1.4}$$

for all probability measures  $\mu_0, \mu_1$  on  $X$ . In particular, (1.4) holds in non-RCD models where there is a semigroup  $P_t$  which satisfies (1.3) for each  $t$ , but with a time-dependent constant  $C(t)$  that is not of the form appearing in (1.2). We discuss several examples and applications in Section 4, including subelliptic diffusions arising in sub-Riemannian geometry, non-symmetric Ornstein–Uhlenbeck operators on Carnot groups, Langevin dynamics driven by Lévy processes, and others.

Furthermore, in this general setting, we are able to show (also in Theorem 3.7) that the converse implication holds as well, so that (1.3) and (1.4) are in fact equivalent. They are also equivalent to a parabolic Harnack inequality. Thus the Hellinger–Kantorovich contraction can be seen as a new aspect of a well-known family of functional inequalities, providing additional tools and motivations for their study.

The key tool in all of this is the dynamic dual formulation of the Hellinger–Kantorovich distance (Definition 3.1), originally introduced in [27], which expresses  $\mathbb{H}_\alpha(\mu_0, \mu_1)$  as the supremum of  $\int \varphi_1 d\mu_1 - \int \varphi_0 d\mu_0$  over a family of time-dependent functions  $\varphi_s : [0, 1] \times X \rightarrow \mathbb{R}$  satisfying a certain Hamilton–Jacobi partial differential inequality in time and space. This formula extends the so-called Kantorovich duality for the Kantorovich–Wasserstein distance, and also includes an expression for the Hellinger distance. Having the distance defined in terms of solutions of a partial differential inequality makes it particularly convenient to relate it to functional inequalities where the gradient appears, as we demonstrate in Theorems 3.4 and 3.7. Indeed, this technique also provides an

extension of the Kuwada duality theorem [25, 27], relating  $L^2$  gradient estimates to a Kantorovich–Wasserstein contraction inequality; see Corollary 3.5.

Pursuing this idea in a different direction, in Section 5 we use a dynamic dual approach to formulate a new family of transportation-cost divergences  $T_{a,b}$  on the space of probability measures, which are of “entropic” type and include the Rényi divergence. In place of reverse Poincaré inequalities, this family is designed to connect with reverse logarithmic Sobolev inequalities of the form

$$Pf|\nabla \ln Pf|^2 \leq C(P(f \ln f) - (Pf) \ln Pf). \tag{1.5}$$

We note that an equality of the type (1.5) is stronger than an inequality of the type (1.3), by using (1.5) with  $1 + \varepsilon f$ . After developing some properties of the  $T_{a,b}$  divergences, we show in Theorem 5.15 that (1.5) is actually equivalent to a family of entropic transportation-cost contraction inequalities for  $T_{a,b}$ , of the form

$$T_{0,\kappa C}(\mu_0 P, \mu_1 P) \leq T_{\kappa,\kappa C}(\mu_0, \mu_1), \quad \kappa > 0. \tag{1.6}$$

These two statements are moreover equivalent to a Wang-type parabolic Harnack inequality, as well as to an integrated Harnack inequality (see Remark 5.16). Thus, the reverse log Sobolev inequality (1.5) also has a “transport” aspect. We discuss in Section 6 how (1.6) can be used, in finite or infinite dimensions, to prove certain quasi-invariance results that were previously obtained via Wang Harnack inequalities or other methods; see for instance [9].

For other applications, and a general overview of reverse Poincaré and log-Sobolev inequalities and of the connections with Harnack type inequalities, we refer to the book [35].

## 2 General setup and notation

Throughout the paper, unless otherwise specified,  $(X, d)$  denotes a complete, proper, separable metric space which is a length space; in particular, path connected. We suppose that  $X$  is equipped with a strong upper gradient  $|\nabla f|$  as defined in [1, Definition 1.2.1]. More precisely, for a measurable function  $f : X \rightarrow \mathbb{R}$  we define

$$|\nabla f|(x) = \lim_{r \rightarrow 0} \sup_{0 < d(x,y) \leq r} \frac{|f(x) - f(y)|}{d(x,y)} \tag{2.1}$$

and denote by  $\text{Lip}_b(X)$  the space of all bounded Lipschitz functions on  $X$ . Then, we have the following result:

**Lemma 2.1** (Proposition 1.11, [17]). *For every  $f \in \text{Lip}_b(X)$ ,  $|\nabla f|$  is a strong upper gradient in the sense that for each rectifiable curve  $\gamma : [0, L] \rightarrow X$  parametrized by arc-length we have*

$$|f(\gamma(L)) - f(\gamma(0))| \leq \int_0^L |\nabla f|(\gamma(s)) ds.$$

One may also verify that  $|\nabla f|$  satisfies the chain rule:

**Lemma 2.2.** *If  $f : X \rightarrow \mathbb{R}$  is Lipschitz in a neighborhood of  $x$  and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is differentiable at  $f(x)$ , then  $|\nabla(\phi \circ f)|(x) = |\phi'(f(x))| |\nabla f|(x)$ .*

Let  $\mathcal{B}_X$  denote the Borel  $\sigma$ -algebra of  $(X, d)$ , and  $\mathcal{P}(X)$  the set of Borel probability measures on  $X$ . We suppose we are given a Markov probability kernel  $P : X \times \mathcal{B}_X \rightarrow [0, 1]$ , and we denote by  $Pf, \mu P$  the usual action of  $P$  on bounded Borel functions  $f$  and Borel probability measures  $\mu$ , i.e.

$$Pf(x) := \int_X f(y)P(x, dy), \quad \mu P(A) := \int_X P(x, A)\mu(dx).$$

In some applications,  $P$  will be taken to be a Markov semigroup  $P_t$ , which may or may not be symmetric with respect to some reference measure. Our setting is similar to [25]. This is more general than the setting of [28], which only considered the symmetric semigroup  $P_t$  generated by the Cheeger energy with respect to the given gradient and a given reference measure.

Given  $\mu_0, \mu_1 \in \mathcal{P}(X)$ , the 2-Kantorovich–Wasserstein distance  $W_2(\mu_0, \mu_1)$  is defined as usual by

$$W_2(\mu_0, \mu_1)^2 := \inf_{\pi} \int d(x_0, x_1)^2 \mu(dx_0, dx_1), \tag{2.2}$$

the infimum taken over all couplings  $\pi \in \mathcal{P}(X \times X)$  of  $\mu_0, \mu_1$ . In particular, for point masses  $\mu_i = \delta_{x_i}$ , we have  $W_2(\delta_{x_0}, \delta_{x_1}) = d(x_0, x_1)$ . We let  $\mathcal{P}_2(X) \subset \mathcal{P}(X)$  denote the Wasserstein space of probability measures  $\mu$  having a finite second moment, i.e. for which  $\int_X d(x, x_0)^2 \mu(dx) < \infty$  for some (equivalently, all)  $x_0 \in X$ .

The 2-Hellinger distance is defined by

$$\text{He}_2(\mu_0, \mu_1)^2 := \int_X \left( \sqrt{\frac{d\mu_1}{dm}} - \sqrt{\frac{d\mu_0}{dm}} \right)^2 dm \tag{2.3}$$

where  $m$  is any measure such that  $\mu_0, \mu_1$  are both absolutely continuous with respect to  $m$ ; the definition is independent of  $m$ . Convergence in Hellinger distance is equivalent to convergence in total variation, and we have  $\text{He}_2(\mu_0, \mu_1)^2 \leq 2$  for all  $\mu_0, \mu_1 \in \mathcal{P}(X)$ , with equality iff  $\mu_0, \mu_1$  are mutually singular.

The stated hypotheses on the space  $X$  are meant to strike a balance between generality and convenience; one may certainly be able to weaken them in various ways. We have preferred to keep the emphasis on the techniques and their applications, rather than on stating the most general abstract theorems. In particular, in Section 6 we already depart from this setting to consider infinite-dimensional examples based on abstract Wiener space, where  $X$  is a separable Banach space (which is not proper), the test functions are taken to be the cylinder functions instead of all bounded Lipschitz functions, and the gradient  $\nabla$  is derived from the Malliavin gradient, whose norm is not an upper gradient with respect to the norm distance on  $X$ . This requires only trivial modifications to the arguments in the earlier sections; we discuss the details in Section 6.

### 3 Hellinger–Kantorovich distances and functional inequalities

#### 3.1 The dynamic dual formulation and basic properties

In this section, we consider the family of Hellinger–Kantorovich distances studied in [26, 27, 28]. We focus on the so-called dynamic dual formulation of these distances, in which they may be defined as the supremum of a difference of integrals over a class of subsolutions of a Hamilton–Jacobi-type equation in time and space variables. This idea is directly descended from a dynamic dual formulation of the Kantorovich–Wasserstein distance, introduced in [30]. Using this formulation of these distances, we will see that Poincaré and reverse Poincaré type inequalities for  $P$  lead directly to contraction results with respect to these distances (Theorem 3.4).

We study the Hellinger–Kantorovich distance via a slightly different parametrization which is more convenient for our purposes. As above, let  $\text{Lip}_b(X)$  denote the Banach space of all bounded Lipschitz functions on  $X$ . We remark for later use that for any finite measure  $\mu$  on  $X$ , we have  $\text{Lip}_b(X)$  dense in  $L^1(\mu)$ , and in particular that for any bounded Borel function  $f$  there is a sequence  $f_n \in \text{Lip}_b(X)$  with  $f_n \rightarrow f$   $\mu$ -a.e. and boundedly.

**Definition 3.1.** Let  $a, b \geq 0$ . We denote by  $\mathcal{A}_{a,b}$  the class of all functions  $\varphi = \varphi_s(x) \in C^1([0, 1], \text{Lip}_b(X))$  satisfying the differential inequality

$$\partial_s \varphi_s + a|\nabla \varphi_s|^2 + b\varphi_s^2 \leq 0. \tag{3.1}$$

Then for probability measures  $\mu_1, \mu_2 \in \mathcal{P}(X)$  we set

$$W_{a,b}(\mu_0, \mu_1) = \sup_{\varphi \in \mathcal{A}_{a,b}} \left[ \int_X \varphi_1 d\mu_1 - \int_X \varphi_0 d\mu_0 \right]. \tag{3.2}$$

To avoid confusion, we note that  $W_{a,b}$  itself is not a distance on (a subset of)  $\mathcal{P}(X)$ , but rather the square of a distance.

**Lemma 3.2.** The squared distances  $W_{a,b}$  satisfy the following basic properties:

1. If  $a \leq a'$  and  $b \leq b'$  then  $W_{a',b'} \leq W_{a,b}$ .
2. For any  $c > 0$ , we have  $W_{ca,cb} = c^{-1}W_{a,b}$ .
3. When  $b > 0$ , we have  $W_{a,b} = b^{-1}HK_{4a/b}^2$ , where  $HK$  is the Hellinger–Kantorovich distance as defined in [28, Definition 2.11].
4.  $W_{1/2,0} = \frac{1}{2}W_2^2$ , where  $W_2$  is the Kantorovich–Wasserstein 2-distance.
5.  $W_{0,1} = \text{He}_2^2$ , where  $\text{He}_2$  is the Hellinger 2-distance.

*Proof.* Item 1 is clear because when  $a \leq a'$  and  $b \leq b'$ , we have  $\mathcal{A}_{a',b'} \subseteq \mathcal{A}_{a,b}$ . Item 2 holds because  $\varphi \in \mathcal{A}_{ca,cb}$  if and only if  $c\varphi \in \mathcal{A}_{a,b}$ . For item 3, in the notation of [28, Eq. (39)] (see also [27, Section 8.4]), we have  $HK_\alpha^2 = W_{\alpha/4,1}$ , and the general statement follows using item 2. Item 4 can be found as Proposition 2.10 of [28], but goes back at least as far as [30, Section 3]; see also other references in [28].

Item 5 is almost Proposition 2.8 of [28], but there is a slight difference because our class of functions  $\mathcal{A}_{0,1}$  is required to be Lipschitz in space, whereas [28, Eq. (32)] uses functions which are only bounded. This is easily handled with a straightforward approximation argument, which we now give for completeness.

Let  $\mathcal{A}_{0,1}^B = \{\varphi \in C^1([0, 1], B(X)) : \partial_s \varphi_s + \varphi_s^2 \leq 0\}$ . The statement of [28, Proposition 2.8] is that

$$\text{He}_2^2(\mu_0, \mu_1) = \sup_{\psi \in \mathcal{A}_{0,1}^B} \int \psi_1 d\mu_1 - \int \psi_0 d\mu_0.$$

It is clear that  $W_{0,1}(\mu_0, \mu_1) \leq \text{He}_2^2(\mu_0, \mu_1)$ , since  $\mathcal{A}_{0,1} \subset \mathcal{A}_{0,1}^B$ . Now given  $\varphi \in \mathcal{A}_{0,1}$ , it is clear from a Grönwall-type argument that we must have  $\varphi_s \leq \varphi_0/(1 + s\varphi_0)$  for all  $s$ ; in particular we must have  $\varphi_0 > -1$  or else  $\varphi$  will have a singularity. Hence

$$W_{0,1}(\mu_0, \mu_1) = \sup \left\{ \int \frac{f}{1+f} d\mu_1 - \int f d\mu_0 : f \in \text{Lip}_b(X), f > -1 \right\}$$

and likewise

$$\text{He}_2^2(\mu_0, \mu_1) = \sup \left\{ \int \frac{f}{1+f} d\mu_1 - \int f d\mu_0 : f \in B_b(X), f > -1 \right\}.$$

Now the result follows by noting that for each  $f \in B_b(X)$  with  $f > -1$ , we can find a sequence of bounded Lipschitz functions  $f_n$  with  $f_n > -1$  and  $f_n \rightarrow f$  boundedly and  $(\mu_0 + \mu_1)$ -almost everywhere. We then have  $\int f_n d\mu_0 \rightarrow \int f d\mu_0$ , and since the sequence  $f_n/(1 + f_n)$  is bounded above by 1, Fatou’s lemma also gives  $\limsup_{n \rightarrow \infty} \int \frac{f_n}{1+f_n} d\mu_1 \geq \int \frac{f}{1+f} d\mu_1$ . From this we conclude that  $W_{0,1}(\mu_0, \mu_1) \geq \text{He}_2^2(\mu_0, \mu_1)$ .  $\square$

Thus, the (squared) distances  $W_{a,b}$  naturally interpolate between the Kantorovich–Wasserstein distance, which is perhaps the most familiar transportation distance, and the Hellinger distance, which metrizes convergence in total variation. As will be seen in the next subsection, this makes it valuable for obtaining inequalities relating these two distances.

