

# Gradient bounds for Kolmogorov type diffusions

Fabrice Baudoin<sup>a,1</sup>, Maria Gordina<sup>a,2,3</sup> and Phanuel Mariano<sup>b,3</sup>

<sup>a</sup>*Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA. E-mail: fabrice.baudoin@uconn.edu; maria.gordina@uconn.edu*  
<sup>b</sup>*Department of Mathematics and Physics, University of New Haven, West Haven, CT 06516, USA. E-mail: pmariano@newhaven.edu*

Received 8 March 2018; revised 16 January 2019; accepted 26 February 2019

**Abstract.** We study gradient bounds and other functional inequalities for the diffusion semigroup generated by Kolmogorov type operators. The focus is on two different methods: coupling techniques and generalized  $\Gamma$ -calculus techniques. The advantages and drawbacks of each of these methods are discussed.

**Résumé.** Nous étudions des bornes de gradient et autres inégalités fonctionnelles pour des semigroupes de diffusion engendrés par des opérateurs de type Kolmogorov. Nous nous concentrons sur deux méthodes : couplage et  $\Gamma$ -calculus. Les avantages et inconvénients de ces méthodes sont discutés.

*MSC:* Primary 60J60; secondary 60J45; 58J65; 35H10

*Keywords:* Coupling; Hypoelliptic diffusion; Kolmogorov diffusion; Curvature-dimension inequality; Gradient estimates

## 1. Introduction

In the last few years, there has been considerable interest in studying gradient bounds for semigroups generated by hypoelliptic diffusion operators. The motivation for such bounds comes from their potential applications to sub-Riemannian geometry (e.g. [14,15]), quasi-invariance of heat kernel measures in infinite dimensions (e.g. [16,35]), functional inequalities such as Poincaré and log-Sobolev type inequalities (e.g. [13,25,46,56]), and the study of convergence to equilibrium for hypocoercive diffusions (e.g. [12,17]). In particular, the gradient bounds we present in this paper might be used to prove the existence of a spectral gap similarly to [18] once one has spectral localization tools. In the present paper we are interested in gradient bounds for Kolmogorov type diffusion operators for which we present and compare two different techniques:  $\Gamma$ -calculus methods and coupling techniques.

The Kolmogorov operator on  $\mathbb{R}^2$  defined as  $L = \frac{1}{2} \frac{\partial^2}{\partial x^2} + x \frac{\partial}{\partial y}$  was initially introduced by A. N. Kolmogorov in [45], where he obtained an explicit expression for the transition density of the diffusion process whose generator is the operator  $L$ . Later L. Hörmander in [41] used this operator as the simplest example of a hypoelliptic second order differential operator. The semigroup generated by  $L$  is Gaussian and thus the corresponding heat kernel may be computed explicitly, as was observed already by A. N. Kolmogorov. However, despite an explicit Gaussian heat kernel, it is somehow challenging to derive relevant functional inequalities for this semigroup. We refer for instance to R. Hamilton’s notes [40], where Riccati type equations are used to prove Li–Yau and parabolic Harnack inequalities. This (classical) Kolmogorov operator is the starting point for our consideration of several hypoelliptic operators.

As we mention above we present two techniques to prove gradient estimates in this setting. The first technique is based on a version of the Bakry–Émery  $\Gamma$ -calculus originally introduced in [7]. Recall that the generator of the Kolmogorov diffusion is hypoelliptic but not elliptic, and therefore we cannot rely on the curvature-dimension inequalities to prove the gradient bounds as presented in [8, Section 3.2.3]. The main novelty in our work is to introduce a family of operators  $\Gamma^{\alpha,\beta}$  and  $\Gamma_2^{\alpha,\beta}$  in Notation 2.2 for which one is able to reproduce the Bakry–Émery arguments with some modifications in

<sup>1</sup>Research was supported in part by NSF Grant DMS-1660031.

<sup>2</sup>Research was supported in part by the Simons Fellowship.

<sup>3</sup>Research was supported in part by NSF Grants DMS-1405169, DMS-1712427.

this degenerate setting. We note that over the last years versions of the  $\Gamma$ -calculus have already been used for hypoelliptic operators (e.g. [13–15]). However, those references consider hypoelliptic operators satisfying the strong Hörmander's condition and the methods developed there do not apply to the Kolmogorov operator. While [12] considers hypoelliptic operators satisfying the weak Hörmander's condition, it is mostly concerned with hypocoercive estimates. The key idea in our work is that to get sharp gradient estimates, we make the parameters  $\alpha$  and  $\beta$  time-dependent as observed in Remark 2.4 and the proof of Theorem 2.5.

The second technique is coupling. The coupling techniques have seen recent progress for such degenerate operators. In [19], the authors were the first to consider couplings of hypoelliptic diffusions, as they prove existence of successful coupling for the Kolmogorov diffusion and Brownian motion on the Heisenberg group. Then in [10], S. Banerjee and W. Kendall used a non-Markovian strategy to couple the iterated Kolmogorov diffusion. The most relevant to our results is [9], where coupling techniques have been used to prove gradient estimates on the Heisenberg group considered as a sub-Riemannian manifold.

The paper is organized as follows. We start by considering Kolmogorov diffusions in Section 2, where we use both generalized  $\Gamma$ -calculus and coupling techniques to prove gradient estimates such as in Proposition 2.5 and Proposition 2.10. This setting provides the first illustration to contrast these two methods: while the coupling method is somewhat simpler, and yields a family of gradient estimates, other functional inequalities such as the reverse Poincaré and the reverse log-Sobolev inequalities for the corresponding semigroup do not seem to be trackable by coupling techniques. But we can prove these inequalities by using the generalized  $\Gamma$ -calculus. Moreover, we are able to use only this approach (not the coupling techniques) to obtain sharper gradient bounds for the relativistic diffusion considered in Section 3. The relativistic diffusion has been introduced by R. Dudley and studied extensively in [3, 5, 26–28, 30, 33, 34, 43, 44, 49]. We refer the reader to [29] for the history of related objects both in mathematics and physics.

In Section 4.3 we use the coupling by parallel translation on Riemannian manifolds. The coupling can be described by a central limit theorem argument for the geodesic random walks as in [52]. It would be interesting to see if such a coupling can be carried out on sub-Riemannian manifolds using the approximation of Brownian motion by random walks as studied in [2, 22, 37]. If such a coupling can be constructed, then our results and techniques would be valid for an even larger class of hypoelliptic diffusions.

## 2. Kolmogorov diffusion in $\mathbb{R}^d \times \mathbb{R}^d$

Our main object in this section is a Kolmogorov diffusion in  $\mathbb{R}^d \times \mathbb{R}^d$  defined by

$$\mathbf{X}_t = \left( B_t, \int_0^t B_s \, ds \right),$$

where  $B_t$  is a Brownian motion in  $\mathbb{R}^d$  with the variance  $\sigma^2$ .

**Definition 2.1.** Let  $f(p, \xi)$ ,  $p \in \mathbb{R}^d$ ,  $\xi \in \mathbb{R}^d$  be a function on  $\mathbb{R}^d \times \mathbb{R}^d$ . For  $\sigma > 0$ , the Kolmogorov operator for  $f \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$  is defined by

$$(Lf)(p, \xi) := \langle p, \nabla_\xi f(p, \xi) \rangle + \frac{\sigma^2}{2} \Delta_p f(p, \xi) = \sum_{j=1}^d p_j \frac{\partial f}{\partial \xi_j}(p, \xi) + \frac{\sigma^2}{2} \Delta_p f(p, \xi),$$

where  $\Delta_p$  is the Laplace operator  $\Delta$  on  $\mathbb{R}^d$  acting on the variable  $p$  and  $\nabla_\xi$  is the gradient on  $\mathbb{R}^d$  acting on the variable  $\xi$ .

Note that for  $d = 1$  and  $\sigma = 1$  this is the original Kolmogorov operator. By Hörmander's theorem in [41], the operator  $L$  is hypoelliptic and generates a Markov process  $X_t$ . It follows then that the process  $X_t$  admits a smooth transition probability density with respect to the Lebesgue measure.

### 2.1. $\Gamma$ -calculus

First we use geometric methods such as generalized  $\Gamma$ -calculus to prove gradient bounds for the semigroup generated by the Kolmogorov operator  $L$ . Moreover, we show that the estimate is sharp. We point out that a generalization of  $\Gamma$ -calculus for the Kolmogorov operator has been carried by F. Y. Wang in [55, pp. 300–303]. However, our methods are different and yield optimal results as we explain in Remark 2.6.

Recall that the *carré du champ* operator for  $L$  is defined by

$$\Gamma(f) := \frac{1}{2} Lf^2 - f Lf,$$

where  $f$  is from an appropriate space of functions which will be specified later. A straightforward computation shows that

$$\Gamma(f) = \frac{1}{2} \sigma^2 \|\nabla_p f\|^2, \quad (2.1)$$

where  $\nabla_p$  is the standard gradient operator on  $\mathbb{R}^d$  acting on the variable  $p$ , and  $\|\cdot\|$  is the  $\mathbb{R}^d$ -norm.

**Notation 2.2.** For  $\alpha \in \mathbb{R}$ ,  $\beta \geq 0$  we define a symmetric first-order differential bilinear form  $\Gamma^{\alpha, \beta} : C^\infty(\mathbb{R}^d \times \mathbb{R}^d) \times C^\infty(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \mathbb{R}$  by

$$\begin{aligned} \Gamma^{\alpha, \beta}(f, g) &:= \sum_{i=1}^d \left( \frac{\partial f}{\partial p_i} - \alpha \frac{\partial f}{\partial \xi_i} \right) \left( \frac{\partial g}{\partial p_i} - \alpha \frac{\partial g}{\partial \xi_i} \right) + \beta \sum_{i=1}^d \frac{\partial f}{\partial \xi_i} \frac{\partial g}{\partial \xi_i} \\ &= \langle \nabla_p f, \nabla_p g \rangle - \alpha \langle \nabla_p f, \nabla_\xi g \rangle - \alpha \langle \nabla_\xi f, \nabla_p g \rangle + (\alpha^2 + \beta) \langle \nabla_\xi f, \nabla_\xi g \rangle, \end{aligned} \quad (2.2)$$

with the usual convention that  $\Gamma^{\alpha, \beta}(f) := \Gamma^{\alpha, \beta}(f, f)$ . We will also consider

$$\Gamma_2^{\alpha, \beta}(f) = \frac{1}{2} L \Gamma^{\alpha, \beta}(f) - \Gamma^{\alpha, \beta}(f, Lf).$$

We start with the following key lemma.

**Lemma 2.3.** For  $f \in C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$

$$\Gamma_2^{\alpha, \beta}(f) \geq \alpha \sum_{i=1}^d \left( \frac{\partial f}{\partial \xi_i} \right)^2 - \sum_{i=1}^d \frac{\partial f}{\partial \xi_i} \frac{\partial f}{\partial p_i} = \alpha \|\nabla_\xi f\|^2 - \langle \nabla_\xi f, \nabla_p f \rangle.$$

**Proof.** Let  $\alpha \in \mathbb{R}$ ,  $\beta \geq 0$ . A computation shows that

$$\begin{aligned} \Gamma_2^{\alpha, \beta}(f) &= \alpha \sum_{i=1}^d \left( \frac{\partial f}{\partial \xi_i} \right)^2 - \sum_{i=1}^d \frac{\partial f}{\partial \xi_i} \frac{\partial f}{\partial p_i} \\ &\quad + \frac{\sigma^2}{2} \sum_{i=1}^d \sum_{j=1}^d \left( \frac{\partial^2 f}{\partial p_i \partial p_j} - \alpha \frac{\partial^2 f}{\partial p_i \partial \xi_j} \right)^2 \\ &\quad + \frac{\sigma^2}{2} \beta \sum_{i=1}^d \sum_{j=1}^d \left( \frac{\partial^2 f}{\partial p_i \partial \xi_j} \right)^2 \\ &\geq \alpha \sum_{i=1}^d \left( \frac{\partial f}{\partial \xi_i} \right)^2 - \sum_{i=1}^d \frac{\partial f}{\partial \xi_i} \frac{\partial f}{\partial p_i}. \end{aligned}$$

□

We are now in position to prove regularization properties for the semigroup  $P_t = e^{tL}$ . But first we have the following remark that will make the proofs in this section easier to read.

