8-6-2018

Functional Inequalities for Hypoelliptic Diffusions Using Probabilistic and Geometric Methods

Phanuel Mariano

University of Connecticut - Storrs, phanuel.mariano@uconn.edu

Functional Inequalities for Hypoelliptic Diffusions Using Probabilistic and Geometric Methods

Phanuel de Andrade Mariano, Ph.D.

University of Connecticut, 2018

ABSTRACT

We study gradient bounds and other functional inequalities related to hypoelliptic diffusions. One of the key techniques in our work is the use of coupling of diffusion processes to prove gradient bounds. We also use generalized Γ-calculus to prove various functional inequalities. In this dissertation we present two research directions; gradient bounds for harmonic functions on the Heisenberg group, and gradient bounds for the heat semigroup generated by Kolmogorov type diffusions.

For the first research direction, we construct a non-Markovian coupling for Brownian motions in the three-dimensional Heisenberg group. We then derive properties of this coupling such as estimates on the coupling rate, estimates for the CDF of the coupling time and upper and lower bounds on the total variation distance between the laws of the Brownian motions. Finally, we use these properties to prove gradient estimates for harmonic functions for the hypoelliptic sub-Laplacian which is the generator of Brownian motion in the Heisenberg group. In particular, we prove the well known Cheng-Yau inequality and a Caccioppoli-type inequality on the Heisenberg group.

Phanuel Mariano, University of Connecticut, 2018

For the second research direction, we study gradient bounds and other functional inequalities for the diffusion semigroup generated by Kolmogorov-type operators. Unlike the first research direction, the focus is on two different methods: coupling techniques and generalized Γ-calculus techniques. We discuss the advantages and drawbacks of each of these methods. For the coupling technique, we use a coupling by parallel transport (or synchronous coupling) to induce a coupling on the Kolmogorov type diffusions. In the Γ-calculus approach, we will prove a new generalized curvature dimension inequality to study various functional inequalities.

Functional Inequalities for Hypoelliptic Diffusions Using Probabilistic and Geometric Methods

Phanuel de Andrade Mariano

M.S. University of Connecticut, 2014

B.A. Western Connecticut State University, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2018

Doctor of Philosophy Dissertation

Functional Inequalities for Hypoelliptic Diffusions Using Probabilistic and Geometric Methods

Presented by

Phanuel de Andrade Mariano, B.A., M.S.

Major Advisor _____
Maria Gordina

Associate Advisor _____
Fabrice Baudoin

Associate Advisor _____
Iddo Ben-Ari

Associate Advisor _____
Alexander Teplyaev

University of Connecticut

2018

To Mamãe and Papai.

ACKNOWLEDGMENTS

The past 5 years have been exciting, challenging, difficult, rewarding, fun and filled with many emotions. I want to thank everyone who have impacted and guided me through this journey. There have been many people who have helped me along the way. The friendships and the companionship I've made at UConn will forever impact my life.

I want to start by thanking my academic advisor, Professor Maria (Masha) Gordina. Ever since my first meeting with Masha in regards to selecting an advisor, she made it very clear that my interests and overall well being at Uconn was very important to her. She took a genuine interest in guiding me through the exhausting life of being a graduate student. I'm grateful for the decision I made that day. Masha has been invaluable in guiding me through my research experience at UConn. I'm truly thankful for all that she has taught me, both in math and about life's challenges. I will always be indebted for her mathematical, emotional and financial support over the years.

I would also like to thank my dissertation committee, Prof. Fabrice Baudoin, Prof. Iddo Ben-Ari and Prof. Alexander (Sasha) Teplyaev. My mathematics will always be shaped by the courses I have taken with them, the research papers I have written with them and from all the insightful conversations we've had.

I would also like to thank Prof. Amit Savkar for all the opportunities I was afforded at UConn due to him. I want to thank Masha, Prof. Luke Rogers and Sasha

kind feedback. I would like to thank Professors Josephine Hamer and Barry Mittag from WestConn. Both had been crucial for my preparation to get into UConn. Thank you both for making me believe that I was good enough to get a PhD.

I want to thank my sister Samela Brown and my brother-in-law Chris Brown. I've always been so proud of my sister for her strong character and accomplishments. Thank you both for always being there for me and keeping me level headed as I completed my Ph.D. Landon and Levi, I can't wait to see you guys grow up.

Last but not least, I want to thank my parents. Wagner and Valeria Mariano, I wouldn't be here if it weren't for you. I want to thank you for all of the hard work you put in to help me get where I am today. Thank you for the part time jobs you took to help pay for college, your constant encouragement, your belief in me and your love. I want to thank you for moving our family across continents in order to provide my sister and I a better life. You both are the main reasons that made earning this Ph.D. possible. This dissertation is just as much yours as it is mine.

Contents

Chapter 1

Introduction

The general goal of this dissertation is to prove functional inequalities via probabilistic and geometric approaches. The main probabilistic tool we use is the concept of coupling. Recall that a coupling of two probability measures μ_1 and μ_2, defined on respective measure spaces $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$, is a measure μ on the product space $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ with marginals μ_1 and μ_2. In this dissertation, we will be interested in coupling of the laws of two Markov processes $(X_t : t \geqslant 0)$ and $(Y_t : t \geqslant 0)$. In the first part of the dissertation, we consider Markov processes that live in the geometric setting of a sub-Riemannian manifold such as the Heisenberg group $\mathbb{H}^3$. In the second part, we consider Kolmogorov type diffusions that live in $M \times \mathbb{R}^k$, where M is either a Riemannian or sub-Riemannian manifold. Couplings have been an extremely useful tool in probability theory and has resulted in establishing deep connections between probability, analysis and geometry.

We start by providing some background on couplings and gradient estimates in our setting. We give an introduction to sub-Riemannian geometry, in particular

the Heisenberg group. We also consider the Kolmogorov diffusion. We provide a introduction to the use of generalized curvature-dimension inequalities in proving functional inequalities.

This dissertation is based on results in [BGM16] and [BGM18].

1.1 Preliminaries

1.1.1 Coupling basics

Consider two probability spaces $(\Omega, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$. We have the following definition.

Definition 1.1.1. A *coupling* of μ_1 and μ_2 is a measure μ on $(\Omega_1 \times \Omega_2, \mathcal{F}_2 \times \mathcal{F}_2)$ with marginals μ_1 and μ_2, respectively. A *coupling* of two Markov processes X and Y is coupling of their laws. The coupling is said to be *successful* if the two processes couple within finite time almost surely, that is, the *coupling time* for X_t and Y_t defined as

$$\tau(X, Y) = \inf\{t \geqslant 0 : X_s = Y_s \text{ for all } s \geqslant t\}.$$

is almost surely finite. We assume that $X_t = Y_t$ for $t > \tau(X, Y)$.

A major application of couplings arises in estimating the *total variation distance* between the laws of two Markov processes at time t which in general is very hard to compute explicitly. Such an estimate can be obtained from the *Aldous' inequality*

$$||\mathcal{L}(X_t) - \mathcal{L}(Y_t)||_{TV} \leqslant \mu\{\tau(X, Y) > t\}, \tag{1.1.1}$$

where μ is the coupling of the Markov processes X and Y, $\mathcal{L}(X_t)$ and $\mathcal{L}(Y_t)$ denote the laws (distributions) of X_t and Y_t respectively, and

$$\|\nu\|_{TV} = \sup\{|\nu(A)| : A \text{ measurable}\}$$

denotes the total variation norm of the measure ν. The proof of Aldous' inequality is rather simple.

Proposition 1.1.2 (Aldous' inequality)**.** *Let τ be the coupling time for two Markov processes X and Y. For any $t > 0$,*

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\|_{TV} \leqslant \mathbb{P}\left(\tau > t\right),$$

where $\mathcal{L}(X_t)$ and $\mathcal{L}(Y_t)$ are the laws of X_t and Y_t, respectively.

Proof. For any Borel set A,

$$
\begin{aligned}
\mathcal{L}(X_t)(A) - \mathcal{L}(Y_t)(A) &= \mathbb{P}\left(X_t \in A\right) - \mathbb{P}\left(Y_t \in A\right) \\
&= \mathbb{P}\left(X_t \in A, X_t = Y_t\right) + \mathbb{P}\left(Y_t \in A, X_t \neq Y_t\right) \\
&\quad - \mathbb{P}\left(Y_t \in A, X_t = Y_t\right) - \mathbb{P}\left(Y_t \in A, X_t \neq Y_t\right) \\
&= \mathbb{P}\left(Y_t \in A, X_t \neq Y_t\right) - \mathbb{P}\left(Y_t \in A, X_t \neq Y_t\right) \\
&\leqslant \mathbb{P}\left(X_t \neq Y_t\right).
\end{aligned}
$$

Similarly one can prove

$$\mathcal{L}(Y_t)(A) - \mathcal{L}(X_t)(A) \leqslant \mathbb{P}\left(X_t \neq Y_t\right).$$

3

Putting these two together we have that

$$|\mathcal{L}(X_t)(A) - \mathcal{L}(Y_t)(A)| \leqslant \mathbb{P}\left(X_t \neq Y_t\right) = \mathbb{P}\left(\tau > t\right).$$

$\square$

This, in turn, can be used to provide sharp rates of convergence of Markov processes to their respective stationary distributions, when they exist (see [LPW09] for some such applications in studying mixing times of Markov chains).

This raises a natural question: how can we couple two Markov processes so that the probability of failing to couple by time t (coupling rate) is minimized (in an appropriate sense) for some, preferably all, t? Griffeath [Gri75] was the first to prove that *maximal couplings*, that is, the couplings for which the Aldous' inequality becomes an equality for each t in the time set of the Markov process, exist for discrete time Markov chains. This was later greatly simplified by Pitman [Pit76] and generalized to non-Markovian processes by Goldstein [Gol79] and continuous time càdlàg processes by Sverchkov and Smirnov [SS90].

These constructions, though extremely elegant, have a major drawback: they are typically very implicit. Thus, it is very hard, if not impossible, to perform detailed calculations and obtain precise estimates using these couplings. Part of the implicitness comes from the fact that these couplings are non-Markovian.

A *Markovian* coupling of two Markov processes X and Y is a coupling where, for any $t \geqslant 0$, the joint process $\{(X_s, Y_s) : s \geqslant t\}$ conditioned on the filtration $\sigma\{(X_s, Y_s) : s \leqslant t\}$ is again a coupling of the laws of X and Y, but now starting from (X_t, Y_t). These are the most widely used couplings in deriving estimates and performing detailed calculations as their constructions are typically explicit. However,

4

these couplings usually do not attain the optimal rates. In fact, it has been shown in [BK17] that the existence of a maximal coupling that is also Markovian imposes enormous constraints on the generator of the Markov process and its state space. Further, [BK16] describes an example using Kolmogorov diffusions defined as a two dimensional diffusion given by a standard Brownian motion along with its running time integral, where for *any Markovian coupling*, the probability of failing to couple by time t does not even attain the same order of decay (with t) as the total variation distance. More precisely, they showed that if the driving Brownian motions start from the same point, then the total variation distance between the corresponding Kolmogorov diffusions decays like $t^{-3/2}$ whereas for any Markovian coupling, the coupling rate is at best of order $t^{-1/2}$.

This brings us to the main subject of the first part of this dissertation: when can we produce non-Markovian couplings that are explicit enough to give us good bounds on the total variation distance between the laws of X_t and Y_t when Markovian couplings fail to do so? And what information can such couplings provide about the geometry of the state space of these Markov processes? In this dissertation, we look at the Heisenberg group which is the simplest example of a sub-Riemannian manifold and Brownian motion on it. The latter is the Markov process whose generator is the sub-Laplacian on the Heisenberg group as described in Section 1.1. We construct an explicit successful non-Markovian coupling of two copies of this process starting from different points in $\mathbb{H}^3$ and use it to derive sharp bounds on the total variation distance between their laws at time t. We also use this coupling to produce gradient estimates for harmonic functions on the Heisenberg group (more details below), thus providing a non-trivial link between probability and geometric analysis in the sub-Riemannian setting.

We note here that successful *Markovian couplings* of Brownian motions on the Heisenberg group have been constructed in [Ken07] and rates of these couplings have been studied in [Ken10]. However, the rates for the coupling we construct are much better. In fact, we show in Remark 2.1.2 that it is impossible to derive the rates we get from Markovian couplings. Moreover, the coupling we consider is efficient, that is, the coupling rate and the total variation distance decay like the same power of t as pointed out in Remark 2.2.3.

In 1986, Lindvall-Rogers [TL86] constructed successful couplings in Euclidean space using the idea of reflection coupling. Suppose the process (X_t) is given by a stochastic differential equation

$$dX_t = \sigma\left(X_t\right)dB_t + b\left(X_t\right)dt,$$

where $(B_t)_{t\geq 0}$ is a Brownian motion. We want to construct a new process (Y_t),

$$dY_t = \sigma\left(Y_t\right)dB_t' + b\left(Y_t\right)dt,$$

for some suitable B_t' on the same probability space, having the same distribution of X_t. Thus the only thing we have to do is choose a suitable Brownian motion B_t'. In order for this coupling to be successful, the suitable Brownian will be

$$dB_t' = H_t dB_t,$$

where H is chosen to be reflection in the plane orthogonal to $\sigma\left(y\right)^{-1}\left(x - y\right)$. As an example let us consider the simplest case when $\sigma \equiv I$ with $b = 0$. That is, when we only consider Brownian motion. Let P be the hyperplane perpendicular to the line

through x and y with $\frac{x+y}{2} \in P$. While simultaneously running $X_t = B_t$, Y_t will be the reflected Brownian motion in P.

1.1.2 Gradient estimates

Now we would like to describe gradient estimates in our geometric settings and how couplings have been used to prove them previously. Let us start with a classical gradient estimate for harmonic functions in $\mathbb{R}^d$. Suppose u is a real-valued function u on $\mathbb{R}^d$ which is harmonic in a ball $B_{2\delta}(x_0)$, then there exists a positive constant C_d (which depends only on the dimension d and not on u) such that

$$\sup_{x \in B_\delta(x_0)} |\nabla u(x)| \leqslant \frac{C_d}{\delta} \sup_{x \in B_{2\delta}(x_0)} |u(x)| \, .$$

In 1975, Cheng and Yau (see [CY75, Yau75, RS94]) generalized the classical gradient estimate to complete Riemannian manifolds M of dimension $d \geqslant 2$ with Ricci curvature bounded below by $-(d-1)K$ for some $K \geqslant 0$. They proved that any positive harmonic function on a Riemannian ball $B_\delta(x_0)$ satisfies

$$\sup_{x \in B_{\delta/2}(x_0)} \frac{|\nabla u(x)|}{u(x)} \leqslant C_d \left(\frac{1}{\delta} + \sqrt{K} \right) \, .$$

Moreover, in addition to such estimates, there is a vast literature on functional inequalities such as heat kernel gradient estimates, Poincaré inequalities, heat kernel estimates, elliptic and parabolic Harnack inequalities etc on Riemannian manifolds or more generally on measure metric spaces. Quite often these results require assumptions such as volume doubling and curvature bounds.

In 1991, M. Cranston in [Cra91] used the method of coupling two diffusion pro-

cesses to obtain a similar gradient estimate for solutions to the equation

$$\frac{1}{2}\Delta u + Zu = 0 \qquad (1.1.2)$$

on a Riemannian manifold (M, g) whose Ricci curvature is bounded below and Z is a bounded vector field. This coupling is known as the Kendall-Cranston coupling as it was based on the techniques in [Ken89]. In particular, M. Cranston proved the following gradient estimate.

Theorem 1.1.3 (Cranston). Suppose (M, g) is a complete d-dimensional Riemannian manifold with distance ρ_M and assume $\mathrm{Ric}_M \geqslant -Kg$. Let Z be a C^1 vector field on M such that $|Z(x)| \leqslant m$ for all $x \in M$. There is a constant $c = c(K, d, m)$ such that whenever $\delta > 0$ and (1.1.2) is satisfied in some Riemannian ball $B_{2\delta}(x_0)$, we have

$$|\nabla u(x)| \leqslant c\left(\frac{1}{\delta} + 1\right) \sup_{x \in B(x_0, 3\delta/2)} |u(x)|, \quad x \in B(x_0, \delta).$$

If (1.1.2) is satisfied on M and u is bounded and positive, then

$$|\nabla u(x)| \leqslant 2\left(\sqrt{K(d-1)} + m\right) \|u\|_\infty.$$

Cranston's approach generalized the coupling of Brownian motions on manifolds of Kendall [Ken86] to couple processes with the generator $L = \frac{1}{2}\Delta + Z$. The methods in that paper required tools from Riemannian geometry such as the Laplacian comparison theorem and the index theorem to obtain estimates on the processes $\rho_M(X_t, Y_t)$ and $\rho_M(X_t, X_0)$ where ρ_M is the Riemannian distance. M. Cranston also proved similar results on $\mathbb{R}^d$ in [Cra92].

8

1.1.3 Sub-Riemannian basics

A sub-Riemannian manifold M can be thought of as a Riemannian manifold where we have a constrained movement. Namely, such a manifold has the structure $(M, \mathcal{H}, \langle \cdot, \cdot \rangle)$, where allowed directions are only the ones in the horizontal distribution, which is a suitable subbundle $\mathcal{H}$ of the tangent bundle TM. For more detail on sub-Riemannian manifolds we refer to [Mon02].