**Proposition 3.3.** *If  $x_0, x_1 \in X$  and  $\delta_{x_0}, \delta_{x_1} \in \mathcal{P}(X)$  are the corresponding Dirac measures, then*

$$W_{a,b}(\delta_{x_0}, \delta_{x_1}) = \frac{1}{b} \left( 2 - 2 \cos \left( \frac{\sqrt{b}}{2\sqrt{a}} d(x_0, x_1) \wedge \frac{\pi}{2} \right) \right) \leq \frac{1}{4a} d(x_0, x_1)^2 \wedge \frac{2}{b}.$$

*Proof.* For  $a = \frac{1}{2}, b = 2$ , this is [27, Eq. (6.31)]; see also [27, Section 8] for the explanation that the LET distance corresponds to  $\mathbb{H}^2$ , which is our  $W_{1/2,2}$ . Other values of  $a$  can be handled by rescaling the distance  $d$ , and general values of  $a, b$  are then covered by Lemma 3.2 2.

We note, however, that the upper bound  $W_{a,b}(\delta_{x_0}, \delta_{x_1}) \leq \frac{1}{4a} d(x_0, x_1)^2 \wedge \frac{2}{b}$  can be shown much more easily, and is comparable to the exact expression up to a universal constant multiple (whose value is something like 1.2). The upper bound  $W_{a,b}(\mu_0, \mu_1) \leq \frac{2}{b}$  is essentially trivial, and can be seen, for instance, by noting

$$W_{a,b} \leq W_{0,b} = \frac{1}{b} W_{0,1} = \frac{1}{b} \text{He}_2^2$$

and that  $\text{He}_2^2(\mu_0, \mu_1) \leq 2$  for all  $\mu_0, \mu_1$ . The upper bound  $W_{a,b}(\delta_{x_0}, \delta_{x_1}) \leq \frac{1}{4a} d(x_0, x_1)^2$  can be seen in a similar way by comparing to the Kantorovich–Wasserstein distance  $W_{1/2,0}$ . But it can also be shown directly from the “dynamic dual” definition of  $W_{a,b}$ . We give the argument here, partly for comparison with Proposition 5.12 below.

Let  $a > 0$  and  $b \geq 0$ . Recall that  $(X, d)$  is assumed to be a complete length space, so there exists a constant speed geodesic  $\gamma : [0, 1] \rightarrow X$  joining  $x_0$  to  $x_1$ : namely,  $\gamma_0 = x_0, \gamma_1 = x_1$ , and  $d(\gamma_s, \gamma_t) = |s - t|d(x_0, x_1)$ . Since  $\nabla$  is a strong upper gradient, for any Lipschitz  $f : X \rightarrow \mathbb{R}$  we have that  $f \circ \gamma$  is absolutely continuous and  $|\frac{d}{ds} f(\gamma_s)| \leq |\nabla f|(\gamma_s)d(x_0, x_1)$ ; see [1, Definition 1.2.1]. Now using the chain rule, we have

$$\begin{aligned} \varphi_1(x_1) - \varphi_0(x_0) &= \int_0^1 \frac{d}{ds} \varphi_s(\gamma_s) ds \\ &\leq \int_0^1 [\partial_s \varphi_s(\gamma_s) + |\nabla \varphi_s|(\gamma_s)d(x_0, x_1)] ds \\ &\leq \int_0^1 [-a|\nabla \varphi_s|(\gamma_s)^2 - b\varphi_s(\gamma_s)^2 + |\nabla \varphi_s|(\gamma_s)d(x_0, x_1)] ds \\ &= \int_0^1 \left[ -a \left( |\nabla \varphi_s(\gamma_s)| - \frac{1}{2a} d(x_0, x_1) \right)^2 + \frac{1}{4a} d(x_0, x_1)^2 - b\varphi_s(\gamma_s)^2 \right] ds \end{aligned}$$

by completing the square. Discarding the two negative terms and taking the supremum over  $\varphi_s \in \mathcal{A}_{a,b}$ , we recover the desired bound.  $\square$

### 3.2 Functional inequalities

Thanks to the form of the dynamic dual definition for  $W_{a,b}$ , one obtains a direct implication between functional inequalities involving the gradient and contractions of Hellinger–Kantorovich distances. This was the key idea in the results of [28]; here we make the implication more explicit and collect several cases into a single statement.

**Theorem 3.4.** *Suppose that for some  $a > 0$  and  $b, \gamma, \delta \geq 0$ , the Markov operator  $P$  satisfies the functional inequality*

$$a|\nabla Pf|^2 + b(Pf)^2 \leq \gamma P|\nabla f|^2 + \delta P(f^2), \quad f \in \text{Lip}_b(X). \quad (3.3)$$

*Then we have the transportation distance contraction*

$$W_{\gamma,\delta}(\mu_0 P, \mu_1 P) \leq W_{a,b}(\mu_0, \mu_1), \quad \mu_0, \mu_1 \in \mathcal{P}(X). \quad (3.4)$$

*Proof.* Since  $a > 0$ , (3.3) implies that the Markovian operator  $P$  is a bounded operator on  $\text{Lip}_b(X)$ . Now let  $\varphi \in \mathcal{A}_{\gamma,\delta}$ . Since  $\varphi \in C^1([0, 1], \text{Lip}_b(X))$ , we have  $P\varphi_s \in C^1([0, 1], \text{Lip}_b(X))$  as well, and  $P\partial_s \varphi_s = \partial_s P\varphi_s$ . Hence

$$\begin{aligned} \partial_s P\varphi_s &= P\partial_s \varphi_s \leq P[-\gamma|\nabla \varphi_s|^2 - \delta \varphi_s^2] \\ &= -\gamma P|\nabla \varphi_s|^2 - \delta P(\varphi_s^2) \\ &\leq -a|\nabla P\varphi_s|^2 - b(P\varphi_s)^2 \end{aligned}$$

where we used the fact that  $P$  is positivity preserving, and the assumed inequality (3.3). This shows that  $P\varphi_s \in \mathcal{A}_{a,b}$ . Thus for  $\mu_0, \mu_1 \in \mathcal{P}(X)$  we have

$$\begin{aligned} W_{\gamma,\delta}(\mu_0 P, \mu_1 P) &= \sup_{\varphi \in \mathcal{A}_{\gamma,\delta}} \int_X P\varphi_1 \, d\mu_1 - \int_X P\varphi_0 \, d\mu_0 \\ &\leq \sup_{\psi \in \mathcal{A}_{a,b}} \int_X \psi_1 \, d\mu_1 - \int_X \psi_0 \, d\mu_0 \\ &= W_{a,b}(\mu_0, \mu_1) \end{aligned}$$

as desired. □

**Corollary 3.5.** *If  $P$  satisfies the gradient estimate  $|\nabla Pf|^2 \leq CP|\nabla f|^2$  for some  $C$ , then for any  $b \geq 0$  we have*

$$W_{1,b}(\mu_0 P, \mu_1 P) \leq W_{C,b}(\mu_0, \mu_1).$$

*In particular, taking  $b = 0$  we recover the Kuwada-type duality*

$$W_2(\mu_0 P, \mu_1 P)^2 \leq CW_2(\mu_0, \mu_1)^2.$$

The case  $C = 1$  of Corollary 3.5 is [27, Theorem 8.24], and when additionally  $b = 0$  it reduces to [25, Proposition 3.7].

*Proof.* Noting that  $(Pf)^2 \leq P(f^2)$  by Jensen’s inequality, the gradient estimate  $|\nabla Pf|^2 \leq CP|\nabla f|^2$  implies that (3.3) holds with  $a = 1, \gamma = C, \delta = b$ . □

**Remark 3.6.** Note that, conversely, the estimate

$$W_2(\mu_0 P, \mu_1 P)^2 \leq CW_2(\mu_0, \mu_1)^2$$

implies the gradient estimate  $|\nabla Pf|^2 \leq CP|\nabla f|^2$ ; see [25].

**Theorem 3.7.** *Let  $C > 0$ . The following are equivalent:*

1. *The reverse Poincaré inequality*

$$|\nabla Pf|^2 \leq C(P(f^2) - (Pf)^2), \quad f \in \text{Lip}_b(X). \quad (\text{RPI})$$

2. *The Hellinger–Kantorovich contraction*

$$\text{He}_2(\mu_0 P, \mu_1 P)^2 \leq \text{HK}_{4/C}(\mu_0, \mu_1)^2 \leq \frac{C}{4} W_2(\mu_0, \mu_1)^2, \quad \mu_0, \mu_1 \in \mathcal{P}(X). \quad (\text{HKC})$$

3. The Harnack type inequality

$$Pf(x) \leq Pf(y) + \sqrt{C}d(x, y)\sqrt{P(f^2)(x)}, \quad x, y \in X, f \in B_b(X), f \geq 0. \quad (\text{HPI})$$

We point out, for future use, that the reverse Poincaré inequality (RPI) is equivalent to the apparently weaker form

$$|\nabla Pf|^2 \leq CP(f^2), \quad f \in \text{Lip}_b(X) \quad (3.5)$$

Indeed, to see that (3.5) self-improves to (RPI), suppose  $f \in \text{Lip}_b(X)$ , fix an arbitrary  $x \in X$ , and let  $g(y) = f(y) - Pf(x)$ . Then apply (3.5) to  $g$  and evaluate at  $x$ .

Before we give the proof of the theorem, we state a lemma interesting in itself.

**Lemma 3.8.** For any  $f \in B_b(X)$ , and  $x, y \in X$ ,

$$|Pf(x) - Pf(y)|^2 \leq 2\text{He}_2(\delta_x P, \delta_y P)^2 (P(f^2)(x) + P(f^2)(y)).$$

*Proof.* Let  $m$  be a Borel measure such that both  $\delta_x P$  and  $\delta_y P$  are absolutely continuous with respect to  $m$ . We denote

$$P_m(x, \cdot) = \frac{d\delta_x P}{dm}, \quad P_m(y, \cdot) = \frac{d\delta_y P}{dm}.$$

We have

$$\begin{aligned} & |Pf(x) - Pf(y)| \\ &= \left| \int P_m(x, z)f(z) dm(z) - \int P_m(y, z)f(z) dm(z) \right| \\ &= \left| \int \sqrt{P_m(x, z)}\sqrt{P_m(x, z)}f(z) dm(z) - \int \sqrt{P_m(y, z)}\sqrt{P_m(y, z)}f(z) dm(z) \right| \\ &\leq \left| \int \sqrt{P_m(x, z)}\sqrt{P_m(x, z)}f(z) dm(z) - \int \sqrt{P_m(x, z)}\sqrt{P_m(y, z)}f(z) dm(z) \right| \\ &\quad + \left| \int \sqrt{P_m(x, z)}\sqrt{P_m(y, z)}f(z) dm(z) - \int \sqrt{P_m(y, z)}\sqrt{P_m(y, z)}f(z) dm(z) \right| \\ &\leq \int \left| \sqrt{P_m(x, z)} - \sqrt{P_m(y, z)} \right| \sqrt{P_m(x, z)}f(z) dm(z) \\ &\quad + \int \left| \sqrt{P_m(x, z)} - \sqrt{P_m(y, z)} \right| \sqrt{P_m(y, z)}f(z) dm(z) \end{aligned}$$

Therefore, by the Cauchy-Schwarz inequality,

$$\begin{aligned} |Pf(x) - Pf(y)|^2 &\leq \text{He}_2(\delta_x P, \delta_y P)^2 \left( \sqrt{P(f^2)(x)} + \sqrt{P(f^2)(y)} \right)^2 \\ &\leq 2\text{He}_2(\delta_x P, \delta_y P)^2 (P(f^2)(x) + P(f^2)(y)). \quad \square \end{aligned}$$

We are now ready for the proof of Theorem 3.7.

*Proof of Theorem 3.7.* (RPI)  $\implies$  (HKC): This follows from Theorem 3.4 with  $\gamma = 0$  and  $b = \delta = C$ . We note again that this direction is the essence of [28, Theorem 5.4].

(HKC)  $\implies$  (RPI): Assume that

$$\text{He}_2(\mu_0 P, \mu_1 P)^2 \leq \frac{C}{4} W_2(\mu_0, \mu_1)^2.$$

Then, for every  $x, y \in X$ ,

$$\text{He}_2(\delta_x P, \delta_y P)^2 \leq \frac{C}{4} d(x, y)^2.$$

Therefore, from Lemma 3.8 one deduces

$$|Pf(x) - Pf(y)|^2 \leq \frac{C}{2}d(x, y)^2(P(f^2)(x) + P(f^2)(y)). \tag{3.6}$$

Similarly, one has

$$\begin{aligned} |P(f^2)(x) - P(f^2)(y)|^2 &\leq \frac{C}{2}d(x, y)^2(P(f^4)(x) + P(f^4)(y)) \\ &\leq Cd(x, y)^2\|f\|_\infty^4, \end{aligned}$$

which implies that  $P(f^2)$  is a continuous function. Since

$$|\nabla Pf|(x) = \lim_{r \rightarrow 0} \sup_{0 < d(x, y) \leq r} \frac{|Pf(x) - Pf(y)|}{d(x, y)},$$

we may divide both sides of (3.6) by  $d(x, y)^2$  and let  $y \rightarrow x$  to obtain

$$|\nabla Pf|(x) \leq CP(f^2)(x)$$

which, as noted above, self-improves to (RPI).

(RPI)  $\implies$  (HPI) and (HPI)  $\implies$  (RPI): The proof follows from Proposition 1.3 in [36] so we omit it for conciseness.  $\square$

## 4 Applications to convergence to equilibrium

In this section, we focus on the applications of the transportation type inequalities proven in Theorem 3.4 as a powerful tool to prove convergence to equilibrium for Markov semigroups. We will mostly focus on the applications of the transportation inequality

$$\text{He}_2(\mu_0 P, \mu_1 P)^2 \leq \frac{C}{4}W_2(\mu_0, \mu_1)^2,$$

which, according to Theorem 3.7, comes from the reverse Poincaré inequality

$$|\nabla Pf|^2 \leq C(P(f^2) - (Pf)^2).$$

The original Kuwada duality proved in Corollary 3.5 relating the transportation inequality

$$W_2^2(\mu_0 P, \mu_1 P)^2 \leq CW_2^2(\mu_0, \mu_1)^2$$

to the gradient bound

$$|\nabla Pf|^2 \leq CP(|\nabla f|^2)$$

was already illustrated as a tool to prove convergence to equilibrium in [3], so we will spend less time on it. Also, our examples will be finite dimensional, though applications could be given in an infinite dimensional framework as in Section 6. In particular, applications to stochastic partial differential equations might be the object of a future work.

