GRADIENT BOUNDS FOR KOLMOGOROV TYPE DIFFUSIONS

FABRICE BAUDOIN[⋆], MARIA GORDINA[†‡], AND PHANUEL MARIANO[†]

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 Γ-calculus techniques. The advantages and drawbacks of each of these methods are discussed.

CONTENTS

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. [10, 11]), quasi-invariance of heat kernel measures in infinite dimensions (e.g. [12, 27]), functional inequalities such as Poincaré and log-Sobolev type inequalities (e.g. [9, 19, 37, 46]), and the study of convergence to equilibrium for hypocoercive diffusions (e.g. [8, 13]). In particular, the gradient bounds we present in this paper might be used to prove a spectral gap existence similarly to [14] 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: Γ-calculus methods and coupling methods.

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 [36], where he obtained an explicit expression for the transition density of the diffusion process whose generator is the operator L. Later L. Hörmander in [32] 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 [31], where Ricatti 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. While more geometric methods have been used for hypoelliptic operators (e.g. [8–11]) in the last years, the coupling techniques have seen recent progress for such degenerate operators. In [15], 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 [6], S. Banerjee and W.Kendall used a non-Markovian strategy to couple the iterated Kolmogorov diffusion. The most relevant to our results is [5], 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 Γ-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 Γ-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 [1, 3, 20, 21, 23, 25, 26, 34, 35, 40], while the history of related objects both in mathematics and physics can be found in [22]. Observe that generalized Γ-calculus gives relatively simple proofs of functional inequalities for the relativistic diffusion compared to previous results.

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 [42]. 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 introduced in [29]. If such a coupling can be constructed, then our results and techniques would be valid for even a 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

$$X_t = \left(B_t, \int_0^t B_s ds \right),$$

where B_t is a Brownian motion in $\mathbb{R}^d$ with the variance σ^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 $f \in C^2\left(\mathbb{R}^d \times \mathbb{R}^d\right)$ 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 Δ_p is the Laplace operator Δ on $\mathbb{R}^d$ acting on the variable p and ∇_ξ is the gradient on $\mathbb{R}^d$ acting on the variable ξ.

Note that for $d = 1$ and $\sigma = 1$ this is the original Kolmogorov operator. By Hörmander's theorem in [32], 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. **Γ-calculus.** First we use geometric methods such as generalized Γ-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 Γ-calculus for the Kolmogorov operator has been carried by F.Y. Wang in [45, 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

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

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

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

where ∇_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 $\alpha \in \mathbb{R}$, $\beta \geqslant 0$ we define a symmetric first-order differential bilinear form $\Gamma^{\alpha,\beta} : C^\infty\left(\mathbb{R}^d \times \mathbb{R}^d\right) \times C^\infty\left(\mathbb{R}^d \times \mathbb{R}^d\right) \to \mathbb{R}$ by*

$$\Gamma^{\alpha,\beta}(f,g) := \sum_{i=1}^d \left(\frac{\partial f}{\partial p_i} - \alpha \frac{\partial f}{\partial \xi_i} \right) \left(\frac{\partial g}{\partial p_i} - \alpha \frac{\partial g}{\partial \xi_i} \right) + \beta \sum_{i=1}^d \frac{\partial f}{\partial \xi_i} \frac{\partial g}{\partial \xi_i}$$

(2.2) $$= \langle \nabla_p f, \nabla_p g \rangle - \alpha \langle \nabla_p f, \nabla_\xi g \rangle - \alpha \langle \nabla_\xi f, \nabla_p g \rangle + \left(\alpha^2 + \beta\right) \langle \nabla_\xi f, \nabla_\xi g \rangle,$$

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 $f \in C^\infty \left(\mathbb{R}^d \times \mathbb{R}^d \right)$*

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

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

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

$\square$

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

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

where ϕ is the functional

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

We are now in position to prove regularization properties for the semigroup $P_t = e^{tL}$.

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

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

and

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

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 [17, pp. 80-85]). We consider then (at a given fixed point (ξ, p)) the functional

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

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

$$\phi'(s) = 2P_s \left(\Gamma_2^{\alpha(s),\beta} \left(P_{t-s}f \right) + \langle \nabla_p \left(P_{t-s}f \right), \nabla_\xi \left(P_{t-s}f \right) f \rangle + s \| \nabla_\xi \left(P_{t-s}f \right) \|^2 \right)$$
$$\geqslant 2P_s \left(\alpha(s) \| \nabla_\xi (P_{t-s}f) \|^2 - \langle \nabla_\xi(P_{t-s}f), \nabla_p(P_{t-s}f) \rangle \right)$$
$$+ \langle \nabla_\xi(P_{t-s}f), \nabla_p(P_{t-s}f) \rangle + s \| \nabla_\xi P_{t-s}f \|^2 = 0.$$

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

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

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

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

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

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 $t \geqslant 0$ we have*

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

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 $t > 0$*

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

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

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

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

$$\phi'(s) = - P_s(\Gamma^{\alpha(s),\beta(s)}(P_{t-s}f))$$
$$+ (t-s) P_s \langle \nabla_p P_{t-s}f, \nabla_\xi P_{t-s}f \rangle + (t-s) \left(2\alpha'(s)\alpha(s) + \beta'(s) \right) P_s \| \nabla_\xi P_{t-s}f \|^2$$
$$\geqslant -P_s(\| \nabla_p P_{t-s}f \|^2) = -\frac{2}{\sigma^2} P_s(\Gamma(P_{t-s}f)).$$

Therefore, we have

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

where we used the fact that ϕ 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

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

$\square$

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* $t > 0$

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

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

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

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

$$\phi'(s) = - P_s((P_{t-s}f)\Gamma^{\alpha(s),\beta(s)}(\ln P_{t-s}f)) + 2(t-s)P_s((P_{t-s}f)\Gamma_2^{\alpha(s),\beta(s)}(\ln P_{t-s}f))$$

$$- 2(t-s)\alpha'(s) \sum_{i=1}^d P_s \left((P_{t-s}f) \frac{\partial \ln P_{t-s}f}{\partial \xi_i} \frac{\partial \ln P_{t-s}f}{\partial p_i} \right)$$

$$+ 2(t-s)\alpha(s)\alpha'(s) P_s[(P_{t-s}f)\| \nabla_\xi \ln P_{t-s}f \|^2]$$

$$+ (t-s)\beta'(s) P_s[(P_{t-s}f)\| \nabla_\xi \ln P_{t-s}f \|^2]$$

$$\geqslant - P_s((P_{t-s}f)\| \nabla_p \ln P_{t-s}f \|^2) = -\frac{2}{\sigma^2} P_s((P_{t-s}f)\Gamma(\ln P_{t-s}f)).$$

Therefore, we have

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

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

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

$\square$

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 [9, 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*

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

where

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

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

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

We can now integrate the previous inequality as in [9, Proposition 3.4] and deduce

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

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

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

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 $\widetilde{X}_t$ are two diffusions in $\mathbb{R}^d$ defined on this space with the same generator L, starting at $x, \widetilde{x} \in \mathbb{R}^d$ respectively. By their coupling we understand a diffusion $\left(X_t, \widetilde{X}_t \right)$ in $\mathbb{R}^d \times \mathbb{R}^d$ such that its law is a coupling of the laws of X_t and $\widetilde{X}_t$. That is, the first and the second d-dimensional (marginal) distributions of $\left(X_t, \widetilde{X}_t \right)$ are given by distributions of X_t and $\widetilde{X}_t$.