Namely, for a smooth connected d-dimensional manifold M with the tangent bundle TM, let $\mathcal{H} \subset TM$ be an m-dimensional smooth sub-bundle such that the sections of $\mathcal{H}$ satisfy *Hörmander's condition* (the *bracket generating condition*) formulated in Assumption 1. We assume that on each fiber of $\mathcal{H}$ there is an inner product $\langle \cdot, \cdot \rangle$ which varies smoothly between fibers. In this case, the triple $(M, \mathcal{H}, \langle \cdot, \cdot \rangle)$ is called a *sub-Riemannian manifold* of rank m, $\mathcal{H}$ is called the *horizontal distribution*, and $\langle \cdot, \cdot \rangle$ is called the *sub-Riemannian metric*. The vectors (resp. vector fields) $X \in \mathcal{H}$ are called *horizontal vectors* (resp. horizontal vector fields), and curves γ in M whose tangent vectors are horizontal, are called *horizontal curves*.

Assumption 1. (Hörmander's condition) We will say that $\mathcal{H}$ satisfies *Hörmander's (bracket generating) condition* if horizontal vector fields with their Lie brackets span the tangent space $T_p M$ at every point $p \in M$.

Hörmander's condition guarantees analytic and topological properties such as hypoellipticity of the corresponding sub-Laplacian and topological properties of the sub-Riemannian manifold M. We explain briefly both aspects below. First we define

the *Carnot-Carathéodory metric* d_{CC} on M by

$$d_{CC}(x, y) = \tag{1.1.3}$$

$$\inf \left\{ \int_0^1 \|\gamma'(t)\|_{\mathcal{H}} \, dt \ \text{ where } \gamma(0) = x, \gamma(1) = y, \gamma \text{ is a horizontal curve} \right\},$$

where as usual $\inf(\emptyset) := \infty$. Here the norm is induced by the inner product on $\mathcal{H}$, namely, $\|v\|_{\mathcal{H}} := (\langle v, v \rangle_p)^{\frac{1}{2}}$ for $v \in \mathcal{H}_p$, $p \in M$. The Chow-Rashevski theorem says that Hörmander's condition is sufficient to ensure that any two points in M can be connected by a finite length horizontal curve. Moreover, the topology generated by the the Carnot-Carathéodory metric coincides with the original topology of the manifold M.

As we are interested in a Brownian motion on a sub-Riemannian manifold $(M, \mathcal{H}, \langle \cdot, \cdot \rangle)$, a natural question is what its generator is. While there is no canonical operator such as the Laplace-Beltrami operator on a Riemannian manifold, there is a notion of a *sub-Laplacian* on sub-Riemannian manifolds. A second order differential operator defined on $C^\infty(M)$ is called a *sub-Laplacian* $\Delta_{\mathcal{H}}$ if for every $p \in M$ there is a neighborhood U of p and a collection of smooth vector fields $\{X_0, X_1, ..., X_m\}$ defined on U such that $\{X_1, ..., X_m\}$ are orthonormal with respect to the sub-Riemannian metric and

$$\Delta_{\mathcal{H}} = \sum_{k=1}^m X_k^2 + X_0.$$

By the classical theorem of L. Hörmander in [Hör67, Theorem 1.1] Hörmander's condition (Assumption 1) guarantees that any sub-Laplacian is hypoelliptic. For more properties of sub-Laplacians which are generators of a Brownian motion on a sub-Riemannian manifold we refer to [GL16].

Finally, the *horizontal gradient* $\nabla_{\mathcal{H}}$ is a horizontal vector field such that for any smooth $f : M \to \mathbb{R}$ we have that for all $X \in \mathcal{H}$,

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

We define the length of the gradient as in [Kuw10]. For a function f on M, let

$$|\nabla_{\mathcal{H}} f|\left(x \right) := \lim_{r \downarrow 0} \sup_{0 < d_{CC}(x, \tilde{x}) \leqslant r} \left| \frac{f\left(x \right) - f\left(\tilde{x} \right)}{d_{CC}\left(x, \tilde{x} \right)} \right|, \tag{1.1.4}$$

and set $\|\nabla_{\mathcal{H}} f\|_{\infty} := \sup_{x \in \mathbb{H}^3} |\nabla_{\mathcal{H}} f|\left(x \right)$.

1.1.4 The Heisenberg group

The Heisenberg group $\mathbb{H}^3$ is the simplest non-trivial example of a sub-Riemannian manifold. Namely, let $\mathbb{H}^3 \cong \mathbb{R}^3$ with the multiplication defined by

$$\left(x_1, y_1, z_1 \right) \star \left(x_2, y_2, z_2 \right) := \left(x_1 + x_2, y_1 + y_2, z_1 + z_2 + \left(x_1 y_2 - x_2 y_1 \right) \right),$$

with the group identity $e = (0, 0, 0)$ and the inverse given by $(x, y, z)^{-1} = (-x, -y, -z)$.

We define $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ as the unique left-invariant vector fields with $\mathcal{X}_e = \partial_x$, $\mathcal{Y}_e = \partial_y$, and $\mathcal{Z}_e = \partial_z$, so that

$$\mathcal{X} = \partial_x - y\partial_z,$$
$$\mathcal{Y} = \partial_y + x\partial_z,$$
$$\mathcal{Z} = \partial_z.$$

The horizontal distribution is defined by $\mathcal{H} = \text{span}\{\mathcal{X}, \mathcal{Y}\}$ fiberwise. Observe that $[\mathcal{X}, \mathcal{Y}] = 2\mathcal{Z}$, so Hörmander's condition is easily satisfied. Moreover, as any iterated Lie bracket of length greater than two vanishes, $\mathbb{H}^3$ is a nilpotent group of step 2. The Lebesgue measure on $\mathbb{R}^3$ is a Haar measure on $\mathbb{H}^3$. We endow $\mathbb{H}^3$ with the sub-Riemannian metric $\langle \cdot, \cdot \rangle$ so that $\{\mathcal{X}, \mathcal{Y}\}$ is an orthonormal frame for the horizontal distribution. As pointed out in [GL16, Example 6.1], the (sum of squares) operator

$$\Delta_{\mathcal{H}} = \frac{1}{2}\left(\mathcal{X}^2 + \mathcal{Y}^2\right) \tag{1.1.5}$$

is a natural sub-Laplacian for the Heisenberg group with this sub-Riemannian structure.

In general it is very cumbersome to compute the Carnot-Carathéodory distance d_{CC} explicitly. In the case of the Heisenberg group an explicit formula for the distance is known. Let $r(\mathbf{x}) = d_{CC}(\mathbf{x}, e)$ be the distance between $\mathbf{x} = (x, y, z) \in \mathbb{H}^3$ and the identity $e = (0, 0, 0)$. In [CTW10] the distance is given by the formula

$$r(\mathbf{x})^2 = \nu(\theta_c)\left(x^2 + y^2 + |z|\right),$$

where θ_c is the unique solution of $\mu(\theta)(x^2 + y^2) = |z|$ in the interval $[0, \pi)$ and $\mu(z) = \frac{z}{\sin^2 z} - \cot z$ and where

$$\nu(z) = \frac{z^2}{\sin^2 z}\frac{1}{1 + \mu(z)} = \frac{z^2}{z + \sin^2 z - \sin z \cos z}, \quad \nu(0) = 2.$$

Since the distance is left-invariant, we have

$$d_{CC}(\mathbf{x}, \tilde{\mathbf{x}}) = d_{CC}\left(\tilde{\mathbf{x}}^{-1} \star \mathbf{x}, e\right)$$

which gives us an explicit expression for d_{CC} on the Heisenberg group. Although ν is not continuous it was shown in [CCG07] that d_{CC} is continuous.

We will not use this explicit expression for d_{CC}. Instead, since $\nu \geqslant 0$ and bounded below and above by positive constants in the interval $[0, \pi)$, it is clear that the Carnot-Carathéodory distance is equivalent to the pseudo-metric

$$\rho\left(\mathbf{x}, \mathbf{y}\right) = \left((x - \tilde{x})^2 + (y - \tilde{y})^2 + |z - \tilde{z} + x\tilde{y} - y\tilde{x}|\right)^{\frac{1}{2}}. \tag{1.1.6}$$

Finally, we can describe Brownian motion whose generator is $\Delta_{\mathcal{H}}/2$ explicitly as follows. Let B_1, B_2 be real-valued independent Brownian motions starting from 0. Define Brownian motion on the Heisenberg group $\mathbf{X}_t : [0, \infty) \times \Omega \to \mathbb{H}$ to be the solution of the following Stratonovich stochastic differential equation (SDE)

$$
\begin{aligned}
d\mathbf{X}_t &= \mathcal{X}\left(\mathbf{X}_t\right) \circ dB_1(t) + \mathcal{Y}\left(\mathbf{X}_t\right) \circ dB_2(t), \\
\mathbf{X}_0 &= (b_1, b_2, a).
\end{aligned}
$$

Letting $\mathbf{X}_t = (X_1(t), X_2(t), X_3(t))$ we see that the SDE reduces to

$$
d\mathbf{X}_t = \begin{pmatrix} 1 \\ 0 \\ -X_2(t) \end{pmatrix} \circ dB_1(t) + \begin{pmatrix} 0 \\ 1 \\ X_1(t) \end{pmatrix} \circ dB_2(t),
$$

so that one needs to solve the following system of equations

$$
\begin{aligned}
dX_1(t) &= dB_1(t) \\
dX_2(t) &= dB_2(t), \\
dX_3(t) &= -X_2(t) \circ dB_1(t) + X_1(t) \circ dB_2(t).
\end{aligned}
$$

Since the covariation of two independent Brownian motions is zero we get that

$$
\begin{aligned}
X_1(t) &= b_1 + B_1(t), \\
X_2(t) &= b_2 + B_2(t), \\
X_3(t) &= a + \int_0^t (B_1(s) + b_1)dB_2(s) - \int_0^t (B_2(s) + b_2)dB_1(s). \qquad (1.1.7)
\end{aligned}
$$

1.1.5   Curvature-dimension inequalities and Γ-calculus

In this section we review the geometric methods that goes back to Bakry-Émery in [BE85] to prove functional inequalities (see [BL06,Bak06]). Consider an $n-$dimensional Riemannian manifold $\mathbb{M}$ with Laplacian Δ. Bakry and Émery developed the functional calculus, now known by many as $\Gamma-$calculus, based on the differential forms

$$
\Gamma(f, f) := \frac{1}{2}\left(\Delta\left(fg\right) - f\Delta g - g\Delta f\right),
$$

and

$$
\Gamma_2\left(f, g\right) := \frac{1}{2}\left(\Delta\Gamma(f, g) - \Gamma\left(f, \Delta g\right) - \Gamma\left(g, \Delta f\right)\right),
$$

for functions $f \in C^\infty (\mathbb{M})$. Note that

$$\Gamma(f) := \Gamma(f, f) = \|\nabla f\|^2 ,$$

where ∇ is the Riemannian gradient and $\|\cdot\|$ is the norm associated to the underlying Riemannian metric. One can also compute

$$\Gamma_2(f) := \Gamma_2 (f, f) = \left\|\nabla^2 f\right\|_2^2 + 2\mathrm{Ric} (\nabla f, \nabla f) ,$$

where $\nabla^2 f$ is the Riemannian Hessian. The computation of $\Gamma_2(f)$ is due to the well known Bochner's formula in terms of Γ

$$\Delta\Gamma (f) = 2 \left\|\nabla^2 f\right\|_2^2 + 2\Gamma (f, \Delta f) + 2\mathrm{Ric} (\nabla f, \nabla f) .$$

We say Δ satisfies the *curvature-dimension inequality* $CD (\rho, n)$ if

$$\Gamma_2 (f) \geq \frac{1}{2} (\Delta f)^2 + \rho\Gamma (f) ,$$

for all $f \in C^\infty (\mathbb{M})$. By a result of Bakry in [Bak94] it was proven that $CD (\rho, n)$ is equivalent to

$$\mathrm{Ric} (\nabla f, \nabla f) \geq \rho \|\nabla f\|^2 .$$

This connection allowed Bakry, Ledoux and others to use this analytic approach to reprove results in differential geometry relating to heat kernels and heat semigroups. In fact it turns out that the following are all equivalent (see [BL06, Bak06, Bak94, BE85, vRS05]):

1. $\mathrm{Ricci}\,(\nabla f, \nabla f) \geq \rho \,|\nabla f|^2$ for all $f \in C_0^\infty\,(\mathbb{M})$,

2. $|\nabla P_t f| \leqslant e^{-\rho t} P_t\,(|\nabla f|)$ for all $f \in C_0^\infty\,(\mathbb{M})$ and $t > 0$,

3. $|\nabla P_t f|^p \leqslant e^{-p\rho t} P_t\,(|\nabla f|^p)$ for all $f \in C_0^\infty\,(\mathbb{M})$ and $t > 0$ and $p \geq 1$.

4. $\|\nabla P_t f\|_\infty \leqslant e^{-\rho t}\,\|\nabla f\|_\infty$ for all $f \in C_0^\infty\,(\mathbb{M})$ and $t > 0$.

5. There exists a function $\rho(t) > 0$ such that $\rho(0) = 1, \rho'(0)$ exists, and

   $$|\nabla P_t f|^2 \leqslant \rho(t) P_t\,\left(|\nabla f|^2\right) \text{ for all } f \in C_0^\infty\,(\mathbb{M}) \text{ and } t > 0.$$

6. $\Gamma_2(f, f) \geq \rho \Gamma(f, f)$ for all $f \in C_0^\infty\,(\mathbb{M})$ and $t > 0$.

7. There exists a coupling $\left(B_t, \tilde{B}_t\right)$ of Brownian motions on $\mathbb{M}$ started at $(x, \tilde{x})$ such that for all $t \geq 0$, $d_{\mathbb{M}}\left(B_t, \tilde{B}_t\right) \leqslant e^{-\rho t/2} d_{\mathbb{M}}(x, \tilde{x})$.

8. For every function $f \in C_0^\infty\,(\mathbb{M})$, and every $t \geq 0$,

   $$P_t\,\left(f^2\right) - (P_t f)^2 \leqslant \frac{1 - e^{-2\rho t}}{\rho} P_t\,(\Gamma(f)).$$

9. For every function $f \in C_0^\infty\,(\mathbb{M})$, and every $t \geq 0$,

   $$\Gamma\,(P_t f) \leqslant \frac{\rho}{e^{2\rho t} - 1}\,\left(P_t\,\left(f^2\right) - (P_t f)^2\right).$$

One can also use Γ-calculus to prove various other inequalities such as logarithm Sovolev inequalities, Sobolev inequalities, isoperimetric inequalities and Harnack inequalities to name a few.

This $\Gamma-$calculus approach has also allowed for careful analysis of various elliptic operators when Δ is replaced with a general Markov diffusion operator L. We provide

two examples of well known one dimensional diffusion processes that satisfy $CD\,(\rho, n)$ for some ρ and n.

Example 1.1.1. Consider the *Ornstein–Uhlenbeck process* with the generator

$$L = \frac{d^2}{dx^2} - \rho x \frac{d}{dx},$$

on $\mathbb{R}$ where $\rho > 0$. We show L satisfies $CD\,(\rho, \infty)$. This shows that the dimension of the process does not have to match the spatial dimension of the process and in fact can be infinite. First, a simple computation shows that

$$\Gamma(f) = (f')^2.$$

We also have that

$$
\begin{aligned}
L\,(f')^2 &= \frac{d^2}{dx^2}\,(f')^2 - \rho x \frac{d}{dx}\,(f')^2 \\
&= 2\frac{d}{dx}\,(f'f'') - 2\rho x f' f'' \\
&= 2\,(f'')^2 + 2f'' f''' - 2\rho x f' f'',
\end{aligned}
$$

and

$$2f'\,(Lf)' = 2f'\,(f''' - 2f' - \rho x f'')\,.$$

This shows that

$$\Gamma_2(f) = (f'')^2 + \rho\,(f')^2.$$

17

If L satisfies $CD\left(\rho, n\right)$ then $\Gamma_2(f, f) \geq \frac{1}{n}\left(Lf\right)^2 + \rho\Gamma\left(f\right)$, which would mean

$$\Gamma_2(f) = \left(f''\right)^2 + \rho\left(f'\right)^2 \geq \frac{1}{n}\left(f'' - \rho x f'\right)^2 + \rho\left(f'\right)^2 = \frac{1}{n}\left(Lf\right)^2 + \rho\Gamma\left(f\right),$$

which only holds if $n = \infty$.

Example 1.1.2. Consider the $d-dimensional$ $Bessel$ $process$ with generator

$$L = \frac{d^2}{dx^2} + \frac{d-1}{x}\frac{d}{dx},$$

on $(0, \infty)$. Similar computations as the example above shows that

$$\Gamma(f) = \left(f'\right)^2 \text{ and } \Gamma_2(f) = \left(f''\right)^2 + \frac{d-1}{x^2}\left(f'\right)^2.$$

Also note

$$
\begin{aligned}
\left(Lf\right)^2 &= \left(f'' + \frac{d-1}{x}f'\right)^2 \\
&= \left(f''\right)^2 + 2\frac{d-1}{x}f'f'' + \frac{(d-1)^2}{x^2}\left(f'\right)^2.
\end{aligned}
\tag{1.1.8}
$$

Now if L satisfies $CD\left(\rho, n\right)$ then $\Gamma_2(f) \geq \frac{1}{n}\left(Lf\right)^2 + \rho\Gamma\left(f\right)$. By Cauchy-Schwarz we have that

$$\frac{2}{n}\frac{d-1}{x}f'f'' \leqslant \frac{d-1}{n}\left(\frac{1}{x^2}\left(f'\right)^2 + \left(f''\right)^2\right).
\tag{1.1.9}$$

Combining (1.1.8) and (1.1.9) we have that

$$\frac{1}{n}\left(Lf\right)^2 \leqslant \frac{d}{n}\left(f''\right)^2 + \frac{1}{n}\frac{d\left(d-1\right)}{x^2}\left(f'\right)^2,$$

18

thus we can see that the optimal values for n and ρ are $n = d$ and $\rho = 0$. Thus L satisfies $CD(0, d)$. Note that the spatial dimension is 1 yet the dimension in the curvature dimension inequality is $d \geq 1$.