### 4.1 Diffusions with $\Gamma_2 \geq 0$

In this section, as an illustration of our general results, we first show how to recover the results of [28]. Let  $\Delta$  be a locally subelliptic diffusion operator (see Section 1.2 in [4] for a definition of local subellipticity) on a smooth manifold  $M$ . For smooth functions  $f, g : M \rightarrow \mathbb{R}$ , we can define the *carré du champ* operator as the symmetric first-order bilinear differential form given by:

$$\Gamma(f, g) := \frac{1}{2}(\Delta(fg) - f\Delta g - g\Delta f). \tag{4.1}$$

We write  $\Gamma(f)$  for  $\Gamma(f, f)$ . (When  $\Delta$  is the Laplacian on  $\mathbb{R}^n$  or on a Riemannian manifold, we have  $\Gamma(f) = |\nabla f|^2$ .) We assume that  $\Delta$  is symmetric with respect to some smooth measure  $\mu$  (not necessarily finite), which means that for every pair of smooth and compactly supported functions  $f, g \in C_0^\infty(M)$ ,

$$\int_M g \Delta f \, d\mu = \int_M f \Delta g \, d\mu.$$

There is an intrinsic distance associated to the operator  $\Delta$  that we now describe. An absolutely continuous curve  $\gamma : [0, T] \rightarrow M$  is said to be subunit for the operator  $L$  if for every smooth function  $f : M \rightarrow \mathbb{R}$  we have  $|\frac{d}{dt} f(\gamma(t))| \leq \sqrt{(\Gamma f)(\gamma(t))}$ . We then define the subunit length of  $\gamma$  as  $\ell_s(\gamma) = T$ . Given  $x, y \in M$ , we indicate then with

$$S(x, y) := \{\gamma : [0, T] \rightarrow M \mid \gamma \text{ is subunit for } \Gamma, \gamma(0) = x, \gamma(T) = y\}$$

and assume that  $S(x, y) \neq \emptyset$  for every  $x, y \in M$ . For instance, if  $L$  is an elliptic operator or if  $L$  is a sum of squares operator that satisfies Hörmander’s condition, then this assumption is satisfied. Under this assumption,

$$d(x, y) := \inf\{\ell_s(\gamma) \mid \gamma \in S(x, y)\} \tag{4.2}$$

defines a distance on  $M$  and  $(M, d)$  is by construction a length space. The carré du champ operator yields a strong upper gradient structure on  $(M, d)$  and from Theorem 1.12 in [4] one has

$$d(x, y) = \sup\{|f(x) - f(y)|, f \in C^\infty(M), \|\Gamma(f)\|_\infty \leq 1\}, \quad x, y \in M.$$

We assume that the metric space  $(M, d)$  is complete. In that case, from Propositions 1.20 and 1.21 in [4], the operator  $\Delta$  is essentially self-adjoint on  $C_0^\infty(M)$ . The semigroup in  $L^2(M, \mu)$  generated by  $\Delta$  will be denoted by  $(P_t)_{t \geq 0}$ . The Bakry  $\Gamma_2$  operator is defined as

$$\Gamma_2(f, g) = \frac{1}{2} (\Delta(\Gamma(f, g)) - \Gamma(f, \Delta g) - \Gamma(g, \Delta f)), \quad f, g \in C^\infty(M).$$

**Theorem 4.1.** *Assume that for every  $f \in C^\infty(M)$ ,  $\Gamma_2(f, f) \geq 0$ . Then, for every  $\nu_1, \nu_2 \in \mathcal{P}_2(M)$  and  $t > 0$ ,*

$$\text{He}_2(\nu_1 P_t, \nu_2 P_t)^2 \leq \frac{1}{8t} W_2(\nu_1, \nu_2)^2.$$

*Therefore, if the invariant measure  $\mu$  is a probability measure which belongs to  $\mathcal{P}_2(M)$ , then for every  $x \in M$  and  $t > 0$ ,*

$$\text{He}_2(\delta_x P_t, \mu)^2 \leq \frac{1}{8t} W_2(\delta_x, \mu)^2$$

*and when  $t \rightarrow +\infty$ ,  $\delta_x P_t$  converges to  $\mu$  in total variation for every  $x \in M$ .*

*Proof.* It follows from Bakry-Émery calculus (see for instance [2, Proposition 3.3 (5)]) that since  $\Gamma_2 \geq 0$  one has the following gradient bound that holds for bounded and Lipschitz functions  $f$ ,

$$\Gamma(P_t f) \leq \frac{1}{2t} (P_t(f^2) - (P_t f)^2), \quad t > 0$$

which yields the conclusion thanks to Theorem 3.7. □

**Example 4.2.** An example where the theorem applies is the case where  $\Delta$  is the Laplace-Beltrami operator on a complete Riemannian manifold of non-negative Ricci curvature. In that case, the invariant measure  $\mu$  is the Riemannian volume measure, and the assumption  $\Gamma_2 \geq 0$  is equivalent to the condition that the Ricci curvature of  $M$  is non-negative.

**Remark 4.3.** More generally, if  $\Delta$  is taken to be the operator generated by the Cheeger energy as in [28], so that  $\Gamma = \text{Ch}$ , then the hypothesis of Theorem 4.1 essentially asks for  $X$  to be an  $RCD(0, \infty)$  space, and the conclusion is included in [28, Theorem 5.2]. Indeed, the  $\Gamma_2 \geq K$  condition was already the key idea of the results of [28]. Our purpose in stating Theorem 4.1 is to draw attention to the consequence that  $\delta_x P_t$  converges in total variation to its equilibrium measure, at a rate no slower than  $1/\sqrt{t}$ .

**Remark 4.4.** If  $\Gamma_2 \geq a$ , then, Bakry–Émery calculus also yields the gradient bound

$$\Gamma(P_t f) \leq e^{-2at} P_t(\Gamma(f)).$$

which therefore implies from Theorem 3.4 the following contraction property in the  $W_2$  distance:

$$W_2(\nu_1 P_t, \nu_2 P_t)^2 \leq e^{-2at} W_2(\nu_1, \nu_2)^2,$$

This appears in [25] and [32].

#### 4.2 Subelliptic operators

The assumption  $\Gamma_2 \geq 0$  requires some form of ellipticity of  $\Delta$ . In order to generalize the previous theorem to truly subelliptic operators, one can make use of the generalized  $\Gamma$ -calculus developed in [8, 5]. In addition to the carré du champ form  $\Gamma$  defined in (4.1), we assume that  $M$  is endowed with another smooth symmetric bilinear differential form, indicated with  $\Gamma^Z$ , satisfying for  $f, g \in C^\infty(M)$

$$\Gamma^Z(fg, h) = f\Gamma^Z(g, h) + g\Gamma^Z(f, h),$$

and  $\Gamma^Z(f) = \Gamma^Z(f, f) \geq 0$ . Let us assume that:

(H.1) There exists an increasing sequence  $h_k \in C_0^\infty(M)$  such that  $h_k \nearrow 1$  on  $M$ , and

$$\|\Gamma(h_k)\|_\infty + \|\Gamma^Z(h_k)\|_\infty \rightarrow 0, \text{ as } k \rightarrow \infty.$$

(H.2) For any  $f \in C^\infty(M)$  one has

$$\Gamma(f, \Gamma^Z(f)) = \Gamma^Z(f, \Gamma(f)).$$

Let us then consider

$$\Gamma_2^Z(f, g) = \frac{1}{2} [\Delta \Gamma^Z(f, g) - \Gamma^Z(f, \Delta g) - \Gamma^Z(g, \Delta f)]. \tag{4.3}$$

As for  $\Gamma$  and  $\Gamma^Z$ , we will freely use the notations  $\Gamma_2(f) = \Gamma_2(f, f)$ ,  $\Gamma_2^Z(f) = \Gamma_2^Z(f, f)$ .

**Theorem 4.5.** Let  $\rho_1 \geq 0, \rho_2 > 0$  and  $\kappa > 0$ . Assume that for every  $f \in C^\infty(M)$  and  $\nu > 0$ ,

$$\Gamma_2(f) + \nu \Gamma_2^Z(f) \geq \left(\rho_1 - \frac{\kappa}{\nu}\right) \Gamma(f) + \rho_2 \Gamma^Z(f). \tag{4.4}$$

Then, for every  $\nu_1, \nu_2 \in \mathcal{P}_2(M)$  and  $t > 0$ ,

$$\text{He}_2(\nu_1 P_t, \nu_2 P_t)^2 \leq \frac{1}{8t} \left(1 + \frac{2\kappa}{\rho_2}\right) W_2(\nu_1, \nu_2)^2.$$

Therefore, if the invariant measure  $\mu$  is a probability measure which belongs to  $\mathcal{P}_2(M)$ , then for every  $x \in M$  and  $t > 0$ ,

$$\text{He}_2(\delta_x P_t, \mu)^2 \leq \frac{1}{8t} \left(1 + \frac{2\kappa}{\rho_2}\right) W_2(\delta_x, \mu)^2$$

and when  $t \rightarrow +\infty$ ,  $\delta_x P_t$  converges to  $\mu$  in total variation for every  $x \in M$ .

*Proof.* It follows from Proposition 3.2 in [5] that

$$\Gamma(P_t f) \leq \frac{1}{2t} \left( 1 + \frac{2\kappa}{\rho_2} \right) (P_t(f^2) - (P_t f)^2)$$

and thus the conclusion follows from Theorem 3.7. □

**Example 4.6.** An example where this theorem applies is the case where  $\Delta$  is the sub-Laplacian operator on a compact H-type sub-Riemannian manifold, see [10]. In that case, the invariant measure  $\mu$  is again the Riemannian volume measure and the assumption (4.4) is equivalent to the fact that the horizontal Ricci curvature of  $M$  is non-negative. This applies for instance to the sub-Laplacian on the special unitary group  $SU(2)$ , as well as to compact quotients of the Heisenberg group  $\mathbb{H}^3$ .

### 4.3 Non symmetric Ornstein–Uhlenbeck semigroups on Carnot groups

In this section, we show that the method also applies to hypoelliptic and non-symmetric diffusion operators. In particular we prove a quantitative rate of convergence for the non-symmetric Ornstein–Uhlenbeck semigroup on a Carnot group.

A Carnot group of step (or depth)  $N$  is a simply connected Lie group  $G$  whose Lie algebra can be written

$$\mathfrak{g} = \mathcal{V}_1 \oplus \dots \oplus \mathcal{V}_N,$$

where

$$[\mathcal{V}_i, \mathcal{V}_j] = \mathcal{V}_{i+j}$$

and

$$\mathcal{V}_s = 0, \text{ for } s > N.$$

From the above properties, it is of course seen that Carnot groups are nilpotent. The number

$$\mathfrak{D} = \sum_{i=1}^N i \dim \mathcal{V}_i$$

is called the homogeneous dimension of  $G$ . On  $\mathfrak{g}$  we can consider the family of linear operators which act by scalar multiplication  $t^i$  on  $\mathcal{V}_i$ . These operators are Lie algebra automorphisms, due to the grading, and induce Lie group automorphisms  $\Delta_t : G \rightarrow G$  which are called the canonical dilations of  $G$ . It is easily seen that there exists on  $G$  a complete and smooth vector field  $D$  such that

$$\Delta_t = e^{(\ln t)D}.$$

This vector field  $D$  is called the dilation vector field on  $G$ . If  $X$  is a left (or right) invariant smooth horizontal vector field on  $G$ , we have for every  $f \in C^\infty(G)$ , and  $t \geq 0$ ,

$$X(f \circ \Delta_t) = tXf \circ \Delta_t.$$

Let us now pick a basis  $V_1, \dots, V_d$  of the vector space  $\mathcal{V}_1$ . The vectors  $V_i$  can be seen as left invariant vector fields on  $G$ . In the sequel, these vector fields shall still be denoted by  $V_1, \dots, V_d$ . The left invariant sub-Laplacian on  $G$  is the operator:

$$\sum_{i=1}^d V_i^2.$$

It is essentially self-adjoint on the space of smooth and compactly supported functions with respect to the Haar measure  $\mu$  of  $G$ . The heat semigroup  $(P_t)_{t \geq 0}$  on  $G$  generated by the sub-Laplacian, defined through the spectral theorem, is then a Markov semigroup.

There are two different operators on  $\mathbb{G}$  which are both commonly referred to as Ornstein–Uhlenbeck operators; see [29] for a thorough comparison of the two types and their properties. We are interested here in the *non-symmetric* Ornstein Uhlenbeck operator defined by

$$L = \sum_{i=1}^d V_i^2 - \alpha D$$

where  $\alpha > 0$ . This operator generates a Markov semigroup  $(Q_t)_{t \geq 0}$  which is given by the Mehler formula

$$Q_t f = P_{\frac{1-e^{-\alpha t}}{\alpha}}(f \circ \Delta_{e^{-\alpha t}}), \quad t \geq 0.$$

It is clear that the probability measure  $\delta_e P_{1/\alpha}$  is invariant by  $Q_t$  where  $e$  denotes the identity element in  $\mathbb{G}$ . Note that  $\delta_e P_{1/\alpha}$  is the heat kernel measure started from  $e$  in  $\mathbb{G}$ . From known heat kernel estimates in Carnot groups (see [31]), one easily sees that the invariant measure  $\delta_e P_{1/\alpha} \in \mathcal{P}_2(\mathbb{G})$ . The next theorem proves exponentially fast convergence to equilibrium for  $Q_t$  with a quantitative rate.

**Theorem 4.7.** *For every  $x \in \mathbb{G}$  and  $t > 0$ ,*

$$\text{He}_2(\delta_x Q_t, \delta_e P_{1/\alpha})^2 \leq \frac{\mathfrak{D}\alpha e^{-2\alpha t}}{2(1 - e^{-\alpha t})} W_2(\delta_x, \delta_e P_{1/\alpha})^2.$$

*Proof.* We denote by  $\nabla_{\mathcal{H}}$  the horizontal gradient on  $\mathbb{G}$  given by

$$\nabla_{\mathcal{H}} f = \sum_{i=1}^d (V_i f) V_i.$$

The following reverse Poincaré inequality was proved in [6]:

$$|\nabla_{\mathcal{H}} P_t f|^2 \leq \frac{\mathfrak{D}}{2t} (P_t(f^2) - (P_t f)^2).$$

Since  $Q_t f = P_{\frac{1-e^{-\alpha t}}{\alpha}}(f \circ \Delta_{e^{-\alpha t}})$ , one has

$$\nabla_{\mathcal{H}} Q_t f = e^{-\alpha t} \nabla_{\mathcal{H}} P_{\frac{1-e^{-\alpha t}}{\alpha}}(f \circ \Delta_{e^{-\alpha t}}).$$

Thus,

$$|\nabla_{\mathcal{H}} Q_t f|^2 \leq \frac{\mathfrak{D}\alpha e^{-2\alpha t}}{2(1 - e^{-\alpha t})} P_{\frac{1-e^{-\alpha t}}{\alpha}}((f \circ \Delta_{e^{-\alpha t}})^2) = \frac{\mathfrak{D}\alpha e^{-2\alpha t}}{2(1 - e^{-\alpha t})} Q_t(f^2)$$

and the conclusion follows as before from Theorem 3.7. □

**Remark 4.8.** The above proof and [6] show that if  $\mathbb{G}$  is an H-type group, then the constant  $\frac{\mathfrak{D}\alpha e^{-2\alpha t}}{2(1 - e^{-\alpha t})}$  can be improved into  $\frac{\mathfrak{D}\alpha e^{-2\alpha t}}{2d(1 - e^{-\alpha t})}$ .