**Remark 2.4.** We will repeatedly use the following simple computation. Suppose  $\alpha(s), \beta(s) \in C^1([0, \infty))$ . Then for  $f \in C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$

$$\begin{aligned} \phi'(s) &= 2P_s \left( \Gamma_2^{\alpha(s), \beta(s)}(P_{t-s} f) \right) \\ &\quad - 2\alpha'(s) P_s \langle \nabla_p P_{t-s} f, \nabla_\xi P_{t-s} f \rangle + (\alpha'(s)\alpha(s) + \beta'(s)) P_s \|\nabla_\xi P_{t-s} f\|^2, \end{aligned} \quad (2.3)$$

where  $\phi$  is the functional

$$\phi(s) := P_s(\Gamma^{\alpha(s), \beta(s)}(P_{t-s}f)), \quad 0 \leq s \leq t,$$

**Proposition 2.5 (Bakry–Émery type estimate).** *Let  $f \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$  be a globally Lipschitz function, then one has*

$$\|\nabla_p P_t f\|^2 \leq \sum_{i=1}^d P_t \left( \frac{\partial f}{\partial p_i} + t \frac{\partial f}{\partial \xi_i} \right)^2,$$

and

$$\|\nabla_\xi P_t f\|^2 \leq P_t \|\nabla_\xi f\|^2.$$

**Proof.** Let  $t > 0$ . We first assume that  $f$  is smooth and rapidly decreasing. In that case, the following computations are easily justified since  $P_t$  has a Gaussian kernel (see [23, pp. 80–85]). We consider then (at a given fixed point  $(\xi, p)$ ) the functional

$$\phi(s) = P_s(\Gamma^{\alpha(s), \beta}(P_{t-s}f)), \quad 0 \leq s \leq t,$$

where  $\alpha(s) = -s$  and  $\beta$  is a non-negative constant. Then by (2.3) and Lemma 2.3

$$\begin{aligned} \phi'(s) &= 2P_s(\Gamma_2^{\alpha(s), \beta}(P_{t-s}f) + \langle \nabla_p(P_{t-s}f), \nabla_\xi(P_{t-s}f) \rangle + s \|\nabla_\xi(P_{t-s}f)\|^2) \\ &\geq 2P_s(\alpha(s) \|\nabla_\xi(P_{t-s}f)\|^2 - \langle \nabla_\xi(P_{t-s}f), \nabla_p(P_{t-s}f) \rangle) \\ &\quad + \langle \nabla_\xi(P_{t-s}f), \nabla_p(P_{t-s}f) \rangle + s \|\nabla_\xi P_{t-s}f\|^2 = 0. \end{aligned}$$

Thus  $\phi$  is increasing, and therefore  $\phi(0) \leq \phi(t)$ , that is,

$$\Gamma^{\alpha(0), \beta}(P_t f) \leq P_t(\Gamma^{\alpha(t), \beta}(f)).$$

The result follows immediately by taking  $\beta = 0$ . Now, if  $f \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$  is a Lipschitz function, then for any  $s > 0$ , the function  $P_s f$  is smooth and rapidly decreasing (again, since  $P_s$  has a Gaussian kernel). Therefore, applying the inequality we have proved to  $P_s f$  yields

$$\|\nabla_p P_{t+s} f\|^2 \leq \sum_{i=1}^d P_t \left( \frac{\partial P_s f}{\partial p_i} + t \frac{\partial P_s f}{\partial \xi_i} \right)^2.$$

Letting  $s \rightarrow 0$  concludes the argument. To justify this limit one can first observe that since  $f$  is Lipschitz then  $P_s f \rightarrow f$  as  $s \rightarrow 0$ . Then one can also show that  $P_s f$  is dominated by  $g_s(p, \xi) = c_1(|p| + |\xi| + \sqrt{s}) + c_2$  for  $0 < s < 1$  since  $f$  is Lipschitz. A dominated convergence argument finishes the proof.  $\square$

**Remark 2.6 (Bakry–Émery type estimate is sharp).** Suppose  $l$  is any linear form on  $\mathbb{R}^d$ , we define the function  $f(p, \xi) := l(\xi)$ . Note that  $f$  is Lipschitz since  $f$  is linear. Then for every  $(p, \xi) \in \mathbb{R}^d \times \mathbb{R}^d$  and  $t \geq 0$  we have

$$P_t f(p, \xi) = \mathbb{E} \left( f \left( B_t + p, \xi + tp + \int_0^t B_s ds \right) \right) = l(\xi) + tl(p).$$

For this choice of  $f$ , one has  $\|\nabla_p P_t f\|^2 = t^2 \|l\|^2$  and

$$\sum_{i=1}^d P_t \left( \frac{\partial f}{\partial p_i} + t \frac{\partial f}{\partial \xi_i} \right)^2 = t^2 \|l\|^2.$$

Similarly, for this choice of  $f$ ,  $\|\nabla_\xi P_t f\|^2 = P_t \|\nabla_\xi f\|^2$ . So the bounds in Proposition 2.5 are sharp.

**Proposition 2.7 (Reverse Poincaré inequality).** *Let  $f \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$  be a bounded function, then for  $t > 0$*

$$\sum_{i=1}^d \left( \frac{\partial P_t f}{\partial p_i} - \frac{1}{2} t \frac{\partial P_t f}{\partial \xi_i} \right)^2 + \frac{t^2}{12} \left( \frac{\partial P_t f}{\partial \xi_i} \right)^2 \leq \frac{1}{\sigma^2 t} (P_t f^2 - (P_t f)^2).$$

**Proof.** Let  $t > 0$ . By using the same argument as in the previous proof, we can assume that  $f$  is smooth and rapidly decreasing. We consider the functional

$$\phi(s) = (t-s) P_s(\Gamma^{\alpha(s), \beta(s)}(P_{t-s} f)), \quad 0 \leq s \leq t,$$

where  $\alpha(s) = \frac{1}{2}(t-s)$  and  $\beta(s) = \frac{1}{12}(t-s)^2$ . By (2.1), (2.2), (2.3) and Lemma 2.3 we have

$$\begin{aligned} \phi'(s) &= -P_s(\Gamma^{\alpha(s), \beta(s)}(P_{t-s} f)) \\ &\quad + (t-s) P_s \langle \nabla_p P_{t-s} f, \nabla_\xi P_{t-s} f \rangle + (t-s)(2\alpha'(s)\alpha(s) + \beta'(s)) P_s \|\nabla_\xi P_{t-s} f\|^2 \\ &\geq -P_s(\|\nabla_p P_{t-s} f\|^2) = -\frac{2}{\sigma^2} P_s(\Gamma(P_{t-s} f)). \end{aligned}$$

Therefore, we have

$$\phi(0) \leq \frac{2}{\sigma^2} \int_0^t P_s(\Gamma(P_{t-s} f)) ds,$$

where we used the fact that  $\phi$  is positive. We now observe that

$$\frac{2}{\sigma^2} \int_0^t P_s(\Gamma(P_{t-s} f)) ds = \frac{1}{\sigma^2} (P_t f^2 - (P_t f)^2).$$

Therefore, we conclude

$$t \Gamma^{\alpha(0), \beta(0)}(P_t f) \leq \frac{1}{\sigma^2} (P_t f^2 - (P_t f)^2). \quad \square$$

**Proposition 2.8 (Reverse log-Sobolev inequality).** *Let  $f \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$  be a non-negative bounded function. One has for  $t > 0$*

$$\sum_{i=1}^d \left( \frac{\partial \ln P_t f}{\partial p_i} - \frac{1}{2} t \frac{\partial \ln P_t f}{\partial \xi_i} \right)^2 + \frac{1}{12} t^2 \left( \frac{\partial \ln P_t f}{\partial \xi_i} \right)^2 \leq \frac{2}{\sigma^2 t P_t f} (P_t(f \ln f) - P_t f \ln P_t f).$$

**Proof.** As before, we can assume that  $f$  is smooth, non-negative and rapidly decreasing. Let  $t > 0$ . We consider the functional

$$\phi(s) = (t-s) P_s((P_{t-s} f) \Gamma^{\alpha(s), \beta(s)}(\ln P_{t-s} f)), \quad 0 \leq s \leq t,$$

where  $\alpha(s) = \frac{1}{2}(t-s)$  and  $\beta(s) = \frac{1}{12}(t-s)^2$ . Similarly to the previous proofs we have

$$\begin{aligned} \phi'(s) &= -P_s((P_{t-s} f) \Gamma^{\alpha(s), \beta(s)}(\ln P_{t-s} f)) + 2(t-s) P_s((P_{t-s} f) \Gamma^{\alpha(s), \beta(s)}(\ln P_{t-s} f)) \\ &\quad - 2(t-s)\alpha'(s) \sum_{i=1}^d P_s \left( (P_{t-s} f) \frac{\partial \ln P_{t-s} f}{\partial \xi_i} \frac{\partial \ln P_{t-s} f}{\partial p_i} \right) \\ &\quad + 2(t-s)\alpha(s)\alpha'(s) P_s[(P_{t-s} f) \|\nabla_\xi \ln P_{t-s} f\|^2] \\ &\quad + (t-s)\beta'(s) P_s[(P_{t-s} f) \|\nabla_\xi \ln P_{t-s} f\|^2] \\ &\geq -P_s((P_{t-s} f) \|\nabla_p \ln P_{t-s} f\|^2) = -\frac{2}{\sigma^2} P_s((P_{t-s} f) \Gamma(\ln P_{t-s} f)). \end{aligned}$$

Therefore, we have

$$\phi(0) \leq \frac{2}{\sigma^2} \int_0^t P_s((P_{t-s}f)\Gamma(\ln P_{t-s}f)) ds.$$

We now observe that

$$2 \int_0^t P_s((P_{t-s}f)\Gamma(\ln P_{t-s}f)) ds = 2(P_t(f \ln f) - P_t f \ln P_t f),$$

and therefore

$$t(P_t f)\Gamma^{\alpha(0),\beta(0)}(\ln P_t f) \leq \frac{2}{\sigma^2}(P_t(f \ln f) - P_t f \ln P_t f). \quad \square$$

The fact that the reverse log-Sobolev inequality implies a Wang–Harnack inequality for general Markov operators is by now well-known (see for instance [13, Proposition 3.4]). We deduce therefore the following functional inequality.

**Theorem 2.9 (Wang–Harnack inequality).** *Let  $f$  be a non-negative Borel bounded function on  $\mathbb{R}^d \times \mathbb{R}^d$ . Then for every  $t > 0$ ,  $(p, \xi), (p', \xi') \in \mathbb{R}^d \times \mathbb{R}^d$  and  $\alpha > 1$  we have*

$$(P_t f)^\alpha(p, \xi) \leq C_\alpha(t, (p, \xi), (p', \xi'))(P_t f^\alpha)(p', \xi'),$$

where

$$\begin{aligned} C_\alpha(t, (p, \xi), (p', \xi')) \\ := \exp\left(\frac{\alpha}{\alpha-1}\left(\frac{6}{\sigma^2 t^3} \sum_{i=1}^d \left(\frac{t}{2}(p'_i - p_i) + (\xi'_i - \xi_i)\right)^2 + \frac{1}{2\sigma^2 t} \sum_{i=1}^d (p'_i - p_i)^2\right)\right). \end{aligned}$$

**Proof.** As before we assume that  $f$  is non-negative and rapidly decreasing. Let  $t > 0$  be fixed and  $(p, \xi), (p', \xi') \in \mathbb{R}^d \times \mathbb{R}^d$ . We observe first that the reverse log-Sobolev inequality in Proposition 2.8 can be rewritten

$$\Gamma^{\frac{1}{2}t, \frac{1}{12}t^2}(\ln P_t f) \leq \frac{2}{t\sigma^2 P_t f}(P_t(f \ln f) - P_t f \ln P_t f).$$

We can now integrate the previous inequality as in [13, Proposition 3.4] and deduce

$$(P_t f)^\alpha(p, \xi) \leq (P_t f^\alpha)(p', \xi') \exp\left(\frac{\alpha}{\alpha-1} \frac{d_t^2((p, \xi), (p', \xi'))}{2\sigma^2 t}\right),$$

where  $d_t$  is the control distance associated to the gradient  $\Gamma^{\frac{t}{2}, \frac{t^2}{12}}$  defined by (2.2). Therefore

$$\begin{aligned} d_t^2((p, \xi), (p', \xi')) \\ = \frac{12}{t^2} \sum_{i=1}^d \left(\frac{1}{2}t(p'_i - p_i) + (\xi'_i - \xi_i)\right)^2 + \sum_{i=1}^d (p'_i - p_i)^2 \\ = 4 \sum_{i=1}^d (p'_i - p_i)^2 + \frac{12}{t} \sum_{i=1}^d (p'_i - p_i)(\xi'_i - \xi_i) + \frac{12}{t^2} \sum_{i=1}^d (\xi'_i - \xi_i)^2 \end{aligned}$$

and the proof is complete.  $\square$

## 2.2. Coupling

In this section, we use coupling techniques to prove Proposition 2.5 under slightly different assumptions. We start by recalling the notion of a coupling. Suppose  $(\Omega, \mathcal{F}, \mathbb{P})$  is a probability space, and  $X_t$  and  $\tilde{X}_t$  are two diffusions in  $\mathbb{R}^d$  defined on this space with the same generator  $L$ , starting at  $x, \tilde{x} \in \mathbb{R}^d$  respectively. By their coupling we understand a

diffusion  $(X_t, \tilde{X}_t)$  in  $\mathbb{R}^d \times \mathbb{R}^d$  such that its law is a coupling of the laws of  $X_t$  and  $\tilde{X}_t$ . That is, the first and the second  $d$ -dimensional (marginal) distributions of  $(X_t, \tilde{X}_t)$  are given by distributions of  $X_t$  and  $\tilde{X}_t$ .

Let  $\mathbb{P}^{(x, \tilde{x})}$  be the distribution of  $(X_t, \tilde{X}_t)$ , so that  $\mathbb{P}^{(x, \tilde{x})}(X_0 = x, \tilde{X}_0 = \tilde{x}) = 1$ . We denote by  $\mathbb{E}^{(x, \tilde{x})}$  the expectation with respect to the probability measure  $\mathbb{P}^{(x, \tilde{x})}$ .

To prove Proposition 2.10, we use the synchronous coupling of Brownian motions in  $\mathbb{R}^d$ . That is, for  $(p, \tilde{p}) \in \mathbb{R}^d \times \mathbb{R}^d$  we let  $B_t^p = p + B_t$  and  $\tilde{B}_t^{\tilde{p}} = \tilde{p} + \tilde{B}_t$ , where  $B_t$  is a standard Brownian motion in  $\mathbb{R}^d$ .