We refer the reader to [BGL14] for a careful treatment of the $\Gamma-$calculus of general Markov diffusion operators.

Unfortunately sub-Riemannian manifolds do not satisfiy $CD(\rho, n)$ for any ρ or n. More generally, hypoelliptic diffusions do not always satisfy $CD(\rho, n)$. In the work of F. Baudoin and N. Garofalo (see [BG17]) the authors introduced the notion of a *generalized curvature dimension-inequality*. These new techniques imply Li-Yau type inequalities, Harnack inequalities, off-diagonal Gaussian upper bounds, Liouville type theorems and Bonnet-Myers type theorem. The authors also show that the generalized curvature dimension inequality is satisfied by a large class of sub-Riemannian manifolds. This class includes Carnot groups of step two, Sasakian manifolds whose horizontal Webster-Tanaka-Ricci curvature is bounded from below. Since the original work of [BG17] there have been several publications or articles in proving generalized curvature-dimension inequalities in other settings (see [BW14, BT18, Bau17a, Bau17b, BB16, Bau16, BBG14, Wan14, BB12]). In the second part of this dissertation, we will use a generalized Γ-calculus to prove functional inequalities on Kolmogorov type diffusions.

1.1.6 The Kolmogorov diffusion

The Kolmogorov diffusion is the Markov process

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

19

where B_t^x is Brownian motion started at x. Its generator is given by

$$L = \frac{1}{2}\frac{\partial^2}{\partial x^2} + x\frac{\partial}{\partial y}.$$

See Proposition B.0.1 for more details. Note that the integrated Brownian motion process $\int_0^t B_s^x ds$ by itself is not a Markov process.

It was first introduced by Kolmogorov in his 1934 Ann. Math. paper [Kol34], where he provided an explicit expression for the transition density:

$$p_t(x, y; u, v) = \frac{\sqrt{3}}{\pi t^2}\exp\left(-\frac{6(v-y)^2}{t^3} + \frac{6(v-y)(u+x)}{t^2}\right.$$
$$\left. -\frac{2(u^2 + ux + x^2)}{t}\right).$$

L. Hörmander used Kolmogorov's operator as the simplest nontrivial example of a hypoelliptic second order differential operator that is not elliptic. Note that the semigroup generated by L is Gaussian from the corresponding explicit heat kernel. However, despite an explicit Gaussian heat kernel, it is somehow challenging to derive relevant functional inequalities for this semigroup.

This operator satisfies the weak *Hörmander's condition* condition since the vector fields $\left\{\frac{\partial}{\partial x}, x\frac{\partial}{\partial y}\right\}$ and its lie brackets span $\mathbb{R}^2$. Thus L is a hypoelliptic operator. Its corresponding *carré du champ* operator is $\Gamma(f, f) = \left(\frac{\partial f}{\partial x}\right)^2$. This operator is a sort of generalized square of the norm of the gradient. We often write the corresponding diffusion process stated at (x, y) as is

$$X_t = \left(x + B_t, y + tx + \int_0^t B_s ds\right),$$

where B_t is a standard Brownian motion.

In Chapter 4 and 5 we use the coupling technique and Γ-calculus to prove gradient estimates on the heat semigroup. There has been interest in extending gradient estimates of the form

$$\sqrt{\Gamma\left(P_t f, P_t f\right)} \leqslant C_p(t) \left(P_t \left(\Gamma\left(f, f\right)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}} \tag{1.1.10}$$

for $p \in [1, \infty)$ to the hypoelliptic case. In fact, Bruce Driver and Tai Melcher in [DM05] proved (1.1.10) on the Heisenberg group for $1 < p < \infty$. They showed that that the best constant $C(t)$ is independent of t. It was then in [Li06] that H.Q. Li extended (1.1.10) to the case $p = 1$. Later, N. Eldredge in [Eld10] proved (1.1.10) for all $p \geq 1$ for H-type groups. Simpler proofs of the L^p gradient inequality on the Heisenberg group were later shown in [BBBC08]. The authors highlighted that the Kolmogorov operator is a degenerate type Hörmander operator. They remarked that unlike in the Heisenberg group, the Poincaré and reverse Poincaré inequalities are not equivalent (See Remark 4.2 of [BBBC08])

In fact (1.1.10) does not hold for the Kolmogorov operator. A counter–example for this can be seen by taking the same function as in [BBBC08]. Let $f(x, y) = y$. We see that $P_t f = y + tx$, $\Gamma\left(P_t f, P_t f\right) = t^2$ and $\Gamma\left(f, f\right) = 0$. But if (1.1.10) were true then we would have $t \leqslant 0$, which is a contradiction. Even though this estimate is not true, we can still prove a sharp Driver-Melcher type inequality. This will be done in later chapters.

1.2   Outline of the dissertation

The outline of the dissertation is as follows.

The first part of the dissertation consists of Chapters 2 and 3. The first part will be concerned with using coupling to prove gradient estimates on the Heisenberg group. In Chapter 2, we describe an efficient non-Markovian coupling and give estimates for the tail of the coupling time. We use this coupling to give sharp estimates for the total variation distance. In Chapter 3, we use the results from Chapter 2 to prove gradient estimates for harmonic functions on the Heisenberg group. In particular, we prove the well known Cheng-Yau inequality and a Caccioppoli type inequality on the Heisenberg group.

The second part of the dissertation consists of Chapters 4,5,6 and 7. In Chapter 4 we introduce the motivating coupling technique that will be used in later chapters to prove gradient estimates. In Chapter 5 we prove various sharp functional inequalities for the Kolmogorov diffusion on $\mathbb{R}^d \times \mathbb{R}^d$. In Chapter 6 we prove functional inequalities for the relativistic diffusion. In Chapter 7 we prove gradient bounds for general Kolmogorov type diffusions.

Chapter 2

Successful non-Markovian coupling in the 3-dimensional Heisenberg group

Let B_1, B_2 be independent real-valued Brownian motions, starting from b_1 and b_2 respectively. We call the process

$$\mathbf{X}_t = \left(B_1(t), B_2(t), a + \int_0^t B_1(s)dB_2(s) - \int_0^t B_2(s)dB_1(s) \right) \qquad (2.0.1)$$

Brownian motion on the Heisenberg group, with driving Brownian motion $\mathbf{B} = (B_1, B_2)$, starting from (b_1, b_2, a). Let $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ be coupled copies of this process starting from (b_1, b_2, a) and $\left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a} \right)$ respectively. Denote the coupling time

$$\tau = \inf \left\{ t \geqslant 0 : \mathbf{X}_s = \widetilde{\mathbf{X}}_s \text{ for all } s \geqslant t \right\}.$$

We will construct a non-Markovian coupling $\left(\mathbf{X}, \widetilde{\mathbf{X}} \right)$ of two Brownian motions on

the Heisenberg group. This, via the Aldous' inequality, will yield an upper bound on the total variation distance between the laws of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$. Before we state and prove the main theorem, we describe the tools required in its proof.

For $T > 0$, let $\left(B^{\mathrm{br}}, \widetilde{B}^{\mathrm{br}}\right)$ be a coupling of standard Brownian bridges defined on the interval $[0, T]$. If $G^{(T)}$ is a Gaussian variable with mean zero and variance T independent of $\left(B^{\mathrm{br}}, \widetilde{B}^{\mathrm{br}}\right)$, a standard covariance computation shows that the assignment

$$
\begin{aligned}
B(t) &= B^{\mathrm{br}}(t) + \frac{t}{T} G^{(T)} \\
\widetilde{B}(t) &= \widetilde{B}^{\mathrm{br}}(t) + \frac{t}{T} G^{(T)}
\end{aligned}
\tag{2.0.2}
$$

gives a non-Markovian coupling of two standard Brownian motions on $[0, T]$ satisfying $B(T) = \widetilde{B}(T)$. This coupling is similar in spirit to the one developed in [BK16]. The usefulness of this coupling strategy arises when we want to couple two copies of the process $((B(t), F([B]_t)) : t \geqslant 0)$, where B is a Brownian motion, $[B]_t$ denotes the whole Brownian path up until time t (thought of as an element of $C[0, t]$), and F is a (possibly random) functional on $C[0, t]$. We first reflection couple the Brownian motions until they meet. Then, by dividing the future time into intervals $[T_n, T_{n+1}]$ (usually of growing length) and constructing a suitable non-Markovian coupling of the Brownian bridges on each such interval, we can obtain a coupling of the Brownian paths by the above recipe in such a way that the corresponding path functionals agree at one of the deterministic times T_n. As by construction, the coupled Brownian motions agree at the times T_n, we achieve a successful coupling of the joint process (B, F). Further, the rate of coupling attained by this non-Markovian strategy is usually significantly better than Markovian strategies, and is often near optimal (see

24

[BK16]).

We will be interested in the particular choice of the random functional, namely,

$$F\left([w]_t\right) = \int_0^t w(s)dB_1(s),$$

where B_1 is a standard Brownian motion and $w \in C[0,t]$. Our coupling strategy for the Brownian bridges on $[0,T]$ will be based on the Karhunen-Loéve expansion which goes back to [Kar47, Loe48] and for examples of such expansions see [Wan08, p.21]. For the Brownian bridge we have

$$B^{\mathrm{br}}(t) = \sqrt{T}\sum_{k=1}^{\infty} Z_k \frac{\sqrt{2}\sin\left(\frac{k\pi t}{T}\right)}{k\pi} = \sqrt{T}\sum_{k=1}^{\infty} Z_k g_{T,k}\left(t\right) \qquad (2.0.3)$$

for $t \in [0,T]$, where Z_k are i.i.d. standard Gaussian random variables. Thus, in order to couple two Brownian bridges on $[0,T]$, we will couple the random variables $\{Z_k\}_{k\geqslant 1}$.

2.1 Preliminary results

We now state and prove the following lemmas.

Lemma 2.1.1 ([BGM16]). *There exists a non-Markovian coupling of the diffusions*

$$\left\{ \left(B_1(t), B_2(t), a + \int_0^t B_2(s)dB_1(s) \right) : t \geqslant 0 \right\},$$

$$\left\{ \left(\widetilde{B}_1(t), \widetilde{B}_2(t), \widetilde{a} + \int_0^t \widetilde{B}_2(s)d\widetilde{B}_1(s) \right) : t \geqslant 0 \right\},$$

$$B_1(0) = \widetilde{B}_1(0) = b_1, B_2(0) = \widetilde{B}_2(0) = b_2, \text{ and } a > \widetilde{a},$$

for which the coupling time τ satisfies

$$\mathbb{P}(\tau > t) \leqslant C \frac{(a - \widetilde{a})}{t}$$

for some constant $C > 0$ that does not depend on the starting points and $t \geqslant (a - \widetilde{a})$.

Proof. We will write $I(t) = a + \int_0^t B_2(s)dB_1(s)$ and $\widetilde{I}(t) = \widetilde{a} + \int_0^t \widetilde{B}_2(s)d\widetilde{B}_1(s)$. From Brownian scaling, it is clear that for any $r \in \mathbb{R}$, the following distributional equality holds

$$\left(\frac{B_1(t)}{r}, \frac{B_2(t)}{r}, \frac{a + \int_0^t B_2(s)dB_1(s)}{r^2} \right) \tag{2.1.1}$$

$$\stackrel{d}{=} \left(B_1'(t/r^2), B_2'(t/r^2), \frac{a}{r^2} + \int_0^{t/r^2} B_2'(s)dB_1'(s) \right),$$

where B_1', B_2' are independent Brownian motions with $B_1'(0) = b_1/r$, $B_2'(0) = b_2/r$. Thus we can assume $a - \widetilde{a} = 1$. For the general case, we can obtain the corresponding coupling by applying the same coupling strategy to the scaled process using (2.1.1) with $r = \sqrt{a - \widetilde{a}}$.

Let us divide the non-negative real line into intervals $[2^n - 1, 2^{n+1} - 1], n \geqslant 0$. We will synchronously couple B_1 and $\widetilde{B}_1$ at all times. Thus, we sample the same Brownian

path for B_1 and $\widetilde{B}_1$. *Conditional on this Brownian path* $\{B_1(t) : t \geqslant 0\}$ we describe the coupling strategy for B_2 and $\widetilde{B}_2$ inductively on successive intervals. Suppose we have constructed the coupling on $[0, 2^n - 1]$ in such a way that the coupled Brownian motions B_2 and $\widetilde{B}_2$ satisfy $B_2(2^n - 1) = \widetilde{B}_2(2^n - 1) = b_2$ and $I(2^n - 1) > \widetilde{I}(2^n - 1)$. Conditional on $\left\{ \left(B_2(t), \widetilde{B}_2(t) \right) : t \leqslant 2^n - 1 \right\}$ and the whole Brownian path B_1, we will construct the coupling of $B_2(t) - b_2$ and $\widetilde{B}_2(t) - b_2$ for $t \in [2^n - 1, 2^{n+1} - 1]$. To this end, we will couple two Brownian bridges B^{br} and $\widetilde{B}^{\mathrm{br}}$ on $[2^n - 1, 2^{n+1} - 1]$, then sample an independent Gaussian random variable $G^{(2^n)}$ with mean zero, variance 2^n and finally use the recipe (4.3.2) to get the coupling of B_2 and $\widetilde{B}_2$ on $[2^n - 1, 2^{n+1} - 1]$.

Let $\left(Z_1^{(n)}, Z_2^{(n)}, \ldots \right)$ and $\left(\widetilde{Z}_1^{(n)}, \widetilde{Z}_2^{(n)}, \ldots \right)$ denote the Gaussian coefficients in the Karhunen-Loève expansion (4.3.3) corresponding to B^{br} and $\widetilde{B}^{\mathrm{br}}$ respectively. Sample i.i.d Gaussians Z_k and set $Z_k^{(n)} = \widetilde{Z}_k^{(n)} = Z_k$ for $k \geqslant 2$. Now we construct the coupling of $Z_1^{(n)}$ and $\widetilde{Z}_1^{(n)}$. Let $W^{(n)}$ be a standard Brownian motion starting from zero, independent of $\left\{ \left(B_2(t), \widetilde{B}_2(t) \right) : t \leqslant 2^n - 1 \right\}$, $\{Z_k\}_{k \geqslant 2}$ and B_1. In what follows we will repeatedly use the following random functional

$$\lambda_n(t) = \frac{2}{\pi} \int_{2^n-1}^t \sqrt{2} \sin \left(\frac{\pi(s - 2^n + 1)}{2^n} \right) dB_1(s), \, 2^n - 1 \leqslant t \leqslant 2^{n+1} - 1. \quad (2.1.2)$$

Define the random time $\sigma^{(n)}$ by

$$\sigma^{(n)} = \begin{cases} \inf \left\{ t \geqslant 0 : W^{(n)}(t) = -\frac{\left(I(2^n-1) - \widetilde{I}(2^n-1) \right)}{\lambda_n(2^{n+1}-1)} \right\}, & \text{if } \lambda_n(2^{n+1} - 1) \neq 0, \\ \infty, & \text{otherwise.} \end{cases}$$

As $\lambda_n \left(2^{n+1} - 1\right)$ is a Gaussian random variable with mean zero and variance

$$\frac{4}{\pi^2} \int_{2^n - 1}^{2^{n+1} - 1} 2 \sin^2 \left(\frac{\pi \left(s - 2^n + 1\right)}{2^n}\right) ds = \frac{2^{n+2}}{\pi^2},$$

the time $\sigma^{(n)}$ is finite for almost every realization of the Brownian path B_1. Now, define $\widetilde{W}^{(n)}$ as follows

$$\widetilde{W}^{(n)}(t) = \begin{cases} -W^{(n)}(t) & \text{if } t \leqslant \sigma^{(n)} \\ W^{(n)}(t) - 2W^{(n)}\left(\sigma^{(n)}\right) & \text{if } t > \sigma^{(n)}. \end{cases}$$

Conditional on $\left\{\left(B_2(t), \widetilde{B}_2(t)\right) : t \leqslant 2^n - 1\right\}$, $\{Z_k\}_{k \geqslant 2}$ and B_1, $\sigma^{(n)}$ is a stopping time for $W^{(n)}$. Thus $\widetilde{W}^{(n)}$ defined above is also a Brownian motion independent of $\left\{\left(B_2(t), \widetilde{B}_2(t)\right) : t \leqslant 2^n - 1\right\}$, $\{Z_k\}_{k \geqslant 2}$ and B_1.

Finally, we set $Z_1^{(n)} = 2^{-n/2} W^{(n)}\left(2^n\right)$ and $\widetilde{Z}_1^{(n)} = 2^{-n/2} \widetilde{W}^{(n)}\left(2^n\right)$. Under this coupling we get

$$I(t) - \widetilde{I}(t) = I\left(2^n - 1\right) - \widetilde{I}\left(2^n - 1\right) + W^{(n)}\left(2^n \wedge \sigma^{(n)}\right) \lambda_n (t), \tag{2.1.3}$$

for $t \in [2^n - 1, 2^{n+1} - 1]$. In particular, $I\left(2^{n+1} - 1\right) - \widetilde{I}\left(2^{n+1} - 1\right) \geqslant 0$ and equals to zero if and only if $\sigma^{(n)} \leqslant 2^n$. If $I\left(2^n - 1\right) - \widetilde{I}\left(2^n - 1\right) = 0$, we synchronously couple $B_2, \widetilde{B}_2$ after time $2^n - 1$. By induction, the coupling is defined for all time.