#### 4.4 Langevin type dynamics driven by Lévy processes

In this subsection, we work in the space  $X = \mathbb{R}^n$  with its usual Euclidean distance and gradient.

Let  $(N_t)_{t \geq 0}$  be a Lévy process in  $\mathbb{R}^n$ , i.e. a càdlàg stochastic process with stationary and independent increments. We assume that  $N_0 = 0$  a.s. and that for every  $T > 0$ ,  $\mathbb{E} \left( \sup_{t \in [0, T]} |N_t|^2 \right) < +\infty$ . In  $\mathbb{R}^n$ , we consider the following stochastic differential equation with additive noise:

$$dX_t^x = -\nabla U(X_t^x) dt + dN_t, \quad X_0^x = x \in \mathbb{R}^n, \tag{4.5}$$

where  $U : \mathbb{R}^n \rightarrow \mathbb{R}$  is a  $C^2$  function. For simplicity, we assume that  $\nabla U$  is a Lipschitz function, so that it is easily proved that (4.5) has a unique solution for any  $x \in \mathbb{R}^n$  which moreover satisfies for every  $T > 0$ ,  $\mathbb{E} \left( \sup_{t \in [0, T]} |X_t^x|^2 \right) < +\infty$ . For  $t \geq 0$ , we denote by  $P_t$  the Markov kernel defined by

$$P_t f(x) = \mathbb{E}(f(X_t^x)),$$

so that  $P_t(x, A) = \mathbb{P}(X_t^x \in A)$ . It is a contraction semigroup in  $L^\infty(\mathbb{R}^n)$ , and from the square integrability we have that for every  $\mu \in \mathcal{P}_2(\mathbb{R}^n)$  and  $t \geq 0$ ,  $\mu P_t \in \mathcal{P}_2(\mathbb{R}^n)$ .

#### 4.4.1 Convergence to equilibrium in the Kantorovich–Wasserstein distance

Let  $\nabla^2 U$  denote the Hessian of  $U$ .

**Theorem 4.9.** *Assume that there exists  $a > 0$  such that  $\nabla^2 U \geq a$  (uniformly in the sense of quadratic forms). Then, there exists a unique probability measure  $\mu$  in the Wasserstein space  $\mathcal{P}_2(\mathbb{R}^n)$  such that for every  $t \geq 0$ ,  $\mu P_t = \mu$ . Moreover, for every  $t \geq 0$ , and  $\nu \in \mathcal{P}_2(\mathbb{R}^n)$  one has,*

$$W_2(\nu P_t, \mu)^2 \leq e^{-2at} W_2(\nu, \mu)^2.$$

*Proof.* We proceed in several steps.

**Step 1:** *Proving the Bakry–Émery type estimate.*

Let  $J_t = \frac{\partial X_t^x}{\partial x}$  be the first variation process associated with equation (4.5). Since  $P_t f(x) = \mathbb{E}(f(X_t^x))$ , by the chain rule we have

$$\nabla P_t f(x) = \mathbb{E}(J_t^* \nabla f(X_t^x)).$$

Therefore, by the Cauchy–Schwarz inequality,

$$|\nabla P_t f(x)|^2 \leq \mathbb{E}(|J_t^*|^2) \mathbb{E}(|\nabla f(X_t^x)|^2).$$

Since  $\mathbb{E}(|\nabla f(X_t^x)|^2) = P_t(|\nabla f|^2)(x)$ , we are left to estimate  $\mathbb{E}(|J_t^*|^2)$ . To this end, we observe that

$$dJ_t = -\nabla^2 U(X_t^x) J_t dt, \quad J_0 = \mathbf{Id}_{\mathbb{R}^n}. \tag{4.6}$$

From the assumption  $\nabla^2 U \geq a$  this yields

$$|J_t^*|^2 \leq e^{-2at}.$$

One concludes  $\mathbb{E}(|J_t^*|^2) \leq e^{-2at}$  and therefore

$$|\nabla P_t f(x)|^2 \leq e^{-2at} P_t(|\nabla f|^2)(x).$$

By Kuwada duality (Corollary 3.5), this yields that for every  $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ ,

$$W_2(\nu_0 P_t, \nu_1 P_t)^2 \leq e^{-2at} W_2(\nu_0, \nu_1)^2. \tag{4.7}$$

**Step 2:** *Proving the existence and uniqueness of the invariant measure.*

Let  $t > 0$ . Thanks to (4.7), the map  $\nu \rightarrow \nu P_t$  is a contraction from  $\mathcal{P}_2(\mathbb{R}^n)$  into itself. Since  $\mathcal{P}_2(\mathbb{R}^n)$  is a complete metric space, one deduces that it admits a unique fixed point; call it  $\mu_t$ . We have then for every  $t > 0$  that  $\mu_t P_t = \mu_t$ . Composing with  $P_s$  yields  $\mu_t P_t P_s = \mu_t P_s$ . Since  $P_t$  is a semigroup, one has  $P_t P_s = P_s P_t$ . Therefore,  $\mu_t P_s P_t = \mu_t P_s$  which means that  $\mu_t P_s$  is invariant for  $P_t$ . By uniqueness this implies  $\mu_t P_s = \mu_t$ . Using now the uniqueness of the invariant measure for  $P_s$  yields  $\mu_t = \mu_s$ . As a conclusion,  $\mu_t$  is independent of  $t$ . We can call it  $\mu$ .

**Step 3:** *Concluding.*

Using (4.7) with  $\nu_0 = \nu$  and  $\nu_1 = \mu$  yields the expected result. □

#### 4.4.2 Convergence to equilibrium in the Hellinger distance

Our next application shows that in the diffusion case one can prove convergence to equilibrium in the Langevin dynamics without assuming coercivity of the Hessian of the potential (i.e.  $\nabla^2 U \geq a > 0$ ). The price to pay is a convergence speed which is not exponential but polynomial. We now assume that  $(N_t)_{t \geq 0}$  is a Brownian motion in  $\mathbb{R}^n$ . In that case, the invariant measure of (4.5) is known explicitly, and is given up to a possible normalization constant by  $e^{-U(x)} dx$ .

**Theorem 4.10.** *Assume that the normalized invariant measure  $d\mu = \frac{1}{Z} e^{-U(x)} dx$  is a probability measure with a finite second moment and that  $\nabla^2 U \geq 0$  ( $U$  convex). Then, for every  $x \in \mathbb{R}^n$*

$$\text{He}_2(\delta_x P_t, \mu)^2 \leq \frac{1}{4t} W_2(\delta_x, \mu)^2.$$

In particular,  $X_t^x$  converges in total variation to  $\mu$  when  $t \rightarrow +\infty$ .

*Proof.* From the Bismut–Elworthy–Li formula [11, 20], we have for every  $v \in \mathbb{R}^n$

$$\langle \nabla P_t f(x), v \rangle = \frac{1}{t} \mathbb{E} \left( f(X_t^x) \int_0^t (J_s v) dN_s \right),$$

where, as before,  $J_t = \frac{\partial X_t^x}{\partial x}$  is the first variation process associated with equation (4.5). From the Cauchy–Schwarz inequality, and the fact that  $\nabla^2 U \geq 0$  implies  $|J_t| \leq 1$  a.s., one has

$$\begin{aligned} \mathbb{E} \left( f(X_t^x) \int_0^t (J_s v) dN_s \right)^2 &\leq \mathbb{E} (f(X_t^x)^2) \mathbb{E} \left( \left( \int_0^t (J_s v) dN_s \right)^2 \right) \\ &\leq \mathbb{E} (f(X_t^x)^2) \mathbb{E} \left( \int_0^t |J_s v|^2 ds \right) \\ &\leq t |v|^2 \mathbb{E} (f(X_t^x)^2) = t |v|^2 P_t(f^2)(x). \end{aligned}$$

One concludes that for every  $v \in \mathbb{R}^n$ ,

$$\langle \nabla P_t f(x), v \rangle^2 = \frac{1}{t} |v|^2 P_t(f^2)(x).$$

This yields

$$|\nabla P_t f(x)|^2 \leq \frac{1}{t} P_t(f^2)(x)$$

which is of the form (3.5). As noted before, this self-improves to (RPI) and thus we have the expected result by Theorem 3.7.  $\square$

**Remark 4.11.** Theorem 4.10 might also be proven using Theorem 4.1 above. However, we wanted to illustrate the use of the Bismut–Elworthy–Li formula as a tool to prove reverse Poincaré inequalities.

## 5 Rényi-type divergences and functional inequalities

### 5.1 The dynamic dual formulation and basic properties

The notions discussed in the previous section can be modified to give a dynamic dual formulation of a family of “entropic” divergences on  $\mathcal{P}(X) \times \mathcal{P}(X)$ , which we will denote by  $T_{a,b}$ . In the same way that the (squared) distances  $W_{a,b}$  included the Hellinger distance, the  $T_{a,b}$  family will include the Rényi divergence (in a different normalization); and where contractions of  $W_{a,b}$  were equivalent to reverse Poincaré inequalities, we will show (Theorem 5.15) that contractions of  $T_{a,b}$  are equivalent to

reverse logarithmic Sobolev inequalities, Wang-type Harnack inequalities, and integrated Harnack inequalities.

**Definition 5.1.** Let  $a, b \geq 0$ . We denote by  $\mathcal{E}_{a,b}$  the class of all positive functions  $\varphi \in C^1([0, 1], \text{Lip}_b(X))$ , bounded and bounded away from 0, satisfying the differential inequality

$$\partial_s \varphi_s + a \varphi_s |\nabla \ln \varphi_s|^2 + b \varphi_s \ln \varphi_s \leq 0. \tag{5.1}$$

Then for probability measures  $\mu_1, \mu_2 \in \mathcal{P}(X)$  we set

$$T_{a,b}(\mu_0, \mu_1) = \sup_{\varphi \in \mathcal{E}_{a,b}} \left[ \int_X \varphi_1 d\mu_1 - \int_X \varphi_0 d\mu_0 \right] \tag{5.2}$$

(We note here a slight abuse of terminology. The functions  $T_{a,b}$  as defined above do not actually satisfy the definition of a statistical divergence, but they have renormalized versions  $\tilde{T}_{a,b}$ , defined in (5.10) below, which are divergences as shown in Proposition 5.8. However, it will be simpler in most cases to work with  $T_{a,b}$  than with  $\tilde{T}_{a,b}$ , and we will continue to use the term “divergence” for either of the two when no confusion will result.)

**Remark 5.2.** By writing  $\varphi_s = e^{\psi_s}$ , we could formulate Definition 5.1 instead as

$$T_{a,b}(\mu_0, \mu_1) = \sup \left\{ \int_X e^{\psi_1} d\mu_1 - \int_X e^{\psi_0} d\mu_0 : \partial_s \psi_s + a |\nabla \psi_s|^2 + b \psi_s \leq 0 \right\}.$$

In this notation the relevant Hamilton–Jacobi differential inequality more closely resembles (3.1), with a 0th order term which is now linear instead of quadratic. However, for comparison to functional inequalities as in Theorem 5.14 below, the original formulation of Definition 5.1 will be more convenient.

**Notation 5.3.** For  $b > 0$ , let  $p = e^b$ , let  $q = p/(p - 1)$  be the conjugate exponent of  $p$ , and set

$$C_b := \frac{1}{q} p^{-q/p} = \frac{1}{q} p^{1-q}. \tag{5.3}$$

We will use this notation throughout the rest of the paper when discussing  $T_{a,b}$ . The reader should keep in mind  $p, q, C_b$  depend implicitly on  $b$ .

The following elementary inequality will be used several times.

**Lemma 5.4.** Let  $b > 0$  and define  $p, q, C_b$  as in Notation 5.3. Suppose  $z, w > 0$ . Then for all  $x > 0$  we have

$$x^{1/p} z - x w \leq C_b \frac{z^q}{w^{q-1}}$$

with equality when  $x = \left(\frac{z}{pw}\right)^q$ .

*Proof.* We can suppose without loss of generality that  $w = 1$ , for applying this case with  $z$  replaced by  $z/w$  and multiplying through by  $w$  yields the general case.

Using Young’s inequality for products  $uv \leq \frac{1}{p} u^p + \frac{1}{q} v^q$ , we have

$$x^{1/p} z - x = (px)^{1/p} (p^{-1/p} z) - x \leq x + \frac{1}{q} p^{-q/p} z^q - x = C_b z^q.$$

Young’s inequality becomes equality precisely when  $u^p = v^q$ , which in this case means  $px = p^{-q/p} z^q$  or  $x = p^{-q/p-1} z^q$ . Since  $-q/p - 1 = -q$  this is the desired expression.