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

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

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

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

Proof. Consider two copies of Kolmogorov diffusions

$$X_t = (B_t^p, Y_t) = \left(p + B_t, \xi + tp + \int_0^t B_t ds \right),$$

$$\widetilde{X}_t = \left(\widetilde{B}_t^p, \widetilde{Y}_t \right) = \left(\widetilde{p} + \widetilde{B}_t, \xi + t\widetilde{p} + \int_0^t \widetilde{B}_t ds \right),$$

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

$$\left| B_t^p - \widetilde{B}_t^p \right| = |p - \widetilde{p}|,$$

$$\left| Y_t - \widetilde{Y}_t \right| = t |p - \widetilde{p}|,$$

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

$$\left| f(X_t) - f\left(\widetilde{X}_t \right) \right|$$

$$= \left| \sum_{i=1}^d \partial_{p_i} f\left(\widetilde{X}_t \right) (p_i - \widetilde{p}_i) + \sum_{i=1}^d t \partial_{\xi_i} f\left(\widetilde{X}_t \right) (p_i - \widetilde{p}_i) + R\left(\widetilde{X}_t \right) \right|$$

$$\leqslant \sum_{i=1}^d \left| \left(\partial_{p_i} f\left(\widetilde{X}_t \right) + t \partial_{\xi_i} f\left(\widetilde{X}_t \right) \right) \right| |p - \widetilde{p}| + \frac{C_f}{2} d^2 (1 + t)^2 |p - \widetilde{p}|^2.$$

Using this estimate and Jensen's inequality we see that

$$\left|P_t f(p, \xi) - P_t f\left(\widetilde{p}, \xi\right)\right| = \left|\mathbb{E}^{((p,\xi),(\widetilde{p},\xi))}\left[f\left(X_t\right) - f\left(\widetilde{X_t}\right)\right]\right|$$

$$\leqslant \mathbb{E}^{((p,\xi),(\widetilde{p},\xi))}\left[\left|f\left(X_t\right) - f\left(\widetilde{X_t}\right)\right|\right]$$

$$\leqslant \sum_{i=1}^{d} \mathbb{E}^{((p,\xi),(\widetilde{p},\xi))}\left[\left|\left(\partial_{p_i} f\left(\widetilde{X_t}\right) + t\partial_{\xi_i} f\left(\widetilde{X_t}\right)\right)\right|^q\right]^{\frac{1}{q}} \left|p - \widetilde{p}\right|$$

$$+ \frac{C_f}{2} d^2 (1+t)^2 \left|p - \widetilde{p}\right|^2$$

$$= \sum_{i=1}^{d} P_t \left(\left|\left(\partial_{p_i} f\left(\widetilde{p}, \xi\right) + t\partial_{\xi_i} f\left(\widetilde{p}, \xi\right)\right)\right|^q\right)^{\frac{1}{q}} \left|p - \widetilde{p}\right|$$

$$+ \frac{C_f}{2} d^2 (1+t)^2 \left|p - \widetilde{p}\right|^2 .$$

Dividing out by $\left|p - \widetilde{p}\right|$ and taking $\widetilde{p} \to p$ we have that

$$\left\|\nabla_p P_t f(p, \xi)\right\| = \limsup_{\widetilde{p} \to p} \frac{\left|P_t f(p, \xi) - P_t f\left(\widetilde{p}, \xi\right)\right|}{\left|p - \widetilde{p}\right|}$$

$$\leqslant \sum_{i=1}^{d} P_t \left(\left|\partial_{p_i} f(p, \xi) + t\partial_{\xi_i} f(p, \xi)\right|^q\right)^{\frac{1}{q}},$$

which proves the statement. $\qquad\square$

Remark 2.11. *When $q = 2$, this coincides with the conclusion of Proposition 2.5. The coupling method here is simpler than the Γ-calculus method and moreover yields a family of inequalities for $q \geqslant 1$. 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 $X_t = (B_t, \int_0^t B_s ds)$, where B_t is a Brownian motion on the d-dimensional hyperbolic space $\mathbb{H}^n$. This is the relativistic Brownian motion introduced by R. Dudley [20] and studied by J. Franchi and Y. Le Jan in [25]. In this section, we will prove functional inequalities for the generator of X_t. Our methods will only involve Γ-calculus through generalized curvature dimension conditions. The emphasis on Γ-calculus in this section will allow us to obtain sharper estimates for the relativistic diffusion. In particular, the estimate (3.6) 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 Γ-calculus and coupling techniques but for a larger class of diffusions.

We follow the notation in [25]. Recall that the Minkowski space is the product $\mathbb{R} \times \mathbb{R}^d$ with $d \geqslant 2$

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

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 $e_0, ..., e_d$. Let $\mathbb{H}^d$ be the positive half of the unit sphere in $\mathbb{R}^{1,d}$, namely,

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

Note that $\mathbb{H}^d$ has a standard parametrization $p = (p_0, \vec{p}) = (\cosh r, \sinh r \, \omega)$ with $r \geqslant 0$, $\omega \in \mathbb{S}^{d-1}$. 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 [26, Proposition 3.5.4]).

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

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

Following the construction in [20], 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 [26, 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 $f \in C^2\left(\mathbb{H}^d \times \mathbb{R}^{1,d}\right)$ is the operator

$$(Lf)(p, \xi) = \langle p, \nabla_\xi f(p, \xi) \rangle + \frac{\sigma^2}{2} \Delta^{\mathbb{H}}_p f(p, \xi) =$$

$$p_0 \frac{\partial f}{\partial \xi_0}(p, \xi) + \sum_{j=1}^d p_j \frac{\partial f}{\partial \xi_j}(p, \xi) + \frac{\sigma^2}{2} \Delta^{\mathbb{H}}_p f(p, \xi),$$

where $\Delta^{\mathbb{H}}_p$ 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 ∇_ξ for the usual Euclidean gradient. Let $\Gamma(f)$ be the *carré du champ operator* for L, while let $\Gamma^{\mathbb{H}}$ be the *carré du champ operator* for $\Delta^{\mathbb{H}}$. 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 ∇_ξ playing a role of the vertical gradient. Namely, we define a symmetric, first-order differential bilinear form $\Gamma^Z : C^\infty\left(\mathbb{H}^d \times \mathbb{R}^{1,d}\right) \times C^\infty\left(\mathbb{H}^d \times \mathbb{R}^{1,d}\right) \to \mathbb{R}$ by

$$(3.4) \qquad\qquad \Gamma^Z(f) := \|\nabla_\xi f\|^2,$$

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

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

$$\Gamma_2(f) \geqslant -\frac{d}{2}\sigma^2\Gamma(f) - \frac{1}{4}\Gamma^Z(f),$$
$$\Gamma_2^Z(f) \geqslant 0.$$

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

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

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 $\Delta_p^{\mathbb{H}}$

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

thus

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

Now we can use an elementary estimate

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

to see that

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

The last term in this inequality is the bilinear form Γ^Z defined by (3.4). 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

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

which concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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 such that $LW \leqslant CW$ for some $C > 0$. Consider the function

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

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

$$W \geqslant 1,$$
$$\|\nabla_\xi W\| + \|\nabla_p W\| \leqslant CW,$$
$$LW \leqslant CW \text{ for some constant } C > 0,$$
$$\{W \leqslant m\} \text{ is compact for every } m.$$

We shall make use of the Lyapunov function W defined by (3.5) to prove the following result.