Now, we claim that the coupling constructed above gives the required bound on the coupling rate. Using Lévy's characterization of Brownian motion and the fact that the $\left\{W^{(n)}\right\}_{n \geqslant 1}$ are independent of the Brownian path B_1, we obtain a Brownian

motion $B^\star$ independent of B_1 such that for all $t \geqslant 0$,

$$\sum_{k=0}^{\infty} \lambda_k \left(2^{k+1} - 1\right) W^{(k)} \left(\left(t - 2^k + 1\right)^+ \wedge 2^k\right) = B^\star \left(T(t)\right),$$

where

$$T(t) = \int_0^t \sum_{k=0}^{\infty} \lambda_k^2 \left(2^{k+1} - 1\right) \mathbb{1} \left(2^k - 1 < s \leqslant 2^{k+1} - 1\right) ds.$$

Note that for any $n \geqslant 0$, the coupling happens after time $2^{n+1} - 1$ if and only if $\sigma^{(k)} > 2^k$ for all $k \leqslant n$, that is, $B^\star(t) > (\widetilde{a} - a) = -1$ for all $t \leqslant T\left(2^{n+1} - 1\right)$. Therefore, if for $y \in \mathbb{R}$, $\tau_y^\star$ denoted the hitting time of level y for the Brownian motion $B^\star$, then we have

$$\mathbb{P} \left(\tau > 2^{n+1} - 1\right) = \mathbb{P} \left(\tau_{-1}^\star > T\left(2^{n+1} - 1\right)\right).$$

By a standard hitting time estimate for Brownian motion, we see that there is a constant $C > 0$ that does not depend on $b_1, b_2, a, \widetilde{a}$ such that

$$\mathbb{P} \left(\tau > 2^{n+1} - 1\right) \leqslant C\mathbb{E} \left[\frac{1}{\sqrt{T\left(2^{n+1} - 1\right)}}\right]. \tag{2.1.4}$$

Thus, we need to obtain an estimate for the right hand side in (4.3.5). Note that $2^{-2n} T\left(2^{n+1} - 1\right)$ has the same distribution as

$$\Psi_n := \frac{4}{\pi^2} \sum_{k=0}^{n} 2^{-2k} U_k^2,$$

where the U_k are i.i.d. standard Gaussian random variables.

For $n \geqslant 1$, $\Psi_n^{-1/2} \leqslant \Psi_1^{-1/2} \leqslant \pi \left(U_0^2 + U_1^2\right)^{-1/2}$. As $\sqrt{U_0^2 + U_1^2}$ has density $re^{-r^2/2} dr$

29

with respect to the Lebesgue measure for $r \geqslant 0$, we conclude that $\mathbb{E}\left[\pi \left(U_0^2 + U_1^2\right)^{-1/2}\right] < \infty$. Thus, for $n \geqslant 1$

$$\mathbb{E}\left[\frac{1}{\sqrt{2^{-2n}T\left(2^{n+1}-1\right)}}\right] = \mathbb{E}\left[\Psi_n^{-1/2}\right] \leqslant \mathbb{E}\left[\Psi_1^{-1/2}\right] \leqslant \mathbb{E}\left[\pi \left(U_0^2 + U_1^2\right)^{-1/2}\right] < \infty.$$

This, along with (4.3.5), implies that there is a positive constant C not depending on $b_1, b_2, a, \widetilde{a}$ such that for $n \geqslant 1$,

$$\mathbb{P}\left(\tau > 2^{n+1} - 1\right) \leqslant \frac{C}{2^n}.$$

It is easy to check that the above inequality implies the lemma. $\qquad \square$

Remark 2.1.2 ([BGM16]). Under the hypothesis of Lemma 2.1.1, it is *not possible to obtain the given rate of decay* of the probability of failing to couple by time t (coupling rate) with *any Markovian coupling*. The proof of this proceeds similar to that of [BK16, Lemma 3.1]. We sketch it here. Under any Markovian coupling μ, a simple Fubini argument shows that there exists a deterministic time $t_0 > 0$ such that $\mu\left(B(t_0) \neq \widetilde{B}(t_0)\right) > 0$. Let τ^B represent the first time when the Brownian motions B and $\widetilde{B}$ meet after time t_0 (which should happen at or before the coupling time of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$). Let $\mathcal{F}_{t_0}$ denote the filtration generated by B and $\widetilde{B}$ up to time t_0 and let $\mathbb{E}_\mu$ denote expectation under the coupling law μ. Then, from the fact that the maximal coupling rate of Brownian motion (equivalently the total variation distance between $B(t)$ and $\widetilde{B}(t)$) decays like $t^{-1/2}$, we deduce that for sufficiently large t

$$\mu(\tau > t) = \mathbb{E}_\mu \mathbb{E}_\mu\left[\tau > t \mid \mathcal{F}_{t_0}\right] \geqslant \mathbb{E}_\mu \mathbb{E}_\mu\left[\tau^B > t \mid \mathcal{F}_{t_0}\right]$$
$$\geqslant C_\mu(t - t_0)^{-1/2} \geqslant C_\mu t^{-1/2},$$

where C_μ denotes a positive constant that depends on the coupling μ. Thus, any Markovian coupling has coupling rate at least $t^{-1/2}$, but the non-Markovian coupling described in Lemma 2.1.1 gives a rate of t^{-1}.

The next proposition gives an estimate for the cumulative distribution function of the coupling time τ for the coupling given in Lemma 2.1.1. The estimate gives a critical value t_0 dependent on the starting points $a, \tilde{a}$ and the constant C that proves when there is a positive probability of failure to couple.

Proposition 2.1.3. Consider the coupling and the coupling time τ given in Lemma 2.1.1. We have that

$$\mathbb{P}\left(\tau \leqslant t\right) \leqslant C\left(t^2 - 1\right) \frac{1}{|a - \tilde{a}|^2}.$$

Moreover if we choose $t < t_0 = \sqrt{\frac{1}{C}|a - \tilde{a}|^2 + 1}$ then $\mathbb{P}\left(\tau < t\right) < 1$ and

$$\mathbb{P}\left(\mathbf{X}_t \neq \tilde{\mathbf{X}}_t \text{ for } 0 < t < t_0\right) > 0.$$

Proof. First thing is note that $B^\star$ is independent of B_1, while $T(t)$ is defined in terms of B_1. Thus $B^\star$ and $T(t)$ are independent of each other. Since

$$\mathbb{P}\left(\tau > 2^{n+1} - 1\right) = \mathbb{P}\left(\tau^\star_{(\tilde{a}-a)/2} > T\left(2^{n+1} - 1\right)\right)$$

then $1 - \mathbb{P}\left(\tau > 2^{n+1} - 1\right) = 1 - \mathbb{P}\left(\tau^\star_{(\tilde{a}-a)/2} > T\left(2^{n+1} - 1\right)\right)$ which means that

$$\mathbb{P}\left(\tau \leqslant 2^{n+1} - 1\right) = \mathbb{P}\left(\tau^\star_{(\tilde{a}-a)/2} \leqslant T\left(2^{n+1} - 1\right)\right).$$

Since $\tau^\star_{(a-\tilde{a})/2}$ is almost surely not zero then

$$
\begin{aligned}
\mathbb{P}\left(\tau \leqslant 2^{n+1} - 1\right) &= \mathbb{P}\left(\tau^\star_{(\tilde{a}-a)/2} \leqslant T\left(2^{n+1} - 1\right)\right) \\
&= \mathbb{P}\left(\frac{T\left(2^{n+1} - 1\right)}{\tau^\star_{(\tilde{a}-a)/2}} \geq 1\right) \\
&\leqslant \mathbb{E}\left[\frac{T\left(2^{n+1} - 1\right)}{\tau^\star_{(\tilde{a}-a)/2}}\right] \\
&= \mathbb{E}\left[T\left(2^{n+1} - 1\right)\right] \mathbb{E}\left[\frac{1}{\tau^\star_{(\tilde{a}-a)/2}}\right].
\end{aligned}
$$

From Brownian motion hitting time estimates the density of $\tau^\star_b$ is $\frac{|b|e^{-b^2/(2t)}}{\sqrt{2\pi t^3}}$ so that $Y = \frac{b^2}{Z^2}$ where $Z \sim N(0,1)$ and $Y \sim \tau^\star_b$. So if $b = (a - \tilde{a})/2$ then

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{\tau^\star_{(\tilde{a}-a)/2}}\right] &= \mathbb{E}\left(\frac{Z^2}{b^2}\right) \\
&= \frac{1}{b^2}\frac{2\sigma^2\sqrt{2\sigma^2}}{\sqrt{\pi}} \\
&= \frac{C}{b^2} \\
&= \frac{C}{|a - \tilde{a}|^2}.
\end{aligned}
$$

Now recall that

$$
T(t) = \int_0^t \sum_{k=0}^\infty \lambda_k^2 \mathbb{I}\left(2^k - 1 < s \leqslant 2^{k+1} - 1\right) ds
$$

where $\lambda_k = \int_{2^k-1}^{2^{k+1}-1} \sqrt{2} \sin\left(\frac{\pi s}{2^k}\right) dB_1(s) \sim \mathcal{N}\left(0, 2^k\right)$. So that

$$
\begin{aligned}
\mathbb{E}\left[T\left(2^{n+1}-1\right)\right] &= \mathbb{E} \int_0^{2^{n+1}-1} \sum_{k=0}^n \lambda_k^2 \mathbb{I}\left(2^k - 1 < s \leqslant 2^{k+1} - 1\right) ds \\
&= \mathbb{E} \sum_{k=0}^n \int_{2^k-1}^{2^{k+1}-1} \lambda_k^2 ds \\
&= \sum_{k=0}^n 2^k \mathbb{E} \lambda_k^2 \\
&= \sum_{k=0}^n 2^{2k} = \sum_{k=0}^n 4^k \\
&= \frac{4\left(1 - 4^{n+1}\right)}{(1-4)} \\
&= \frac{4}{3}\left(\left(2^{n+1}\right)^2 - 1\right) \\
&= \frac{4}{3}\left(\left(2^{n+1}\right) - 1\right)\left(\left(2^{n+1}\right) + 1\right) \\
&= \frac{4}{3}\left(\left(2^{n+1}\right) - 1\right)\left(\left(2^{n+1}\right) - 1 + 2\right).
\end{aligned}
$$

Thus

$$
\mathbb{P}\left(\tau \leqslant 2^{k+1} - 1\right) \leqslant C\left(\left(2^{n+1}\right)^2 - 1\right) \frac{1}{|a - \tilde{a}|^2}.
$$

So that if $2^n \lambda \leqslant t \leqslant 2^n$ for $\lambda < 1$ then

$$
\begin{aligned}
\mathbb{P}\left(\tau \leqslant t\right) &\leqslant C\left(\left(2^n\right)^2 - 1\right) \frac{1}{|a - \tilde{a}|^2} \\
&\leqslant C\left(2\frac{1}{\lambda}t^2 - 1\right) \frac{1}{|a - \tilde{a}|^2}.
\end{aligned}
$$

Taking $\lambda \to 1$ we have that

$$
\mathbb{P}\left(\tau \leqslant t\right) \leqslant Ct\left(t + 2\right) \frac{1}{|a - \tilde{a}|^2}.
$$

Note that we can compute the constant C explicilty above. By choosing $Ct(t+2)\frac{1}{|a-\tilde{a}|^2} < 1$ then $C(t^2+2t) < |a-\tilde{a}|^2$ so that $t < \sqrt{\frac{1}{C}|a-\tilde{a}|^2+1}$. Thus we know that there is a chance that this the Brownian motions doesn't couple if $t < \sqrt{\frac{1}{C}|a-\tilde{a}|^2+1}$. $\square$

The next lemma gives an estimate of the tail of the law of the stochastic integral $\int_0^t B_2(s)dB_1(s)$ run until the first time B_2 hits zero.

Lemma 2.1.4 ([BGM16]). *Let B_1, B_2 be independent Brownian motions with $B_2(0) = b > 0$. For $z \in \mathbb{R}$, let τ_z denote the hitting time of level z by B_2. Then*

$$\mathbb{P}\left(\int_0^{\tau_0} B_2(s)dB_1(s) > y\right) \leqslant \frac{2b}{\sqrt{y}} \quad \text{for } y \geqslant b^2.$$

Proof. For any level $z \geqslant b$, we can write

$$\mathbb{P}\left(\int_0^{\tau_0} B_2(s)dB_1(s) > y\right) =$$
$$\mathbb{P}\left(\int_0^{\tau_0} B_2(s)dB_1(s) > y, \tau_z < \tau_0\right) + \mathbb{P}\left(\int_0^{\tau_0} B_2(s)dB_1(s) > y, \tau_z \geqslant \tau_0\right) \leqslant$$
$$\mathbb{P}\left(\tau_z < \tau_0\right) + \frac{\mathbb{E}\left[\int_0^{\tau_0 \wedge \tau_z} B_2^2(s)ds\right]}{y^2} \leqslant$$
$$\mathbb{P}\left(\tau_z < \tau_0\right) + \frac{z^2}{y^2}\mathbb{E}\left[\tau_0 \wedge \tau_z\right],$$

where the second step follows from Chebyshev's inequality. From standard estimates for Brownian motion, $\mathbb{P}\left(\tau_z < \tau_0\right) = b/z$ and $\mathbb{E}\left[\tau_0 \wedge \tau_z\right] = b(z-b) \leqslant bz$. Using these in the above, we get

$$\mathbb{P}\left(\int_0^{\tau_0} B_2(s)dB_1(s) > y\right) \leqslant \frac{b}{z} + \frac{bz^3}{y^2}.$$

As this bound holds for arbitrary $z \geqslant b$, the result follows by choosing $z = \sqrt{y}$. $\square$

34

Consider two coupled Brownian motions $\left(\mathbf{X}, \widetilde{\mathbf{X}} \right)$ on the Heisenberg group starting from (b_1, b_2, a) and $\left(b_1, \widetilde{b}_2, \widetilde{a} \right)$ respectively. A key object in our coupling construction for Brownian motions on the Heisenberg group $\mathbb{H}^3$ will be the *invariant difference of stochastic areas given* by

$$A(t) = (a - \widetilde{a}) + \left(\int_0^t B_1(s) dB_2(s) - \int_0^t B_2(s) dB_1(s) \right) \tag{2.1.5}$$
$$- \left(\int_0^t \widetilde{B}_1(s) d\widetilde{B}_2(s) - \int_0^t \widetilde{B}_2(s) d\widetilde{B}_1(s) \right) + B_1(t)\widetilde{B}_2(t) - B_2(t)\widetilde{B}_1(t).$$

Note that the Lévy stochastic area is invariant under rotations of coordinates. If the Brownian motions B_1 and $\widetilde{B}_1$ are synchronously coupled at all times, then as the covariation between B_1 and B_2 (and between B_1 and $\widetilde{B}_2$) is zero,

$$A(t) - A(0) = -2 \int_0^t B_2(s) dB_1(s) + 2 \int_0^t \widetilde{B}_2(s) dB_1(s), \tag{2.1.6}$$

where

$$A(0) = a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1, \tag{2.1.7}$$

for $t \geqslant 0$. The next lemma establishes a control on the invariant difference evaluated at the time when the Brownian motions B_2 and $\widetilde{B}_2$ first meet, provided they are reflection coupled up to that time.