Alternatively, one can let  $y = -pw/z < 0$ ,  $f(x) = -px^{1/p}$ , and write

$$\sup_{x>0} x^{1/p} z - x w = \frac{z}{p} \sup_{x>0} (xy - f(x)) = \frac{z}{p} f^*(y)$$

where  $f^*$  denotes the Legendre transformation or Fenchel conjugate of the convex function  $f$ . It is known that  $f^*(y) = -(-y)^{1-q}/(1-q)$  [13, Table 3.1] and this yields the desired statement. We thank the anonymous referee for this observation.  $\square$

**Lemma 5.5.** *Suppose  $a \geq 0, b > 0$ . Let  $f : X \rightarrow \mathbb{R}$  be bounded and 1-Lipschitz. Then for every  $k < \frac{1}{4a}$  we have*

$$T_{a,b}(\mu_0, \mu_1) \geq C_b \frac{\left(\int_X \exp\left(\frac{kb}{e^b - 4ak} f^2\right) d\mu_1\right)^q}{\left(\int_X \exp\left(\frac{kb}{1-4ak} f^2\right) d\mu_0\right)^{q-1}}. \tag{5.4}$$

*Proof.* We consider a function  $\varphi$  of the form  $\varphi_s = \exp(\alpha(s)f^2 + \beta(s))$ . In order to have  $\varphi_s \in \mathcal{E}_{a,b}$  we require

$$\begin{aligned} 0 &\geq \partial_s \varphi_s + a\varphi_s |\nabla \ln \varphi_s|^2 + b\varphi_s \ln \varphi_s \\ &= \varphi_s \cdot (\alpha'(s)f^2 + \beta'(s) + a\alpha(s)^2 |\nabla f^2|^2 + b\alpha(s)f^2 + b\beta(s)). \end{aligned}$$

Since  $|\nabla f| \leq 1$ , we have  $|\nabla f^2|^2 = (2f|\nabla f|)^2 \leq 4f^2$ , so it suffices to have

$$\begin{aligned} \alpha'(s) + 4a\alpha(s)^2 + b\alpha(s) &= 0 \\ \beta'(s) + b\beta(s) &= 0 \end{aligned}$$

which is satisfied by

$$\begin{aligned} \alpha(s) &= \frac{kb}{e^{bs} - 4ak} \\ \beta(s) &= \beta_0 e^{-bs} \end{aligned}$$

for any  $k < \frac{1}{4a}$  and any  $\beta_0 \in \mathbb{R}$ . So with this choice of  $\varphi$ , we have

$$\begin{aligned} T_{a,b}(\mu_0, \mu_1) &\geq \int_X \varphi_1 d\mu_1 - \int_X \varphi_0 d\mu_0 \\ &= e^{\beta_0} e^{-b} \int_X \exp\left(\frac{kb}{e^b - 4ak} f^2\right) d\mu_1 - e^{\beta_0} \int_X \exp\left(\frac{kb}{1-4ak} f^2\right) d\mu_0 \\ &= C_b \frac{\left(\int_X \exp\left(\frac{kb}{e^b - 4ak} f^2\right) d\mu_1\right)^q}{\left(\int_X \exp\left(\frac{kb}{1-4ak} f^2\right) d\mu_0\right)^{q-1}} \end{aligned}$$

when we make an optimal choice of  $\beta_0$  as described in Lemma 5.4, with  $x = e^{\beta_0}$  and noting that  $1/p = e^{-b}$ .  $\square$

In the previous lemma, when  $a = 0$ , the gradient terms vanish, and we can instead consider a function  $\varphi_s$  of the form  $\varphi_s = \exp(\alpha(s)f + \beta(s))$  where  $f$  need only be bounded and Lipschitz. This yields the following improvement:

**Corollary 5.6.** *Suppose  $b > 0$ , and let  $f \in \text{Lip}_b(X)$ . Then*

$$T_{0,b}(\mu_0, \mu_1) \geq C_b \frac{\left(\int_X e^{e^{-b} f} d\mu_1\right)^q}{\left(\int_X e^f d\mu_0\right)^{q-1}}. \tag{5.5}$$

**Lemma 5.7.** *Let  $b > 0, r \geq 0$ . Suppose  $y : [0, 1] \rightarrow (0, \infty)$  is absolutely continuous and satisfies the differential inequality*

$$y' \leq ry - by \ln y \quad \text{a.e.}, \quad y(0) = y_0 > 0. \tag{5.6}$$

Then

$$y(s) \leq \exp\left(\frac{r}{b}(1 - e^{-bs})\right) y_0^{e^{-bs}} \tag{5.7}$$

for all  $0 \leq s \leq 1$ , and in particular, following Notation 5.3,

$$y(1) \leq \exp\left(\frac{r}{qb}\right) y_0^{1/p}. \tag{5.8}$$

*Proof.* Let  $\zeta(s) = e^{bs} \ln y(s)$ ; then  $\zeta(s)$  satisfies  $\zeta'(s) \leq re^{bs}$  a.e. Integrating from 0 to  $s$  yields

$$\zeta(s) \leq \ln y_0 + \frac{r}{b}(e^{bs} - 1) \tag{5.9}$$

which rearranges to (5.7). □

**Proposition 5.8.** For all  $a \geq 0$  and  $b > 0$ , we have  $T_{a,b}(\mu_0, \mu_1) \geq C_b$ , with equality iff  $\mu_0 = \mu_1$ . Thus

$$\tilde{T}_{a,b}(\mu_0, \mu_1) := \ln \frac{1}{C_b} T_{a,b}(\mu_0, \mu_1) \tag{5.10}$$

is a statistical divergence on  $\mathcal{P}(X)$ ; that is,  $\tilde{T}_{a,b}(\mu_0, \mu_1) \geq 0$  with equality iff  $\mu_0 = \mu_1$ .

*Proof.* The lower bound  $T_{a,b}(\mu_0, \mu_1) \geq C_b$  follows from Lemma 5.5 with  $f = 0$ .

To show equality holds when  $\mu_0 = \mu_1 = \mu$ , let  $\varphi \in \mathcal{E}_{a,b}$ . Note that in particular,  $\varphi$  satisfies  $\partial_s \varphi_s + b\varphi_s \ln \varphi_s \leq 0$ ; that is,  $\mathcal{E}_{a,b} \subseteq \mathcal{E}_{0,b}$ . So for each  $x$ ,  $y(s) = \varphi_s(x)$  satisfies (5.6) with  $r = 0$ , and so by Lemma 5.7 and Lemma 5.4 we have

$$\varphi_1(x) - \varphi_0(x) \leq \varphi_0(x)^{1/p} - \varphi_0(x) \leq C_b.$$

Thus  $\int_X (\varphi_1 - \varphi_0) d\mu \leq C_b$  and taking the supremum over  $\varphi_s \in \mathcal{E}_{a,b}$  we have  $T_{a,b}(\mu, \mu) \leq C_b$ .

Conversely, suppose  $\mu_0, \mu_1$  satisfy  $T_{a,b}(\mu_0, \mu_1) = C_b$ . Let  $f : X \rightarrow \mathbb{R}$  be bounded and 1-Lipschitz. Lemma 5.5 then implies

$$\left(\int_X \exp\left(\frac{kb}{e^b - 4ak} f^2\right) d\mu_1\right)^q \leq \left(\int_X \exp\left(\frac{kb}{1 - 4ak} f^2\right) d\mu_0\right)^{q-1}$$

for every  $k < \frac{1}{4a}$ . When  $k = 0$ , both sides equal 1, so we differentiate the inequality at  $k = 0$  to obtain

$$qbe^{-b} \int_X f^2 d\mu_1 \leq (q-1)b \int_X f^2 d\mu_0$$

which rearranges to

$$\int_X f^2 d\mu_1 \leq \int_X f^2 d\mu_0$$

since  $q/(q-1) = p = e^b$ . The rest is a density argument. Replacing  $f$  by  $f + c$  for an arbitrary constant  $c \in \mathbb{R}$  and expanding, we get

$$\int_X f^2 d\mu_1 + 2c \int_X f d\mu_1 + c^2 \leq \int_X f^2 d\mu_0 + 2c \int_X f d\mu_0 + c^2$$

Letting  $c \rightarrow \pm\infty$ , we see this implies  $\int_X f d\mu_1 = \int_X f d\mu_0$  for all bounded 1-Lipschitz  $f$ , and by scaling, the same holds for all bounded Lipschitz  $f$ . This implies  $\mu_0 = \mu_1$ . □

We now show that when  $a = 0$ ,  $\tilde{T}_{0,b}$  recovers the Rényi divergence, whose definition we recall:

**Definition 5.9.** For  $\mu_0, \mu_1 \in \mathcal{P}(X)$  and  $r > 1$ , the Rényi divergence of order  $r$  is given by

$$D_r(\mu_1 \parallel \mu_0) := \frac{1}{r-1} \ln \int_X \left( \frac{d\mu_1}{d\mu_0} \right)^r d\mu_0 = \frac{1}{r-1} \ln \int_X \left( \frac{d\mu_1}{d\mu_0} \right)^{r-1} d\mu_1$$

if  $\mu_1$  is absolutely continuous with respect to  $\mu_0$ , and  $D_r(\mu_1 \parallel \mu_0) = \infty$  otherwise.

**Lemma 5.10.** Let  $b > 0$ . If  $T_{0,b}(\mu_0, \mu_1) < \infty$  then  $\mu_1$  is absolutely continuous with respect to  $\mu_0$ .

*Proof.* Let  $A \subset X$  be a Borel set for which  $\mu_0(A) = 0$ . We can then find a sequence of bounded nonpositive Lipschitz functions  $f_n$  such that  $f_n(x) \rightarrow 0$  for a.e.  $x \in A$  (with respect to  $\mu_0 + \mu_1$ ), and  $f_n(x) \rightarrow -\infty$  for a.e.  $x \in A^c$ . Applying Corollary 5.6 to  $f_n$ , we have

$$\left( \int_X e^{e^{-b} f_n} d\mu_1 \right)^q \leq \frac{T_{0,b}(\mu_0, \mu_1)}{C_b} \left( \int_X e^{f_n} d\mu_0 \right)^{q-1}.$$

Letting  $n \rightarrow \infty$  and using dominated convergence, this becomes

$$\mu_1(A)^q \leq \frac{T_{0,b}(\mu_0, \mu_1)}{C_b} \mu_0(A)^{q-1} = 0. \quad \square$$

**Proposition 5.11.** Let  $b > 0$ . For all  $\mu_0, \mu_1 \in \mathcal{P}(X)$  we have

$$T_{0,b}(\mu_0, \mu_1) = \begin{cases} C_b \int_X \left( \frac{d\mu_1}{d\mu_0} \right)^q d\mu_0, & \mu_1 \ll \mu_0 \\ \infty, & \mu_1 \not\ll \mu_0 \end{cases} \quad (5.11)$$

so that

$$\tilde{T}_{0,b}(\mu_0, \mu_1) = (q-1)D_q(\mu_1 \parallel \mu_0). \quad (5.12)$$

*Proof.* The case  $\mu_1 \not\ll \mu_0$  is the contrapositive of Lemma 5.10, so suppose  $\mu_1 \ll \mu_0$  and let  $\varrho = \frac{d\mu_1}{d\mu_0}$ . We show  $T_{0,b} = C_b \int_X \varrho^q d\mu_0$ .

To show  $T_{0,b} \leq C_b \int_X \varrho^q d\mu_0$ , let  $\varphi \in \mathcal{E}_{0,b}$ . Taking  $y(s) = \varphi_s(x)$  and  $r = 0$  in Lemma 5.7, we have  $\varphi_1 \leq \varphi_0^{1/p}$  pointwise. Hence

$$\begin{aligned} \int_X \varphi_1 d\mu_1 - \int_X \varphi_0 d\mu_0 &= \int_X (\varphi_1 \varrho - \varphi_0) d\mu_0 \\ &\leq \int_X (\varphi_0^{1/p} \varrho - \varphi_0) d\mu_0 \\ &\leq C_b \int_X \varrho^q d\mu_0 \end{aligned}$$

by Lemma 5.4. Taking the supremum over  $\varphi \in \mathcal{E}_{0,b}$  yields the desired upper bound.

For the lower bound, if  $\varrho \in L^q(\mu_0)$ , take a sequence  $f_n \in \text{Lip}_b(X)$  such that  $e^{f_n} \rightarrow \varrho^q$ ,  $\mu_0$ -almost everywhere (hence also  $\mu_1$ -almost everywhere) and in  $L^1(\mu_0)$ . Then Corollary 5.6 gives

$$C_b \left( \int_X e^{e^{-b} f_n} d\mu_1 \right)^q \leq T_{0,b}(\mu_0, \mu_1) \left( \int_X e^{f_n} d\mu_0 \right)^{q-1}.$$

Pass to the limit, applying Fatou's lemma on the left and  $L^1$  convergence on the right, to obtain

$$C_b \left( \int_X \varrho^{qe^{-b}} d\mu_1 \right)^q \leq T_{0,b}(\mu_0, \mu_1) \left( \int_X \varrho^q d\mu_0 \right)^{q-1}. \quad (5.13)$$

Now observe that  $qe^{-b} = q/p = q - 1$  and so  $\int_X \varrho^{qe^{-b}} d\mu_1 = \int_X \varrho^{q-1} d\mu_1 = \int_X \varrho^q d\mu_0$ . Hence (5.13) rearranges to

$$C_b \int_X \varrho^q d\mu_0 \leq T_{0,b}(\mu_0, \mu_1)$$

as desired.

If  $\varrho \notin L^q(\mu_0)$ , then we need to show  $T_{0,b}(\mu_0, \mu_1) = \infty$ . Let  $m \geq 0$  and  $A_m = \{\varrho \leq m\}$ . Choose  $f_n \in \text{Lip}_b(X)$  with  $e^{f_n} \rightarrow \varrho^q 1_{A_m}$ ,  $\mu_0$ -a.e. and in  $L^1(\mu_0)$ . Then proceeding as in the previous case, we obtain

$$C_b \int_{A_m} \varrho^q d\mu_0 \leq T_{0,b}(\mu_0, \mu_1).$$

Letting  $m \rightarrow \infty$  and applying the monotone convergence theorem, we conclude that  $T_{0,b}(\mu_0, \mu_1) = +\infty$ .  $\square$

Finally, we estimate the value of  $T_{a,b}$  for point masses.