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

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

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

$$\varphi_1 (x, s) := \Gamma \left(P_{t-s} f \right)(x),$$
$$\varphi_2 (x, s) := \Gamma^Z \left(P_{t-s} f \right)(x).$$

Then

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

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

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

Then by Theorem 3.1 we have

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

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 (3.5) and a cutoff argument as in [7, Theorem 7.3], we deduce from a parabolic comparison principle

$$P_t \left(\varphi \left(\cdot, t \right) \right)(x) \geqslant \varphi (x, 0).$$

Observe that

$$\varphi\left(x,0\right)=a\left(0\right)\varphi_1\left(x,0\right)+b\left(0\right)\varphi_2\left(x,0\right)=2d\sigma^2\Gamma\left(P_tf\right)\left(x\right)+\Gamma^Z\left(P_tf\right)\left(x\right),$$

$$P_t\left(\varphi\left(\cdot,t\right)\right)\left(x\right)=a\left(t\right)P_t\left(\Gamma\left(f\right)\right)\left(x\right)+b\left(t\right)P_t\left(\Gamma^Z\left(f\right)\right)\left(x\right)=$$

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

therefore

$$2d\sigma^2\Gamma\left(P_tf\right)\left(x\right)+\Gamma^Z\left(P_tf\right)\left(x\right)\leqslant e^{d\sigma^2t}\left(2d\sigma^2P_t\left(\Gamma\left(f\right)\right)\left(x\right)+P_t\left(\Gamma^Z\left(f\right)\right)\left(x\right)\right).$$

$\square$

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

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

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

$$\int_0^t P_s\left(2d\sigma^2\Gamma\left(P_{t-s}f\right)+\Gamma^Z\left(P_{t-s}f\right)\right)ds$$

$$\geqslant\int_0^t P_s\left(2d\sigma^2\Gamma\left(P_{t-s}f\right)\right)ds=d\sigma^2\left(P_t\left(f^2\right)-\left(P_tf\right)^2\right).$$

By Theorem 3.2 we have that

$$\int_0^t P_s\left(2d\sigma^2\Gamma\left(P_{t-s}d\right)+\Gamma^Z\left(P_{t-s}f\right)\right)ds$$

$$\leqslant\int_0^t e^{d\sigma^2(t-s)}P_s\left(2d\sigma^2P_{t-s}\left(\Gamma\left(f\right)\right)+P_{t-s}\left(\Gamma^Z\left(f\right)\right)\right)ds$$

$$=\left(2d\sigma^2P_t\left(\Gamma\left(f\right)\right)+P_t\left(\Gamma^Z\left(f\right)\right)\right)\int_0^t e^{d\sigma^2(t-s)}ds$$

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

This implies

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

$\square$

The next corollary gives us an equivalent estimate to the one in Theorem 3.2. The estimate (3.6) will be similar to the one we will obtain in Theorem 4.8 in a more general setting.

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

$$2d\sigma^2\Gamma\left(P_tf\right)+\Gamma^Z\left(P_tf\right)\leqslant e^{d\sigma^2t}\left(2d\sigma^2P_t\left(\Gamma(f)\right)+P_t\left(\Gamma^Z(f)\right)\right),$$

is equivalent to

$$(3.6)\qquad\Gamma(P_tf)\leqslant e^{d\sigma^2t}P_t\left(\Gamma\left(f\right)\right)+\frac{e^{d\sigma^2t}-1}{2d\sigma^2}P_t\left(\Gamma^Z\left(f\right)\right).$$

Moreover, one has

$$\Gamma^Z\left(P_t f\right) \leqslant P_t\left(\Gamma^Z\left(f\right)\right).$$

Proof. Recall that

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

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

$$\int_0^t P_s\left(2d\sigma^2\Gamma\left(P_{t-s}f\right) + \Gamma^Z\left(P_{t-s}f\right)\right)ds$$

$$\geqslant \int_0^t P_s\left(2d\sigma^2\Gamma\left(P_{t-s}f\right) - 2d\sigma^2\Gamma(P_{t-s}f) - 4\Gamma_2\left(P_{t-s}f\right)\right)ds$$

$$= -2\left(P_t\left(\Gamma(f)\right) - \left(\Gamma(P_t f)\right)\right).$$

On the other hand we have

$$\int_0^t P_s\left(2d\sigma^2\Gamma\left(P_{t-s}f\right) + \Gamma^Z\left(P_{t-s}f\right)\right)ds$$

$$\leqslant \int_0^t e^{d\sigma^2(t-s)}P_s\left(2d\sigma^2 P_{t-s}\left(\Gamma\left(f\right)\right) + P_{t-s}\left(\Gamma^Z\left(f\right)\right)\right)ds$$

$$= \left(2d\sigma^2 P_t\left(\Gamma\left(f\right)\right) + P_t\left(\Gamma^Z\left(f\right)\right)\right)\int_0^t e^{d\sigma^2(t-s)}ds$$

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

Putting these together we have

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

A rearranging of this inequality gives us

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

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

$$2d\sigma^2\Gamma(P_t f) + \Gamma^Z\left(P_t f\right)$$

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

$$= e^{d\sigma^2 t}\left(2d\sigma^2 P_t\left(\Gamma(f)\right) + P_t\left(\Gamma^Z(f)\right)\right) + \Gamma^Z\left(P_t f\right) - P_t\left(\Gamma^Z\left(f\right)\right)$$

$$\leqslant e^{d\sigma^2 t}\left(2d\sigma^2 P_t\left(\Gamma(f)\right) + P_t\left(\Gamma^Z(f)\right)\right) + 0.$$

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

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

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

4. Gradient bounds for a general Kolmogorov diffusion

We now study the diffusions of the type $\mathbf{X}_t = (X_t, \int_0^t \sigma(X_s)ds)$ for $\sigma : \mathbb{R}^k \to \mathbb{R}^k$ where X_t is a Markov process on $\mathbb{R}^k$. We will show that a generalized curvature dimension condition for the generator of $\mathbf{X}_t$ is satisfied as in Theorem 3.1. In section 4.1, we prove gradient bounds for a Kolmogorov type diffusions on $\mathbb{R}^k \times \mathbb{R}^k$ using a Γ-calculus approach. In section 4.2, we show that the results in section 4.1 are applicable to a large class of diffusions. In section 4.3, we prove gradient bounds when X_t is assumed to live on a Riemannian manifold using coupling techniques. 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 assumed to live in the Heisenberg group.