Lemma 2.1.5 ([BGM16])**.** *Let B_1 be a real-valued Brownian motion starting from b_1, and let $B_2, \widetilde{B}_2$ be reflection coupled one-dimensional Brownian motions starting from b_2 and $\widetilde{b}_2$ respectively. Consider the invariant difference of stochastic areas given by (2.1.5) with $B_1 = \widetilde{B}_1$. Define $T_1 = \inf \left\{ t \geqslant 0 : B_2(t) = \widetilde{B}_2(t) \right\}.$ Then there exists*

a positive constant C that does not depend on $b_1, b_2, \widetilde{b}_2, a, \widetilde{a}$ such that for any

$$t \geqslant \max\left\{ \left|b_2 - \widetilde{b}_2\right|^2, 2\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right| \right\},$$

we have the estimate

$$\mathbb{E}\left[\frac{|A(T_1)|}{t} \wedge 1\right] \leqslant C\left(\frac{\left|b_2 - \widetilde{b}_2\right|}{\sqrt{t}} + \frac{\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|}{t}\right).$$

Proof. In the proof, C, C' will denote generic positive constants that do not depend on $b_1, b_2, \widetilde{b}_2, a, \widetilde{a}$, whose values might change from line to line. For any $t > 0$,

$$\mathbb{E}\left[\frac{|A(T_1)|}{t} \wedge 1\right] \leqslant \sum_{k=0}^{\infty} \mathbb{E}\left[\frac{|A(T_1)|}{t} \wedge 1; 2^{-k-1}t < |A(T_1)| \leqslant 2^{-k}t\right] + \mathbb{P}\left(|A(T_1)| > t\right)$$

$$\leqslant \sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(2^{-k-1}t < |A(T_1)| \leqslant 2^{-k}t\right) + \mathbb{P}\left(|A(T_1)| > t\right)$$

$$\leqslant \sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right) + \mathbb{P}\left(|A(T_1)| > t\right). \tag{2.1.8}$$

As B_2 and $\widetilde{B}_2$ are reflection coupled, we can rewrite (2.1.6) as

$$A(t) - A(0) = -2\int_0^t \left(B_2(s) - \widetilde{B}_2(s)\right) dB_1(s)$$

where $\frac{1}{2}\left(B_2 - \widetilde{B}_2\right)$ is a Brownian motion starting from $\frac{1}{2}\left(b_2 - \widetilde{b}_2\right)$ and independent

of B_1. By Lemma 2.1.4, for $t \geqslant \max\left\{ \left| b_2 - \widetilde{b}_2 \right|^2, 2\left| A(0) \right| \right\}$,

$$\mathbb{P}\left(\left| A\left(T_1\right)\right| > t\right) \leqslant \mathbb{P}\left(\left| A\left(T_1\right) - A\left(0\right)\right| > t - \left| A\left(0\right)\right|\right)$$

$$\leqslant \mathbb{P}\left(\left| A\left(T_1\right) - A\left(0\right)\right| > \frac{t}{2}\right) \leqslant C\frac{\left| b_2 - \widetilde{b}_2 \right|}{\sqrt{t}}. \qquad (2.1.9)$$

Further, for $t \geqslant \max\left\{ \left| b_2 - \widetilde{b}_2 \right|^2, 2\left| A(0) \right| \right\}$,

$$\sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(\left| A\left(T_1\right)\right| \geqslant 2^{-k-1}t\right)$$

$$= \sum_{k:2^{-k-1}t\leqslant\max\left\{|b_2-\widetilde{b}_2|^2,2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(\left| A\left(T_1\right)\right| \geqslant 2^{-k-1}t\right)$$

$$+ \sum_{k:2^{-k-1}t>\max\left\{|b_2-\widetilde{b}_2|^2,2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(\left| A\left(T_1\right)\right| \geqslant 2^{-k-1}t\right). \quad (2.1.10)$$

To estimate the first term on the right hand side of (2.1.10), let k_0 be the smallest integer k such that $2^{-k-1}t \leqslant \max\left\{ \left| b_2 - \widetilde{b}_2 \right|^2, 2\left| A(0) \right| \right\}$. Then,

$$\sum_{k:2^{-k-1}t\leqslant\max\left\{|b_2-\widetilde{b}_2|^2,2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(\left| A\left(T_1\right)\right| \geqslant 2^{-k-1}t\right)$$

$$\leqslant \sum_{k=k_0}^{\infty} 2^{-k} = 2^{-k_0+1} = \frac{4}{t}2^{-k_0-1}t \leqslant \frac{4}{t}\max\left\{ \left| b_2 - \widetilde{b}_2 \right|^2, 2\left| A(0) \right| \right\}$$

$$\leqslant 8\left(\frac{\left| b_2 - \widetilde{b}_2 \right|^2}{t} + \frac{\left| A(0) \right|}{t}\right) \leqslant 8\left(\frac{\left| b_2 - \widetilde{b}_2 \right|}{\sqrt{t}} + \frac{\left| a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1 \right|}{t}\right), \quad (2.1.11)$$

where we used the facts that $\frac{|b_2-\widetilde{b}_2|^2}{t} \leqslant \frac{|b_2-\widetilde{b}_2|}{\sqrt{t}}$ for $t \geqslant \left| b_2 - \widetilde{b}_2 \right|^2$ and $A(0) = a - \widetilde{a} +$

$b_1\widetilde{b}_2 - b_2\widetilde{b}_1$ to get the last inequality.

To estimate the second term on the right hand side of (2.1.10), we use Lemma 2.1.4 to get

$$
\sum_{k:2^{-k-1}t>\max\left\{\left|b_2-\widetilde{b}_2\right|^2,2|A(0)|\right\}} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right)
$$

$$
\leqslant \frac{C}{\sqrt{t}} \sum_{k:2^{-k-1}t>\max\left\{\left|b_2-\widetilde{b}_2\right|^2,2|A(0)|\right\}} 2^{-k/2}\left|b_2-\widetilde{b}_2\right|
$$

$$
\leqslant \frac{C\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}} \sum_{k=0}^{\infty} 2^{-k/2} \leqslant C'\frac{\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}}. \quad (2.1.12)
$$

Using (2.1.11) and (2.1.12) in (2.1.10),

$$
\sum_{k=0}^{\infty} 2^{-k}\mathbb{P}\left(|A(T_1)| \geqslant 2^{-k-1}t\right) \leqslant C\left(\frac{\left|b_2-\widetilde{b}_2\right|}{\sqrt{t}} + \frac{\left|a-\widetilde{a}+b_1\widetilde{b}_2-b_2\widetilde{b}_1\right|}{t}\right). \quad (2.1.13)
$$

Using (2.1.9) and (2.1.13) in (2.1.8), we complete the proof of the lemma. $\square$

2.2   Main result

Now, we state and prove our main theorem on coupling of Brownian motions on the Heisenberg group $\mathbb{H}^3$.

Theorem 2.2.1 ([BGM16])**.** There exists a non-Markovian coupling $\left(\mathbf{X}, \widetilde{\mathbf{X}}\right)$ of two Brownian motions on the Heisenberg group starting from (b_1, b_2, a) and $\left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right)$ respectively, and a constant $C > 0$ which does not depend on the starting points such

that the coupling time τ satisfies

$$\mathbb{P}\left(\tau > t\right) \leqslant C\left(\frac{\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|}{\sqrt{t}} + \frac{\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|}{t}\right)$$

for $t \geqslant \max\left\{\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|^2, 2\left|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1\right|\right\}$. Here $\mathbf{b} = (b_1, b_2)$ and $\widetilde{\mathbf{b}} = \left(\widetilde{b}_1, \widetilde{b}_2\right)$.

Proof. We will explicitly construct the non-Markovian coupling. In the proof, C will denote a generic positive constant that does not depend on the starting points.

Since the Lévy stochastic area is invariant under rotations of coordinates, it suffices to consider the case when $b_1 = \widetilde{b}_1$. Recall the *invariant difference of stochastic areas* A defined by (2.1.5). We will synchronously couple the Brownian motions B_1 and $\widetilde{B}_1$ at all times. Recall that under this setup, the invariant difference takes the form (2.1.6). The coupling comprises the following two steps.

Step 1. We use a reflection coupling for B_2 and $\widetilde{B}_2$ until the first time they meet. Let $T_1 = \inf\left\{t \geqslant 0 : B_2(t) = \widetilde{B}_2(t)\right\}$.

Step 2. After time T_1 we apply the coupling strategy described in Lemma 2.1.1 to the diffusions

$$\left\{\left(B_1(t), B_2(t), A(T_1) + \int_{T_1}^{t} B_2(s)dB_1(s)\right) : t \geqslant T_1\right\},$$
$$\left\{\left(\widetilde{B}_1(t), \widetilde{B}_2(t), \int_{T_1}^{t} \widetilde{B}_2(s)d\widetilde{B}_1(s)\right) : t \geqslant T_1\right\}.$$

By standard estimates for the Brownian hitting time we have

$$\mathbb{P}\left(T_1 > t\right) \leqslant \frac{C\left|b_2 - \widetilde{b}_2\right|}{\sqrt{t}} \tag{2.2.1}$$

39

for $t \geqslant \left| b_2 - \widetilde{b}_2 \right|^2$. By Lemma 2.1.1 and Lemma 2.1.5, for $t \geqslant \max \left\{ \left| b_2 - \widetilde{b}_2 \right|^2, 2 \left| A(0) \right| \right\}$,

$$\mathbb{P} \left(\tau - T_1 > t \right) \leqslant C \mathbb{E} \left[\frac{\left| A \left(T_1 \right) \right|}{t} \wedge 1 \right]$$

$$\leqslant C \left(\frac{\left| b_2 - \widetilde{b}_2 \right|}{\sqrt{t}} + \frac{\left| a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1 \right|}{t} \right). \quad (2.2.2)$$

Equations (2.2.1) and (2.2.2) together yield the required tail bound on the coupling time probability stated in the theorem. $\qquad \square$

An interesting observation to note from Theorem 2.2.1 is that, if the Brownian motions start from the same point, then the coupling rate is significantly faster.

The above coupling can be used to get sharp estimates on the total variation distance between the laws of two Brownian motions on the Heisenberg group starting from distinct points.

Theorem 2.2.2 ([BGM16]). *If d_{TV} denotes the total variation distance between probability measures, and $\mathcal{L} \left(\mathbf{X}_t \right), \mathcal{L} \left(\widetilde{\mathbf{X}}_t \right)$ denote the laws of Brownian motions on the Heisenberg group starting from (b_1, b_2, a) and $\left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a} \right)$ respectively, then there exists positive constants C_1, C_2 not depending on the starting points such that*

$$d_{TV} \left(\mathcal{L} \left(\mathbf{X}_t \right), \mathcal{L} \left(\widetilde{\mathbf{X}}_t \right) \right) \leqslant C_1 \left(\frac{\left| \mathbf{b} - \widetilde{\mathbf{b}} \right|}{\sqrt{t}} + \frac{\left| a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1 \right|}{t} \right)$$

$$d_{TV} \left(\mathcal{L} \left(\mathbf{X}_t \right), \mathcal{L} \left(\widetilde{\mathbf{X}}_t \right) \right) \geqslant C_2 \left(\frac{\left| \mathbf{b} - \widetilde{\mathbf{b}} \right|}{\sqrt{t}} \mathbb{1} (\mathbf{b} \neq \widetilde{\mathbf{b}}) + \frac{\left| a - \widetilde{a} \right|}{t} \mathbb{1} (\mathbf{b} = \widetilde{\mathbf{b}}) \right)$$

for $t \geqslant \max \left\{ \left| \mathbf{b} - \widetilde{\mathbf{b}} \right|^2, 2 \left| a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1 \right| \right\}$.

Proof. The upper bound on the total variation distance follows from Theorem 2.2.1 and the Aldous' inequality (1.1.1).

To prove the lower bound, we first address the case $\mathbf{b} \neq \widetilde{\mathbf{b}}$. It is straightforward to see from the definition of the total variation distance that

$$d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) \geqslant d_{TV}\left(\mathcal{L}\left(\mathbf{B}_t\right), \mathcal{L}\left(\widetilde{\mathbf{B}}_t\right)\right).$$

Thus, when $\mathbf{b} \neq \widetilde{\mathbf{b}}$, the lower bound in the theorem follows from the standard estimate on the total variation distance between the laws of Brownian motions using the reflection principle

$$d_{TV}\left(\mathcal{L}\left(\mathbf{B}_t\right), \mathcal{L}\left(\widetilde{\mathbf{B}}_t\right)\right) = \mathbb{P}\left(|N(0,1)| \leqslant \frac{\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|}{2\sqrt{t}}\right) \geqslant \frac{1}{\sqrt{2\pi e}}\frac{\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|}{\sqrt{t}}.$$

where $N(0,1)$ denotes a standard Gaussian variable.

Now, we deal with the case $\mathbf{b} = \widetilde{\mathbf{b}}$. As the generator of Brownian motion on the Heisenberg group is hypoelliptic, the law of Brownian motion starting from (u, v, w) has a density with respect to the Lebesgue measure on $\mathbb{R}^3$ which coincides with the Haar measure on $\mathbb{H}^3$. We denote by $p_t^{(u,v,w)}(\cdot, \cdot, \cdot)$ this density (the heat kernel) at time t. The heat kernel $p_t^{(u,v,w)}(x, y, z)$ is a symmetric function of $((u, v, w), (x, y, z)) \in \mathbb{H}^3 \times \mathbb{H}^3$ and is invariant under left multiplication, that is, $p_t^{(u,v,w)}(x, y, z) = p_t^e((u, v, w)^{-1}(x, y, z)) = p_t^e((x, y, z)(u, v, w)^{-1})$. Using the fact that $(u, v, w)^{-1} = (-u, -v, -w)$ we see that

$$p_t^{(u,v,w)}(x, y, z) = p_t^e(x - u, y - v, z - w - uy + vx), \text{ where } e = (0, 0, 0). \qquad (2.2.3)$$

41

Then

$$d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right),\mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) = \int_{\mathbb{R}^3} \left| p_t^{(b_1,b_2,a)}(x,y,z) - p_t^{(b_1,b_2,\widetilde{a})}(x,y,z) \right| dxdydz$$

$$= \int_{\mathbb{R}^3} |p_t^e(x - b_1, y - b_2, z - a - b_1 y + b_2 x)$$

$$- p_t^e(x - b_1, y - b_2, z - \widetilde{a} - b_1 y + b_2 x)| \, dxdydz$$

$$= \int_{\mathbb{R}^3} |p_t^e(x,y,z-a) - p_t^e(x,y,z-\widetilde{a})| \, dxdydz$$

$$\geqslant \int_{\mathbb{R}} |f_t(z-a) - f_t(z-\widetilde{a})| \, dz,$$

where f_t denotes the density with respect to the Lebesgue measure of the Lévy stochastic area at time t when the driving Brownian motion starts at the origin. The third equality above follows by a simple change of variable formula and the last step follows from two applications of the inequality $\left| \int_{\mathbb{R}} f(x)dx \right| \leqslant \int_{\mathbb{R}} |f(x)|dx$ for real-valued measurable f.

From Brownian scaling, it is easy to see that

$$f_t(z) = \frac{1}{t} f_1\left(\frac{z}{t}\right), \quad z \in \mathbb{R}.$$

Substituting this in the above and using the change of variable formula again, we get

$$d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right),\mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) \geqslant \int_{\mathbb{R}} \left| f_1\left(z - \frac{a}{t}\right) - f_1\left(z - \frac{\widetilde{a}}{t}\right) \right| dz$$

$$= \int_{\mathbb{R}} \left| f_1\left(z - \frac{a - \widetilde{a}}{t}\right) - f_1\left(z\right) \right| dz$$

$$\geqslant \int_{|z|\geqslant 1} \left| f_1\left(z - \frac{a - \widetilde{a}}{t}\right) - f_1\left(z\right) \right| dz.$$

The explicit form of f_1 is well-known (see, for example, [Yor91] or [Neu96, p. 32])

$$f_1(z) = \frac{1}{\cosh \pi z}, \quad z \in \mathbb{R}.$$

Without loss of generality, we assume $a > \widetilde{a}$. By the mean value theorem and the assumption made in the theorem that $\frac{a-\widetilde{a}}{t} \leqslant \frac{1}{2}$,

$$\left| f_1\left(z - \frac{a-\widetilde{a}}{t}\right) - f_1(z) \right| \geqslant \frac{a-\widetilde{a}}{t} \inf_{\zeta \in \left[z - \frac{a-\widetilde{a}}{t}, z\right]} |f_1'(\zeta)|$$

$$\geqslant \frac{a-\widetilde{a}}{t} \inf_{\zeta \in \left[z - \frac{1}{2}, z\right]} |f_1'(\zeta)|.$$

We can explicitly compute

$$|f_1'(\zeta)| = \frac{2\pi |e^{\pi\zeta} - e^{-\pi\zeta}|}{(e^{\pi\zeta} + e^{-\pi\zeta})^2}.$$

This is an even function which is strictly decreasing for $\zeta \geqslant 1/2$. Thus, for $|z| \geqslant 1$,

$$\inf_{\zeta \in \left[z - \frac{1}{2}, z\right]} |f_1'(\zeta)| \geqslant |f_1'(3z/2)|.$$

Thus,

$$d_{TV}\left(\mathcal{L}\left(\mathbf{X}_t\right), \mathcal{L}\left(\widetilde{\mathbf{X}}_t\right)\right) \geqslant \int_{|z|\geqslant 1} \left| f_1\left(z - \frac{a-\widetilde{a}}{t}\right) - f_1(z) \right| dz$$

$$\geqslant \frac{|a-\widetilde{a}|}{t} \int_{|z|\geqslant 1} |f_1'(3z/2)| dz = C_2 \frac{|a-\widetilde{a}|}{t},$$

which completes the proof of the theorem. $\qquad\square$

Several remarks are in order.

Remark 2.2.3 ([BGM16]). Theorem 2.2.2 shows that the non-Markovian coupling strategy we constructed is, in fact, an *efficient coupling strategy* in the sense that the coupling rate decays according to the same power of t as the total variation distance between the laws of the Brownian motions $\mathbf{X}$ and $\widetilde{\mathbf{X}}$. We refer to [BK16, Definition 1] for the precise notion of efficiency.

Remark 2.2.4 ([BGM16]). Although we have stated our results without any quantitative bounds on the constants appearing in the coupling time and total variation estimates, it is possible to track concrete numerical bounds from the proofs presented above.

We need the following elementary fact. For any $x \geqslant 0$ and $0 \leqslant y \leqslant 1$

$$x + y \leqslant \sqrt{2} \left(x^2 + y \right)^{\frac{1}{2}}. \tag{2.2.4}$$

Indeed,

$$(x + y)^2 \leqslant 2x^2 + 2y^2 \leqslant 2 \left(x^2 + y \right),$$

since $y \leqslant 1$. This immediately gives us the following result.