**Proposition 5.12.** *Suppose  $a, b > 0$ ,  $x_0, x_1 \in X$ . Then*

$$C_b \exp\left(\frac{bq}{4a(p-1)}d(x_0, x_1)^2\right) \leq T_{a,b}(\delta_{x_0}, \delta_{x_1}) \leq C_b \exp\left(\frac{1}{4ab}d(x_0, x_1)^2\right) \tag{5.14}$$

or in terms of  $\tilde{T}_{a,b}$ ,

$$\frac{bq}{4a(p-1)}d(x_0, x_1)^2 \leq \tilde{T}_{a,b}(\delta_{x_0}, \delta_{x_1}) \leq \frac{1}{4ab}d(x_0, x_1)^2. \tag{5.15}$$

*Proof.* For the upper bound, suppose  $\varphi \in \mathcal{E}_{a,b}$ , and as in the proof of the upper bound in Proposition 3.3, let  $\gamma : [0, 1] \rightarrow X$  be a constant speed geodesic joining  $x_0$  to  $x_1$ . Using the chain rule, we have

$$\begin{aligned} \frac{d}{ds}\varphi_s(\gamma_s) &\leq \partial_s\varphi_s(\gamma_s) + |\nabla\varphi_s|(\gamma_s)d(x_0, x_1) \\ &\leq -\frac{a}{\varphi_s(\gamma_s)}|\nabla\varphi_s|(\gamma_s)^2 - b\varphi_s(\gamma_s)\ln\varphi_s(\gamma_s) + |\nabla\varphi_s|(\gamma_s)d(x_0, x_1) \\ &\leq \frac{d(x_0, x_1)^2}{4a}\varphi_s(\gamma_s) - b\varphi_s(\gamma_s)\ln\varphi_s(\gamma_s) \end{aligned}$$

by completing the square. So  $y(s) = \varphi_s(\gamma_s)$  satisfies the differential inequality (5.6) with  $r = d(x_0, x_1)^2/4a$ , and by Lemma 5.7 and Lemma 5.4 we have

$$\begin{aligned} \varphi_1(x_1) - \varphi_0(x_0) &\leq \exp\left(\frac{1}{4abq}d(x_0, x_1)^2\right)\varphi_0(x_0)^{1/p} - \varphi_0(x_0) \\ &\leq C_b \exp\left(\frac{1}{4ab}d(x_0, x_1)^2\right). \end{aligned}$$

For the lower bound, apply Lemma 5.5 with  $\mu_i = \delta_{x_i}$  and  $f(x) = d(x_0, x) \wedge d(x_0, x_1)$ , which is bounded and 1-Lipschitz. Since  $f(x_0) = 0$ , the  $d\mu_0$  integral in (5.4) equals 1, and we obtain

$$T_{a,b}(\delta_{x_0}, \delta_{x_1}) \geq C_b \exp\left(\frac{qkb}{e^b - 4ak}d(x_0, x_1)^2\right)$$

for any  $k < 1/4a$ . Letting  $k \uparrow 1/4a$  and recalling that  $e^b = p$ , we have the desired inequality.  $\square$

**Corollary 5.13.** For  $\mu_0, \mu_1 \in \mathcal{P}(X)$ , we have

$$T_{a,b}(\mu_0, \mu_1) \leq C_b \inf_{\pi} \int_{X \times X} \exp\left(\frac{1}{4ab} d(x, y)^2\right) \pi(dx, dy). \quad (5.16)$$

where the infimum is taken over all couplings  $\pi \in \mathcal{P}(X \times X)$  of  $\mu_0, \mu_1$ . As a special case, we have

$$T_{a,b}(\delta_x, \mu) \leq C_b \int_X \exp\left(\frac{1}{4ab} d(x, y)^2\right) \mu(dy). \quad (5.17)$$

*Proof.* Let  $\varphi \in \mathcal{E}_{a,b}$  and let  $\pi$  be a coupling of  $\mu_0, \mu_1$ . Then we have

$$\begin{aligned} \int_X \varphi_1 d\mu_1 - \int_X \varphi_0 d\mu_0 &= \int_{X \times X} (\varphi_1(y) - \varphi_0(x)) \pi(dx, dy) \\ &\leq \int_{X \times X} T_{a,b}(\delta_x, \delta_y) \pi(dx, dy) \\ &\leq C_b \int_{X \times X} \exp\left(\frac{1}{4ab} d(x, y)^2\right) \pi(dx, dy) \end{aligned}$$

and (5.16) follows by taking the supremum over  $\varphi$  and the infimum over  $\pi$ . □

## 5.2 Functional inequalities

In the same way that the Hellinger-Kantorovich contraction property was equivalent to a reverse Poincaré inequality, it turns out that a similar contraction property for  $T_{a,b}$  is equivalent to a reverse logarithmic Sobolev inequality, as well as to a Wang-type Harnack inequality.

For one direction of this equivalence, the key tool is the following general statement, analogous to Theorem 3.4.

**Theorem 5.14.** Let  $a, b, \gamma, \delta \geq 0$ . Suppose that for all  $f \in \text{Lip}_b(X)$  with  $f > 0$ , we have

$$a(Pf)|\nabla \ln Pf|^2 + b(Pf) \ln Pf \leq \gamma P(f|\nabla \ln f|^2) + \delta P(f \ln f). \quad (5.18)$$

Then for all  $\mu_0, \mu_1 \in \mathcal{P}(X)$  we have

$$T_{\gamma,\delta}(\mu_0 P, \mu_1 P) \leq T_{a,b}(\mu_0, \mu_1). \quad (5.19)$$

*Proof.* Suppose that  $\varphi \in \mathcal{E}_{\gamma,\delta}$ . Then we have

$$\begin{aligned} \partial_s P\varphi_s + aP\varphi_s|\nabla \ln P\varphi_s|^2 + bP\varphi_s \ln P\varphi_s &\leq \partial_s P\varphi_s + \gamma P(\varphi_s|\nabla \ln \varphi_s|^2) + \delta P(\varphi_s \ln \varphi_s) \\ &= P(\partial_s \varphi_s + \gamma \varphi_s|\nabla \ln \varphi_s|^2 + \delta \varphi_s \ln \varphi_s) \\ &\leq 0 \end{aligned}$$

since  $\varphi_s \in \mathcal{E}_{\gamma,\delta}$  and  $P$  is positivity preserving. Thus  $P\varphi_s \in \mathcal{E}_{a,b}$ , and so

$$\int \varphi_1 d(\mu_1 P) - \int \varphi_0 d(\mu_0 P) = \int P\varphi_1 d\mu_1 - \int P\varphi_0 d\mu_0 \leq T_{a,b}(\mu_0, \mu_1).$$

Taking the supremum over  $\varphi \in \mathcal{E}_{\gamma,\delta}$  we have  $T_{\gamma,\delta}(\mu_0 P, \mu_1 P) \leq T_{a,b}(\mu_0, \mu_1)$ . □

**Theorem 5.15.** Let  $C > 0$ . The following are equivalent:

1. The reverse logarithmic Sobolev inequality

$$Pf|\nabla \ln Pf|^2 \leq C(P(f \ln f) - (Pf) \ln Pf), \quad f \in \text{Lip}_b(X), f > 0. \quad (\text{rLSI})$$

2. The family of entropic transportation-cost inequalities

$$T_{0,\kappa C}(\mu_0 P, \mu_1 P) \leq T_{\kappa,\kappa C}(\mu_0, \mu_1), \quad \mu_0, \mu_1 \in \mathcal{P}(X), \quad \kappa > 0. \quad (\text{ETI})$$

3. The Wang-type Harnack inequality

$$Pf(x)^p \leq \exp\left(\frac{p}{p-1} \frac{Cd(x,y)^2}{4}\right) P(f^p)(y), \quad (\text{WHI})$$

$$p > 1, \quad f \in \text{Lip}_b(X), \quad f > 0, \quad x, y \in X.$$

*Proof.* (rLSI)  $\implies$  (ETI): Apply Theorem 5.14 with  $a = \kappa$ ,  $b = \delta = \kappa C$ ,  $\gamma = 0$ .

(ETI)  $\implies$  (WHI): Fix  $x, y \in X$ , and let  $m$  be some finite reference measure so that  $\delta_x P, \delta_y P$  are absolutely continuous with respect to  $m$ , with densities  $p_x, p_y$  respectively. Let  $\kappa$  be arbitrary and let  $p = e^{\kappa C}$ ,  $q = p/(p-1)$ . By Proposition 5.11 and Proposition 5.12 with  $a = \kappa$ ,  $b = \kappa C$ , taking  $\mu_0 = \delta_y, \mu_1 = \delta_x$  to match notation with other papers, we have that (ETI) implies the integrated Harnack inequality

$$\int \left(\frac{p_x}{p_y}\right)^q p_y dm = \int \left(\frac{p_x}{p_y}\right)^{q-1} p_x dm \leq \exp\left(\frac{1}{\kappa^2 C^2} \frac{Cd(x,y)^2}{4}\right). \quad (5.20)$$

Now  $q-1 = 1/(p-1)$  so this may be rewritten as

$$\left(\int \left(\frac{p_x}{p_y}\right)^{1/(p-1)} p_x dm\right)^{p-1} \leq \exp\left(\frac{p-1}{(\log p)^2} \frac{Cd(x,y)^2}{4}\right), \quad p > 1. \quad (5.21)$$

Since  $\kappa$  was arbitrary, (5.21) holds for all  $p > 1$ . By an application of Hölder’s inequality (see [9, Lemma 2.11]), (5.21) implies the Wang-type Harnack inequality

$$Pf(x)^p \leq \exp\left(\frac{p-1}{(\log p)^2} \frac{Cd(x,y)^2}{4}\right) P(f^p)(y), \quad p > 1. \quad (5.22)$$

To recover the more usual form of the Wang Harnack inequality (WHI), we would like to have (5.22) with  $\frac{p}{p-1}$  in the exponent in place of  $\frac{p-1}{(\log p)^2}$ . To this end, fix  $\epsilon > 0$ . As  $p \rightarrow 1$ , we have  $\frac{p}{p-1} \sim \frac{p-1}{(\log p)^2}$ , so for all sufficiently small  $p' > 1$  we have  $\frac{p'-1}{(\log p')^2} \leq (1+\epsilon)\frac{p'}{p'-1}$  and thus

$$Pf(x)^{p'} \leq \exp\left(\frac{p'}{p'-1} \frac{(1+\epsilon)Cd(x,y)^2}{4}\right) P(f^{p'})(y). \quad (5.23)$$

In particular this holds for  $p' = p^{1/n}$  for sufficiently large  $n$ . From [34, Proposition 2.1], with  $(1+\epsilon)C$  in place of  $C$ , it follows that (5.23) holds for  $p$  in place of  $p'$ , and letting  $\epsilon \rightarrow 0$  we obtain (WHI).

(WHI)  $\implies$  (rLSI): This is shown, in essence, in [33, Theorem 2.1]. The proof there is in the setting of a manifold with bounded curvature, and requires some minor changes to apply in this setting, so we give the details.

Let  $f$  be a positive bounded Lipschitz function which is bounded away from 0. Observe first that (WHI) implies that  $Pf$  is continuous. To see this, fix  $x \in X$  and  $p > 1$ . Letting  $y \rightarrow x$  in (WHI), we see that  $Pf(x)^p \leq \liminf_{y \rightarrow x} P(f^p)(y)$ . Now as  $p \rightarrow 1$  we have  $f^p \rightarrow f$  uniformly, and since  $P$  is Markovian we also have  $P(f^p) \rightarrow Pf$  uniformly. So we can pass to the limit to conclude  $Pf(x) \leq \liminf_{y \rightarrow x} Pf(y)$ . For the other direction, apply (WHI) with  $x$  and  $y$  interchanged. Let  $y \rightarrow x$  to obtain  $(\limsup_{y \rightarrow x} Pf(y))^p \leq P(f^p)(x)$ , and then let  $p \rightarrow 1$ .

Now, by definition of  $|\nabla Pf|$  there exists a sequence  $y_n \rightarrow x$ , with  $y_n \neq x$ , such that  $\frac{Pf(y_n) - Pf(x)}{d(y_n, x)} \rightarrow \pm |\nabla Pf|(x)$ . Suppose first that we can choose  $y_n$  so that we have the

negative sign, i.e.  $\frac{Pf(y_n)-Pf(x)}{d(y_n,x)} \rightarrow -|\nabla Pf|(x)$ . Set  $r_n = d(y_n, x)$  for convenience, and let  $\delta > 0$  be arbitrary. Then (WHI) with  $y = y_n$  and  $p = 1 + r_n\delta$  reads

$$Pf(x)^{1+r_n\delta} \leq \exp\left(\frac{C}{4\delta}(r_n + r_n^2\delta)\right) P(f^{1+r_n\delta})(y_n).$$

Subtracting  $Pf(x)$ , dividing by  $r_n$ , and breaking up the right side, we have

$$\begin{aligned} \frac{Pf(x)^{1+r_n\delta} - Pf(x)}{r_n} &\leq \frac{\exp\left(\frac{C}{4\delta}(r_n + r_n^2\delta)\right) - 1}{r_n} P(f^{1+r_n\delta})(y_n) \\ &\quad + P\left(\frac{f^{1+r_n\delta} - f}{r_n}\right)(y_n) + \frac{Pf(y_n) - Pf(x)}{r_n}. \end{aligned}$$

We now pass to the limit. Since  $f$  is continuous, bounded, and bounded away from 0, we have  $f^{1+r_n\delta} \rightarrow f$  and  $\frac{1}{r_n}(f^{1+r_n\delta} - f) \rightarrow \delta f \ln f$  uniformly, and so the same is true when  $P$  is applied. We obtain

$$\delta Pf(x) \ln Pf(x) \leq \frac{C}{4\delta} Pf(x) + \delta P(f \ln f)(x) - |\nabla Pf|(x) \tag{5.24}$$

and now optimizing over  $\delta$  and rearranging yields (rLSI).

Otherwise, there exists a sequence  $y_n \rightarrow x$  such that  $\frac{Pf(y_n)-Pf(x)}{d(y_n,x)} \rightarrow +|\nabla Pf|(x)$ . We apply (WHI) with  $x$  and  $y$  interchanged and proceed as before to obtain

$$\begin{aligned} \frac{Pf(y_n)^{1+r_n\delta} - Pf(y_n)}{r_n} &\leq \frac{\exp\left(\frac{C}{4\delta}(r_n + r_n^2\delta)\right) - 1}{r_n} P(f^{1+r_n\delta})(x) \\ &\quad + P\left(\frac{f^{1+r_n\delta} - f}{r_n}\right)(x) + \frac{Pf(x) - Pf(y_n)}{r_n}. \end{aligned}$$

Passing to the limit again yields (5.24). On the left side, we use the fact that since  $Pf$  is continuous, bounded, and bounded away from 0, we have  $\frac{1}{r_n}((Pf)^{1+r_n\delta} - Pf) \rightarrow \delta Pf \ln Pf$  uniformly.  $\square$

**Remark 5.16.** The Wang Harnack inequality (WHI) is also known to be equivalent to the integrated Harnack inequality

$$\int_X \left(\frac{p_x}{p_y}\right)^{1/(p-1)} p_x \, dm \leq \exp\left(\frac{p}{(p-1)^2} \frac{Cd(x,y)^2}{4}\right), \quad p > 1 \tag{IHI}$$

where as above  $p_x, p_y$  are the densities of  $\delta_x P, \delta_y P$  with respect to some reference measure  $m$ ; see [9, Lemma 2.11]. Hence (IHI) is also equivalent to (ETI) and (rLSI). In the proof of Theorem 5.15, we obtained (5.21) which is infinitesimally weaker than (IHI); the self-improvement comes via the application of [34, Proposition 2.1], applying (WHI) along a sequence of points between  $x$  and  $y$ .

A different application of Theorem 5.14 relates a gradient bound for  $P$  to another type of contraction inequality for  $T_{a,b}$ , analogous to Theorem 3.5.

**Proposition 5.17.** *Suppose that for some  $C$ , the operator  $P$  satisfies the  $L^1 \ln L$ -type gradient estimate*

$$Pf|\nabla \ln Pf|^2 \leq CP(f|\nabla \ln f|^2), \quad f \in \text{Lip}_b(X), f > 0. \tag{5.25}$$

Then for every  $\kappa, \epsilon > 0$  we have

$$T_{\kappa C, \epsilon}(\mu_0 P, \mu_1 P) \leq T_{\kappa, \epsilon}(\mu_0, \mu_1), \quad \mu_0, \mu_1 \in \mathcal{P}(X). \tag{5.26}$$

In particular, this holds if we have the stronger  $L^1$ -type gradient estimate

$$|\nabla Pf| \leq C^{1/2} P|\nabla f|, \quad f \in \text{Lip}_b(X). \tag{5.27}$$

*Proof.* By Jensen’s inequality we have  $(Pf) \ln Pf \leq P(f \ln f)$ , and combining this with (5.25) we have that (5.18) holds with  $a = \kappa$ ,  $\gamma = \kappa C$ ,  $b = \delta = \epsilon$ . The conclusion then follows from Theorem 5.14.