**Proposition 2.10.** *Let  $f \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$  with bounded second derivatives. If  $1 \leq q < \infty$  then for  $t \geq 0$ ,*

$$\|\nabla_p P_t f\|^q \leq \sum_{i=1}^d P_t \left( \left| \frac{\partial f}{\partial p_i} + t \frac{\partial f}{\partial \xi_i} \right|^q \right).$$

**Proof.** Consider two copies of Kolmogorov diffusions

$$\begin{aligned} \mathbf{X}_t &= (B_t^p, Y_t) = \left( p + B_t, \xi + tp + \int_0^t B_s ds \right), \\ \tilde{\mathbf{X}}_t &= (\tilde{B}_t^{\tilde{p}}, \tilde{Y}_t) = \left( \tilde{p} + \tilde{B}_t, \xi + t\tilde{p} + \int_0^t \tilde{B}_s ds \right), \end{aligned}$$

where  $B_t$  and  $\tilde{B}_t$  are two Brownian motions started at 0. Note that  $\mathbf{X}_t$  starts at  $(p, \xi)$  and  $\tilde{\mathbf{X}}_t$  starts at  $(\tilde{p}, \xi)$ . In order to construct a coupling of  $(\mathbf{X}_t, \tilde{\mathbf{X}}_t)$  it suffices to couple  $(B_t, \tilde{B}_t)$ . Let us synchronously couple  $(B_t, \tilde{B}_t)$  for all time so that

$$\begin{aligned} |B_t^p - \tilde{B}_t^{\tilde{p}}| &= |p - \tilde{p}|, \\ |Y_t - \tilde{Y}_t| &= t|p - \tilde{p}|, \end{aligned}$$

for all  $t \geq 0$ . By using an estimate on the remainder  $R$  of Taylor's approximation to  $f$  and the assumption that  $f \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$  has bounded second derivatives, there exists a  $C_f \geq 0$  such that

$$\begin{aligned} &|f(\mathbf{X}_t) - f(\tilde{\mathbf{X}}_t)| \\ &= \left| \sum_{i=1}^d \partial_{p_i} f(\tilde{\mathbf{X}}_t)(p_i - \tilde{p}_i) + \sum_{i=1}^d t \partial_{\xi_i} f(\tilde{\mathbf{X}}_t)(p_i - \tilde{p}_i) + R(\tilde{\mathbf{X}}_t) \right| \\ &\leq \sum_{i=1}^d |(\partial_{p_i} f(\tilde{\mathbf{X}}_t) + t \partial_{\xi_i} f(\tilde{\mathbf{X}}_t))| |p - \tilde{p}| + \frac{C_f}{2} d^2 (1+t)^2 |p - \tilde{p}|^2. \end{aligned}$$

Using this estimate and Jensen's inequality we see that

$$\begin{aligned} &|P_t f(p, \xi) - P_t f(\tilde{p}, \xi)| \\ &= |\mathbb{E}^{((p, \xi), (\tilde{p}, \xi))}[f(\mathbf{X}_t) - f(\tilde{\mathbf{X}}_t)]| \\ &\leq \mathbb{E}^{((p, \xi), (\tilde{p}, \xi))}[|f(\mathbf{X}_t) - f(\tilde{\mathbf{X}}_t)|] \\ &\leq \sum_{i=1}^d \mathbb{E}^{((p, \xi), (\tilde{p}, \xi))}[|(\partial_{p_i} f(\tilde{\mathbf{X}}_t) + t \partial_{\xi_i} f(\tilde{\mathbf{X}}_t))|^q]^{\frac{1}{q}} |p - \tilde{p}| \\ &\quad + \frac{C_f}{2} d^2 (1+t)^2 |p - \tilde{p}|^2 \\ &= \sum_{i=1}^d P_t \left( |(\partial_{p_i} f(\tilde{p}, \xi) + t \partial_{\xi_i} f(\tilde{p}, \xi))|^q \right)^{\frac{1}{q}} |p - \tilde{p}| \\ &\quad + \frac{C_f}{2} d^2 (1+t)^2 |p - \tilde{p}|^2. \end{aligned}$$

Dividing out by  $|p - \tilde{p}|$  and taking  $\tilde{p} \rightarrow p$  we have that

$$\begin{aligned} \|\nabla_p P_t f(p, \xi)\| &= \limsup_{\tilde{p} \rightarrow p} \frac{|P_t f(p, \xi) - P_t f(\tilde{p}, \xi)|}{|p - \tilde{p}|} \\ &\leq \sum_{i=1}^d P_t (|\partial_{p_i} f(p, \xi) + t \partial_{\xi_i} f(p, \xi)|^q)^{\frac{1}{q}}, \end{aligned}$$

which proves the statement.  $\square$

**Remark 2.11.** When  $q = 2$ , this coincides with the conclusion of Proposition 2.5. The coupling method here is simpler than the  $\Gamma$ -calculus method and moreover yields a family of inequalities for  $q \geq 1$ . However, on the other hand, it appears difficult to prove the reverse Poincaré and the reverse log-Sobolev inequalities for the semigroup by using coupling techniques.

### 3. Relativistic diffusion

In this section we consider the diffusion  $\mathbf{X}_t = (B_t, \int_0^t B_s ds)$ , where  $B_t$  is a Brownian motion on the  $d$ -dimensional hyperbolic space  $\mathbb{H}^d$ . This is the relativistic Brownian motion introduced by R. Dudley [26] and studied by J. Franchi and Y. Le Jan in [33]. In this section, we will prove functional inequalities for the generator of  $\mathbf{X}_t$ . Our methods will only involve  $\Gamma$ -calculus through generalized curvature dimension conditions. The emphasis on  $\Gamma$ -calculus in this section will allow us to obtain sharper estimates for the relativistic diffusion. In particular, the estimate (3.3) in Corollary 3.4 is sharper than the ones given in Theorems 4.3 and 4.8. In the following sections we will prove similar theorems using both  $\Gamma$ -calculus and coupling techniques but for a larger class of diffusions.

We follow the notation in [33]. Recall that the Minkowski space is the product  $\mathbb{R} \times \mathbb{R}^d$  with  $d \geq 2$

$$\mathbb{R}^{1,d} = \{\xi = (\xi_0, \vec{\xi}) \in \mathbb{R} \times \mathbb{R}^d\}$$

equipped with the Lorentzian norm  $q(\xi, \xi) := \xi_0^2 - \|\vec{\xi}\|^2$ . The standard basis in  $\mathbb{R}^{1,d}$  is denoted by  $e_0, \dots, e_d$ . Let  $\mathbb{H}^d$  be the positive half of the unit sphere in  $\mathbb{R}^{1,d}$ , namely,

$$\mathbb{H}^d := \{p \in \mathbb{R}^{1,d} : p_0 > 0, q(p, p) = 1\}.$$

Note that  $\mathbb{H}^d$  has a standard parametrization  $p = (p_0, \vec{p}) = (\cosh r, \sinh r \omega)$  with  $r \geq 0$ ,  $\omega \in \mathbb{S}^{d-1}$ . In these coordinates the hyperbolic metric is given by  $dr^2 + \sinh^2 r d\omega^2$ , where  $d\omega$  is the metric on the sphere  $\mathbb{S}^{d-1}$ , and the volume element is

$$\int_{\mathbb{H}^d} f(\Omega) d\Omega = \int_0^\infty \int_{\mathbb{S}^{d-1}} f(r, \omega) \sinh^{d-1} r dr d\omega.$$

Finally, the corresponding Laplace–Beltrami operator  $\mathbb{H}^d$  can be written in these coordinates as follows (see [34, Proposition 3.5.4]).

$$\Delta^{\mathbb{H}} f(r, \omega) := \frac{\partial^2 f}{\partial r^2}(r, \omega) + (d-1) \coth r \frac{\partial f}{\partial r}(r, \omega) + \frac{1}{\sinh^2 r} \Delta_{\mathbb{S}^{d-1}}^\omega f(r, \omega),$$

where  $\Delta_{\mathbb{S}^{d-1}}^\omega$  is the Laplace operator on  $\mathbb{S}^{d-1}$  acting on the variable  $\omega$ . We denote by  $\nabla^{\mathbb{H}}$  the gradient on  $\mathbb{H}^d$  viewed as a Riemannian manifold.

Following the construction in [26], we consider a stochastic process with values in the unitary tangent bundle  $T^1 \mathbb{R}^{1,d}$  of the Minkowski space-time  $\mathbb{R}^{1,d}$ . We identify the unit tangent bundle with  $\mathbb{H}^d \times \mathbb{R}^{1,d}$ . Then the relativistic Brownian motion is the process  $X_t := (g_t, \xi_t)$ , where  $g_t$  is a Brownian motion in  $\mathbb{H}^d$  starting at  $e_0$ , and the second process is the time integral of  $g_t$

$$\xi_t := \int_0^t g_s ds.$$

By [34, Theorem VII.6.1] the process  $X_t$  is a Markov Lorentz-invariant diffusion whose generator is the relativistic Laplacian defined as follows. For  $\sigma > 0$ , the relativistic Laplacian for  $f \in C^2(\mathbb{H}^d \times \mathbb{R}^{1,d})$  is the operator

$$\begin{aligned} (Lf)(p, \xi) &= \langle p, \nabla_\xi f(p, \xi) \rangle + \frac{\sigma^2}{2} \Delta_p^{\mathbb{H}} f(p, \xi) \\ &= p_0 \frac{\partial f}{\partial \xi_0}(p, \xi) + \sum_{j=1}^d p_j \frac{\partial f}{\partial \xi_j}(p, \xi) + \frac{\sigma^2}{2} \Delta_p^{\mathbb{H}} f(p, \xi), \end{aligned}$$

where  $\Delta_p^{\mathbb{H}}$  is the Laplace–Beltrami operator  $\Delta^{\mathbb{H}}$  on  $\mathbb{H}^d$  acting on the variable  $p$ . The operator  $L$  is hypoelliptic and generates the Markov process  $X_t$ . Let  $P_t$  be the heat semigroup with the operator  $L$  being its generator.

We consider functions on  $\mathbb{H}^d \times \mathbb{R}^{1,d}$  with  $f(p, \xi)$ ,  $p \in \mathbb{H}^d$ ,  $\xi \in \mathbb{R}^{1,d}$ . Recall that operators  $\nabla^{\mathbb{H}}$  and  $\Delta^{\mathbb{H}}$  act on the variable  $p$  for  $f(p, \xi)$ . We use  $\nabla_\xi$  for the usual Euclidean gradient. Let  $\Gamma(f)$  be the *carré du champ* operator for  $L$ . Recall that we view  $\mathbb{H}^d$  as a Riemannian manifold with  $\Delta^{\mathbb{H}}$  being the Laplace–Beltrami operator.

Our main result of this section is a generalized curvature-dimension inequality for  $\mathbb{H}^d \times \mathbb{R}^{1,d}$  with the operator  $L$  and  $\nabla_\xi$  playing a role of the vertical gradient. Namely, we define a symmetric, first-order differential bilinear form  $\Gamma^Z : C^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d}) \times C^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d}) \rightarrow \mathbb{R}$  by

$$\Gamma^Z(f) := \|\nabla_\xi f\|^2, \quad (3.1)$$

for any  $f \in C^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d})$ .

**Theorem 3.1 (Curvature-dimension condition).** *The operator  $L$  satisfies the following generalized curvature-dimension condition for any  $f \in C^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d})$*

$$\Gamma_2(f) \geq -\frac{d}{2} \sigma^2 \Gamma(f) - \frac{1}{4} \Gamma^Z(f),$$

$$\Gamma_2^Z(f) \geq 0.$$

**Proof.** A simple calculation of the *carré du champ* operator for  $L$  is given by

$$\Gamma(f) := \frac{1}{2} (Lf^2 - 2f Lf) = \frac{\sigma^2}{2} \|\nabla_p^{\mathbb{H}} f\|^2,$$

where as before  $\nabla_p^{\mathbb{H}}$  is the Riemannian gradient on  $\mathbb{H}^d$ . Straightforward computations show that the iterated *carré du champ* operator

$$\Gamma_2(f) := \frac{1}{2} (L\Gamma(f) - 2\Gamma(f, Lf))$$

is given by

$$\Gamma_2(f) = \frac{\sigma^4}{4} \Gamma_2^{\mathbb{H}}(f) - \frac{\sigma^2}{2} \langle \nabla_p^{\mathbb{H}} f, \nabla_\xi f \rangle,$$

where  $\Gamma_2^{\mathbb{H}}(f)$  is the iterated *carré du champ* operator for  $\Delta_p^{\mathbb{H}}$ . Recall that we view  $\mathbb{H}^d$  as a Riemannian manifold with  $\Delta^{\mathbb{H}}$  being the Laplace–Beltrami operator, therefore we can use Bochner’s formula for  $\Delta_p^{\mathbb{H}}$

$$\Gamma_2^{\mathbb{H}}(f) \geq -(d-1) \|\nabla_p^{\mathbb{H}} f\|^2,$$

thus

$$\Gamma_2(f) \geq -\frac{d-1}{2} \sigma^2 \Gamma(f) - \frac{\sigma^2}{2} \langle \nabla_p^{\mathbb{H}} f, \nabla_\xi f \rangle.$$

Now we can use an elementary estimate

$$-\frac{\sigma^2}{2} \langle \nabla_p^{\mathbb{H}} f, \nabla_\xi f \rangle \geq -\frac{\sigma^4}{4} \|\nabla_p^{\mathbb{H}} f\|^2 - \frac{1}{4} \|\nabla_\xi f\|^2 = -\frac{\sigma^2}{2} \Gamma(f) - \frac{1}{4} \|\nabla_\xi f\|^2$$

to see that

$$\Gamma_2(f) \geq -\frac{d}{2}\sigma^2\Gamma(f) - \frac{1}{4}\|\nabla_\xi f\|^2.$$

The last term in this inequality is the bilinear form  $\Gamma^Z$  defined by (3.1). Its iterated form is

$$\Gamma_2^Z(f) := \frac{1}{2}(L\Gamma^Z(f) - 2\Gamma^Z(f, Lf)),$$

for which another routine computation shows that

$$\Gamma_2^Z(f) = \frac{\sigma^2}{2}\|\nabla_\xi \nabla_p^H f\|^2 \geq 0,$$

which concludes the proof.  $\square$

For later use, our first task is to construct a convenient Lyapunov function for the operator  $L$ . A *Lyapunov function* on  $\mathbb{H}^d \times \mathbb{R}^{1,d}$  for the operator  $L$  is a smooth function  $W$  such that  $LW \leq CW$  for some  $C > 0$ . Consider the function

$$W(p, \xi) := 1 + \xi_0^2 + \|\vec{\xi}\|^2 + d_R(p_0, p)^2, \quad p \in \mathbb{H}^d, \xi \in \mathbb{R}^{1,d}, \quad (3.2)$$

where  $p_0$  is a fixed point in  $\mathbb{H}^d$  and  $d_R$  is the Riemannian distance in  $\mathbb{H}^d$ .

We observe that  $W$  is smooth since  $d_R(p_0, \cdot)^2$  is (on the hyperbolic space the exponential map at  $p_0$ , is a diffeomorphism). Using the Laplacian comparison theorem on  $\mathbb{H}^d$ , one can see that  $W$  has the following properties

$$\begin{aligned} W &\geq 1, \\ \|\nabla_\xi W\| + \|\nabla_p W\| &\leq CW, \\ LW &\leq CW \quad \text{for some constant } C > 0, \\ \{W \leq m\} &\text{ is compact for every } m. \end{aligned}$$

We shall make use of the Lyapunov function  $W$  defined by (3.2) to prove the following result.