Proposition 2.2.5 ([BGM16]). *Assume that* $\left| a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1 \right| < 1.$ *Then there exists a constant $C > 0$ such that*

$$\mathbb{P} \left(\tau > t \right) \leqslant \frac{C}{\sqrt{t}} d_{CC} \left((b_1, b_2, a), \left(\widetilde{b_1}, \widetilde{b_2}, \widetilde{a} \right) \right)$$

for $t \geqslant \max \left\{ \left| \mathbf{b} - \widetilde{\mathbf{b}} \right|^2, 2 \left| a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1 \right|, 1 \right\}.$

44

Proof. Since $t > 1$, then $\frac{1}{t} \leqslant \frac{1}{\sqrt{t}}$, so by Theorem 2.2.1

$$\mathbb{P}\left(\tau > t\right) \leqslant C \left(\frac{\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|}{\sqrt{t}} + \frac{\left|a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1\right|}{t} \right)$$

$$\leqslant \frac{C}{\sqrt{t}} \left(\left|\mathbf{b} - \widetilde{\mathbf{b}}\right| + \left|a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1\right| \right)$$

$$\leqslant \frac{C}{\sqrt{t}} \left(\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|^2 + \left|a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1\right| \right)^{\frac{1}{2}}$$

where we used (2.2.4) in the last inequality. Now we consider

$$\rho\left((b_1, b_2, a), \left(\widetilde{b_1,}, \widetilde{b}_2, \widetilde{a}\right) \right) = \left(\left|\mathbf{b} - \widetilde{\mathbf{b}}\right|^2 + \left|a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1\right| \right)^{\frac{1}{2}},$$

as defined by (1.1.6). Recall from Section 1.1 that this pseudo-metric is equivalent to the Carnot-Carathéodory distance $d_{CC}\left((b_1, b_2, a), \left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right) \right)$. This gives us the desired inequality. $\qquad\square$

Liouville type theorems have been known for the Heisenberg group and other types of Carnot groups (e.g. [BLU07, Theorem 5.8.1]). Using the coupling we constructed, we derive a functional inequality (a form of which appeared as [BBBC08, Equation (24)]) which consequently gives us the Liouville property rather easily.

In the following, for any bounded measurable function $u : \mathbb{H}^3 \rightarrow \mathbb{R}$ and any $x \in \mathbb{H}^3$, we define

$$P_t u(x) = \mathbb{E} u\left(\mathbf{X}_t^x\right),$$

where $\mathbf{X}^x$ is a Brownian motion on the Heisenberg group starting from x. By $\|\cdot\|_\infty$ we denote the sup norm.

Corollary 2.2.6 ([BGM16]). For any bounded $u \in C^\infty(\mathbb{H}^3)$ there exists a positive constant C, which does not depend on u, such that for any $t \geqslant 1$

$$\|\nabla_{\mathcal{H}} P_t u\|_\infty \leqslant \frac{C}{\sqrt{t}} \|u\|_\infty. \tag{2.2.5}$$

Consequently, if $\Delta_{\mathcal{H}} u = 0$, then u is a constant.

Proof. Fix $t \geqslant 1$. Take two distinct points (b_1, b_2, a) and $\left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right)$ in $(\mathbb{H}^3, d_{CC})$ sufficiently close to (b_1, b_2, a) with respect to the distance d_{CC} in such a way that

$$\max \left\{ \left|\mathbf{b} - \widetilde{\mathbf{b}}\right|^2, 2 \left|a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 \widetilde{b}_1\right| \right\} \leqslant 1.$$

Then, using the coupling $(\mathbf{X}, \widetilde{\mathbf{X}})$ constructed in Theorem 2.2.1 and by Proposition 2.2.5, we get

$$\begin{aligned}
\left| P_t u \left(b_1, b_2, a\right) - P_t u \left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right) \right| &= \left| \mathbb{E} \left(u\left(\mathbf{X}_t\right) - u\left(\widetilde{\mathbf{X}}_t\right) : \tau > t \right) \right| \\
&\leqslant 2 \|u\|_\infty \mathbb{P}\left(\tau > t\right) \leqslant \frac{2C}{\sqrt{t}} \|u\|_\infty d_{CC} \left((b_1, b_2, a), \left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right)\right).
\end{aligned}$$

Dividing by $d_{CC}\left((b_1, b_2, a), \left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right)\right)$ on both sides above and taking a supremum over all points $\left(\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}\right) \neq (b_1, b_2, a)$, we get (2.2.5).

Finally if $\Delta_{\mathcal{H}} u = 0$, then $P_t u = u$ for all $t \geqslant 0$. Taking $t \to \infty$ in (2.2.5), we get $\nabla_{\mathcal{H}} u \equiv 0$ and hence $u \in C^\infty(\mathbb{H}^3)$ is constant by [BLU07, Proposition 1.5.6]. $\qquad \square$

Chapter 3

Gradient estimates in the 3-dimensional Heisenberg group

The goal of this chapter is to prove gradient estimates using the coupling construction introduced earlier. Let $x = (b_1, b_2, a)$ and $\widetilde{x} = (\widetilde{b_1}, \widetilde{b_2}, \widetilde{a})$. We let $(\mathbf{X}, \widetilde{\mathbf{X}})$ be the non-Markovian coupling of two Brownian motions $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ on the Heisenberg group starting from x and $\widetilde{x}$ respectively as described in Theorem 2.2.1. For a set Q, define the exit time of a process $\mathbf{X}_t$ from this set by

$$\tau_Q(\mathbf{X}) = \inf\{t > 0 : \mathbf{X}_t \notin Q\}.$$

The oscillation of a function over a set Q is defined by

$$\operatorname*{osc}_Q u \equiv \sup_Q u - \inf_Q u.$$

47

3.1 Preliminary results

Before we can formulate and prove the main results of this section, Theorems 3.2.1 and 3.2.3, we need two preliminary results. Lemma 3.1.1 gives second moment estimates for $\sup_{t \leqslant \tau \wedge 1} |\int_0^t (B_2(s) - b_2) dB_1(s)|$, $\sup_{t \leqslant \tau \wedge 1} |B_1(t) - b_1|$ and $\sup_{t \leqslant \tau \wedge 1} |B_2(t) - b_2|$ under the coupling constructed above, when the coupled Brownian motions start from the same point (b_1, b_2). It would be natural to want to apply here Burkholder-Davis-Gundy (BDG) inequalities such as [KS91, p. 163]) which give sharp estimates of moments of $\sup_{t \leqslant T} |M_t|$ for any continuous local martingale M in terms of the moments of its quadratic variation $\langle M \rangle_T$ when T is a stopping time. But the coupling time τ is *not a stopping time* with respect to the filtration generated by (B_1, B_2), and therefore we *can not apply* these inequalities to get the moment estimates.

Lemma 3.1.1 ([BGM16]). *Consider the coupling of the diffusions*

$$\left\{ \left(B_1(t), B_2(t), a + \int_0^t B_2(s) dB_1(s) \right) : t \geqslant 0 \right\}$$
$$\left\{ \left(\widetilde{B}_1(t), \widetilde{B}_2(t), \widetilde{a} + \int_0^t \widetilde{B}_2(s) d\widetilde{B}_1(s) \right) : t \geqslant 0 \right\},$$

described in Lemma 2.1.1, with $B_1(0) = \widetilde{B}_1(0) = b_1$, $B_2(0) = \widetilde{B}_2(0) = b_2$ and $a > \widetilde{a}$, with coupling time τ. Then there exists a positive constant C not depending on $b_1, b_2, a, \widetilde{a}$ such that we have the following

(i) $\mathbb{E} \left(\sup_{t \leqslant \tau \wedge 1} \left| \int_0^t (B_2(s) - b_2) dB_1(s) \right| \right)^2 \leqslant C \mathbb{E}(\tau \wedge 1)^2,$

(ii) $\mathbb{E} \left(\sup_{t \leqslant \tau \wedge 1} |B_1(t) - b_1| \right)^4 \leqslant C \mathbb{E}(\tau \wedge 1)^2,$

(iii) $\mathbb{E} \left(\sup_{t \leqslant \tau \wedge 1} |B_2(t) - b_2| \right)^4 \leqslant C \mathbb{E}(\tau \wedge 1)^2.$

Proof. In this proof, C will denote a generic positive constant whose value does not depend on $b_1, b_2, a, \widetilde{a}$. Our basic strategy will be to find appropriate enlargements of the natural filtration generated by (B_1, B_2) under which τ becomes a stopping time, and then use the Burkholder-Davis-Gundy inequality.

It suffices to prove the statement for $b_1 = b_2 = 0$. Moreover, using scaling of Brownian motion, it is straightforward to check that it is sufficient to prove the statement with $a - \widetilde{a} = 1$ and $\tau \wedge 1$ replaced by $\tau \wedge M$ (for arbitrary $M > 0$). We write $B_2(t) = Y_1(t) + Y_2(t)$, where

$$Y_1(t) = \sum_{n=0}^{\infty} 2^{n/2} Z_1^{(n)} g_{n,1}((t - 2^n + 1)^+ \wedge 2^n)$$

$$Y_2(t) = \sum_{n=0}^{\infty} 2^{n/2} \left(\frac{(t - 2^n + 1)^+ \wedge 2^n}{2^n} Z_0^{(n)} + \sum_{k=2}^{\infty} Z_k^{(n)} g_{n,k}((t - 2^n + 1)^+ \wedge 2^n) \right)$$

$$(3.1.1)$$

with $g_{n,k}(t) = g_{2^n,k}(t)$ as defined in the Karhunen-Loève expansion (4.3.3) and $Z_0^{(n)} = 2^{-n/2} G^{(2^n)}$ for a a Gaussian variable with mean zero and variance 2^n as we used in (4.3.2).

Consider the filtration

$$\mathcal{F}_t^* = \sigma \left(\{B_1(s) : s \leqslant t\} \cup \{W^{(n)}(s) : n \geqslant 0, 0 \leqslant s \leqslant \infty\} \cup \{Z_k^{(n)} : n \geqslant 0, k \geqslant 2\} \right).$$

We assume without loss of generality that $\{\mathcal{F}_t^*\}_{t \geqslant 0}$ is augmented, in the sense that all the null sets of $\mathcal{F}_\infty^*$ and their subsets lie in $\mathcal{F}_0^*$. We claim that τ is a stopping time under the above filtration. To see this, recall that by the definition of coupling time, the coupled processes must evolve together after the coupling time and thus, by the

coupling construction given in Lemma 2.1.1 (in particular, see (2.1.3)),

$$\mathbb{P}[\tau \in \{2^{n+1} - 1 : n \geqslant 0\}] = 1. \tag{3.1.2}$$

Thus, to show that τ is a stopping time with respect to $\mathcal{F}_t^*$, it suffices to show that $\{\tau > 2^{n+1} - 1\}$ is measurable with respect to $\mathcal{F}_{2^{n+1}-1}^*$ for each $n \geqslant 0$. This is because, for $t \in [2^{n+1} - 1, 2^{n+2} - 1)$ $(n \geqslant 0)$,

$$\{\tau > t\} = \{\tau > 2^{n+1} - 1\}$$

almost surely with respect to the coupling measure $\mathbb{P}$, by (3.1.2). Note that for any $n \geqslant 0$,

$$\{\tau > 2^{n+1} - 1\} = \bigcap_{m=0}^{n} \{\sigma^{(m)} > 2^m\}.$$

Recall that

$$\sigma^{(m)} = \inf \left\{ t \geqslant 0 : W^{(m)}(t) = \right.$$
$$\left. - \left(I(2^m - 1) - \widetilde{I}(2^m - 1) \right) \middle/ \left(2 \int_{2^m-1}^{2^{m+1}-1} g_{m,1}(s - 2^m + 1) dB_1(s) \right) \right\}$$

and on the event $\{\tau > 2^{m+1} - 1\}$,

$$B_2(s) - \widetilde{B}_2(s) = Y_1(s) - \widetilde{Y}_1(s) = 2Y_1(s), \quad \text{for all } 0 \leqslant s \leqslant 2^{m+1} - 1.$$

As $\{Y_1(t) : 0 \leqslant t \leqslant 2^{m+1} - 1\}$ depends measurably on $\{Z_1^{(k)} : 0 \leqslant k \leqslant m\}$ and hence on $\{W^{(k)}(s) : k \geqslant 0, 0 \leqslant s < \infty\}$, the above representation for $\sigma^{(m)}$ implies that the event $\{\sigma^{(m)} > 2^m\}$ is measurable with respect to $\mathcal{F}_{2^{m+1}-1}^*$. Thus, for each

50

$n \geqslant 0$, $\{\tau > 2^{n+1} - 1\}$ is measurable with respect to $\mathcal{F}^*_{2^{n+1}-1}$ and hence, τ is indeed a stopping time with respect to $\{\mathcal{F}^*_t\}_{t \geqslant 0}$.

Also, note that $\left(\int_0^t B_2(s) dB_1(s) \right)_{t \geqslant 0}$ remains a continuous martingale under this enlarged filtration. Thus, by the Burkholder-Davis-Gundy inequality, we get

$$\mathbb{E} \left(\sup_{t \leqslant \tau \wedge M} \left| \int_0^t B_2(s) dB_1(s) \right| \right)^2 \leqslant C \mathbb{E} \left(\int_0^{\tau \wedge M} B_2^2(s) ds \right) \leqslant$$
$$C \mathbb{E} \left(\left(\sup_{t \leqslant \tau \wedge M} |B_2(t)| \right)^2 (\tau \wedge M) \right)$$

Now, by the Cauchy-Schwarz inequality

$$\mathbb{E} \left(\left(\sup_{t \leqslant \tau \wedge M} |B_2(t)| \right)^2 (\tau \wedge M) \right) \leqslant \left(\mathbb{E} \left(\sup_{t \leqslant \tau \wedge M} |B_2(t)| \right)^4 \right)^{1/2} \left(\mathbb{E}(\tau \wedge M)^2 \right)^{1/2}.$$

Thus, to complete the proof (i) and (iii), it suffices to show that

$$\mathbb{E} \left(\sup_{t \leqslant \tau \wedge M} |B_2(t)| \right)^4 \leqslant C \mathbb{E}(\tau \wedge M)^2.$$

To show this, define the Brownian motion

$$W(t) = \sum_{n=0}^{\infty} W^{(n)} \left((t - 2^n + 1)^+ \wedge 2^n \right)$$

and the following (augmented) filtration

$$\mathcal{F}^{**}_t = \sigma \left(\{ (B_1(s), W(s)) : s \leqslant t \} \cup \{ Z_k^{(n)} : n \geqslant 0, k \geqslant 2 \} \right).$$

Exactly as before, we can check that τ is a stopping time with respect to this new

51

filtration and W is a Brownian motion (hence a continuous martingale) under it. From the representation (3.1.1), note that

$$\sup_{t \leqslant \tau \wedge M} |Y_1(t)| = \frac{\sqrt{2}}{\pi} \sup_{n:2^{n+1}-1 \leqslant \tau \wedge M} |W(2^{n+1}-1) - W(2^n - 1)| \leqslant \frac{2\sqrt{2}}{\pi} \sup_{t \leqslant \tau \wedge M} |W(t)|.$$

Thus, by the the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\left(\sup_{t \leqslant \tau \wedge M} |Y_1(t)|\right)^4 \leqslant \frac{64}{\pi^4} \mathbb{E}\left(\sup_{t \leqslant \tau \wedge M} |W(t)|\right)^4 \leqslant C\mathbb{E}(\tau \wedge M)^2. \tag{3.1.3}$$

To estimate $\sup_{t \leqslant \tau \wedge M} |Y_2(t)|$, note that Y_2 and τ are independent. Thus, by a conditioning argument, it suffices to show that for fixed $T > 0$,

$$\mathbb{E}\left(\sup_{t \leqslant T} |Y_2(t)|\right)^4 \leqslant CT^2. \tag{3.1.4}$$

To see this, observe that $Y_2(t) = B_2(t) - Y_1(t)$ for each $t \geqslant 0$ and thus

$$\sup_{t \leqslant T} |Y_2(t)| \leqslant \sup_{t \leqslant T} |B_2(t)| + \sup_{t \leqslant T} |Y_1(t)|.$$

Again by the the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\left(\sup_{t \leqslant T} |B_2(t)|\right)^4 \leqslant CT^2.$$

By exactly the same argument as the one used to estimate the supremum of Y_1, but now applied to a fixed time T, we get

$$\mathbb{E}\left(\sup_{t \leqslant T} |Y_1(t)|\right)^4 \leqslant CT^2.$$

The two estimates above yield (3.1.4), and hence complete the proof of (i) and (iii).

Similarly, (ii) follows from the fact that B_1 is a Brownian motion under the filtration $\{\mathcal{F}_t^*\}_{t \geqslant 0}$ and the Burkholder-Davis-Gundy inequality. $\quad\square$

The next lemma estimates $\mathbb{E}(\tau \wedge 1)^2$.

Lemma 3.1.2 ([BGM16]). *Under the coupling of Lemma 2.1.1, there exists a positive constant C not depending on $b_1, b_2, a, \widetilde{a}$ such that*

$$\mathbb{E}(\tau \wedge 1)^2 \leqslant C(|a - \widetilde{a}| \wedge 1).$$

Proof. Without loss of generality, we assume $|a - \widetilde{a}| \leqslant 1$. We can write

$$\mathbb{E}(\tau \wedge 1)^2 = \int_0^1 \mathbb{P}(\tau > \sqrt{t})dt$$
$$\leqslant |a - \widetilde{a}|^2 + \int_{|a-\widetilde{a}|^2}^1 \mathbb{P}(\tau > \sqrt{t})dt.$$

From Lemma 2.1.1, we get a constant C that does not depend on $b_1, b_2, a, \widetilde{a}$ such that for $t > |a - \widetilde{a}|^2$,

$$\mathbb{P}(\tau > \sqrt{t}) \leqslant C \frac{|a - \widetilde{a}|}{\sqrt{t}}.$$

Using this we get

$$\mathbb{E}(\tau \wedge 1)^2 \leqslant |a - \widetilde{a}|^2 + C|a - \widetilde{a}| \int_0^1 \frac{1}{\sqrt{t}}dt \leqslant (1 + 2C)|a - \widetilde{a}|,$$

which proves the lemma. $\quad\square$

3.2 The gradient estimate

Let $D \subset \mathbb{H}^3$ be a domain. Later in Theorem 3.2.3 we give gradient estimates for harmonic functions in D, but we start by a result on the coupling time τ. Define the Heisenberg ball of radius $r > 0$ with respect to the distance ρ

$$B(x,r) = \{y \in \mathbb{H}^3 : \rho(x,y) < r\}.$$

Recall that ρ is the pseudo-metric equivalent to d_{CC} defined by (1.1.6). For $x \in D$, let $\delta_x = \rho(x, D^c)$.