To see that (5.27) implies (5.25), using the former together with the bivariate Jensen inequality for the convex function  $\psi(x, y) = x^2/y$ , we obtain

$$\frac{|\nabla Pf|^2}{Pf} \leq C \frac{P(|\nabla f|^2)}{Pf} \leq CP \left( \frac{|\nabla f|^2}{f} \right)$$

which is equivalent to (5.25) thanks to the chain rule (Lemma 2.2). □

**Remark 5.18.** It might seem more natural to take  $\epsilon = 0$  in (5.26), but in fact that statement would have no content, as one can show that  $T_{a,0}(\mu_0, \mu_1) = +\infty$  for all  $\mu_0 \neq \mu_1$ .

## 6 Applications to quasi-invariance

The reverse logarithmic Sobolev inequality (rLSI) has been the object of significant study in the literature, although not nearly as much as the “forward” logarithmic Sobolev inequality. One particularly interesting area of application is in proving absolute continuity of heat kernel measures; especially in the presence of group structure, where it can be used to show *quasi-invariance* of a heat kernel measure under group translation. Such results are commonly obtained through the use of the Wang Harnack inequality (WHI), which as noted in Section 5.2 is equivalent to (rLSI). In this section, we consider some examples and show how the entropic transportation-cost inequality (ETI) provides an alternate route to these conclusions.

Although in this paper we limit our attention to a few specific known results, there are many other situations where similar questions about absolute continuity could be considered, especially in stochastic PDE, see [35]. The techniques developed in this paper may be useful in the study of these problems, and we hope to address this in future work.

### 6.1 Subelliptic heat kernels on finite-dimensional Lie groups

Let  $\mathbb{G}$  be a finite-dimensional connected real Lie group with identity element  $e$ , and suppose that  $\mathbb{G}$  is equipped with a left-invariant sub-Riemannian geometry: a bracket-generating left-invariant sub-bundle  $\mathcal{H} \subset T\mathbb{G}$ , and a sub-Riemannian metric  $g$  which is a left-invariant inner product on  $\mathcal{H}$ . We denote by  $\nabla$  the horizontal sub-gradient, and  $|\nabla f| := \sqrt{g(\nabla f, \nabla f)}$ . Let  $d$  be the Carnot–Carathéodory distance on  $\mathbb{G}$ ; by the Chow–Rashevskii theorem, the bracket-generating condition implies that  $d(x, y) < \infty$  for all  $x, y \in \mathbb{G}$ . Let  $L$  be the left-invariant sub-Laplacian induced by  $g$ ,  $P_t = e^{tL}$  the heat semigroup generated by  $L$ , and  $\mu_t = \delta_e P_t$  the heat kernel measure.

Under these conditions, Hörmander’s theorem implies that  $L$  is subelliptic and hence  $\mu_t$  is a smooth measure for all  $t > 0$ . Our purpose here is to remark that at least part of this conclusion can be recovered using our techniques instead, if one has a reverse log Sobolev inequality.

Recall that in general, a Borel probability measure  $\mu$  on a topological group  $\mathbb{G}$  is said to be *quasi-invariant* under left translation by an element  $x \in \mathbb{G}$  if  $\mu$  and its left translation  $\mu^x(A) = \mu(x^{-1}A)$  are mutually absolutely continuous. If this holds for every  $x$  in some subgroup  $H \subset \mathbb{G}$ , we say  $\mu$  is quasi-invariant under left translation by  $H$ .

**Proposition 6.1.** *Suppose, under the above assumptions, that  $P_t$  satisfies the reverse logarithmic Sobolev inequality*

$$P_t f |\nabla P_t f|^2 \leq C(t) (P_t(f \ln f) - (P_t f) \ln P_t f). \tag{6.1}$$

Then for all  $t > 0$ ,  $\mu_t$  is quasi-invariant under translation by every  $x \in \mathbb{G}$ . As a consequence,  $\mu_t$  is absolutely continuous with respect to left Haar measure and has full support.

*Proof.* By Theorem 5.15, (6.1) implies the entropic transportation-cost inequality

$$T_{0,\kappa C(t)}(\mu_0 P_t, \mu_1 P_t) \leq T_{\kappa,\kappa C(t)}(\mu_0, \mu_1), \quad \kappa > 0.$$

Taking  $\mu_0 = \delta_e$ ,  $\mu_1 = \delta_x$  and applying Lemma 5.12 to bound  $T_{\kappa,\kappa C(t)}(\delta_e, \delta_x)$ , we find that  $T_{0,b}(\mu_t, \mu_t^x) < \infty$ , and so Lemma 5.10 implies that  $\mu_t^x \ll \mu_t$ ; the opposite relation  $\mu_t \ll \mu_t^x$  follows by symmetry.

The consequence that  $\mu_t$  is absolutely continuous with respect to left Haar measure is a standard fact about locally compact groups; see for instance [14, Ch. 7, §1.9, Proposition 11].  $\square$

By the results in [5], the reverse log Sobolev inequality holds in sub-Riemannian manifolds satisfying a generalized curvature-dimension inequality of the type introduced in [8]. It was shown in [8] that such inequalities hold for step two Carnot groups and the three-dimensional model groups  $\mathbb{S}\mathbb{U}(2)$  and  $\mathbb{S}\mathbb{L}(2)$ , and in [7] for three-dimensional solvable groups.

## 6.2 Abstract Wiener space

The phenomenon of quasi-invariance is more interesting in groups that are not locally compact, such as infinite dimensional vector spaces or Lie groups. Here, the smoothness of a measure cannot be described in terms of absolute continuity to Haar measure, since Haar measure does not exist, and so quasi-invariance provides a more “intrinsic” notion of regularity.

In this subsection, we consider the very classical example of abstract Wiener space. As this and similar infinite-dimensional models do not fit exactly into the setting defined in Section 2, we shall briefly discuss how to adapt the results of Sections 3 and 5 in this case, as a prototype for later examples. We give basic definitions here to fix notation; for further background on abstract Wiener space and Gaussian measures on infinite-dimensional spaces, we refer to [12, 24].

An *abstract Wiener space* consists of a real separable Banach space  $W$  equipped with a centered non-degenerate Gaussian Borel measure  $\mu$ . We denote by  $H \subset W$  the associated dense *Cameron–Martin space*, into which the continuous dual  $W^*$  is naturally embedded. A smooth *cylinder function* is a function  $F : W \rightarrow \mathbb{R}$  of the form  $F(x) = \varphi(f_1(x), \dots, f_n(x))$  for some  $n$ , where  $\varphi \in C_b^\infty(\mathbb{R}^n)$  is a smooth function with all partial derivatives bounded, and  $f_1, \dots, f_n \in W^* \subset H$ ; unless otherwise specified, we assume without loss of generality that  $f_1, \dots, f_n$  are orthonormal in  $H$ . We let  $Cyl(W)$  denote the space of all such functions; this will be used in place of  $Lip_b(W)$  as a space of test functions. It is a standard fact that  $Cyl(W)$  is dense in  $L^p(\mu)$  for  $1 \leq p < \infty$ .

The *Malliavin gradient*  $DF : W \rightarrow H$  of a cylinder function is defined by  $(DF)(x) = \sum_{i=1}^n (\partial_i \varphi)(f_1(x), \dots, f_n(x)) f_i$ , so that when the  $f_i$  are orthonormal in  $H$  we have

$$\|DF(x)\|_H^2 = \sum_{i=1}^n |\partial_i \varphi(f_1(x), \dots, f_n(x))|^2 = |\nabla \varphi(f_1(x), \dots, f_n(x))|^2.$$

Note that  $\|DF\|_H$  is not a strong upper gradient on  $W$  with respect to the distance induced by its norm  $\|\cdot\|_W$ .

The heat semigroup  $P_t$  on  $W$  is the convolution semigroup induced by the rescaled measure  $\mu$ , namely  $P_t F(x) = \int F(x + \sqrt{t}y) \mu(dy)$ . When  $F$  is a cylinder function of the

form  $F(x) = \varphi(f_1(x), \dots, f_n(x))$ , we have  $P_t F(x) = p_t \varphi(f_1(x), \dots, f_n(x))$  where  $p_t$  is the standard heat semigroup on  $\mathbb{R}^n$ ; in particular,  $P_t F$  is again a cylinder function.

We recall that  $p_t$  satisfies the reverse Poincaré inequality

$$|\nabla p_t \varphi|^2 \leq \frac{1}{t}(p_t \varphi^2 - (p_t \varphi)^2), \quad \varphi \in C_b^\infty(\mathbb{R}^n) \tag{6.2}$$

and the reverse logarithmic Sobolev inequality

$$p_t \varphi |\nabla \ln p_t \varphi|^2 \leq \frac{2}{t}(p_t(\varphi \ln \varphi) - p_t \varphi \ln p_t \varphi), \quad \varphi \in C_b^\infty(\mathbb{R}^n), \varphi > 0. \tag{6.3}$$

These follow, for instance, by standard  $\Gamma$ -calculus from the elementary commutation  $\nabla p_t \varphi = p_t \nabla \varphi$ . See for instance [2, Proposition 3.3], taking  $\rho = 0$ . Note that the constants in these inequalities are dimension-independent. As such, evaluating at  $(f_1(x), \dots, f_n(x))$ ,  $x \in W$ , we obtain the corresponding inequalities for  $P_t$  on  $(W, \mu)$ :

$$\|DP_t F\|_H^2 \leq \frac{1}{t}(P_t F^2 - (P_t F)^2), \quad F \in Cyl(W) \tag{6.4}$$

$$P_t F \|D \ln P_t F\|_H^2 \leq \frac{2}{t}(P_t(F \ln F) - P_t F \ln P_t F), \quad F \in Cyl(W), F > 0. \tag{6.5}$$

We modify Definitions 3.1 and 5.1 and by taking our class of test functions to be smooth cylinder functions of space and time, e.g. functions  $F_s : [0, 1] \times W \rightarrow \mathbb{R}$  of the form  $F_s = \varphi(s, f_1(x), \dots, f_n(x))$ ,  $\varphi \in C_b^\infty([0, 1] \times \mathbb{R}^n)$ . Let  $Cyl([0, 1] \times W)$  denote the space of such functions. Then we redefine

$$\begin{aligned} \mathcal{A}_{a,b} &= \{F_s \in Cyl([0, 1] \times W) : \partial_s F_s + a \|DF_s\|_H^2 + b F_s^2 \leq 0\} \\ \mathcal{E}_{a,b} &= \{F_s \in Cyl([0, 1] \times W) : F > 0, \partial_s F_s + a F_s \|D \ln F_s\|_H^2 + b F_s^2 \leq 0\} \end{aligned}$$

and define  $W_{a,b}, T_{a,b}$  accordingly on  $\mathcal{P}(W)$ . We have  $W_{0,b}$  and  $T_{0,b}$  related to Hellinger and Rényi divergences in the same way as before. Moreover we can follow the proof of the upper bound in Proposition 5.12, taking  $\gamma(s) = sx_1 + (1-s)x_0$  and noting  $|\frac{d}{ds} F(\gamma(s))| \leq \|DF(\gamma(s))\|_H \|x_1 - x_0\|_H$ , to conclude

$$T_{a,b}(\delta_{x_0}, \delta_{x_1}) \leq C_b \exp\left(\frac{1}{4ab} \|x_0 - x_1\|_H^2\right). \tag{6.6}$$

Now Theorem 5.15 allows us to recover the classical Cameron–Martin quasi-invariance theorem [16]. For  $t > 0$ , let  $\mu_t = \mu(t^{-1/2} \cdot) = \delta_0 P_t$  be the rescaling of the Gaussian measure  $\mu$ , and for  $h \in H$  let  $\mu_t^h = \mu(t^{-1/2}(\cdot - h)) = \delta_h P_t$  be its translation by  $h$ . We then obtain:

**Proposition 6.2** (Cameron–Martin theorem). *For all  $t > 0$  and  $h \in H$ , the measures  $\mu_t, \mu_t^h$  are mutually absolutely continuous.*

*Proof.* The logic is the same as in the proof of Proposition 6.1. Since the reverse logarithmic Sobolev inequality holds, Theorem 5.15 and (6.6) imply that for any  $\kappa > 0$ , we have

$$T_{0,2\kappa/t}(\mu_t, \mu_t^h) \leq T_{1,2\kappa/t}(\delta_0, \delta_h) \leq C_{2\kappa/t} \exp\left(\frac{t}{8\kappa^2} \|h\|_H^2\right) < \infty \tag{6.7}$$

Thus by Lemma 5.10 we have  $\mu_t^h \ll \mu_t$ , and the reverse statement  $\mu_t \ll \mu_t^h$  follows by symmetry.  $\square$

We also obtain a quantitative estimate on the  $L^p$  norm of the density  $d\mu_t^h/d\mu_t$ , which is perhaps most convenient to consider in the form of (5.21):

$$\left(\int_W \left(\frac{d\mu_t}{d\mu_t^h}\right)^{1/(p-1)} d\mu_t\right)^{p-1} \leq \exp\left(\frac{p-1}{(\log p)^2} \frac{\|h\|_H^2}{2t}\right). \tag{6.8}$$

As the left side is known to exactly equal  $\exp\left(\frac{p}{p-1} \frac{\|h\|_H^2}{2t}\right)$  (so that (IHI) is sharp), (6.8) becomes sharp as  $p \rightarrow 1$ .

One may also apply the reverse logarithmic Sobolev inequality for the Ornstein–Uhlenbeck  $Q_s$ , which is the symmetric Markov semigroup on  $L^2(\mu)$  generated by the Dirichlet form  $\mathcal{E}(F, F) = \int_W \|DF\|_H^2 d\mu$ . It satisfies

$$Q_s F \|D \ln Q_s F\|_H^2 \leq \frac{2}{e^{2s} - 1} (Q_s(F \ln F) - Q_s F \ln Q_s F). \tag{6.9}$$

See for instance [2, Section 3], noting that the carré du champ of  $Q_s$  is  $\Gamma(F, F) = \|DF\|_H^2$ , without a factor of  $\frac{1}{2}$ . Carrying out the above computations with  $Q_s$  and noting that  $\delta_h Q_s = \mu_{1-e^{-2s}}^{e^{-s}h}$ , one obtains exactly the same results for  $t < 1$ .