**Theorem 3.2 (Gradient estimate).** *Consider the operator  $L$  and its corresponding heat semigroup  $P_t$ . For any  $f \in C_0^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d})$  and  $t \geq 0$*

$$2d\sigma^2\Gamma(P_t f)(x) + \Gamma^Z(P_t f)(x) \leq e^{d\sigma^2 t} (2d\sigma^2 P_t(\Gamma(f))(x) + P_t(\Gamma^Z(f))(x)).$$

**Proof.** We fix  $t > 0$  throughout the proof. For  $0 < s < t$ ,  $x \in \mathbb{H}^d \times \mathbb{R}^{1,d}$  we denote

$$\begin{aligned} \varphi_1(x, s) &:= \Gamma(P_{t-s} f)(x), \\ \varphi_2(x, s) &:= \Gamma^Z(P_{t-s} f)(x). \end{aligned}$$

Then

$$\begin{aligned} L\varphi_1 + \frac{\partial \varphi_1}{\partial s} &= 2\Gamma_2(P_{t-s} f), \\ L\varphi_2 + \frac{\partial \varphi_2}{\partial s} &= 2\Gamma_2^Z(P_{t-s} f). \end{aligned}$$

Now we would like to find two non-negative smooth functions  $a(s)$  and  $b(s)$  such that for

$$\varphi(x, s) := a(s)\varphi_1(x, s) + b(s)\varphi_2(x, s),$$

we have

$$L\varphi + \frac{\partial \varphi}{\partial s} \geq 0.$$

Then by Theorem 3.1 we have

$$\begin{aligned}
L\varphi + \frac{\partial \varphi}{\partial s} &= a'(s)\Gamma(P_{t-s}f) + b'(s)\Gamma^Z(P_{t-s}f) + 2a(s)\Gamma_2(P_{t-s}f) + 2b(s)\Gamma_2^Z(P_{t-s}f) \\
&\geq a'(s)\Gamma(P_{t-s}f) + b'(s)\Gamma^Z(P_{t-s}f) + 2a(s)\left(-\frac{d}{2}\sigma^2\Gamma(P_{t-s}f) - \frac{1}{4}\Gamma^Z(P_{t-s}f)\right) \\
&= (a' - ad\sigma^2)\Gamma(P_{t-s}f) + \left(b' - \frac{a}{2}\right)\Gamma^Z(P_{t-s}f).
\end{aligned}$$

One can easily see that if we choose  $b(s) = e^{\alpha s}$  and  $a(s) = ke^{\alpha s}$  with  $\alpha = d\sigma^2$  and  $k = 2d\sigma^2$ , then the last expression is 0. Using the existence of the Lyapunov function  $W$  as defined by (3.2) and a cutoff argument as in [11, Theorem 7.3], we deduce from a parabolic comparison principle

$$P_t(\varphi(\cdot, t))(x) \geq \varphi(x, 0).$$

Observe that

$$\begin{aligned}
\varphi(x, 0) &= a(0)\varphi_1(x, 0) + b(0)\varphi_2(x, 0) = 2d\sigma^2\Gamma(P_t f)(x) + \Gamma^Z(P_t f)(x), \\
P_t(\varphi(\cdot, t))(x) &= a(t)P_t(\Gamma(f))(x) + b(t)P_t(\Gamma^Z(f))(x) \\
&= e^{d\sigma^2 t}(2d\sigma^2 P_t(\Gamma(f))(x) + P_t(\Gamma^Z(f))(x)),
\end{aligned}$$

therefore

$$2d\sigma^2\Gamma(P_t f)(x) + \Gamma^Z(P_t f)(x) \leq e^{d\sigma^2 t}(2d\sigma^2 P_t(\Gamma(f))(x) + P_t(\Gamma^Z(f))(x)). \quad \square$$

**Corollary 3.3 (Poincaré type inequality).** *For any  $f \in C_0^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d})$  and  $t \geq 0$*

$$P_t(f^2) - (P_t f)^2 \leq \frac{e^{d\sigma^2 t} - 1}{(d\sigma^2)^2}(2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))).$$

**Proof.** Since  $\Gamma^Z(f) := \|\nabla_\xi f\|^2 \geq 0$  and  $P_t(f^2) - (P_t f)^2 = 2 \int_0^t P_s(\Gamma(P_{t-s}f)) ds$ , then for  $\sigma > 0$ ,

$$\begin{aligned}
\int_0^t P_s(2d\sigma^2\Gamma(P_{t-s}f) + \Gamma^Z(P_{t-s}f)) ds \\
\geq \int_0^t P_s(2d\sigma^2\Gamma(P_{t-s}f)) ds = d\sigma^2(P_t(f^2) - (P_t f)^2).
\end{aligned}$$

By Theorem 3.2 we have that

$$\begin{aligned}
\int_0^t P_s(2d\sigma^2\Gamma(P_{t-s}f) + \Gamma^Z(P_{t-s}f)) ds \\
\leq \int_0^t e^{d\sigma^2(t-s)} P_s(2d\sigma^2 P_{t-s}(\Gamma(f)) + P_{t-s}(\Gamma^Z(f))) ds \\
= (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))) \int_0^t e^{d\sigma^2(t-s)} ds \\
= \frac{e^{d\sigma^2 t} - 1}{d\sigma^2}(2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))).
\end{aligned}$$

This implies

$$P_t(f^2) - (P_t f)^2 \leq \frac{e^{d\sigma^2 t} - 1}{(d\sigma^2)^2}(2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))). \quad \square$$

The next corollary gives us an equivalent estimate to the one in Theorem 3.2. The estimate (3.3) will be similar to the one we will obtain in Theorem 4.8 in a more general setting.

**Corollary 3.4.** *For any  $f \in C_0^\infty(\mathbb{H}^d \times \mathbb{R}^{1,d})$ , the gradient estimate*

$$2d\sigma^2\Gamma(P_t f) + \Gamma^Z(P_t f) \leq e^{d\sigma^2 t} (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))),$$

is equivalent to

$$\Gamma(P_t f) \leq e^{d\sigma^2 t} P_t(\Gamma(f)) + \frac{e^{d\sigma^2 t} - 1}{2d\sigma^2} P_t(\Gamma^Z(f)). \quad (3.3)$$

Moreover, one has

$$\Gamma^Z(P_t f) \leq P_t(\Gamma^Z(f)).$$

**Proof.** Recall that

$$P_t(\Gamma(f)) - \Gamma(P_t f) = 2 \int_0^t P_s(\Gamma_2(P_{t-s} f)) ds.$$

Using the curvature dimension inequality  $\Gamma^Z(f) \geq -2d\sigma^2\Gamma(f) - 4\Gamma_2(f)$  we have

$$\begin{aligned} & \int_0^t P_s(2d\sigma^2\Gamma(P_{t-s} f) + \Gamma^Z(P_{t-s} f)) ds \\ & \geq \int_0^t P_s(2d\sigma^2\Gamma(P_{t-s} f) - 2d\sigma^2\Gamma(P_{t-s} f) - 4\Gamma_2(P_{t-s} f)) ds \\ & = -2(P_t(\Gamma(f)) - (\Gamma(P_t f))). \end{aligned}$$

On the other hand we have

$$\begin{aligned} & \int_0^t P_s(2d\sigma^2\Gamma(P_{t-s} f) + \Gamma^Z(P_{t-s} f)) ds \\ & \leq \int_0^t e^{d\sigma^2(t-s)} P_s(2d\sigma^2 P_{t-s}(\Gamma(f)) + P_{t-s}(\Gamma^Z(f))) ds \\ & = (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))) \int_0^t e^{d\sigma^2(t-s)} ds \\ & = \frac{e^{d\sigma^2 t} - 1}{d\sigma^2} (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))). \end{aligned}$$

Putting these together we have

$$\Gamma(P_t f) - P_t(\Gamma(f)) \leq \frac{e^{d\sigma^2 t} - 1}{2d\sigma^2} (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))).$$

A rearranging of this inequality gives us

$$\Gamma(P_t f) \leq e^{d\sigma^2 t} P_t(\Gamma(f)) + \frac{e^{d\sigma^2 t} - 1}{2d\sigma^2} P_t(\Gamma^Z(f)).$$

Conversely, assume  $\Gamma(P_t f) \leq e^{d\sigma^2 t} P_t(\Gamma(f)) + \frac{e^{d\sigma^2 t} - 1}{2d\sigma^2} P_t(\Gamma^Z(f))$  then

$$\begin{aligned} & 2d\sigma^2\Gamma(P_t f) + \Gamma^Z(P_t f) \\ & \leq 2d\sigma^2 \left( e^{d\sigma^2 t} P_t(\Gamma(f)) + \frac{e^{d\sigma^2 t} - 1}{2d\sigma^2} P_t(\Gamma^Z(f)) \right) + \Gamma^Z(P_t f) \end{aligned}$$

$$\begin{aligned}
&= e^{d\sigma^2 t} (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))) + \Gamma^Z(P_t f) - P_t(\Gamma^Z(f)) \\
&\leq e^{d\sigma^2 t} (2d\sigma^2 P_t(\Gamma(f)) + P_t(\Gamma^Z(f))) + 0.
\end{aligned}$$

The last inequality is due to  $\Gamma^Z(P_t f) \leq P_t(\Gamma^Z(f))$ . To see this, consider the functional  $\phi(s) = P_s(\Gamma^Z(P_{t-s} f))$  for  $0 \leq s \leq t$ . A calculation shows that

$$\Phi'(s) = 2P_s(\Gamma_2^Z(P_{t-s} f)) \geq 0,$$

which shows  $\phi(s)$  is increasing, so that  $0 \leq \phi(t) - \phi(0) = P_t(\Gamma^Z(f)) - \Gamma^Z(P_t f)$ .  $\square$

#### 4. Gradient bounds for a general Kolmogorov diffusion

We now study generalizations of Kolmogorov type diffusions and prove gradient bounds using  $\Gamma$ -calculus and coupling techniques. We give examples and give other generalizations along the way.

We start with an outline of this section. In Section 4.1, we prove gradient bounds for diffusions of the type  $\mathbf{X}_t = (X_t, \int_0^t \sigma(X_s) ds)$  for  $\sigma : \mathbb{R}^k \rightarrow \mathbb{R}^k$ , where  $X_t$  is a Markov process on  $\mathbb{R}^k$  satisfying a  $\Gamma_2$  lower bound. We show that a generalized curvature-dimension condition for the generator of  $\mathbf{X}_t$  is satisfied similarly to Theorem 3.1. In Section 4.2, we show that the results in Section 4.1 are valid when  $X_t$  is a Brownian motion on a complete Riemannian manifold isometrically embedded in  $\mathbb{R}^k$  for some  $k$ . In Section 4.3, we use coupling techniques to prove gradient bounds when  $X_t$  is assumed to live in a Riemannian manifold  $M$ . In this section  $M$  is not necessarily embedded in some  $\mathbb{R}^k$ . In Section 4.4, we generalize the results in Section 4.3 to iterated Kolmogorov diffusions. Finally in Section 4.5, we prove gradient bounds when  $X_t$  is a hypoelliptic Brownian motion on the Heisenberg group.

##### 4.1. $\Gamma$ -calculus

We now study the diffusion  $\mathbf{X}_t = (X_t, \int_0^t \sigma(X_s) ds)$ , where  $X_t$  is a Markov process in  $\mathbb{R}^k$  whose generator is given by

$$L = \sum_{i=1}^k V_i^2 + V_0.$$

Here  $V_i, i = 0, \dots, k$  are smooth vector fields, and we assume that  $\sigma : \mathbb{R}^k \rightarrow \mathbb{R}^k$  is a  $C^1$  map such that

$$C_\sigma := \sup_{p \in \mathbb{R}^k} \left( \sum_{i,j=1}^d (V_i \sigma_j(p))^2 \right)^{\frac{1}{2}} < \infty. \quad (4.1)$$

We consider functions on  $\mathbb{R}^k \times \mathbb{R}^k$  with  $f(p, \xi)$ ,  $p, \xi \in \mathbb{R}^k$ . The generator for  $\mathbf{X}_t$  is given by

$$\mathcal{L} = L + \sum_{i=1}^k \sigma_i(p) \frac{\partial}{\partial \xi_i}.$$

We first prove a generalized curvature-dimension inequality for  $\mathcal{L}$  given some assumptions on  $L$ . Let  $\Gamma(f)$  be the *carré du champ* operator for  $\mathcal{L}$ , while  $\Gamma^L(f)$  will be associated with  $L$ . Let  $\Gamma_2(f)$  and  $\Gamma_2^L(f)$  be the corresponding iterated *carré du champ* operators.

We define a symmetric, first-order differential bilinear form  $\Gamma^Z : C^\infty(\mathbb{R}^k \times \mathbb{R}^k) \times C^\infty(\mathbb{R}^k \times \mathbb{R}^k) \rightarrow \mathbb{R}$  by

$$\Gamma^Z(f) = \|\nabla_\xi f\|^2$$

for any  $f \in C^\infty(\mathbb{R}^k \times \mathbb{R}^k)$ .

**Theorem 4.1 (Curvature-dimension inequality).** *If the operator  $L$  satisfies*

$$\Gamma_2^L(f) \geq \rho \Gamma^L(f),$$

then the operator  $\mathcal{L}$  satisfies the following generalized curvature-dimension inequality for any  $f \in C^\infty(\mathbb{R}^k \times \mathbb{R}^k)$ ,

$$\Gamma_2(f) \geq \left( \rho - \frac{C_\sigma}{2} \right) \Gamma(f) - \frac{C_\sigma}{2} \Gamma^Z(f),$$

$$\Gamma_2^Z(f) \geq 0.$$

**Proof.** A simple calculation of the *carré du champ* of  $\mathcal{L}$  and  $L$  shows that

$$\Gamma(f) := \frac{1}{2}(\mathcal{L}f^2 - 2f\mathcal{L}f) = \sum_{j=1}^k (V_i f)^2,$$

$$\Gamma^L(f) := \frac{1}{2}(Lf^2 - 2fLf) = \sum_{j=1}^k (V_i f)^2.$$

More computations of the iterated carré du champ  $\Gamma_2(f) := \frac{1}{2}(\mathcal{L}\Gamma(f) - 2\Gamma(f, \mathcal{L}f))$  show that

$$\Gamma_2(f) = \Gamma_2^L(f) - \sum_{i=1}^k \sum_{j=1}^k (V_i f)(V_i \sigma_j) \frac{\partial f}{\partial \xi_j}.$$

By the assumption on  $\Gamma_2^L(f)$  we have

$$\Gamma_2(f) \geq \rho \Gamma(f) - \sum_{i=1}^k \sum_{j=1}^k (V_i f)(V_i \sigma_j) \frac{\partial f}{\partial \xi_j}.$$

Using the Cauchy–Schwarz inequality, the bound on  $\sigma$  and the elementary estimate  $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ , we see that

$$\begin{aligned} \sum_{i,j=1}^k (V_i f)(V_i \sigma_j) \frac{\partial f}{\partial \xi_j} &\leq \left( \sum_{i,j=1}^k (V_i \sigma_j)^2 \right)^{\frac{1}{2}} \left( \sum_{i,j=1}^k (V_i f)^2 \left( \frac{\partial f}{\partial \xi_j} \right)^2 \right)^{\frac{1}{2}} \\ &\leq C_\sigma (\Gamma(f))^{\frac{1}{2}} (\Gamma^Z(f))^{\frac{1}{2}} \\ &\leq \frac{C_\sigma}{2} (\Gamma(f)) + \frac{C_\sigma}{2} (\Gamma^Z(f)). \end{aligned}$$

Using this inequality with the previous one give us the desired first curvature-dimension inequality. The second inequality we want to prove is a lower bound on

$$\Gamma_2^Z(f) := \frac{1}{2}(\mathcal{L}\Gamma^Z(f) - 2\Gamma^Z(f, \mathcal{L}f)),$$

for which routine computations shows that

$$\Gamma_2^Z(f) = \sum_{i,j=1}^k \left( V_i \frac{\partial f}{\partial \xi_j} \right)^2 \geq 0,$$

as needed.  $\square$

In order to prove a gradient bound for the heat semigroup we must make the following assumption on the existence of a Lyapunov function for the operator  $\mathcal{L}$ . As in Section 3, we say that a smooth function  $W : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$  is a *Lyapunov function* on  $\mathbb{R}^k$  for  $\mathcal{L}$  if

$$\mathcal{L}W \leq CW,$$

for some  $C > 0$ . The existence of a Lyapunov function immediately implies that  $\mathcal{L}$  is the generator of a Markov semigroup  $(P_t)_{t \geq 0}$  that uniquely solves the heat equation in  $L^\infty$ .