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

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

where the V_i for $i = 0, \ldots, k$ are smooth vector fields. Here we assume $\sigma : \mathbb{R}^k \to \mathbb{R}^k$ is a C^1 map such that

$$(4.7) \qquad C_\sigma := \left(\sum_{i,j=1}^{d} (V_i \sigma_j)^2 \right)^{\frac{1}{2}} < \infty.$$

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 $\Gamma^Z : C^\infty \left(\mathbb{R}^k \times \mathbb{R}^k \right) \times C^\infty \left(\mathbb{R}^k \times \mathbb{R}^k \right) \to \mathbb{R}$ by

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\sum_{i,j=1}^{k}\left(V_if\right)\left(V_i\sigma_j\right)\frac{\partial f}{\partial \xi_j} \leqslant \left(\sum_{i,j=1}^{k}\left(V_i\sigma_j\right)^2\right)^{\frac{1}{2}}\left(\sum_{i,j=1}^{k}\left(V_if\right)^2\left(\frac{\partial f}{\partial \xi_j}\right)^2\right)^{\frac{1}{2}}$$

$$\leqslant C_\sigma\left(\Gamma(f)\right)^{\frac{1}{2}}\left(\Gamma^Z(f)\right)^{\frac{1}{2}}$$

$$\leqslant \frac{C_\sigma}{2}\left(\Gamma(f)\right) + \frac{C_\sigma}{2}\left(\Gamma^Z(f)\right).$$

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

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

for which routine computations shows that

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

as needed.   □

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 $W : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ is a *Lyapunov function* on $\mathbb{R}^k$ for $\mathcal{L}$ if

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

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^∞.

Throughout this section, we will need the following assumption.

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

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

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

$$\Gamma\left(P_t f\right)(x) + \frac{C_\sigma}{C_\sigma - 2\rho} \Gamma^Z\left(P_t f\right)(x)$$

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

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

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

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

Then

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

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

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

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

Then by Theorem 4.1 we have

$$\mathcal{L}\varphi + \frac{\partial \varphi}{\partial s} =$$

$$a'(s)\Gamma\left(P_{t-s} f\right) + b'(s)\Gamma^Z\left(P_{t-s} f\right) + 2a(s)\Gamma_2\left(P_{t-s} f\right) + 2b(s)\Gamma_2^Z\left(P_{t-s} f\right) \geqslant$$

$$a'(s)\Gamma\left(P_{t-s} f\right) + b'(s)\Gamma^Z\left(P_{t-s} f\right) + 2a(s)\left(\left(\rho - \frac{C_\sigma}{2}\right)\Gamma\left(P_{t-s} f\right) - \frac{C_\sigma}{2}\Gamma^Z\left(P_{t-s} f\right)\right) =$$

$$\left(a'(s) + a(s)(2\rho - C_\sigma)\right)\Gamma\left(P_{t-s} f\right) + \left(b'(s) - a(s)C_\sigma\right)\Gamma^Z\left(P_{t-s} f\right).$$

One can easily see that if

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

the last expression is 0. Using the existence of the Lyapunov function W and a cutoff argument as in [7, Theorem 7.3], we deduce from a parabolic comparison principle,

$$P_t\left(\varphi(\cdot, t)\right)(x) \geqslant \varphi(x, 0).$$

Observe that

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

while

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

$\square$

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

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

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

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

By Theorem 4.3 we have that

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

So that

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

$\square$

4.2. **Examples.** To illustrate the results in Section 4.1 we study a large class of examples. 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 $X_t=\left(B_t,\int_0^t\sigma\left(B_s\right)ds\right)$ where $\sigma:M\to\mathbb{R}^k$ satisfies (4.7) and

$$|\sigma(p)-\sigma(\tilde{p})|\leqslant C_\sigma d_M\left(p,\tilde{p}\right),$$

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 [33, p. 77]). 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 $\mathrm{Ric} \geqslant \rho$ for some $\rho \in \mathbb{R}$. Then it is known from the Li-Yau upper and lower bounds in [39] that the heat kernel $p(x, y, t)$ of M satisfies the following Gaussian estimates. Namely, for some $\tau > 0$

$$\frac{c_1}{Vol(B(p_0, \sqrt{\tau}))} \exp\left(-\frac{c_2 d_M(p_0, p_1)^2}{\tau}\right) \leqslant p(p_0, p_1, \tau)$$

$$\leqslant \frac{c_3}{Vol(B(p_0, \sqrt{\tau}))} \exp\left(-\frac{c_4 d_M(p_0, p_1)^2}{\tau}\right),$$

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

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

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

Lemma 4.5. *The function W defined in (4.8) is smooth and satisfies the following properties,*

$$W \geqslant 1,$$
$$\|\nabla_\xi W\| + \|\nabla_p W\| \leqslant CW,$$
$$\mathcal{L}W \leqslant CW \text{ for some constant } C > 0,$$
$$\{W \leqslant m\} \text{ is compact for every } m.$$

Here ∇_p is the Riemannian gradient on M and ∇_ξ is the Euclidean gradient on $\mathbb{R}^k$.

Proof. From estimates for logarithmic derivatives of the heat kernel in [30, 39], one has for some constants $C_1, C_2 > 0$

(4.9) $$\|\nabla_p \ln p(p_0, p, \tau)\|^2 \leqslant C_1 + C_2 d_M(p_0, p)^2,$$

(4.10) $$\Delta_p(-\ln p(p_0, p, \tau)) \leqslant C_1 + C_2 d_M(p_0, p)^2.$$

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

$$d_M(p_0, p)^2 \leqslant -\frac{\tau}{c_4} \ln\left(\frac{\mathrm{Vol}(B(p_0, \sqrt{\tau}))}{c_3} p(p_0, p, \tau)\right)$$

(4.11) $$\leqslant C(K - \ln(p(p_0, p, \tau)))$$

for a fixed $\tau \geqslant 0$ and a constant $C > 0$. Hence, $\|\nabla_\xi W\| + \|\nabla_p W\| \leqslant CW$ can be shown using (4.9), (4.11) and the inequality $(1 + x)^{\frac{1}{2}} \leqslant 1 + cx$ for $x \geqslant 0$ and $c \geqslant \frac{1}{2}$. On the other hand, $\mathcal{L}W \leqslant CW$ can be shown using (4.10), (4.11), the

Cauchy-Schwarz inequality, and the Lipschitz property of σ. Finally, the fact that $\{W \leqslant m\}$ 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 (4.8) is a Lyapunov function satisfying Assumption 4.2. As a consequence, Theorem 4.3 can be applied to complete Riemannian manifolds with $\mathrm{Ric} \geqslant \rho$ since the condition $\mathrm{Ric} \geqslant \rho$ is equivalent to

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

4.3. **Coupling.** Let (M, g) be a complete connected $d-$dimensional Riemannian manifold which is embedded in $\mathbb{R}^k$. We consider the process

$$(4.12) \qquad\qquad X_t = \left(B_t, \int_0^t \sigma (B_s)\, ds \right),$$

where B_t is Brownian motion on M. We assume the map $\sigma : M \to \mathbb{R}^k$ is a globally C_σ-Lipschitz map in the sense that

$$(4.13) \qquad\qquad |\sigma (p) - \sigma (\widetilde{p})| \leqslant C_\sigma d_M (p, \widetilde{p}),$$

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

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 ∇_p and Δ_p act on the variable p for $f (p, \xi)$, where Δ_p is the Laplace-Beltrami operator. We use ∇_ξ 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 $f \in C_0^\infty (M \times \mathbb{R}^k)$.