Finally, we remark that the Cameron–Martin quasi-invariance theorem can also be obtained using the Hellinger–Kantorovich contraction property of Theorem 3.7. Indeed, the reverse Poincaré inequality (6.4) for  $P_t$  implies

$$\text{He}_2(\mu_t, \mu_t^h)^2 \leq \frac{1}{4t} \|h\|_H^2 \tag{6.10}$$

since the Kantorovich–Wasserstein distance between point masses in this setting corresponds to the Cameron–Martin distance; this can be checked directly from the dynamic dual definition as in Proposition 3.3. Unfortunately, (6.10) has no content unless  $\frac{1}{4t} \|h\|_H^2 < 2$ , so to work around this, choose an integer  $n$  so large that  $n^{-2} \frac{1}{4t} \|h\|_H^2 < 2$ . Applying (6.10) with  $h/n$  in place of  $h$ , we conclude that  $\text{He}_2^2(\mu_t, \mu_t^{h/n}) < 2$  and in particular that  $\mu_t, \mu_t^{h/n}$  are not mutually singular. By the Feldman–Hájek dichotomy theorem for Gaussian measures [22, 21, 23, 15], they must therefore be mutually absolutely continuous, which we denote by  $\mu_t \sim \mu_t^{h/n}$ . Repeating this argument  $n$  times, we have  $\mu_t \sim \mu_t^{h/n} \sim \mu_t^{2h/n} \sim \dots \sim \mu_t^h$ , and since  $\sim$  is an equivalence relation, we have  $\mu_t \sim \mu_t^h$  as desired.

Although this argument uses only the reverse Poincaré inequality, which is *a priori* weaker than the reverse logarithmic Sobolev inequality used in Proposition 6.2, the conclusion is also weaker as it does not yield any quantitative information about the distance between the measures  $\mu_t, \mu_t^h$ .

We note that some proofs of the Feldman–Hájek dichotomy theorem, including Feldman’s original proof [22, 21], make use of the Cameron–Martin quasi-invariance theorem, which would seem to make the above argument circular. However, it is possible to prove the dichotomy theorem directly, without assuming quasi-invariance—see for example [15]—and this breaks the cycle.

### 6.3 Infinite dimensional Heisenberg-like groups

The ideas of the previous two subsections come together in the study of infinite-dimensional groups where the semigroup in question is not elliptic. In [9], the authors considered infinite-dimensional Heisenberg-like groups, introduced in [19], with their hypoelliptic heat kernels and corresponding heat semigroups. These groups carry a natural sub-Riemannian geometry analogous to the Heisenberg group and other Carnot groups of step two. They use generalized curvature-dimension inequalities to show that these spaces satisfy a reverse logarithmic Sobolev inequality. From this, they derive a Wang-type Harnack inequality, and use this to show quasi-invariance of the heat kernel measure under the group translation. In this section, we show that as in the case of Gaussian measures, transport inequalities provide an alternate route from reverse log Sobolev to quasi-invariance in this setting. We only sketch the argument here, as the details are closely analogous to those for the Gaussian case.

We follow the notation of [9] and refer the reader there for complete definitions, background, and further references. Let  $(W, H, \mu)$  be an abstract Wiener space and  $\mathbf{C}$  a finite-dimensional inner product space. Suppose that  $\mathfrak{g} = W \times \mathbf{C}$  is equipped with a continuous Lie bracket  $[\cdot, \cdot]$  satisfying  $[W, W] = \mathbf{C}$  and  $[\mathfrak{g}, \mathbf{C}] = 0$ . The corresponding Banach Lie group  $\mathbb{G}$  is given by  $\mathbb{G} = W \times \mathbf{C}$  equipped with the nonabelian group operation  $g_1 \cdot g_2 = g_1 + g_2 + \frac{1}{2}[g_1, g_2]$  defined by the Baker–Campbell–Hausdorff formula. Then  $\mathfrak{g}_{CM} = H \times \mathbf{C}$  is a dense Lie subalgebra of  $\mathfrak{g}$ , called the Cameron–Martin Lie subalgebra, and likewise  $\mathbb{G}_{CM} = H \times \mathbf{C} \subset \mathbb{G}$  is a dense subgroup of  $\mathbb{G}$ .

If  $B_t$  is a standard Brownian motion on  $(W, \mu)$ , we may define a left-invariant Brownian motion  $g_t$  on  $\mathbb{G}$  by the formula  $g_t = \left( B_t, \frac{1}{2} \int_0^t [B_s, dB_s] \right)$ . Let  $\nu_t = \text{Law}(g_{2t})$  be the heat kernel measure induced by  $g_t$ . By analogy with the finite-dimensional Heisenberg group, one expects the measure  $\nu_t$  to be “smooth” in some sense. One cannot express this smoothness in terms of a density with respect to Lebesgue or Haar measure because the latter do not exist in infinite dimensions, but another reasonable notion of smoothness would be for  $\nu_t$  to be quasi-invariant under left translation by elements of the Cameron–Martin subgroup  $\mathbb{G}_{CM}$ . The main result of [9] is that this is in fact the case. (We also mention [18] where the same statement was shown through different means, by producing a density of  $\nu_t$  with respect to the measure  $\mu \times m$ , where  $\mu$  is the Gaussian measure on  $W$  and  $m$  is Lebesgue measure on  $\mathbf{C}$ .)

It is shown in [9] that the group  $\mathbb{G}$  can be approximated by finite-dimensional projection groups  $\mathbb{G}_P$ , each of which is a nilpotent Lie group of step 2. This leads to a notion of smooth cylinder functions  $F : \mathbb{G} \rightarrow \mathbb{R}$  which can be differentiated in directions  $X \in \mathfrak{g}_{CM}$ , and thus a horizontal gradient  $\nabla_H F : \mathbb{G} \rightarrow H$  can be defined for such functions. If  $\gamma : [0, 1] \rightarrow \mathbb{G}_{CM}$  is an absolutely continuous horizontal path, then its derivative  $\gamma'$  can be identified as a curve in  $H$ , and we have the chain rule  $\frac{d}{ds} F(\gamma(s)) = \langle \nabla_H F(\gamma(s)), \gamma'(s) \rangle_H$ . Moreover,  $\mathbb{G}_{CM}$  is a length space with respect to the horizontal distance  $d_{CM}$ , and so the estimates on  $W_{a,b}(\delta_0, \delta_g), T_{a,b}(\delta_0, \delta_g)$  from Propositions 3.3 and 5.12 go through for  $g \in \mathbb{G}_{CM}$ , with  $d = d_{CM}$ .

Now [9, Proposition 4.8] shows, by means of generalized curvature-dimension inequalities as introduced in [8], that each projection group  $\mathbb{G}_P$  satisfies a reverse logarithmic Sobolev inequality, with a uniform constant of the form  $C/t$  where  $C$  depends only on the structure of  $\mathbb{G}$ , and not on the projection. This can be restated as the following reverse logarithmic Sobolev inequality for cylinder functions on  $\mathbb{G}$ :

$$P_t F \|\nabla \ln P_t F\| \leq \frac{C}{t} (P_t(F \ln F) - P_t f \ln P_t f) \tag{6.11}$$

and so as in Proposition 6.2 above, we recover a version of the main quasi-invariance result of [9] and [18]:

**Proposition 6.3.** *For each  $t > 0$ , the heat kernel measure  $\nu_t$  on  $\mathbb{G}$  is quasi-invariant under left translation by elements of  $\mathbb{G}_{CM}$ .*

Moreover, the bounds on  $T_{0,b}(\nu_t, \nu_t^g)$  in (ETI) yield  $L^q$  bounds on the Radon–Nikodym derivative  $d\nu_t^g/d\nu$ , as in the proof of Theorem 5.15, which are asymptotically equivalent to the integrated Harnack inequalities of [9, Section 5.2] as  $p \rightarrow 1$  and  $q \rightarrow \infty$ .

## References

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR-2401600
- [2] Dominique Bakry, *Functional inequalities for Markov semigroups*, Probability measures on groups: recent directions and trends, Tata Inst. Fund. Res., Mumbai, 2006, pp. 91–147. MR-2213477

- [3] Fabrice Baudoin, *Bakry-Émery meet Villani*, J. Funct. Anal. **273** (2017), no. 7, 2275–2291. MR-3677826
- [4] Fabrice Baudoin, *Geometric inequalities on Riemannian and sub-Riemannian manifolds by heat semigroups techniques*, arXiv:1801.05702, to appear in Levico Summer School Lecture notes. (2018).
- [5] Fabrice Baudoin and Michel Bonnefont, *Log-Sobolev inequalities for subelliptic operators satisfying a generalized curvature dimension inequality*, J. Funct. Anal. **262** (2012), no. 6, 2646–2676. MR-2885961
- [6] Fabrice Baudoin and Michel Bonnefont, *Reverse Poincaré inequalities, isoperimetry, and Riesz transforms in Carnot groups*, Nonlinear Anal. **131** (2016), 48–59. MR-3427969
- [7] Fabrice Baudoin and Matthew Cecil, *The subelliptic heat kernel on the three-dimensional solvable Lie groups*, Forum Math. **27** (2015), no. 4, 2051–2086. MR-3365789
- [8] Fabrice Baudoin and Nicola Garofalo, *Curvature-dimension inequalities and Ricci lower bounds for sub-Riemannian manifolds with transverse symmetries*, J. Eur. Math. Soc. (JEMS) **19** (2017), no. 1, 151–219. MR-3584561
- [9] Fabrice Baudoin, Maria Gordina, and Tai Melcher, *Quasi-invariance for heat kernel measures on sub-Riemannian infinite-dimensional Heisenberg groups*, Trans. Amer. Math. Soc. **365** (2013), no. 8, 4313–4350. MR-3055697
- [10] Fabrice Baudoin, Erlend Grong, Gianmarco Molino, and Luca Rizzi, *H-type foliations*, 2018, arXiv:1812.02563.
- [11] Jean-Michel Bismut, *Large deviations and the Malliavin calculus*, Progress in Mathematics, vol. 45, Birkhäuser Boston Inc., Boston, MA, 1984. MR-755001
- [12] Vladimir I. Bogachev, *Gaussian measures*, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, Providence, RI, 1998. MR-1642391
- [13] Jonathan M. Borwein and Adrian S. Lewis, *Convex analysis and nonlinear optimization*, second ed., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, vol. 3, Springer, New York, 2006, Theory and examples. MR-2184742
- [14] Nicolas Bourbaki, *Integration. II. Chapters 7–9*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 2004, Translated from the 1963 and 1969 French originals by Sterling K. Berberian. MR-2098271
- [15] E. J. Brody, *An elementary proof of the Gaussian dichotomy theorem*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **20** (1971), 217–226. MR-334324
- [16] R. H. Cameron and W. T. Martin, *Transformations of Wiener integrals under translations*, Ann. of Math. (2) **45** (1944), 386–396. MR-0010346
- [17] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal. **9** (1999), no. 3, 428–517. MR-1708448
- [18] Bruce K. Driver, Nathaniel Eldredge, and Tai Melcher, *Hypoelliptic heat kernels on infinite-dimensional Heisenberg groups*, Trans. Amer. Math. Soc. **368** (2016), no. 2, 989–1022. MR-3430356
- [19] Bruce K. Driver and Maria Gordina, *Heat kernel analysis on infinite-dimensional Heisenberg groups*, J. Funct. Anal. **255** (2008), no. 9, 2395–2461. MR-2473262
- [20] K. D. Elworthy and X.-M. Li, *Formulae for the derivatives of heat semigroups*, J. Funct. Anal. **125** (1994), no. 1, 252–286. MR-1297021
- [21] J. Feldman, *Correction to “Equivalence and perpendicularity of Gaussian processes”*, Pacific J. Math. **9** (1959), 1295–1296. MR-108852
- [22] Jacob Feldman, *Equivalence and perpendicularity of Gaussian processes*, Pacific J. Math. **8** (1958), 699–708. MR-102760
- [23] Yaroslav Gaek, *On a property of normal distribution of any stochastic process*, Czechoslovak Math. J. **8(83)** (1958), 610–618. MR-104290
- [24] Hui Hsiung Kuo, *Gaussian measures in Banach spaces*, Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin, 1975. MR-0461643
- [25] Kazumasa Kuwada, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), no. 11, 3758–3774. MR-2606871

- [26] Matthias Liero, Alexander Mielke, and Giuseppe Savaré, *Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves*, SIAM J. Math. Anal. **48** (2016), no. 4, 2869–2911. MR-3542003
- [27] Matthias Liero, Alexander Mielke, and Giuseppe Savaré, *Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures*, Invent. Math. **211** (2018), no. 3, 969–1117. MR-3763404
- [28] Giulia Luise and Giuseppe Savaré, *Contraction and regularizing properties of heat flows in metric measure spaces*, Discrete Contin. Dyn. Syst. Ser. S **14** (2021), no. 1, 273–297. MR-4186212
- [29] Françoise Lust-Piquard, *Ornstein-Uhlenbeck semi-groups on stratified groups*, J. Funct. Anal. **258** (2010), no. 6, 1883–1908. MR-2578458
- [30] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400. MR-1760620
- [31] N. Th. Varopoulos, Laurent Saloff-Coste, and Thierry Coulhon, *Analysis and geometry on groups*, Cambridge Tracts in Mathematics, vol. 100, Cambridge University Press, Cambridge, 1992. MR-1218884
- [32] Max-K. von Renesse and Karl-Theodor Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), no. 7, 923–940. MR-2142879
- [33] Feng-Yu Wang, *Dimension-free Harnack inequality and its applications*, Front. Math. China **1** (2006), no. 1, 53–72. MR-2225399
- [34] Feng-Yu Wang, *Harnack inequalities on manifolds with boundary and applications*, J. Math. Pures Appl. (9) **94** (2010), no. 3, 304–321. MR-2679029
- [35] Feng-Yu Wang, *Harnack inequalities for stochastic partial differential equations*, Springer-Briefs in Mathematics, Springer, New York, 2013. MR-3099948
- [36] Feng-Yu Wang, *Derivative formula and gradient estimates for Gruschin type semigroups*, J. Theoret. Probab. **27** (2014), no. 1, 80–95. MR-3174217

**Acknowledgments.** The authors are grateful for helpful discussions with Maria Gordina, Martin Hairer, Ronan Herry, Kazumasa Kuwada, Xue-Mei Li, and Giuseppe Savaré. We also thank the anonymous referee for their careful reading and useful suggestions. This article was completed during a sabbatical visit by author N. Eldredge to the Department of Mathematics at the University of Connecticut; he would like to thank the Department and especially Maria Gordina for their hospitality, especially in view of the difficult circumstances created by the COVID-19 pandemic.