Throughout this section, we will need the following assumption.

**Assumption 4.2.** There exists a Lyapunov function  $W : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$  such that  $W \geq 1$ ,  $\sqrt{\Gamma(W)} + \sqrt{\Gamma^Z(W)} \leq CW$ , for some constant  $C > 0$  and  $\{W \leq m\}$  is compact for every  $m$ . Here  $\Gamma$  is applied to the first coordinate of  $W$  while  $\Gamma^Z$  is applied to the second coordinate.

We are now ready to prove the main result of this section.

**Theorem 4.3 (Gradient estimate).** Suppose Assumption 4.2 holds and let  $P_t$  be the heat semigroup associated to  $\mathcal{L}$ . If  $C_\sigma > 2\rho$  and the operator  $L$  satisfies

$$\Gamma_2^L(f) \geq \rho \Gamma^L(f),$$

then for any  $f \in C_0^\infty(\mathbb{R}^k \times \mathbb{R}^k)$ ,  $t \geq 0$  and  $x \in \mathbb{R}^k \times \mathbb{R}^k$

$$\begin{aligned} \Gamma(P_t f)(x) + \frac{C_\sigma}{C_\sigma - 2\rho} \Gamma^Z(P_t f)(x) \\ \leq e^{(C_\sigma - 2\rho)t} \left( P_t(\Gamma(f))(x) + \frac{C_\sigma}{C_\sigma - 2\rho} P_t(\Gamma^Z(f))(x) \right). \end{aligned}$$

**Proof.** We fix  $t > 0$  throughout the proof. For  $0 < s < t$  and  $x = (p, \xi) \in \mathbb{R}^k \times \mathbb{R}^k$  we denote

$$\varphi_1(x, s) := \Gamma(P_{t-s} f)(x),$$

$$\varphi_2(x, s) := \Gamma^Z(P_{t-s} f)(x).$$

Then

$$\mathcal{L}\varphi_1 + \frac{\partial \varphi_1}{\partial s} = 2\Gamma_2(P_{t-s} f),$$

$$\mathcal{L}\varphi_2 + \frac{\partial \varphi_2}{\partial s} = 2\Gamma_2^Z(P_{t-s} f).$$

Now we would like to find two non-negative smooth functions  $a(s)$  and  $b(s)$  such that for

$$\varphi(x, s) := a(s)\varphi_1(x, s) + b(s)\varphi_2(x, s),$$

we have

$$\mathcal{L}\varphi + \frac{\partial \varphi}{\partial s} \geq 0.$$

Then by Theorem 4.1 we have

$$\begin{aligned} \mathcal{L}\varphi + \frac{\partial \varphi}{\partial s} \\ = a'(s)\Gamma(P_{t-s} f) + b'(s)\Gamma^Z(P_{t-s} f) + 2a(s)\Gamma_2(P_{t-s} f) + 2b(s)\Gamma_2^Z(P_{t-s} f) \\ \geq a'(s)\Gamma(P_{t-s} f) + b'(s)\Gamma^Z(P_{t-s} f) + 2a(s) \left( \left( \rho - \frac{C_\sigma}{2} \right) \Gamma(P_{t-s} f) - \frac{C_\sigma}{2} \Gamma^Z(P_{t-s} f) \right) \\ = (a'(s) + a(s)(2\rho - C_\sigma))\Gamma(P_{t-s} f) + (b'(s) - a(s)C_\sigma)\Gamma^Z(P_{t-s} f). \end{aligned}$$

One can easily see that if

$$a(s) = e^{(C_\sigma - 2\rho)s} \quad \text{and} \quad b(s) = \frac{C_\sigma}{C_\sigma - 2\rho} e^{(C_\sigma - 2\rho)s},$$

the last expression is 0. Using the existence of the Lyapunov function  $W$  and a cutoff argument as in [11, Theorem 7.3], we deduce from a parabolic comparison principle,

$$P_t(\varphi(\cdot, t))(x) \geq \varphi(x, 0).$$

Observe that

$$\varphi(x, 0) = a(0)\varphi_1(x, 0) + b(0)\varphi_2(x, 0) = \Gamma(P_t f)(x) + \frac{C_\sigma}{C_\sigma - 2\rho} \Gamma^Z(P_t f)(x),$$

while

$$\begin{aligned} P_t(\varphi(\cdot, t))(x) &= a(t)P_t(\Gamma(f))(x) + b(t)P_t(\Gamma^Z(f))(x) \\ &= e^{(C_\sigma - 2\rho)t} \left( P_t(\Gamma(f))(x) + \frac{C_\sigma}{C_\sigma - 2\rho} P_t(\Gamma^Z(f))(x) \right). \end{aligned}$$

□

**Corollary 4.4 (Poincaré type inequality).** *If  $C_\sigma > 2\rho$  then for any  $f \in C_0^\infty(\mathbb{R}^k \times \mathbb{R}^k)$  and  $t \geq 0$*

$$P_t(f^2) - (P_t f)^2 \leq 2 \frac{e^{(C_\sigma - 2\rho)t} - 1}{C_\sigma - 2\rho} \left( P_t(\Gamma(f)) + \frac{C_\sigma}{C_\sigma - 2\rho} P_t(\Gamma^Z(f)) \right).$$

**Proof.** Since  $\Gamma^Z(f) := \|\nabla_\xi f\|^2 \geq 0$  and  $P_t(f^2) - (P_t f)^2 = 2 \int_0^t P_s(\Gamma(P_{t-s} f)) ds$ , then

$$\begin{aligned} &\int_0^t P_s \left( \Gamma(P_{t-s} f) + \frac{C_\sigma}{C_\sigma - 2\rho} \Gamma^Z(P_{t-s} f) \right) ds \\ &\geq \frac{1}{2} \int_0^t 2P_s(\Gamma(P_{t-s} f)) ds = \frac{1}{2} (P_t(f^2) - (P_t f)^2). \end{aligned}$$

By Theorem 4.3 we have that

$$\begin{aligned} &\int_0^t P_s \left( \Gamma(P_{t-s} f) + \frac{C_\sigma}{C_\sigma - 2\rho} \Gamma^Z(P_{t-s} f) \right) ds \\ &\leq \int_0^t e^{(-2\rho + C_\sigma)(t-s)} P_s \left( P_{t-s}(\Gamma(f)) + \frac{C_\sigma}{C_\sigma - 2\rho} P_{t-s}(\Gamma^Z(f)) \right) ds \\ &= \left( P_t(\Gamma(f)) + \frac{C_\sigma}{C_\sigma - 2\rho} P_t(\Gamma^Z(f)) \right) \int_0^t e^{(C_\sigma - 2\rho)(t-s)} ds \\ &= \frac{e^{(C_\sigma - 2\rho)t} - 1}{C_\sigma - 2\rho} \left( P_t(\Gamma(f)) + \frac{C_\sigma}{C_\sigma - 2\rho} P_t(\Gamma^Z(f)) \right). \end{aligned}$$

So that

$$P_t(f^2) - (P_t f)^2 \leq 2 \frac{e^{(C_\sigma - 2\rho)t} - 1}{C_\sigma - 2\rho} \left( P_t(\Gamma(f)) + \frac{C_\sigma}{C_\sigma - 2\rho} P_t(\Gamma^Z(f)) \right).$$

□

#### 4.2. Application to Riemannian manifolds

To illustrate the results in Section 4.1 we study a large class of examples based on Brownian motion in a Riemannian manifold. For general background on general theory of stochastic analysis on manifolds we refer to [20, 32, 42]. Consider a complete Riemannian manifold  $(M, g)$  of dimension  $d$  which is isometrically embedded in  $\mathbb{R}^k$  for some  $k$ . Let  $B_t$  be a Brownian motion on  $M$  and consider the process  $\mathbf{X}_t = (B_t, \int_0^t \sigma(B_s) ds)$  where  $\sigma : M \rightarrow \mathbb{R}^k$  satisfies (4.1) and

$$|\sigma(p) - \sigma(\tilde{p})| \leq C_\sigma d_M(p, \tilde{p}),$$

for all  $p, \tilde{p} \in M$  where  $d_M$  is the intrinsic Riemannian distance on  $M$ . We can write the generator of  $B_t$  as

$$\Delta_p = \sum_{i=1}^k P_i^2,$$

for some vector fields  $P_i$  on  $\mathbb{R}^k$  (see for instance [42, Theorem 3.1.4]). The generator of  $X_t$  is

$$\mathcal{L} = \Delta_p + \sum_{i=1}^k \sigma_i(p) \frac{\partial}{\partial \xi_i},$$

for functions  $f(p, \xi) \in M \times \mathbb{R}^k$  where  $p \in M, \xi \in \mathbb{R}^k$ .

To apply Theorem 4.3 we first need to construct an appropriate Lyapunov function  $W$  for the operator  $\mathcal{L}$  satisfying Assumption 4.2. Once we construct  $W$ , we will spend the rest of the section verifying Assumption 4.2 for  $W$ . For this, we assume that the Ricci curvature  $\text{Ric} \geq \rho$  for some  $\rho \in \mathbb{R}$ . Then it is known from the Li–Yau upper and lower bounds in [48] that the heat kernel  $p(x, y, t)$  of  $M$  satisfies the following Gaussian estimates. Namely, for some  $\tau > 0$

$$\begin{aligned} & \frac{c_1}{\text{Vol}(B(p_0, \sqrt{\tau}))} \exp\left(-\frac{c_2 d_M(p_0, p_1)^2}{\tau}\right) \\ & \leq p(p_0, p_1, \tau) \leq \frac{c_3}{\text{Vol}(B(p_0, \sqrt{\tau}))} \exp\left(-\frac{c_4 d_M(p_0, p_1)^2}{\tau}\right), \end{aligned}$$

where  $d_M$  is the Riemannian distance in  $M$  and  $p_0, p_1 \in M$ . Consider now the smooth Lyapunov function

$$W(p, \xi) := K + \|\xi\|^2 - \ln p(p_0, p, \tau), \quad p \in M, \xi \in \mathbb{R}^k, \quad (4.2)$$

where  $p_0$  is an arbitrary fixed point in  $M$ , and  $K$  is a constant large enough so that  $W \geq 1$ .

**Lemma 4.5.** *The function  $W$  defined in (4.2) is smooth and satisfies the following properties,*

$$W \geq 1,$$

$$\|\nabla_\xi W\| + \|\nabla_p W\| \leq CW,$$

$$\mathcal{L}W \leq CW \quad \text{for some constant } C > 0,$$

$$\{W \leq m\} \quad \text{is compact for every } m.$$

Here  $\nabla_p$  is the Riemannian gradient on  $M$  and  $\nabla_\xi$  is the Euclidean gradient on  $\mathbb{R}^k$ .

**Proof.** From estimates for logarithmic derivatives of the heat kernel in [39, 48], one has for some constants  $C_1, C_2 > 0$

$$\|\nabla_p \ln p(p_0, p, \tau)\|^2 \leq C_1 + C_2 d_M(p_0, p)^2, \quad (4.3)$$

$$\Delta_p(-\ln p(p_0, p, \tau)) \leq C_1 + C_2 d_M(p_0, p)^2. \quad (4.4)$$

We can then conclude with the Li–Yau upper and lower Gaussian bounds. To see this note that the Gaussian bounds can be rearranged as

$$\begin{aligned} d_M(p_0, p)^2 & \leq -\frac{\tau}{c_4} \ln\left(\frac{\text{Vol}(B(p_0, \sqrt{\tau}))}{c_3} p(p_0, p, \tau)\right) \\ & \leq C(K - \ln(p(p_0, p, \tau))) \end{aligned} \quad (4.5)$$

for a fixed  $\tau \geq 0$  and a constant  $C > 0$ . Hence,  $\|\nabla_\xi W\| + \|\nabla_p W\| \leq CW$  can be shown using (4.3), (4.5) and the inequality  $(1+x)^{\frac{1}{2}} \leq 1 + cx$  for  $x \geq 0$  and  $c \geq \frac{1}{2}$ . On the other hand,  $\mathcal{L}W \leq CW$  can be shown using (4.4), (4.5), the Cauchy–Schwarz inequality, and the Lipschitz property of  $\sigma$ . Finally, the fact that  $\{W \leq m\}$  is compact for every  $m$  also follows from the Li–Yau upper and lower Gaussian bounds.  $\square$

Lemma 4.5 proves that  $W$  defined by (4.2) is a Lyapunov function satisfying Assumption 4.2. As a consequence, Theorem 4.3 can be applied to complete Riemannian manifolds with  $\text{Ric} \geq \rho$  since the condition  $\text{Ric} \geq \rho$  is equivalent to

$$\Gamma_2^\Delta(f) \geq \rho \Gamma(f).$$

### 4.3. Coupling

Let  $(M, g)$  be a complete connected  $d$ -dimensional Riemannian manifold. In this section  $M$  is not necessarily embedded in  $\mathbb{R}^k$ . We assume the existence of a map  $\sigma : M \rightarrow \mathbb{R}^k$  for some  $k \geq 1$  that is globally  $C_\sigma$ -Lipschitz map in the sense that

$$|\sigma(p) - \sigma(\tilde{p})| \leq C_\sigma d_M(p, \tilde{p}), \quad (4.6)$$

for all  $p, \tilde{p} \in M$ . Here we denote by  $d_M$  the Riemannian distance on  $M$ , and by  $d_E$  we denote the Euclidean metric in  $\mathbb{R}^k$ . We will consider the joint process

$$\mathbf{X}_t = \left( B_t, \int_0^t \sigma(B_s) ds \right), \quad (4.7)$$

on the product space  $M \times \mathbb{R}^k$  where  $B_t$  is Brownian motion on  $M$ .