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

In [42–44] it has been shown that if we assume $\mathrm{Ric} (M) \geqslant K$ for some $K \in \mathbb{R}$, then there exists a Markovian coupling of Brownian motions $(B_t)_{t \geqslant 0}$ and $\left(\widetilde{B}_t \right)_{t \geqslant 0}$ on M starting at p and $\widetilde{p}$ such that

$$(4.14) \qquad\qquad d_M \left(B_t, \widetilde{B}_t \right) \leqslant e^{-Kt/2} d_M (p, \widetilde{p})$$

for all $t \geqslant 0$, $\mathbb{P}^{(p, \widetilde{p})}$-almost surely. This construction is known as a *coupling by parallel transport*. This coupling can be constructed using stochastic differential equations as in [18, 44], or by a central limit theorem argument for the geodesic random walks as in [42]. It turns out that the existence of the coupling satisfying (4.14) is equivalent to

$$(4.15) \qquad\qquad \|\nabla P_t f\| \leqslant e^{-Kt} P_t (\|\nabla f\|),$$

for all $f \in C_0^\infty (M)$ and all $t > 0$. We also point out that in [41], M. Pascu and I. Popescu constructed explicit Markovian couplings where equality in (4.14) is attained for $t \geqslant 0$ given some extra geometric assumptions.

The coupling by parallel transport that gives (4.14) is in the elliptic setting. In this section, we will use the coupling by parallel transport to induce a coupling for (4.12) 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*

$$(4.16) \qquad \lim_{r \to 0} \sup_{\widetilde{p}: 0 < d_M(p,\widetilde{p}) \leqslant r} \frac{|f(p) - f(\widetilde{p})|}{d_M (p, \widetilde{p})} = \|\nabla f(p)\|.$$

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

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

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

$$\lim_{r \to 0} \sup_{\widetilde{p}: 0 < d_M(p,\widetilde{p}) \leqslant r} \frac{|f(p) - f(\widetilde{p})|}{d_M (p, \widetilde{p})} \leqslant \|\nabla f(p)\|.$$

On the other hand, find a unit speed geodesic $\gamma : (-\epsilon, \epsilon) \to M$ 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

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

which means we have that the left hand side of (4.16) must be at least $\|\nabla f(p)\|$. This proves (4.16).

$\square$

The following lemma gives an estimate for $\left| f(p, \xi) - f(\widetilde{p}, \widetilde{\xi}) \right|$ 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 ξ. If $f \in C_0^2 \left(M \times \mathbb{R}^k \right)$,*

then there exists a $C_f > 0$ depending on a bound on the Hessian of f such that

$$\left| f(p,\xi) - f(\widetilde{p},\widetilde{\xi}) \right| \leqslant \left\| \nabla_p f\left(\widetilde{p},\widetilde{\xi}\right) \right\| d_M\left(p,\widetilde{p}\right) + \left\| \nabla_\xi f\left(\widetilde{p},\widetilde{\xi}\right) \right\| d_E\left(\xi,\widetilde{\xi}\right)$$
$$+ C_f \left(d_M\left(p,\widetilde{p}\right) + d_E\left(\xi,\widetilde{\xi}\right) \right)^2$$

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

Proof. Let $p, \widetilde{p} \in M$ with $T_1 = d_M\left(p,\widetilde{p}\right)$ and consider a unit speed geodesic $\gamma :$ $[0,T_1] \to M$ such that $\gamma(0) = \widetilde{p}$ and $\gamma(T_1) = p$. Let $\xi, \widetilde{\xi} \in \mathbb{R}^k$ with $T_2 = d_E\left(\xi,\widetilde{\xi}\right)$ and consider $\beta(s) = \frac{s}{d_E(\xi,\widetilde{\xi})}\left(\xi - \widetilde{\xi}\right) + \widetilde{\xi}$ on $-\infty \leqslant s \leqslant T_2$ such that $\beta(0) = \widetilde{\xi}$ and $\beta(T_2) = \xi$. Extend γ 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

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

Now by the chain rule we have

$$F_t(0,0) = \frac{d}{dt}\left[f\left(\gamma(t), \beta(0)\right)\right]_{t=0} = \langle \nabla_p f\left(\gamma(0), \beta(0)\right), \gamma'(0) \rangle$$
$$\leqslant \|\nabla_p f\left(\gamma(0), \beta(0)\right)\| = \left\| \nabla_p f\left(\widetilde{p},\widetilde{\xi}\right) \right\|.$$

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

$$\left| f(p,\xi) - f(\widetilde{p},\widetilde{\xi}) \right| = |F(T_1, T_2) - F(0,0)|$$
$$\leqslant \left\| \nabla_p f\left(\widetilde{p},\widetilde{\xi}\right) \right\| T_1 + \left\| \nabla_\xi f\left(\widetilde{p},\widetilde{\xi}\right) \right\| T_2 + C_f (T_1 + T_2)^2,$$

as needed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We are now ready to state and prove the main theorem of this section. We start by considering the coupling of Brownian motions $\left(B_t, \widetilde{B}_t\right)$ starting at $(p, \widetilde{p})$ by parallel transport satisfying (4.14), as introduced in [42,43] . This coupling induces a coupling $\mathbb{P}^{(\mathbf{x}, \widetilde{\mathbf{x}})}$ on $\left(M \times \mathbb{R}^d\right) \times \left(M \times \mathbb{R}^d\right)$ for two Kolmogorov type diffusions

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

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

Theorem 4.8 (Bakry-Émery type estimate). *Let M be a complete connected Riemannian manifold such that $Ric(M) \geqslant K$ for some $K \in \mathbb{R}$. Let σ be a $C_\sigma-Lipschitz$ map as in (4.13) and $f \in C^2\left(M \times \mathbb{R}^k\right)$ with a bounded Hessian. Then for every $q \geqslant 1$ and $t \geqslant 0$,*

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

where

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

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 $\widetilde{\mathbf{x}} = (\widetilde{p}, \xi) \in M \times \mathbb{R}^k$. If $K \neq 0$, we consider the coupling by parallel transport of Brownian motions $\left(B_t, \widetilde{B}_t\right)$ starting at $(p, \widetilde{p})$. This coupling gives us that

$$(4.17) \qquad d_M\left(B_t, \widetilde{B}_t\right) \leqslant e^{-Kt/2} d_M\left(p, \widetilde{p}\right),$$