Consider the coupling of two Brownian motions on the Heisenberg group $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ starting from points $x, \widetilde{x} \in D$ respectively as described by Theorem 2.2.1. We choose these points in such a way that $\rho(x, \widetilde{x})$ is small enough compared to δ_x. The following theorem estimates the probability (as a function of δ_x and $\rho(x, \widetilde{x})$) that one of the processes exits the ball $B(x, \delta_x)$ before coupling happens. This turns out to be pivotal in proving the gradient estimate.

Theorem 3.2.1 ([BGM16]). Let $x = (b_1, b_2, a) \in D$, $\widetilde{x} = (\widetilde{b}_1, \widetilde{b}_2, \widetilde{a}) \in D$ such that $\rho(x, \widetilde{x}) < \delta_x/32$, $|\mathbf{b} - \widetilde{\mathbf{b}}| \leqslant 1$ and $|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1| \leqslant 1/2$. Then, under the same coupling of Theorem 2.2.1, there exists a constant $C > 0$ that does not depend on $x, \widetilde{x}$ such that

$$\mathbb{P}\left(\tau > \tau_{B(x,\delta_x)}(\mathbf{X}) \wedge \widetilde{\tau}_{B(x,\delta_x)}(\widetilde{\mathbf{X}})\right) \leqslant C\left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1 + \delta_x)^3}{\delta_x^4}\right)\rho(x, \widetilde{x}).$$

Proof. In this proof, C will denote a generic positive constant (whose value might change from line to line) that does not depend on $x, \widetilde{x}$.

Let $\hat{b}_i = \frac{b_i + \widetilde{b}_i}{2}$ for $i = 1, 2$ and $\hat{a} = \frac{a + \widetilde{a}}{2}$. We define the Heisenberg cube by

$$Q = \left\{ (y_1, y_2, y_3) \in \mathbb{R}^3 : \max_{i=1,2} \left| y_i - \hat{b}_i \right| \leq \frac{\delta_x}{8}, \left| \hat{a} - y_3 + \hat{b}_1 y_2 - \hat{b}_2 y_1 \right| \leq \frac{\delta_x^2}{16} \right\}.$$

Write $\hat{x} = (\hat{b}_1, \hat{b}_2, \hat{a})$. It is straightforward to check that $\rho(x, \hat{x}) \leq \rho(x, \widetilde{x})/\sqrt{2} < \delta_x / 32\sqrt{2}$. Moreover, for $y \in Q$

$$\rho(\hat{x}, y) = \left(|y_1 - \hat{b}_1|^2 + |y_2 - \hat{b}_2|^2 + \left| \hat{a} - y_3 + \hat{b}_1 y_2 - \hat{b}_2 y_1 \right| \right)^{1/2}$$
$$\leq |y_1 - \hat{b}_1| + |y_2 - \hat{b}_2| + \left| \hat{a} - y_3 + \hat{b}_1 y_2 - \hat{b}_2 y_1 \right|^{1/2} \leq \delta_x/2.$$

Thus, by the triangle inequality, for any $y \in Q$

$$\rho(x, y) \leq \rho(x, \hat{x}) + \rho(\hat{x}, y) < \delta_x$$

and hence, $Q \subset B(x, \delta_x)$. Note that we can write $Q = Q_1 \cap Q_2$ where

$$Q_1 = \left\{ (y_1, y_2, y_3) \in \mathbb{R}^3 : \max_{i=1,2} \left| y_i - \hat{b}_i \right| \leq \frac{\delta_x}{8} \right\},$$
$$Q_2 = \left\{ (y_1, y_2, y_3) \in \mathbb{R}^3 : \left| \hat{a} - y_3 + \hat{b}_1 y_2 - \hat{b}_2 y_1 \right| \leq \frac{\delta_x^2}{16} \right\}.$$

As the Lévy stochastic area is invariant under rotations of coordinates, it suffices to assume that $b_1 = \widetilde{b}_1$. We define

$$U(t) = a - \hat{a} + \int_0^t B_1(s)dB_2(s) - \int_0^t B_2(s)dB_1(s) + B_1(t)\hat{b}_2 - B_2(t)\hat{b}_1.$$

Note that

$$dU(t) = (B_1(t) - \hat{b}_1)dB_2(t) - (B_2(t) - \hat{b}_2)dB_1(t).$$

55

Writing

$$\sigma_u = \inf\{t \geqslant 0 : |U(t)| > u\},$$

we observe that $\tau_{Q_2}(\mathbf{X}) = \sigma_{\delta_x^2/16}$ and hence, $\tau_Q(\mathbf{X}) = \tau_{Q_1}(\mathbf{X}) \wedge \tau_{Q_2}(\mathbf{X}) = \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}$. We can write

$$\mathbb{P}\left(\tau > \tau_{B(x,\delta_x)}(\mathbf{X}) \wedge \widetilde{\tau}_{B(x,\delta_x)}\left(\widetilde{\mathbf{X}}\right)\right) \leqslant \mathbb{P}(\tau > \tau_Q(\mathbf{X}) \wedge \tau_Q(\widetilde{\mathbf{X}}))$$
$$\leqslant \mathbb{P}(\tau > \tau_Q(\mathbf{X})) + \mathbb{P}(\tau > \tau_Q(\widetilde{\mathbf{X}})).$$

Now we estimate $\mathbb{P}(\tau > \tau_Q(\mathbf{X}))$, the second term in the inequality above can be estimated similarly. First we define

$$Q_1^* = \left\{(y_1, y_2, y_3) \in \mathbb{R}^3 : \max_{i=1,2}\left|y_i - \hat{b}_i\right| \leqslant \frac{\delta_x}{16}\right\}.$$

We have

$$\mathbb{P}(\tau > \tau_Q(\mathbf{X})) = \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16})$$
$$\leqslant \mathbb{P}(T_1 > \tau_{Q_1^*}(\mathbf{X})) + \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}))$$
$$\leqslant \mathbb{P}(T_1 > \tau_{Q_1^*}(\mathbf{X})) + \mathbb{P}(\sigma_{\delta_x^2/32} \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))$$
$$+ \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}). \qquad (3.2.1)$$

It follows from a computation involving standard Brownian estimates (see, for example, the proof of [Cra92, Theorem 1]) that

$$\mathbb{P}(T_1 > \tau_{Q_1^*}(\mathbf{X})) \leqslant C\frac{|\mathbf{b} - \widetilde{\mathbf{b}}|}{\delta_x}. \qquad (3.2.2)$$

56

To estimate the second term in (3.2.1), note that

$$\mathbb{P}(\sigma_{\delta_x^2/32} \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) = \mathbb{P}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t)| > \frac{\delta_x^2}{32}\right).$$

Now, as $T_1 \wedge \tau_{Q_1^*}(\mathbf{X})$ is a stopping time with respect to the natural filtration generated by (B_1, B_2), by the the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t) - U(0)|\right)^2$$

$$\leqslant C\mathbb{E}\left(\int_0^{T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |\mathbf{B}(s) - \hat{\mathbf{b}}|^2 ds\right)$$

$$\leqslant C\mathbb{E}\left(\int_0^{T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} \delta_x^2 ds\right)$$

$$\leqslant C\delta_x^2 \mathbb{E}(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})).$$

We can again appeal to standard Brownian estimates (e.g. see the proof of [Cra92, Theorem 1]) to see that

$$\mathbb{E}\left(T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))\right) \leqslant C\delta_x|\mathbf{b} - \hat{\mathbf{b}}|. \tag{3.2.3}$$

Using this estimate gives us

$$\mathbb{E}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t)|\right)^2 \leqslant 2\mathbb{E}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t) - U(0)|\right)^2 + 2|U(0)|^2$$

$$\leqslant C\delta_x^3|\mathbf{b} - \hat{\mathbf{b}}| + 2|a - \hat{a} + b_1\hat{b}_2 - b_2\hat{b}_1|^2 \leqslant \frac{C}{2}\delta_x^3|\mathbf{b} - \widetilde{\mathbf{b}}| + \frac{1}{2}|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1|^2.$$

By assumption $|a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1| < 1$, and therefore

$$\mathbb{E}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t)|\right)^2 \leqslant C(1 + \delta_x)^3(|\mathbf{b} - \widetilde{\mathbf{b}}| + |a - \widetilde{a} + b_1\widetilde{b}_2 - b_2\widetilde{b}_1|)$$

$$\leqslant C(1 + \delta_x)^3 \rho(x, \widetilde{x}),$$

where the last inequality follows from (2.2.4). Thus, by the Chebyshev inequality

$$\mathbb{P}\left(\sup_{t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |U(t)| > \frac{\delta_x^2}{32}\right) \leqslant C\frac{(1 + \delta_x)^3}{\delta_x^4}\rho(x, \widetilde{x}),$$

which, in turn, gives us

$$\mathbb{P}(\sigma_{\delta_x^2/32} \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) \leqslant C\frac{(1 + \delta_x)^3}{\delta_x^4}\rho(x, \widetilde{x}). \tag{3.2.4}$$

To estimate the last term in (3.2.1), we write

$$\mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}) \leqslant \mathbb{P}(\tau - T_1 > 1)$$

$$+ \mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}, \tau - T_1 \leqslant 1). \tag{3.2.5}$$

By Lemma 2.1.1, we get

$$\mathbb{P}(\tau - T_1 > 1) \leqslant C\mathbb{E}|A(T_1) \wedge 1|,$$

where A is the invariant difference of stochastic areas defined in (2.1.5).

Applying Lemma 2.1.5 with $t = 1$ and appealing to our assumption that $|\mathbf{b} - \widetilde{\mathbf{b}}| \leqslant 1$

and $|a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 b_1| \leqslant 1/2$, we have

$$\mathbb{E}|A(T_1) \wedge 1| \leqslant C(|\mathbf{b} - \widetilde{\mathbf{b}}| + |a - \widetilde{a} + b_1 \widetilde{b}_2 - b_2 b_1|) \leqslant C\rho(x, \widetilde{x}).$$

which gives

$$\mathbb{P}(\tau - T_1 > 1) \leqslant C\rho(x, \widetilde{x}). \tag{3.2.6}$$

Finally, we need to estimate $\mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}, \tau - T_1 \leqslant 1)$. Note that

$$\mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}, \tau - T_1 \leqslant 1)$$

$$\leqslant \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| \geqslant \delta_x/16 \right) +$$

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_2(t) - B_2(T_1)| \geqslant \delta_x/16 \right)$$

$$+ \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U(t) - U(T_1)| \geqslant \delta_x^2/32, \right.$$

$$\left. \sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \right). \tag{3.2.7}$$

By the strong Markov property applied at T_1, along with parts (ii) and (iii) of Lemma 3.1.1 and the Chebyshev inequality, we get

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_i(t) - B_i(T_1)| \geqslant \delta_x/16 \right) \leqslant C\frac{\mathbb{E}((\tau - T_1) \wedge 1)^2}{\delta_x^4}$$

for $i = 1, 2$. From the explicit construction of the coupling strategy given in Theorem

2.2.1 and Lemma 3.1.2 and Lemma 2.1.5, we obtain

$$\mathbb{E}((\tau - T_1) \wedge 1)^2 \leqslant \mathbb{E}|A(T_1) \wedge 1| \leqslant C\rho(x, \widetilde{x}).$$

and thus,

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_i(t) - B_i(T_1)| \geqslant \delta_x/16\right) \leqslant C\frac{\rho(x, \widetilde{x})}{\delta_x^4}. \qquad (3.2.8)$$

for $i = 1, 2$. To handle the last term in (3.2.7), define

$$U^*(t) = U(t) - (B_1(t) - \hat{b}_1)(B_2(t) - \hat{b}_2).$$

Note that

$$dU^*(t) = -2(B_2(t) - \hat{b}_2)dB_1(t).$$

and $U^*(T_1) = U(T_1)$ as $B_2(T_1) = \hat{b}_2$. Further, observe that

$$\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U(t) - U(T_1)| \leqslant$$

$$\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U^*(t) - U^*(T_1)| + \sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - \hat{b}_1||B_2(t) - \hat{b}_2|.$$

Using this, we can bound the last term in (3.2.7) as

$$
\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U(t) - U(T_1)| \geqslant \delta_x^2/32,\right.
$$

$$
\left.\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right)
$$

$$
\leqslant \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U^*(t) - U^*(T_1)| \geqslant \delta_x^2/64\right)
$$

$$
+ \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - \hat{b}_1||B_2(t) - \hat{b}_2| \geqslant \delta_x^2/64,\right.
$$

$$
\left.\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right).
$$

$$(3.2.9)$$

By conditioning at time T_1 and part (i) of Lemma 3.1.1, followed by applications of Lemma 3.1.2 and Lemma 2.1.5, we obtain

$$
\mathbb{E}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U^*(t) - U^*(T_1)|\right)^2 \leqslant
$$

$$
4\mathbb{E}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} \left|\int_{T_1}^t (B_2(s) - \hat{b}_2) dB_1(s)\right|\right)^2 \leqslant
$$

$$
C\mathbb{E}((\tau - T_1) \wedge 1)^2 \leqslant \mathbb{E}|A(T_1) \wedge 1| \leqslant C\rho(x, \widetilde{x}).
$$

Consequently, by the Chebyshev inequality

$$
\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |U^*(t) - U^*(T_1)| \geqslant \delta_x^2/64\right) \leqslant C\frac{\rho(x, \widetilde{x})}{\delta_x^4}. \qquad (3.2.10)
$$

61

Moreover,

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - \hat{b}_1||B_2(t) - \hat{b}_2| \geqslant \delta_x^2/64,\right.$$

$$\left.\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right)$$

$$\leqslant \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_2(t) - \hat{b}_2| \geqslant \delta_x/8\right)$$

$$+ \mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right.$$

$$\left.\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right).$$

$$(3.2.11)$$

We use the fact $B_2(T_1) = \hat{b}_2$ and proceed exactly along the lines of the proof of (3.2.8) to obtain

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1 + (\tau - T_1) \wedge 1} |B_2(t) - \hat{b}_2| \geqslant \delta_x/8\right) \leqslant C\frac{\rho(x, \widetilde{x})}{\delta_x^4}. \qquad (3.2.12)$$

The second probability appearing on the right hand side of (3.2.11) can be bounded

as follows

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right.$$

$$\left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right)$$

$$\leqslant \mathbb{P}\left(\sup_{(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) \leqslant t \leqslant (T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))+(\tau-(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})))\wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right.$$

$$\left.\sup_{(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) \leqslant t \leqslant (T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))+(\tau-(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})))\wedge 1} |B_1(t) - B_1(T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))| < \delta_x/16\right)$$

$$\leqslant \mathbb{P}\left(|B_1(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) - \hat{b}_1| > \delta_x/16\right). \quad (3.2.13)$$

We will use the fact that $b_1 = \hat{b}_1$. By an application of the Chebyshev inequality followed by the Burkholder-Davis-Gundy inequality, and using (3.2.3), we get

$$\mathbb{P}\left(|B_1(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) - \hat{b}_1| > \delta_x/16\right) \leqslant C\frac{\mathbb{E}|B_1(T_1 \wedge \tau_{Q_1^*}(\mathbf{X})) - \hat{b}_1|^2}{\delta_x^2}$$

$$\leqslant C\frac{\mathbb{E}\sup_{0 \leqslant t \leqslant T_1 \wedge \tau_{Q_1^*}(\mathbf{X})} |B_1(t) - b_1|^2}{\delta_x^2} \leqslant C\frac{\mathbb{E}(T_1 \wedge \tau_{Q_1^*}(\mathbf{X}))}{\delta_x^2} \leqslant C\frac{|\mathbf{b} - \hat{\mathbf{b}}|}{\delta_x}.$$

Using this in (3.2.13),

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - \hat{b}_1| \geqslant \delta_x/8,\right.$$

$$\left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right) \leqslant C\frac{|\mathbf{b} - \hat{\mathbf{b}}|}{\delta_x}.$$

$$(3.2.14)$$

Using (3.2.12) and (3.2.14) in (3.2.11), we obtain

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - \hat{b}_1||B_2(t) - \hat{b}_2| \geqslant \delta_x^2/64,\right.$$
$$\left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right)$$
$$\leqslant C\left(\frac{1}{\delta_x} + \frac{1}{\delta_x^4}\right)\rho(x, \widetilde{x}). \quad (3.2.15)$$

Finally, using (3.2.10) and (3.2.15) in (3.2.9),

$$\mathbb{P}\left(\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |U(t) - U(T_1)| \geqslant \delta_x^2/32,\right.$$
$$\left.\sup_{T_1 \leqslant t \leqslant T_1+(\tau-T_1)\wedge 1} |B_1(t) - B_1(T_1)| < \delta_x/16, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X})\right)$$
$$\leqslant C\left(\frac{1}{\delta_x} + \frac{1}{\delta_x^4}\right)\rho(x, \widetilde{x}). \quad (3.2.16)$$

Using the estimates from (3.2.8) and (3.2.16) in (3.2.7), we get

$$\mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}, \tau - T_1 \leqslant 1)$$
$$\leqslant C\left(\frac{1}{\delta_x} + \frac{1}{\delta_x^4}\right)\rho(x, \widetilde{x}). \quad (3.2.17)$$

Using (3.2.6) and (3.2.17) in (3.2.5), we get

$$\mathbb{P}(\tau > \tau_{Q_1}(\mathbf{X}) \wedge \sigma_{\delta_x^2/16}, T_1 \leqslant \tau_{Q_1^*}(\mathbf{X}) \wedge \sigma_{\delta_x^2/32}) \leqslant C\left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4}\right)\rho(x, \widetilde{x}). \quad (3.2.18)$$

Using the estimates (3.2.2), (3.2.4) and (3.2.18) in (3.2.1), we obtain

$$\mathbb{P}(\tau > \tau_Q(\mathbf{X})) \leqslant C \left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1+\delta_x)^3}{\delta_x^4} \right) \rho(x, \widetilde{x}). \qquad (3.2.19)$$

The same estimate for $\mathbb{P}(\tau > \tau_Q(\widetilde{\mathbf{X}}))$ is obtained by interchanging the roles of x and $\widetilde{x}$. This completes the proof of the theorem. $\qquad \square$

Remark 3.2.2. Theorem 3.2.1 and its proof remain unchanged if we replace δ_x by $\alpha\delta_x$ for any $\alpha \in (0, 1]$.