Let  $P_t$  be the associated heat semigroup. We consider functions on  $M \times \mathbb{R}^k$  with  $f(p, \xi)$ ,  $p \in M, \xi \in \mathbb{R}^k$ . Recall that the operators  $\nabla_p$  and  $\Delta_p$  act on the variable  $p$  for  $f(p, \xi)$ , where  $\Delta_p$  is the Laplace–Beltrami operator. We use  $\nabla_\xi$  for the usual Euclidean gradient. Given a Riemannian metric  $g$ , for all  $p \in M$  and  $v \in T_p M$  we denote  $\|v\| = g_p(v, v)^{\frac{1}{2}}$ . Our main result of this section is a bound on  $\|\nabla_p P_t f\|$  for functions  $f \in C^\infty(M \times \mathbb{R}^k)$  with bounded Hessian. This will be a generalization of the result obtained in Section 2.2.

Let us recall the notion of a coupling of diffusions on a manifold  $M$ . Suppose  $X_t$  and  $\tilde{X}_t$  are  $M$ -valued diffusions starting at  $x, \tilde{x} \in M$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Then by a coupling of  $X_t$  and  $\tilde{X}_t$  we call a  $C(\mathbb{R}_+, M \times M)$ -valued random variable  $(X_t, \tilde{X}_t)$  on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that the marginal processes for  $(X_t, \tilde{X}_t)$  have the same laws as  $X_t$  and  $\tilde{X}_t$ . Let  $\mathbb{P}^{(x, \tilde{x})}$  be the distribution of  $(X_t, \tilde{X}_t)$ , so that  $\mathbb{P}^{(x, \tilde{x})}(X_0 = x, \tilde{X}_0 = \tilde{x}) = 1$ . We denote by  $\mathbb{E}^{(x, \tilde{x})}$  the expectation with respect to the probability measure  $\mathbb{P}^{(x, \tilde{x})}$ .

In [52–54] it has been shown that if we assume  $\text{Ric}(M) \geq K$  for some  $K \in \mathbb{R}$ , then there exists a Markovian coupling of Brownian motions  $(B_t)_{t \geq 0}$  and  $(\tilde{B}_t)_{t \geq 0}$  on  $M$  starting at  $p$  and  $\tilde{p}$  such that

$$d_M(B_t, \tilde{B}_t) \leq e^{-Kt/2} d_M(p, \tilde{p}) \quad (4.8)$$

for all  $t \geq 0$ ,  $\mathbb{P}^{(p, \tilde{p})}$ -almost surely. This construction is known as a *coupling by parallel transport*. This coupling can be constructed using stochastic differential equations as in [24, 54], or by a central limit theorem argument for the geodesic random walks as in [52]. It turns out that the existence of the coupling satisfying (4.8) is equivalent to

$$\|\nabla P_t f\| \leq e^{-Kt} P_t(\|\nabla f\|), \quad (4.9)$$

for all  $f \in C_0^\infty(M)$  and all  $t > 0$ . We also point out that in [50, 51], M. Pascu and I. Popescu constructed explicit Markovian couplings where equality in (4.8) is attained for  $t \geq 0$  dependent on  $K$  and given some extra geometric assumptions.

The coupling by parallel transport that gives (4.8) is in the elliptic setting. In this section, we will use the coupling by parallel transport to induce a coupling for (4.7) in the hypoelliptic setting. We will then use this coupling to prove gradient bounds for  $(P_t)_{t \geq 0}$ . Before stating the result on the gradient bound, we have the following proposition.

**Proposition 4.6.** *Let  $(M, g)$  be a Riemannian manifold. If  $f \in C^1(M)$  then*

$$\lim_{r \rightarrow 0} \sup_{\tilde{p} : 0 < d_M(p, \tilde{p}) \leq r} \frac{|f(p) - f(\tilde{p})|}{d_M(p, \tilde{p})} = \|\nabla f(p)\|. \quad (4.10)$$

**Proof.** Let  $p, \tilde{p} \in M$  with  $T = d_M(p, \tilde{p})$  and consider a unit speed geodesic  $\gamma : [0, T] \rightarrow M$  such that  $\gamma(0) = \tilde{p}$  and  $\gamma(T) = p$ . Then

$$\begin{aligned} |f(p) - f(\tilde{p})| &= \left| \int_0^{d(p, \tilde{p})} g(\nabla f(\gamma(s)), \gamma'(s)) ds \right| \\ &\leq \int_0^{d(p, \tilde{p})} |g(\nabla f(\gamma(s)), \gamma'(s))| ds \\ &\leq \max_{0 \leq s \leq d(p, \tilde{p})} \|\nabla f(\gamma(s))\| \cdot d(p, \tilde{p}), \end{aligned}$$

where we used the Cauchy–Schwarz inequality. Since  $p, \tilde{p}$  are arbitrary, dividing out both sides by  $d(p, \tilde{p})$  we have that

$$\lim_{r \rightarrow 0} \sup_{\tilde{p}: 0 < d_M(p, \tilde{p}) \leq r} \frac{|f(p) - f(\tilde{p})|}{d_M(p, \tilde{p})} \leq \|\nabla f(p)\|.$$

On the other hand, find a unit speed geodesic  $\gamma: (-\epsilon, \epsilon) \rightarrow M$  such that  $\gamma(0) = p$  and  $\gamma'(0) = \nabla f(p)/\|\nabla f(p)\|$ . Define  $F(s) = f(\gamma(s))$ . Since  $F'(s) = g(\nabla f(\gamma(s)), \gamma'(s))$ , then

$$F'(0) = g\left(\nabla f(p), \frac{\nabla f(p)}{\|\nabla f(p)\|}\right) = \|\nabla f(p)\|.$$

Now by the definition of the derivative we have that

$$\lim_{h \rightarrow 0} \frac{F(h) - F(0)}{h} \rightarrow \|\nabla f(p)\|,$$

which means we have that the left hand side of (4.10) must be at least  $\|\nabla f(p)\|$ . This proves (4.10).  $\square$

The following lemma gives an estimate for  $|f(p, \xi) - f(\tilde{p}, \tilde{\xi})|$  on  $M \times \mathbb{R}^k$ .

**Lemma 4.7.** *Let  $(M, g)$  be a complete Riemannian manifold which is assumed to be embedded in  $\mathbb{R}^k$ . For a function  $f(p, \xi)$  we denote by  $\nabla_p f$  the Riemannian gradient acting on  $p$ , and by  $\nabla_\xi f$  the Euclidean gradient acting on  $\xi$ . If  $f \in C^2(M \times \mathbb{R}^k)$  with a bounded Hessian, then there exists a  $C_f > 0$  depending on a bound on the Hessian of  $f$  such that*

$$\begin{aligned} |f(p, \xi) - f(\tilde{p}, \tilde{\xi})| &\leq \|\nabla_p f(\tilde{p}, \tilde{\xi})\| d_M(p, \tilde{p}) + \|\nabla_\xi f(\tilde{p}, \tilde{\xi})\| d_E(\xi, \tilde{\xi}) \\ &\quad + C_f (d_M(p, \tilde{p}) + d_E(\xi, \tilde{\xi}))^2 \end{aligned}$$

for any  $(p, \xi), (\tilde{p}, \tilde{\xi}) \in M \times \mathbb{R}^k$ .

**Proof.** Let  $p, \tilde{p} \in M$  with  $T_1 = d_M(p, \tilde{p})$  and consider a unit speed geodesic  $\gamma: [0, T_1] \rightarrow M$  such that  $\gamma(0) = \tilde{p}$  and  $\gamma(T_1) = p$ . Let  $\xi, \tilde{\xi} \in \mathbb{R}^k$  with  $T_2 = d_E(\xi, \tilde{\xi})$  and consider  $\beta(s) = \frac{s}{d_E(\xi, \tilde{\xi})}(\xi - \tilde{\xi}) + \tilde{\xi}$  on  $-\infty \leq s \leq T_2$  such that  $\beta(0) = \tilde{\xi}$  and  $\beta(T_2) = \xi$ . Extend  $\gamma$  to  $[-\epsilon, T_1]$  for some  $\epsilon > 0$  and define  $F(t, s) = f(\gamma(t), \beta(s))$ . By an estimate on the remainder of Taylor's approximation there exists a  $C_f > 0$  depending only on a bound on the Hessian of  $f$  such that

$$|F(t, s) - F(0, 0)| \leq |F_t(0, 0)t + F_s(0, 0)s| + C_f(t + s)^2.$$

By the chain rule we have

$$\begin{aligned} F_t(0, 0) &= \frac{d}{dt} [f(\gamma(t), \beta(0))]_{t=0} = \langle \nabla_p f(\gamma(0), \beta(0)), \gamma'(0) \rangle \\ &\leq \|\nabla_p f(\gamma(0), \beta(0))\| = \|\nabla_p f(\tilde{p}, \tilde{\xi})\|. \end{aligned}$$

Similarly  $F_s(0, 0) = \frac{d}{ds} [f(\gamma(0), \beta(s))]_{s=0} \leq \|\nabla_\xi f(\tilde{p}, \tilde{\xi})\|$ . Then

$$\begin{aligned} |f(p, \xi) - f(\tilde{p}, \tilde{\xi})| &= |F(T_1, T_2) - F(0, 0)| \\ &\leq \|\nabla_p f(\tilde{p}, \tilde{\xi})\| T_1 + \|\nabla_\xi f(\tilde{p}, \tilde{\xi})\| T_2 + C_f(T_1 + T_2)^2, \end{aligned}$$

as needed.  $\square$

We are now ready to state and prove the main theorem of this section. We start by considering the coupling of Brownian motions  $(B_t, \tilde{B}_t)$  starting at  $(p, \tilde{p})$  by parallel transport satisfying (4.8), as introduced in [52,53]. This coupling induces a coupling  $\mathbb{P}^{(x, \tilde{x})}$  on  $(M \times \mathbb{R}^d) \times (M \times \mathbb{R}^d)$  for two Kolmogorov type diffusions

$$\mathbf{X}_t = \left( B_t, \xi + \int_0^t \sigma(B_s) ds \right) \quad \text{and} \quad \tilde{\mathbf{X}}_t = \left( \tilde{B}_t, \tilde{\xi} + \int_0^t \sigma(\tilde{B}_s) ds \right),$$

started at  $\mathbf{x} = (p, \xi)$  and  $\tilde{\mathbf{x}} = (\tilde{p}, \tilde{\xi})$  respectively.

**Theorem 4.8 (Bakry–Émery type estimate).** *Let  $M$  be a complete connected Riemannian manifold such that  $\text{Ric}(M) \geq K$  for some  $K \in \mathbb{R}$ . Let  $\sigma$  be a  $C_\sigma$ -Lipschitz map as in (4.6) and  $f \in C^2(M \times \mathbb{R}^k)$  with a bounded Hessian. Then for every  $q \geq 1$  and  $t \geq 0$ ,*

$$\|\nabla_p P_t f\|^q \leq P_t((K_1(t)\|\nabla_p f\| + K_2(t)\|\nabla_\xi f\|)^q),$$

where

$$K_1(t) = e^{-Kt/2} \quad \text{and} \quad K_2(t) = \begin{cases} C_\sigma t, & K = 0, \\ C_\sigma \frac{1-e^{-Kt/2}}{K/2}, & K \neq 0. \end{cases}$$

**Proof.** As before let  $d_M$  be the Riemannian distance on  $M$ , and let  $d_E$  be the Euclidean distance on  $\mathbb{R}^k$ . Take  $\mathbf{x} = (p, \xi) \in M \times \mathbb{R}^k$  and  $\tilde{\mathbf{x}} = (\tilde{p}, \xi) \in M \times \mathbb{R}^k$ . If  $K \neq 0$ , we consider the coupling by parallel transport of Brownian motions  $(B_t, \tilde{B}_t)$  starting at  $(p, \tilde{p})$ . This coupling gives us that

$$d_M(B_t, \tilde{B}_t) \leq e^{-Kt/2} d_M(p, \tilde{p}), \quad (4.11)$$

for all  $t \geq 0$ . Denote  $Y_t = \xi + \int_0^t \sigma(B_s) ds$  and  $\tilde{Y}_t = \xi + \int_0^t \sigma(\tilde{B}_s) ds$ . If  $K \neq 0$  then

$$\begin{aligned} d_E(Y_t, \tilde{Y}_t) &\leq \int_0^t |\sigma(B_s) - \sigma(\tilde{B}_s)| ds \leq C_\sigma \int_0^t d_M(B_s, \tilde{B}_s) ds \\ &\leq C_\sigma d_M(p, \tilde{p}) \int_0^t e^{-Ks/2} ds = C_\sigma \left( \frac{1 - e^{-Kt/2}}{K/2} \right) d_M(p, \tilde{p}), \end{aligned} \quad (4.12)$$

where we used (4.6) and (4.8). If  $K = 0$ , we consider the same coupling for the Brownian motions  $(B_t, \tilde{B}_t)$  starting at  $(p, \tilde{p})$  so that

$$d_M(B_t, \tilde{B}_t) \leq d_M(p, \tilde{p}), \quad (4.13)$$

for all  $t \geq 0$ . A similar computation as in (4.12) gets us the estimate

$$d_E(Y_t, \tilde{Y}_t) \leq C_\sigma t d_M(p, \tilde{p}), \quad (4.14)$$

from (4.13). Combining (4.11) and (4.13) we get

$$d_M(B_t, \tilde{B}_t) \leq K_1(t) d_M(p, \tilde{p}), \quad (4.15)$$

while combining (4.12) and (4.14) we have

$$d_E(Y_t, \tilde{Y}_t) \leq K_2(t) d_M(p, \tilde{p}), \quad (4.16)$$

for all  $t \geq 0$ , where all of these inequalities hold  $\mathbb{P}^{(\mathbf{x}, \tilde{\mathbf{x}})}$ -almost surely. By Lemma 4.7, there exists a  $C_f \geq 1$  depending on a bound on the Hessian of  $f \in C^2(M \times \mathbb{R}^k)$  such that

$$\begin{aligned} &|f(B_t, Y_t) - f(\tilde{B}_t, \tilde{Y}_t)| \\ &\leq \|\nabla_p f(\tilde{B}_t, \tilde{Y}_t)\| d_M(B_t, \tilde{B}_t) + \|\nabla_\xi f(\tilde{B}_t, \tilde{Y}_t)\| d_E(Y_t, \tilde{Y}_t) \\ &\quad + C_f (d_M(B_t, \tilde{B}_t) + d_E(Y_t, \tilde{Y}_t))^2, \end{aligned} \quad (4.17)$$

for all  $t \geq 0$ ,  $\mathbb{P}^{(\mathbf{x}, \tilde{\mathbf{x}})}$ -almost surely.