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

$$d_E\left(Y_t, \widetilde{Y}_t\right) \leqslant \int_0^t \left|\sigma\left(B_s\right) - \sigma\left(\widetilde{B}_s\right)\right| ds \leqslant C_\sigma \int_0^t d_M\left(B_s, \widetilde{B}_s\right) ds$$

$$(4.18) \qquad \leqslant C_\sigma d_M\left(p, \widetilde{p}\right) \int_0^t e^{-Ks/2} ds = C_\sigma \left(\frac{1 - e^{-Kt/2}}{K/2}\right) d_M\left(p, \widetilde{p}\right),$$

where we used (4.13) and (4.14) . If $K = 0$, we consider the same coupling for the Brownian motions $\left(B_t, \widetilde{B}_t\right)$ starting at $(p, \widetilde{p})$ so that

$$(4.19) \qquad d_M\left(B_t, \widetilde{B}_t\right) \leqslant d_M\left(p, \widetilde{p}\right),$$

Âfor all $t \geqslant 0$. A similar computation as in (4.18) gets us the estimate

$$(4.20) \qquad d_E\left(Y_t, \widetilde{Y}_t\right) \leqslant C_\sigma t d_M\left(p, \widetilde{p}\right),$$

from (4.19). Combining (4.17) and (4.19) we get

$$(4.21) \qquad d_M\left(B_t, \widetilde{B}_t\right) \leqslant K_1(t) d_M\left(p, \widetilde{p}\right),$$

while combining (4.18) and (4.20) we have

$$(4.22) \qquad d_E\left(Y_t, \widetilde{Y}_t\right) \leqslant K_2(t) d_M\left(p, \widetilde{p}\right),$$

for all $t \geqslant 0$, where all of these inequalities hold $\mathbb{P}^{(\mathbf{x}, \widetilde{\mathbf{x}})}$−almost surely. By Lemma 4.7, there exists a $C_f \geqslant 1$ depending on a bound on the Hessian of $f \in C_0^2\left(M \times \mathbb{R}^k\right)$ such that

$$\left|f\left(B_t, Y_t\right) - f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right| \leqslant \left\|\nabla_p f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right\| d_M\left(B_t, \widetilde{B}_t\right) + \left\|\nabla_\xi f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right\| d_E\left(Y_t, \widetilde{Y}_t\right)$$

$$(4.23) \qquad + C_f \left(d_M\left(B_t, \widetilde{B}_t\right) + d_E\left(Y_t, \widetilde{Y}_t\right)\right)^2,$$

for all $t \geqslant 0$, $\mathbb{P}^{(\mathbf{x}, \widetilde{\mathbf{x}})}$−almost surely.

Using inequalities (4.21), (4.22) and (4.23), we have that for $f \in C_0^2\left(M \times \mathbb{R}^k\right)$

$$\left|P_t f\left(p, \xi\right) - P_t f\left(\widetilde{p}, \xi\right)\right| = \left|\mathbb{E}^{(\mathbf{x}, \widetilde{\mathbf{x}})}\left[f\left(B_t, Y_t\right) - f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right]\right|$$

$$\leqslant \mathbb{E}^{(\mathbf{x}, \widetilde{\mathbf{x}})}\left[\left\|\nabla_p f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right\| d_M\left(B_t, \widetilde{B}_t\right) + \left\|\nabla_\xi f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right\| d_E\left(Y_t, \widetilde{Y}_t\right)\right]$$

$$+ C_f \mathbb{E}^{(\mathbf{x}, \widetilde{\mathbf{x}})}\left[d_M\left(B_t, \widetilde{B}_t\right) + d_E\left(Y_t, \widetilde{Y}_t\right)\right]^2$$

$$\leqslant \mathbb{E}^{(\mathbf{x}, \widetilde{\mathbf{x}})}\left[K_1(t)\left\|\nabla_p f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right\| + K_2(t)\left\|\nabla_\xi f\left(\widetilde{B}_t, \widetilde{Y}_t\right)\right\|\right] d_M\left(p, \widetilde{p}\right)$$

$$+ C_f \left(K_1(t) + K_2(t)\right)^2 d_M\left(p, \widetilde{p}\right)^2.$$

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

$$|P_t f(p, \xi) - P_t f(\widetilde{p}, \xi)|$$
$$\leqslant \left(\mathbb{E}^{(\mathbf{x}, \widetilde{\mathbf{x}})} \left[\left(K_1(t) \left\| \nabla_p f\left(\widetilde{B}_t, \widetilde{Y}_t \right) \right\| + K_2(t) \left\| \nabla_\xi f\left(\widetilde{B}_t, \widetilde{Y}_t \right) \right\| \right)^q \right] \right)^{\frac{1}{q}} d_M(p, \widetilde{p})$$
$$+ C_f \left(K_1(t) + K_2(t) \right)^2 d_M(p, \widetilde{p})^2.$$

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

$$\frac{|P_t f(p, \xi) - P_t f(\widetilde{p}, \xi)|}{d_M(p, \widetilde{p})} \leqslant \left[P_t \left((K_1(t) \|\nabla_p f\| + K_2(t) \|\nabla_\xi f\|)^q \right) (\widetilde{p}, \xi) \right]^{\frac{1}{q}}$$
$$+ C_f \left(K_1(t) + K_2(t) \right)^2 d_M(p, \widetilde{p}).$$

Since

$$\lim_{r \to 0} \sup_{\widetilde{p} : 0 < d_M(p, \widetilde{p}) \leqslant r} \frac{|P_t f(p, \xi) - P_t f(\widetilde{p}, \xi)|}{d_M(p, \widetilde{p})} = \|\nabla_p P_t f(p, \xi)\|$$

by Proposition 4.6, we have the desired result. □

Remark 4.9. *The constants obtained in Theorem 4.8 using the coupling technique are sharper than the constants in Theorem 4.3 using Γ-calculus. The trade off here being that the Γ-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 is $M = \mathbb{R}^d$. 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 [2] and further studied in [13]. When $\mathcal{M} = \mathbb{R}^{d+1}$ and $\sigma = \kappa = 1$ the diffusion is of the form $X_t = (B_t, \int_0^t B_s ds)$ 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, $\mathbb{S}^d = \left\{ \mathbf{x} \in \mathbb{R}^{d+1} \mid |\mathbf{x}| = 1 \right\}$. 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

$$(4.24) \qquad\qquad\qquad d_E(\mathbf{x}, \mathbf{y}) \leqslant d_{\mathbb{S}^d}(\mathbf{x}, \mathbf{y}),$$

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 (4.24) shows that $\sigma : \mathbb{S}^d \to \mathbb{R}^{d+1}$ is a $C_\sigma = 1$-Lipschtiz map. Thus we can apply Theorem 4.8 to the manifold $M = \mathbb{S}^d$, since $Ric = (d-1)g$ where g is the Riemannian metric.