Using inequalities (4.15), (4.16) and (4.17), we have that for  $f \in C^2(M \times \mathbb{R}^k)$

$$\begin{aligned} &|P_t f(p, \xi) - P_t f(\tilde{p}, \xi)| \\ &= |\mathbb{E}^{(\mathbf{x}, \tilde{\mathbf{x}})}[f(B_t, Y_t) - f(\tilde{B}_t, \tilde{Y}_t)]| \\ &\leq \mathbb{E}^{(\mathbf{x}, \tilde{\mathbf{x}})}[\|\nabla_p f(\tilde{B}_t, \tilde{Y}_t)\| d_M(B_t, \tilde{B}_t) + \|\nabla_\xi f(\tilde{B}_t, \tilde{Y}_t)\| d_E(Y_t, \tilde{Y}_t)] \\ &\quad + C_f \mathbb{E}^{(\mathbf{x}, \tilde{\mathbf{x}})}[d_M(B_t, \tilde{B}_t) + d_E(Y_t, \tilde{Y}_t)]^2 \end{aligned}$$

$$\begin{aligned} &\leq \mathbb{E}^{(\mathbf{x}, \tilde{\mathbf{x}})} [K_1(t) \|\nabla_p f(\tilde{B}_t, \tilde{Y}_t)\| + K_2(t) \|\nabla_\xi f(\tilde{B}_t, \tilde{Y}_t)\|] d_M(p, \tilde{p}) \\ &\quad + C_f (K_1(t) + K_2(t))^2 d_M(p, \tilde{p})^2. \end{aligned}$$

Using Jensen's inequality for  $q \geq 1$  we have

$$\begin{aligned} &|P_t f(p, \xi) - P_t f(\tilde{p}, \xi)| \\ &\leq (\mathbb{E}^{(\mathbf{x}, \tilde{\mathbf{x}})} [(K_1(t) \|\nabla_p f(\tilde{B}_t, \tilde{Y}_t)\| + K_2(t) \|\nabla_\xi f(\tilde{B}_t, \tilde{Y}_t)\|)^q])^{\frac{1}{q}} d_M(p, \tilde{p}) \\ &\quad + C_f (K_1(t) + K_2(t))^2 d_M(p, \tilde{p})^2. \end{aligned}$$

Dividing the last inequality out by  $d_M(p, \tilde{p})$  we have that

$$\begin{aligned} \frac{|P_t f(p, \xi) - P_t f(\tilde{p}, \xi)|}{d_M(p, \tilde{p})} &\leq [P_t ((K_1(t) \|\nabla_p f\| + K_2(t) \|\nabla_\xi f\|)^q)(\tilde{p}, \xi)]^{\frac{1}{q}} \\ &\quad + C_f (K_1(t) + K_2(t))^2 d_M(p, \tilde{p}). \end{aligned}$$

Since

$$\lim_{r \rightarrow 0} \sup_{\tilde{p}: 0 < d_M(p, \tilde{p}) \leq r} \frac{|P_t f(p, \xi) - P_t f(\tilde{p}, \xi)|}{d_M(p, \tilde{p})} = \|\nabla_p P_t f(p, \xi)\|$$

by Proposition 4.6, we have the desired result.  $\square$

**Remark 4.9.** The constants obtained in Theorem 4.8 using the coupling technique are sharper than the constants in Theorem 4.3 using  $\Gamma$ -calculus. The trade off here being that the  $\Gamma$ -calculus approach allows for the result to be proven for a wider class of Kolmogorov type diffusions.

**Remark 4.10.** We note that when applying the triangle inequality to the right hand sides of the inequalities in Propositions 2.5, 2.10, we recover Theorem 4.8 when the manifold  $M = \mathbb{R}^d$ . Here we have  $k = d$ ,  $\sigma(\mathbf{x}) = \mathbf{x}$  and  $C_\sigma = 1$ .

**Example 4.11 (Velocity spherical Brownian motion).** The velocity spherical Brownian is a diffusion process which takes values in  $T^1 \mathcal{M}$ , the unit tangent bundle of a Riemannian manifold of finite volume. The generator is of the form

$$L = \frac{\sigma^2}{2} \Delta_v + \kappa \xi.$$

It was introduced in [4] and further studied in [17]. When  $\mathcal{M} = \mathbb{R}^{d+1}$  and  $\sigma = \kappa = 1$  the diffusion is of the form  $\mathbf{X}_t = (B_t, \int_0^t B_s ds)$  where  $B_t$  is a Brownian motion on the  $d$ -dimensional sphere  $\mathbb{S}^d$ . Here we take  $\mathbb{S}^d$  to have the usual embedding in  $\mathbb{R}^{d+1}$ , that is,  $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^{d+1} \mid |\mathbf{x}| = 1\}$ . Let  $d_{\mathbb{S}^d}$  be the spherical distance and  $d_E(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$  is the Euclidean distance in  $\mathbb{R}^{d+1}$ . The explicit spherical distance is given by

$$d_{\mathbb{S}^d}(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x} \cdot \mathbf{y}),$$

for  $\mathbf{x}, \mathbf{y} \in \mathbb{S}^d$ , where the standard Euclidean inner product is used. It is easy to see that

$$d_E(\mathbf{x}, \mathbf{y}) \leq d_{\mathbb{S}^d}(\mathbf{x}, \mathbf{y}), \tag{4.18}$$

for all  $\mathbf{x}, \mathbf{y} \in \mathbb{S}^d$  since the Riemannian structure of  $\mathbb{S}^d$  is induced by the Euclidean structure of the ambient space  $\mathbb{R}^{d+1}$ . Inequality (4.18) shows that  $\sigma : \mathbb{S}^d \rightarrow \mathbb{R}^{d+1}$  is a  $C_\sigma = 1$ -Lipschitz map. Thus we can apply Theorem 4.8 to the manifold  $M = \mathbb{S}^d$ , since  $\text{Ric} = (d-1)g$  where  $g$  is the Riemannian metric.

**Example 4.12.** Let  $k = 1$  and fix a  $p_0 \in M$ . We consider the map  $\sigma : M \rightarrow \mathbb{R}$  defined by

$$\sigma(p) = d_M(p, p_0).$$

Note that this map is globally 1-Lipschitz since

$$|\sigma(p) - \sigma(\tilde{p})| = |d_M(p, p_0) - d_M(\tilde{p}, p_0)| \leq d_M(p, \tilde{p}),$$

for all  $p, \tilde{p} \in M$ . Thus we can always apply Theorem 4.8 to the process

$$\mathbf{X}_t = \left( B_t, \int_0^t d_M(B_s, p_0) ds \right),$$

where  $B_t$  is Brownian motion on  $M$ .

#### 4.4. Iterated Kolmogorov diffusions

Our technique can also be applied in studying *iterated Kolmogorov diffusions*. These processes have been studied recently by S. Banerjee and W. Kendall in [10] and K. Habermann in [38] as they provide a natural class of diffusions satisfying a weak Hörmander condition.

An iterated Kolmogorov diffusion is of the form  $\mathbf{X}_t = (B_t, I_1(t), \dots, I_n(t))$ , where

$$\begin{aligned} I_0(t) &= \sigma(B_t), \\ I_r(t) &= \int_0^t I_{r-1}(s) ds \quad \text{for } r = 1, \dots, n, \end{aligned}$$

where  $B_t$  is a Brownian motion on a manifold  $M$  and  $\sigma : M \rightarrow \mathbb{R}^k$  is  $C_\sigma$ -Lipschitz. Let  $P_t$  be the heat semigroup corresponding to the diffusion

$$\mathbf{X}_t = (B_t, I_1(t), \dots, I_n(t)).$$

Using an argument similar to the proof of Theorem 4.8, we get the following result.

**Theorem 4.13.** *Let  $M$  be a complete connected Riemannian manifold such that  $\text{Ric}(M) \geq K$  for some  $K \in \mathbb{R}$ . When  $K = 0$  and  $f \in C_0^\infty(M \times \mathbb{R}^k \times \dots \times \mathbb{R}^k)$  with  $f(p, \xi_1, \dots, \xi_n)$ ,  $p \in M$ ,  $\xi_1, \dots, \xi_n \in \mathbb{R}^k$  we have the following gradient bound for the iterated Kolmogorov diffusion semigroup  $P_t$ ,*

$$\|\nabla_p P_t f\|^q \leq P_t \left( \left( \|\nabla_p f\| + C_\sigma t \|\nabla_{\xi_1} f\| + \dots + C_\sigma \frac{t^n}{n!} \|\nabla_{\xi_n} f\| \right)^q \right),$$

for  $q \geq 1$ . When  $K \neq 0$ , we have

$$\|\nabla_p P_t f\|^q \leq P_t \left( (\|\nabla_p f\| + K_1(t) \|\nabla_{\xi_1} f\| + \dots + K_n(t) \|\nabla_{\xi_n} f\|)^q \right),$$

for  $q \geq 1$ , where

$$\begin{aligned} K_1(t) &= C_\sigma \frac{1 - e^{-Kt/2}}{K/2}, \\ K_r(t) &= \int_0^t K_{r-1}(s) ds \quad \text{for } r = 2, \dots, n. \end{aligned}$$

#### 4.5. Heisenberg group

The Heisenberg group is the simplest nontrivial example of a sub-Riemannian manifold. The 3-dimensional Heisenberg group is  $\mathbb{G} = \mathbb{R}^3$  with the group law defined by

$$(x_1, y_1, z_1) \star (x_2, y_2, z_2) := \left( x_1 + x_2, y_1 + y_2, z_1 + z_2 + \frac{1}{2}(x_1 y_2 - x_2 y_1) \right).$$

The identity element is  $e = (0, 0, 0)$  with the inverse given by  $(x, y, z)^{-1} = (-x, -y, -z)$ . We define the following left-invariant vector fields by

$$\begin{aligned} \mathcal{X} &:= \partial_x - \frac{y}{2} \partial_z, \\ \mathcal{Y} &:= \partial_y - \frac{x}{2} \partial_z, \\ \mathcal{Z} &:= \partial_z. \end{aligned}$$

The horizontal distribution is defined by  $\mathcal{H} = \text{span}\{\mathcal{X}, \mathcal{Y}\}$ , fiberwise. Vectors in  $\mathcal{H}$  are said to be *horizontal*. We endow  $\mathbb{G}$  with the sub-Riemannian metric  $g(\cdot, \cdot)$  so that  $\{\mathcal{X}, \mathcal{Y}\}$  forms an orthogonal frame for the horizontal distribution  $\mathcal{H}$ . With this metric we can define norms on vectors by  $\|v\| = (g_p(v, v))^{\frac{1}{2}}$  for  $v \in \mathcal{H}_p$ ,  $p \in \mathbb{G}$ . The Lebesgue measure on  $\mathbb{R}^3$  is a Haar measure on the Heisenberg group. The distance associated to  $\mathcal{H}$  is the Carnot–Carathéodory distance  $d_{\text{CC}}$ . The *horizontal gradient*  $\nabla_{\mathcal{H}}$  is a horizontal vector field such that for any smooth  $f : \mathbb{G} \rightarrow \mathbb{R}$  we have that for all  $X \in \mathcal{H}$

$$g(\nabla_{\mathcal{H}} f, X) = X(f).$$

The operator

$$\Delta_{\mathcal{H}} = \frac{1}{2}(\mathcal{X}^2 + \mathcal{Y}^2)$$

is a natural sub-Laplacian for the Heisenberg as pointed out in [1,37] and also in [36, Example 6.1]. Brownian motion on the Heisenberg group is defined to be the diffusion process  $\{B_t^p\}_{t \geq 0}$  starting at  $p = (x, y, z) \in \mathbb{R}^3$  whose infinitesimal generator is  $\Delta_{\mathcal{H}}$ . Explicitly the process is given by

$$B_t^p = \left( B_1(t), B_2(t), z + \frac{1}{2} \int_0^t B_1(s) dB_2(s) - \frac{1}{2} \int_0^t B_2(s) dB_1(s) \right),$$

where  $(B_1, B_2)$  is a Brownian motion starting at  $(x, y)$ .

Gradient bounds of Bakry–Émery type were studied for the Heisenberg group in [6,25,31,47]. In particular, the  $L^1$ -gradient bounds for the heat semigroup have been proven first in [47] and also in [6]. As pointed out in [46], Kuwada's duality between  $L^1$ -gradient bounds and  $L^\infty$ -Wasserstein control shows that for each  $t > 0$ , and  $p, \tilde{p} \in \mathbb{G}$ , there exists a coupling  $(B_t^p, \tilde{B}_t^p)$  of Brownian motions on the Heisenberg group such that

$$d_{\text{CC}}(B_t^p, \tilde{B}_t^p) \leq K d_{\text{CC}}(p, \tilde{p}), \quad (4.19)$$

almost surely for some constant  $K \geq 1$  that does not depend on  $p, \tilde{p}, t$ . We remark that in [21], the authors show that any coupling that satisfies (4.19) on  $\mathbb{G}$  must be non-Markovian. This further highlights the need for more non-Markovian coupling techniques as in [9,10].

Consider the Kolmogorov diffusion  $\mathbf{X}_t = (B_t^p, \xi + \int_0^t \sigma(B_s^p) ds)$  on  $\mathbb{G} \times \mathbb{R}^3$ , where  $\sigma : \mathbb{G} \rightarrow \mathbb{R}^3$  satisfies (4.6) and let  $P_t$  be the heat semigroup associated with  $X_t$ . Using a similar argument as in Lemma 4.7 with the sub-Riemannian metric  $g$  and the horizontal gradient  $\nabla_{\mathcal{H}}$ , we can get an estimate

$$\begin{aligned} |f(p, \xi) - f(\tilde{p}, \xi)| &\leq \|\nabla_{\mathcal{H}} f(\tilde{p}, \xi)\| d_{\text{CC}}(p, \tilde{p}) + \|\nabla_\xi f(\tilde{p}, \xi)\| d_E(\xi, \tilde{\xi}) \\ &\quad + C_f (d_{\text{CC}}(p, \tilde{p}) + d_E(\xi, \tilde{\xi}))^2, \end{aligned} \quad (4.20)$$

for functions  $f \in C_0^\infty(\mathbb{G} \times \mathbb{R}^3)$ , where  $C_f \geq 0$ . The argument in Theorem 4.8 can be used to prove gradient bounds for  $P_t$  when  $B_t^p$  is a Brownian motion on a sub-Riemannian manifold once we have a synchronous coupling and an estimate similar to (4.20). Thus using (4.19) and (4.20) for the Heisenberg group we obtain the following result.

**Theorem 4.14.** *For all  $q \geq 1$  and  $f \in C_0^\infty(\mathbb{G} \times \mathbb{R}^3)$ ,*

$$\|\nabla_{\mathcal{H}} P_t f\|^q \leq K^q P_t ((\|\nabla_{\mathcal{H}} f\| + C_\sigma t \|\nabla_\xi f\|)^q). \quad (4.21)$$

The best constant  $K$  in (4.19) and (4.21) is not known. The best known estimate for  $K$  as of this writing is  $K \geq \sqrt{2}$  (see [25, Proposition 2.7]). In fact the best constant  $K$  is conjectured to be  $\sqrt{2}$  in [6, Remark 3.2].

**Example 4.15.** Consider for  $p = (x, y, z) \in \mathbb{G}$  the map  $\sigma : \mathbb{G} \rightarrow \mathbb{R}^3$  defined by  $\sigma(p) = (x, y, 0)$  and the diffusion  $\mathbf{X}_t = (B_t^p, \xi + \int_0^t \sigma(B_s^p) ds)$ . A straightforward computation shows that

$$\sqrt{x^2 + y^2} \leq d_{\text{CC}}(e, p),$$

so that by the left-invariance of  $d_{\text{CC}}$  we have that  $\sigma$  is 1-Lipschitz in the sense of (4.6). Thus Theorem 4.14 can be applied to  $\mathbf{X}_t$ .

## Acknowledgements

The authors would like to thank Sayan Banerjee, Bruce Driver and Tai Melcher for helpful discussions and insights. We would also like to thank two anonymous referees for their careful review of the paper and whose suggestions greatly improved the present paper.

## References

- [1] A. Agrachev, U. Boscain, J.-P. Gauthier and F. Rossi. The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups. *J. Funct. Anal.* **256** (8) (2009) 2621–2655. [MR2502528](#)
- [2] A. Agrachev, U. Boscain, R. Neel and L. Rizzi. Intrinsic random walks in Riemannian and sub-Riemannian geometry via volume sampling. *ESAIM Control Optim. Calc. Var.* **24** (3) (2018) 1075–1105. [MR3877194](#) <https://doi.org/10.1051/cocv/2017037>
- [3] J. Angst. Poisson boundary of a relativistic diffusion in curved space-times: An example. *ESAIM Probab. Stat.* **19** (2015) 502–514. [MR3423304](#)
- [4] J. Angst, I. Bailleul and C. Tardif. Kinetic Brownian motion on Riemannian manifolds. *Electron. J. Probab.* **20** (110) (2015) 40. [MR3418542](#)
- [5] I. Bailleul. Poisson boundary of a relativistic diffusion. *Probab. Theory Related Fields* **141** (1–2) (2008) 283–329. [MR2372972](#)
- [6] D. Bakry, F. Baudoin, M. Bonnefont and D. Chafaï. On gradient bounds for the heat kernel on the Heisenberg group. *J. Funct. Anal.* **255** (8) (2008) 1905–1938. [MR2462581](#)
- [7] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités, XIX 1983/84* 177–206. *Lecture Notes in Math.* **1123** Springer, Berlin, 1985. [MR0889476](#)
- [8] D. Bakry, I. Gentil and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Springer, Cham, 2014. [MR3155209](#)
- [9] S. Banerjee, M. Gordina and P. Mariano. Coupling in the Heisenberg group and its applications to gradient estimates. *Ann. Probab.* **46** (2018) 3275–3312. [MR3857856](#)
- [10] S. Banerjee and W. S. Kendall. Coupling the Kolmogorov diffusion: Maximality and efficiency considerations. *Adv. in Appl. Probab.* **48** (2016) 15–35. [MR3539295](#)
- [11] F. Baudoin. Sub-Laplacians and hypoelliptic operators on totally geodesic Riemannian foliations. In *Geometry, Analysis and Dynamics on Sub-Riemannian Manifolds. Vol. I* 259–321. *EMS Ser. Lect. Math.* Eur. Math. Soc., Zürich, 2016. [MR3587668](#)
- [12] F. Baudoin, Bakry–Émery meet Villani. *J. Funct. Anal.* **273** (7) (2017) 2275–2291. [MR3677826](#)
- [13] F. Baudoin and M. Bonnefont. Log-Sobolev inequalities for subelliptic operators satisfying a generalized curvature dimension inequality. *J. Funct. Anal.* **262** (6) (2012) 2646–2676. [MR2885961](#)
- [14] F. Baudoin, M. Bonnefont and N. Garofalo. A sub-Riemannian curvature-dimension inequality, volume doubling property and the Poincaré inequality. *Math. Ann.* **358** (3–4) (2014) 833–860. [MR3175142](#)
- [15] F. Baudoin and N. Garofalo. Curvature-dimension inequalities and Ricci lower bounds for sub-Riemannian manifolds with transverse symmetries. *J. Eur. Math. Soc. (JEMS)* **19** (1) (2017) 151–219. [MR3584561](#)
- [16] F. Baudoin, M. Gordina and T. Melcher. Quasi-invariance for heat kernel measures on sub-Riemannian infinite-dimensional Heisenberg groups. *Trans. Amer. Math. Soc.* **365** (8) (2013) 4313–4350. [MR3055697](#)
- [17] F. Baudoin and C. Tardif. Hypocoercive estimates on foliations and velocity spherical Brownian motion. *Kinet. Relat. Models* **11** (1) (2018) 1–23. [MR3708179](#)
- [18] F. Baudoin and J. Wang. Curvature dimension inequalities and subelliptic heat kernel gradient bounds on contact manifolds. *Potential Anal.* **40** (2) (2014) 163–193. [MR3152160](#)
- [19] G. Ben Arous, M. Cranston and W. S. Kendall. Coupling constructions for hypoelliptic diffusions: Two examples. In *Stochastic Analysis 193–212. Ithaca, NY, 1993. Proc. Sympos. Pure Math.* **57**. Amer. Math. Soc., Providence, RI, 1995. [MR1335472](#)
- [20] J.-M. Bismut. *Large Deviations and the Malliavin Calculus. Progress in Mathematics* **45**. Birkhäuser Boston, Inc., Boston, MA, 1984. [MR0755001](#)
- [21] M. Bonnefont and N. Juillet. Couplings in  $L^p$  distance of two Brownian motions and their Lévy area. Arxiv preprint, 2018.
- [22] U. Boscain, R. Neel and L. Rizzi. Intrinsic random walks and sub-Laplacians in sub-Riemannian geometry. *Adv. Math.* **314** (2017) 124–184. [MR3658714](#)
- [23] O. Calin, D.-C. Chang, K. Furutani and C. Iwasaki. *Methods and Techniques. Heat Kernels for Elliptic and Sub-Elliptic Operators, Applied and Numerical Harmonic Analysis*. Birkhäuser/Springer, New York, 2011. [MR2723056](#)
- [24] M. Cranston. Gradient estimates on manifolds using coupling. *J. Funct. Anal.* **99** (1) (1991) 110–124. [MR1120916](#)
- [25] B. K. Driver and T. Melcher. Hypoelliptic heat kernel inequalities on the Heisenberg group. *J. Funct. Anal.* **221** (2005) 340–365. [MR2124868](#) <https://doi.org/10.1016/j.jfa.2004.06.012>
- [26] R. M. Dudley. Lorentz-invariant Markov processes in relativistic phase space. *Ark. Mat.* **6** (1966) 241–268. [MR0198540](#)
- [27] R. M. Dudley. A note on Lorentz-invariant Markov processes. *Ark. Mat.* **6** (1967) 575–581. [MR0216567](#)
- [28] R. M. Dudley. Asymptotics of some relativistic Markov processes. *Proc. Natl. Acad. Sci. USA* **70** (1973) 3551–3555. [MR0339344](#)
- [29] J. Dunkel. Relativistic Brownian motion and diffusion processes. Ph.D. thesis, Universität Augsburg, 2008.
- [30] J. Dunkel and P. Hägggi. Relativistic Brownian motion. *Phys. Rep.* **471** (1) (2009) 1–73. [MR2503143](#)
- [31] N. Eldredge. Gradient estimates for the subelliptic heat kernel on  $H$ -type groups. *J. Funct. Anal.* **258** (2) (2010) 504–533. [MR2557945](#)
- [32] K. D. Elworthy. *Stochastic Differential Equations on Manifolds. London Mathematical Society Lecture Note Series* **70**. Cambridge University Press, Cambridge, 1982. [MR0675100](#)
- [33] J. Franchi and Y. Le Jan. Relativistic diffusions and Schwarzschild geometry. *Comm. Pure Appl. Math.* **60** (2) (2007) 187–251. [MR2275328](#)
- [34] J. Franchi and Y. Le Jan. *Hyperbolic Dynamics and Brownian Motion: an Introduction. Oxford Mathematical Monographs*. Oxford University Press, Oxford, 2012. [MR2977883](#)
- [35] M. Gordina. An application of a functional inequality to quasi-invariance in infinite dimensions. In *Convexity and Concentratio* 251–266. Springer, New York, New York, 2017. [MR3837273](#)
- [36] M. Gordina and T. Laetsch. Sub-Laplacians on Sub-Riemannian Manifolds. *Potential Anal.* **44** (4) (2016) 811–837. [MR3490551](#)

- [37] M. Gordina and T. Laetsch. A convergence to Brownian motion on sub-Riemannian manifolds. *Trans. Amer. Math. Soc.* **369** (9) (2017) 6263–6278. [MR3660220](#)
- [38] K. Habermann. Small-time fluctuations for the bridge in a model class of hypoelliptic diffusions of weak Hörmander type, 2018.
- [39] R. S. Hamilton. A matrix Harnack estimate for the heat equation. *Comm. Anal. Geom.* **1** (1) (1993) 113–126. [MR1230276](#)
- [40] R. S. Hamilton. Li–Yau estimates and their Harnack inequalities. In *Geometry and Analysis. No. 1* 329–362. *Adv. Lect. Math. (ALM)* **17**. Int. Press, Somerville, MA, 2011. [MR2882429](#)
- [41] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.* **119** (1967) 147–171. [MR0222474](#)
- [42] E. P. Hsu. *Stochastic Analysis on Manifolds. Graduate Studies in Mathematics* **38**. American Mathematical Society, Providence, RI, 2002. [MR1882015](#)
- [43] N. Ikeda and H. Matsumoto. The Kolmogorov operator and classical mechanics. In *Memoriam Marc Yor – Séminaire de Probabilités XLVII* 497–504. *Lecture Notes in Math.* **2137**. Springer, Cham, 2015. [MR3444311](#)
- [44] K. M. Jansons and P. D. Metcalfe. Optimally coupling the Kolmogorov diffusion, and related optimal control problems. *LMS J. Comput. Math.* **10** (2007) 1–20. [MR2291679](#)
- [45] A. Kolmogoroff. Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). *Ann. of Math.* (2) **35** (1) (1934) 116–117. [MR1503147](#)
- [46] K. Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.* **258** (11) (2010) 3758–3774. [MR2606871](#)
- [47] H.-Q. Li. Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg. *J. Funct. Anal.* **236** (2) (2006) 369–394. [MR2240167](#)
- [48] P. Li and S.-T. Yau. On the parabolic kernel of the Schrödinger operator. *Acta Math.* **156** (3–4) (1986) 153–201. [MR0834612](#)
- [49] H. P. McKean Jr. A winding problem for a resonator driven by a white noise. *J. Math. Kyoto Univ.* **2** (1963) 227–235. [MR0156389](#)
- [50] M. N. Pascu and I. Popescu. Shy and fixed-distance couplings of Brownian motions on manifolds. *Stochastic Process. Appl.* **126** (2) (2016) 628–650. [MR3434996](#) <https://doi.org/10.1016/j.spa.2015.09.014>
- [51] M. N. Pascu and I. Popescu. Couplings of Brownian motions of deterministic distance in model spaces of constant curvature. *J. Theoret. Probab.* **31** (4) (2018) 2005–2031. [MR3866605](#)
- [52] M.-K. von Renesse. Intrinsic coupling on Riemannian manifolds and polyhedra. *Electron. J. Probab.* **9** (14) (2004) 411–435 (electronic). [MR2080605](#)
- [53] M.-K. von Renesse and K.-T. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.* **58** (7) (2005) 923–940. [MR2142879](#)
- [54] F.-Y. Wang. On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups. *Probab. Theory Related Fields* **108** (1) (1997) 87–101. [MR1452551](#)
- [55] F.-Y. Wang. *Analysis for Diffusion Processes on Riemannian Manifolds. Advanced Series on Statistical Science & Applied Probability* **18**. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014. [MR3154951](#)
- [56] F.-Y. Wang. Derivative formulas and Poincaré inequality for Kohn–Laplacian type semigroups. *Sci. China Math.* **59** (2) (2016) 261–280. [MR3454046](#)