4.4. **Iterated Kolmogorov diffusions.** Our technique can also be applied in studying *iterated Kolmogorov diffusions* similar to those studied by Banerjee and Kendall in [6]. An iterated Kolmogorov diffusion is of the form $X_t = (B_t, I_1(t), \ldots, I_n(t))$ where

$$I_0(t) = \sigma\left(B_t\right),$$

$$I_r(t) = \int_0^t I_{r-1}(s)ds, \quad \text{for } r = 1, \ldots, n,$$

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

$$X_t = (B_t, I_1(t), \ldots, I_n(t)).$$

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

Theorem 4.12. *Let M be a complete connected Riemannian manifold such that* $\mathrm{Ric}(M) \geqslant K$ *for some $K \in \mathbb{R}$. When $K = 0$ and $f \in C_0^\infty\left(M \times \mathbb{R}^k \times \cdots \times \mathbb{R}^k\right)$ with $f\left(p, \xi_1, \ldots, \xi_n\right), p \in M, \xi_1, \ldots, \xi_n \in \mathbb{R}^k$ we have the following gradient bound for the iterated Kolmogorov diffusion semigroup P_t,*

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

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

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

for $q \geqslant 1$, where

$$K_1(t) = C_\sigma \frac{1 - e^{-Kt/2}}{K/2},$$

$$K_r(t) = \int_0^t K_{r-1}(s)ds, \quad \text{for } r = 2, \ldots, n.$$

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

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

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

$$\mathcal{X} := \partial_x - \frac{y}{2}\partial_z,$$

$$\mathcal{Y} := \partial_y - \frac{x}{2}\partial_z,$$

$$\mathcal{Z} := \partial_z.$$

The horizontal distribution is defined by $\mathcal{H} = \mathrm{span}\{\mathcal{X}, \mathcal{Y}\}$, fiberwise. Vectors in $\mathcal{H}$ are said to be *horizontal*. We endow $\mathbb{G}$ with the sub-Riemannian metric $g\left(\cdot, \cdot\right)$ 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 $\|v\| = \left(g_p\left(v, v\right)\right)^{\frac{1}{2}}$ 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_{CC}. The *horizontal gradient*

$\nabla_{\mathcal{H}}$ is a horizontal vector field such that for any smooth $f : \mathbb{G} \to R$ we have that for all $X \in \mathcal{H}$

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

The operator

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

is a natural sub-Laplacian for the Heisenberg as pointed out in [28, Example 6.1]. Brownian motion on the Heisenberg group is defined to be the diffusion process $\{B_t^p\}_{t \geqslant 0}$ starting at $p = (x, y, z) \in \mathbb{R}^3$ whose infinitesimal generator is $\Delta_{\mathcal{H}}$. Explicitly the process is given by

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

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 [4, 19, 24, 38]. In particular, the L^1-gradient bounds for the heat semigroup have been proven in [4] and [38]. As pointed out in [37], Kuwada's duality between L^1-gradient bounds and L^∞-Wasserstein control shows that for each $t > 0$, and $p, \widetilde{p} \in \mathbb{G}$, there exists a coupling $\left(B_t^p, \widetilde{B}_t^p\right)$ of Brownian motions on the Heisenberg group such that

$$(4.25) \qquad d_{CC}\left(B_t^p, \widetilde{B}_t^p\right) \leqslant K d_{CC}\left(p, \widetilde{p}\right),$$

almost surely for some constant $K \geqslant 1$ that does not depend on $p, \widetilde{p}, t$. We remark that in [16], the authors show that any coupling that satisfy (4.25) on $\mathbb{G}$ must be non-Markovian. This further highlights the need for more non-Markovian coupling techniques as in [5, 6].

Consider the Kolmogorov diffusion $X_t = \left(B_t^p, \xi + \int_0^t \sigma(B_s^p)ds\right)$ on $\mathbb{G} \times \mathbb{R}^3$, where $\sigma : \mathbb{G} \to \mathbb{R}^3$ satisfies 4.13 and let P_t be the heat semigroup associated with X_t. Using a similar argument as in Proposition 4.7 with the sub-Riemannian metric g and the horizontal gradient $\nabla_{\mathcal{H}}$, we can get an estimate

$$|f\left(p, \xi\right) - f\left(\widetilde{p}, \xi\right)| \leqslant \|\nabla_{\mathcal{H}}f(\widetilde{p}, \xi)\| \, d_{CC}\left(p, \widetilde{p}\right) + \|\nabla_\xi f(\widetilde{p}, \xi)\| \, d_E\left(\xi, \widetilde{\xi}\right)$$

$$(4.26) \qquad\qquad + C_f\left(d_{CC}\left(p, \widetilde{p}\right) + d_E\left(\xi, \widetilde{\xi}\right)\right)^2,$$

for functions $f \in C_0^\infty\left(\mathbb{G} \times \mathbb{R}^3\right)$, where $C_f \geqslant 0$. 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 (4.26). Thus using (4.25) and (4.26) for the Heisenberg group we obtain the following result.

Theorem 4.13. *For all $q \geqslant 1$ and $f \in C_0^\infty\left(\mathbb{G} \times \mathbb{R}^3\right)$,*

$$(4.27) \qquad \|\nabla_{\mathcal{H}}P_t f\|^q \leqslant K^q P_t\left((\|\nabla_{\mathcal{H}}f\| + C_\sigma t \|\nabla_\xi f\|)^q\right).$$

The best constant K in (4.27) is not known. The best known estimate for K as of this writing is $K \geqslant \sqrt{2}$ (see [19, Proposition 2.7]).

Example 4.14. *Consider for $p = (x, y, z) \in \mathbb{G}$ the map $\sigma : \mathbb{G} \to \mathbb{R}^3$ defined by $\sigma(p) = (x, y, 0)$ and the diffusion $X_t = \left(B_t^p, \xi + \int_0^t \sigma(B_s^p) \, ds \right)$. A straightforward computation shows that*

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

so that by the left-invariance of d_{CC} we have that σ is 1-Lipschitz in the sense of (4.13). Thus Theorem 4.13 can be applied to X_t.

References

[1] Jürgen Angst, *Poisson boundary of a relativistic diffusion in curved space-times: an example*, ESAIM Probab. Stat. **19** (2015), 502–514. MR 3423304

[2] Jürgen Angst, Ismaël Bailleul, and Camille Tardif, *Kinetic Brownian motion on Riemannian manifolds*, Electron. J. Probab. **20** (2015), no. 110, 40. MR 3418542

[3] Ismael Bailleul, *Poisson boundary of a relativistic diffusion*, Probab. Theory Related Fields **141** (2008), no. 1-2, 283–329. MR 2372972 (2008m:60155)

[4] Dominique Bakry, Fabrice Baudoin, Michel Bonnefont, and Djalil Chafaï, *On gradient bounds for the heat kernel on the Heisenberg group*, J. Funct. Anal. **255** (2008), no. 8, 1905–1938. MR 2462581 (2010m:35534)

[5] Sayan Banerjee, Maria Gordina, and Phanuel Mariano, *Coupling in the Heisenberg group and its applications to gradient estimates*, to appear in the Annals of Probability, 2018.

[6] Sayan Banerjee and Wilfrid S. Kendall, *Coupling the Kolmogorov diffusion: maximality and efficiency considerations*, Adv. in Appl. Probab. **48** (2016), no. A, 15–35. MR 3539295

[7] Fabrice Baudoin, *Sub-Laplacians and hypoelliptic operators on totally geodesic Riemannian foliations*, Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. 1, EMS Ser. Lect. Math., Eur. Math. Soc., Zürich, 2016, pp. 259–321. MR 3587668

[8] ———, *Bakry-Émery meet Villani*, J. Funct. Anal. **273** (2017), no. 7, 2275–2291. MR 3677826

[9] 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

[10] Fabrice Baudoin, Michel Bonnefont, and Nicola Garofalo, *A sub-Riemannian curvature-dimension inequality, volume doubling property and the Poincaré inequality*, Math. Ann. **358** (2014), no. 3-4, 833–860. MR 3175142

[11] 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

[12] 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

[13] Fabrice Baudoin and Camille Tardif, *Hypocoercive estimates on foliations and velocity spherical Brownian motion*, Kinet. Relat. Models **11** (2018), no. 1, 1–23. MR 3708179

[14] Fabrice Baudoin and Jing Wang, *Curvature dimension inequalities and subelliptic heat kernel gradient bounds on contact manifolds*, Potential Anal. **40** (2014), no. 2, 163–193. MR 3152160

[15] Gérard Ben Arous, Michael Cranston, and Wilfrid S. Kendall, *Coupling constructions for hypoelliptic diffusions: two examples*, Stochastic analysis (Ithaca, NY, 1993), Proc. Sympos. Pure Math., vol. 57, Amer. Math. Soc., Providence, RI, 1995, pp. 193–212. MR 1335472

[16] Michel Bonnefont and Nicolas Juillet, *Couplings in L^p distance of two Brownian motions and their Lévy area*, Arxiv preprint, 2018.

[17] Ovidiu Calin, Der-Chen Chang, Kenro Furutani, and Chisato Iwasaki, *Heat kernels for elliptic and sub-elliptic operators*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2011, Methods and techniques. MR 2723056 (2011i:58037)

[18] M. Cranston, *Gradient estimates on manifolds using coupling*, J. Funct. Anal. **99** (1991), no. 1, 110–124. MR 1120916 (93a:58175)

[19] Bruce K. Driver and Tai Melcher, *Hypoelliptic heat kernel inequalities on the Heisenberg group*, J. Funct. Anal. **221** (2005), 340–365.

[20] R. M. Dudley, *Lorentz-invariant Markov processes in relativistic phase space*, Ark. Mat. **6** (1966), 241–268. MR 0198540 (33 #6695)

[21] _____, *A note on Lorentz-invariant Markov processes*, Ark. Mat. **6** (1967), 575–581 (1967). MR 0216567 (35 #7398)

[22] Jörn Dunkel, *Relativistic Brownian motion and diffusion processes*, Ph.D. thesis, Universität Augsburg, 2008.

[23] Jörn Dunkel and Peter Hänggi, *Relativistic Brownian motion*, Phys. Rep. **471** (2009), no. 1, 1–73. MR 2503143

[24] Nathaniel Eldredge, *Gradient estimates for the subelliptic heat kernel on H-type groups*, J. Funct. Anal. **258** (2010), no. 2, 504–533. MR 2557945 (2011d:35217)

[25] Jacques Franchi and Yves Le Jan, *Relativistic diffusions and Schwarzschild geometry*, Comm. Pure Appl. Math. **60** (2007), no. 2, 187–251. MR 2275328 (2007m:58044)

[26] _____, *Hyperbolic dynamics and Brownian motion*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2012, An introduction. MR 2977883

[27] Maria Gordina, *An application of a functional inequality to quasi-invariance in infinite dimensions*, pp. 251–266, Springer New York, New York, NY, 2017.

[28] Maria Gordina and Thomas Laetsch, *Sub-Laplacians on Sub-Riemannian Manifolds*, Potential Anal. **44** (2016), no. 4, 811–837. MR 3490551

[29] _____, *A convergence to Brownian motion on sub-Riemannian manifolds*, Trans. Amer. Math. Soc. **369** (2017), no. 9, 6263–6278, In print: September 2017. MR 3660220

[30] Richard S. Hamilton, *A matrix Harnack estimate for the heat equation*, Comm. Anal. Geom. **1** (1993), no. 1, 113–126. MR MR1230276 (94g:58215)

[31] _____, *Li-Yau estimates and their Harnack inequalities*, Geometry and analysis. No. 1, Adv. Lect. Math. (ALM), vol. 17, Int. Press, Somerville, MA, 2011, pp. 329–362. MR 2882429

[32] Lars Hörmander, *Hypoelliptic second order differential equations*, Acta Math. **119** (1967), 147–171. MR 0222474 (36 #5526)

[33] Elton P. Hsu, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics, vol. 38, American Mathematical Society, Providence, RI, 2002. MR 1882015 (2003c:58026)

[34] Nobuyuki Ikeda and Hiroyuki Matsumoto, *The Kolmogorov operator and classical mechanics*, In memoriam Marc Yor—Séminaire de Probabilités XLVII, Lecture Notes in Math., vol. 2137, Springer, Cham, 2015, pp. 497–504. MR 3444311

[35] Kalvis M. Jansons and Paul D. Metcalfe, *Optimally coupling the Kolmogorov diffusion, and related optimal control problems*, LMS J. Comput. Math. **10** (2007), 1–20. MR 2291679

[36] A. Kolmogoroff, *Zufällige Bewegungen (zur Theorie der Brownschen Bewegung)*, Ann. of Math. (2) **35** (1934), no. 1, 116–117. MR 1503147

[37] Kazumasa Kuwada, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), no. 11, 3758–3774. MR 2606871 (2011d:35109)

[38] Hong-Quan Li, *Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg*, J. Funct. Anal. **236** (2006), no. 2, 369–394. MR MR2240167 (2007d:58045)

[39] Peter Li and Shing-Tung Yau, *On the parabolic kernel of the Schrödinger operator*, Acta Math. **156** (1986), no. 3-4, 153–201. MR 834612 (87f:58156)

[40] H. P. McKean, Jr., *A winding problem for a resonator driven by a white noise*, J. Math. Kyoto Univ. **2** (1963), 227–235. MR 0156389

[41] Mihai N. Pascu and Ionel Popescu, *Shy and fixed-distance couplings of Brownian motions on manifolds*, Stochastic Process. Appl. **126** (2016), no. 2, 628–650.

[42] Max-K. von Renesse, *Intrinsic coupling on Riemannian manifolds and polyhedra*, Electron. J. Probab. **9** (2004), no. 14, 411–435 (electronic). MR 2080605

[43] 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 (2006j:53048)

[44] Feng-Yu Wang, *On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups*, Probab. Theory Related Fields **108** (1997), no. 1, 87–101. MR MR1452551 (98h:58184)

[45] _____, *Analysis for diffusion processes on Riemannian manifolds*, Advanced Series on Statistical Science & Applied Probability, vol. 18, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014. MR 3154951

[46] _____, *Derivative formulas and Poincaré inequality for Kohn-Laplacian type semigroups*, Sci. China Math. **59** (2016), no. 2, 261–280. MR 3454046

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CONNECTICUT, STORRS, CT 06269, U.S.A.
Email address: `fabrice.baudoin@uconn.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CONNECTICUT, STORRS, CT 06269, U.S.A.
Email address: `maria.gordina@uconn.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CONNECTICUT, STORRS, CT 06269, U.S.A.
Email address: `phanuel.mariano@uconn.edu`