The above theorem yields the gradient estimate formulated in Theorem 3.2.3. Before we can formulate our result, we explain the argument in the proof of [Kuw10, Proposition 4.1] that leads to (3.2.20).

Recall that $\Delta_{\mathcal{H}}$ denotes the sub-Laplacian which is the generator of the Brownian motion on $\mathbb{H}^3$, and for any function f on $\mathbb{H}^3$, $|\nabla_{\mathcal{H}} f|$ denotes the associated length of the horizontal gradient of f defined by (1.1.4). As before $\|\cdot\|_{\mathcal{H}}$ denotes the norm induced by the sub-Riemannian metric on horizontal vectors. We can use the fact that $\{\mathcal{X}, \mathcal{Y}\}$ is an orthonormal frame for the horizontal distribution, therefore for any Lipschitz continuous function u defined on a domain D in $\mathbb{H}^3$,

$$\|\nabla_{\mathcal{H}} u\|_{\mathcal{H}}^2 = (\mathcal{X}u)^2 + (\mathcal{Y}u)^2$$

holds in D (where $\mathcal{X}u$ and $\mathcal{Y}u$ are interpreted in the distributional sense). Now we can use [HK00, Theorem 11.7] for the vector fields $\{\mathcal{X}, \mathcal{Y}\}$ in $\mathbb{H}^3$ identified with $\mathbb{R}^3$. We need to check some assumptions in this theorem. First, if u is Lipschitz continuous

on $\overline{D}$, it is clear that

$$|\nabla_{\mathcal{H}} u|\,(x) \leqslant \sup_{z,\tilde{z}\in\overline{D},z\neq\tilde{z}} \frac{|u(z)-u(\tilde{z})|}{d_{CC}(z,\tilde{z})} < \infty,$$

for all $x \in \overline{D}$, and hence $|\nabla_{\mathcal{H}} u|$ is locally integrable. In addition, as u is Lipschitz continuous, $|\nabla_{\mathcal{H}} u|$ is an upper gradient of u by [Kuw10, Lemma 2.1], so [HK00, Theorem 11.7] is applicable and we have that

$$\|\nabla_{\mathcal{H}} u\|_{\mathcal{H}} \leqslant |\nabla_{\mathcal{H}} u|, \tag{3.2.20}$$

a.e. with respect to the Lebesgue measure.

Theorem 3.2.3 ([BGM16]). *Suppose u satisfies $\Delta_{\mathcal{H}} u = 0$ on $D \subset \mathbb{H}^3$. Fix any constant $\alpha \in (0,1)$. There exists a constant $C > 0$ that does not depend on $u, \delta_x, x, D, \alpha$ such that for every $x \in D$*

$$\|\nabla_{\mathcal{H}} u(x)\|_{\mathcal{H}} \leqslant |\nabla_{\mathcal{H}} u|\,(x) \leqslant C \left(1 + \frac{1}{\alpha\delta_x} + \frac{1}{(\alpha\delta_x)^4} + \frac{(1+\alpha\delta_x)^3}{(\alpha\delta_x)^4}\right) \operatorname{osc}_{B(x,\alpha\delta_x)} u. \tag{3.2.21}$$

Proof. Recall that by hypoellipticity we know that if $\Delta_{\mathcal{H}} u = 0$ then u must be smooth. Fix $\alpha \in (0,1)$. Since u is continuous on $\overline{B(x,\alpha\delta_x)}$, $\operatorname{osc}_{B(x,\alpha\delta_x)} u < \infty$. Let $x = (b_1, b_2, a) \in D$, $\tilde{x} = (\tilde{b}_1, \tilde{b}_2, \tilde{a}) \in D$ such that $\rho(x, \tilde{x}) < \alpha\delta_x/32$, $|\mathbf{b} - \tilde{\mathbf{b}}| \leqslant 1$ and $|a - \tilde{a} + b_1\tilde{b}_2 - b_2\tilde{b}_1| \leqslant 1/2$. Consider the coupling from Theorem 2.2.1 of two Brownian motions, $\mathbf{X}$ and $\widetilde{\mathbf{X}}$, on the Heisenberg group starting from the points x and $\tilde{x}$ respectively.

By Theorem 3.2.1, Remark 3.2.2 and the equivalence of the Carnot-Carathéodory

metric d_{CC} and the pseudo-metric ρ, we have

$$
\mathbb{P}\left(\tau > \tau_{B(x,\alpha\delta_x)}\left(\mathbf{X}\right) \wedge \widetilde{\tau}_{B(x,\alpha\delta_x)}\left(\widetilde{\mathbf{X}}\right)\right)
$$
$$
\leqslant C\left(1 + \frac{1}{\alpha\delta_x} + \frac{1}{(\alpha\delta_x)^4} + \frac{(1+\alpha\delta_x)^3}{(\alpha\delta_x)^4}\right) d_{CC}\left(x,\widetilde{x}\right),
$$

where C is a constant independent of $x, \tilde{x}, u, \delta_x, D$ and α.

Using the coupling from Theorem 2.2.1 and Itô's formula we have that

$$
\left|u\left(x\right) - u\left(\widetilde{x}\right)\right| = \left|\mathbb{E}\left[u\left(\mathbf{X}_{\tau_{B(x,\alpha\delta_x)}(\mathbf{X})}\right) - u\left(\widetilde{\mathbf{X}}_{\widetilde{\tau}_{B(x,\alpha\delta_x)}(\widetilde{\mathbf{X}})}\right)\right]\right|
$$
$$
\leqslant \mathbb{E}\left[\left|u\left(\mathbf{X}_{\tau_{B(x,\alpha\delta_x)}(\mathbf{X})}\right) - u\left(\widetilde{\mathbf{X}}_{\widetilde{\tau}_{B(x,\alpha\delta_x)}(\widetilde{\mathbf{X}})}\right)\right|\right]
$$
$$
\leqslant \left(\operatorname{osc}_{B(x,\alpha\delta_x)} u\right) \cdot \mathbb{P}\left(\tau > \tau_{B(x,\alpha\delta_x)}\left(\mathbf{X}\right) \wedge \widetilde{\tau}_{B(x,\alpha\delta_x)}\left(\widetilde{\mathbf{X}}\right)\right)
$$
$$
\leqslant C\left(\operatorname{osc}_{B(x,\alpha\delta_x)} u\right)\left(1 + \frac{1}{\alpha\delta_x} + \frac{1}{(\alpha\delta_x)^4} + \frac{(1+\alpha\delta_x)^3}{(\alpha\delta_x)^4}\right) d_{CC}\left(x,\widetilde{x}\right).
$$

Since u is continuously differentiable on $\overline{B\left(x,\alpha\delta_x\right)}$, (3.2.20) holds for every $x \in D$. Dividing out by $d_{CC}\left(x,\widetilde{x}\right)$ and using (3.2.20) we have that for every $x \in D$,

$$
\left\|\nabla_{\mathcal{H}}u(x)\right\|_{\mathcal{H}} \leqslant \left|\nabla_{\mathcal{H}}u\right|\left(x\right) = \lim_{r\downarrow 0} \sup_{0 < d_{CC}(x,\tilde{x}) \leqslant r} \frac{\left|u\left(x\right) - u\left(\widetilde{x}\right)\right|}{d_{CC}\left(x,\widetilde{x}\right)}
$$
$$
\leqslant C\left(1 + \frac{1}{\alpha\delta_x} + \frac{1}{(\alpha\delta_x)^4} + \frac{(1+\alpha\delta_x)^3}{(\alpha\delta_x)^4}\right) \operatorname{osc}_{B(x,\alpha\delta_x)} u,
$$

as needed. $\qquad\square$

3.3 Applications of the gradient estimate

In this section, we prove some corollaries and applications of the gradient estimate in Theorem 3.2.3. One of which is the well known Cheng-Yau estimate as given in [CY75, Yau75]. We also prove a Caccioppoli-type inequality on $\mathbb{H}^3$.

Corollary 3.3.1 ([BGM16])**.** Let u be a non-negative solution to $\Delta_{\mathcal{H}} u = 0$ on $D \subset \mathbb{H}^3$. There exists a constant $C > 0$ that does not depend on u, δ_x, x, D such that

$$\|\nabla_{\mathcal{H}} u(x)\|_{\mathcal{H}} \leqslant |\nabla_{\mathcal{H}} u|(x) \leqslant C \left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1 + \delta_x)^3}{\delta_x^4} \right) u(x)$$

for every $x \in D$.

Proof. By [BLU07, Corollary 5.7.3] we have the following Harnack inequality

$$\sup_{B(x,\alpha^* \delta_x)} u \leqslant C \inf_{B(x,\alpha^* \delta_x)} u, \tag{3.3.1}$$

for $x \in D \subset \mathbb{H}^3$, where $\alpha^* \in (0,1)$ and $C > 0$ are constants not depending on u, x, δ_x, D. Using equations (3.2.21) and (3.3.1) and absorbing α^* in C gives the desired result. $\qquad\square$

Let us recall that we say a function $u : D \to \mathbb{R}$ is *harmonic* on $D \subset \mathbb{H}^3$ if $\Delta_{\mathcal{H}} u = 0$ on D. We can use Corollary 3.3.1 and the stratified structure of $\mathbb{H}^3$ to prove the Cheng-Yau gradient estimate. In particular, this recovers the fact that non-negative harmonic functions on the Heisenberg group must be constant. We thank F. Baudoin for pointing out the connection between the gradient estimate in Corollary 3.3.1 and the Cheng-Yau inequality.

Corollary 3.3.2 ([BGM16]). If u is any positive harmonic function in a ball $B(x_0, 2r) \subset$ $\mathbb{H}^3$, then there exists a universal constant $C > 0$ not dependent on u and x_0 such that

$$\sup_{x \in B(x_0, r)} \|\nabla_{\mathcal{H}} \log u(x)\|_{\mathcal{H}} \leqslant \frac{C}{r}.$$

Moreover, if u is any positive harmonic function on $\mathbb{H}^3$, then u must be a constant.

Proof. Suppose $u > 0$ is harmonic in $B(0, 2)$. Writing $\delta_x = \rho(x, (B(0, 2))^c)$ for $x \in B(0, 2)$, we obtain by Corollary 3.3.1

$$\frac{\|\nabla_{\mathcal{H}} u(x)\|_{\mathcal{H}}}{u(x)} \leqslant C' = C \sup_{x \in B(0,1)} \left(1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1 + \delta_x)^3}{\delta_x^4}\right), \quad x \in B(0, 1), \quad (3.3.2)$$

where C is the same constant as in Corollary 3.3.1. This implies that

$$\sup_{B(0,1)} \|\nabla_{\mathcal{H}} \log u(x)\|_{\mathcal{H}} \leqslant C'. \tag{3.3.3}$$

Now suppose that $u > 0$ is harmonic in $B(x_0, 2r)$ for $r > 0$. By left invariance and the dilation properties of $\mathbb{H}^3$ we see that (3.3.3) implies

$$\sup_{B(x_0, r)} \|\nabla_{\mathcal{H}} \log u(x)\|_{\mathcal{H}} \leqslant \frac{C'}{r}.$$

If u is harmonic on all of $\mathbb{H}^3$, taking $r \to \infty$ gives us that u must be constant. $\square$

We refer the reader to [CGL93] for a Caccioppoli-type inequality similar to the one below.

Corollary 3.3.3 (Caccioppoli-type inequality). Take $u \in L^2_{Loc}(B(x, R))$ such that

$\Delta_{\mathbb{H}} u = 0$ and $u > 0$ on $B(x,R) \subset \mathbb{H}$. We have that for all $r < R$,

$$\int_{B(x,r)} \|\nabla_{\mathcal{H}} u(y)\|_{\mathcal{H}}^2 \, dy \leqslant C(r,R)^2 \int_{B(x,r)} u(y)^2 dy,$$

where

$$C(r,R) = c_1 \left(1 + \frac{1}{R-r} + \frac{1}{(R-r)^4} + \frac{(1+R-r)^3}{(R-r)^4} \right).$$

Proof. Consider $u \in L^2_{Loc}(B(x,R))$. Let $\delta_{y,D} = dist(D^c, y)$. Fix $x \in \mathbb{H}$ and consider $y \in B(x,r)$. Using the gradient estimate from Corollary 3.3.1 we have

$$\|\nabla_{\mathcal{H}} u(y)\|_{\mathcal{H}} \leqslant C \left(1 + \frac{1}{\delta_{y,D}} + \frac{1}{(\delta_{y,D})^4} + \frac{(1+\delta_{y,D})^3}{(\delta_{y,D})^4} \right) u(y).$$

If we use $D = B\left(y, \frac{R-r}{2}\right)$ in $\delta_{y,D}$ then

$$\|\nabla_{\mathcal{H}} u(y)\|_{\mathcal{H}} \leqslant c_1 \left(1 + \frac{1}{R-r} + \frac{1}{(R-r)^4} + \frac{(1+R-r)^3}{(R-r)^4} \right) u(y).$$

Integrating both sides we have that

$$
\begin{aligned}
\int_{B(x,r)} \|\nabla_{\mathcal{H}} u(y)\|_{\mathcal{H}}^2 \, dy &\leqslant \int_{B(x,r)} c_1^2 \left(1 + \frac{1}{R-r} + \frac{1}{(R-r)^4} + \frac{(1+R-r)^3}{(R-r)^4} \right)^2 u(y)^2 dy \\
&\leqslant C(r,R)^2 \int_{B(x,r)} u(y)^2 dy,
\end{aligned}
$$

where

$$C(r,R) = c_1 \left(1 + \frac{1}{R-r} + \frac{1}{(R-r)^4} + \frac{(1+R-r)^3}{(R-r)^4} \right).$$

$\square$

Chapter 4

Coupling techniques for heat semigroups

In this chapter, we will start by presenting new proofs to known results on the gradient bounds for heat semigroup on $\mathbb{R}^n$ and on the Heisenberg group $\mathbb{H}$. We will also present a new gradient bound for the Kolmogorov diffusion. The result will be improved in a later chapter (see Proposition 5.1.10). These new proofs will employ the use of coupling techniques. The coupling we use is synchronous coupling of Brownian motions. The techniques shown in this chapter will be the motivation for the coupling techniques used in Chapters 5 and 7.

We start by recalling the notion of a coupling. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and B_t and $\tilde{B}_t$ are two Brownian motions in $\mathbb{R}^n$ defined on this space, starting at $x, \tilde{x} \in \mathbb{R}^n$ respectively. By their coupling we mean a diffusion $\left(B_t, \widetilde{B}_t\right)$ in $\mathbb{R}^n \times \mathbb{R}^n$ such that its law is a coupling of the laws of B_t and $\widetilde{B}_t$. By that we mean that the first and second $n-$dimensional marginal distributions of $\left(B_t, \widetilde{B}_t\right)$ are given by distributions of B_t and $\widetilde{B}_t$.

Let $\mathbb{P}^{(x,\tilde{x})}$ be the distributions of $\left(B_t, \widetilde{B}_t\right)$, so that $\mathbb{P}^{(x,\tilde{x})}\left(B_0 = x, \tilde{B}_0 = \tilde{x}\right) = 1$. We denote by $\mathbb{E}^{(x,\tilde{x})}$ the expectation with respect to the probability measure $\mathbb{P}^{(x,\tilde{x})}$.

Definition 4.0.1. We say that a coupling $\left(B_t, \widetilde{B}_t\right)$ in $\mathbb{R}^n \times \mathbb{R}^n$ is a *synchronous* coupling if for $(x, \tilde{x}) \in \mathbb{R}^n \times \mathbb{R}^n$ we let

$$B_t^x = x + B_t,$$

$$\widetilde{B}_t^{\tilde{x}} = \tilde{x} + B_t,$$

where B_t is a standard Brownian motion in $\mathbb{R}^n$.

In particular here we have that synchronous coupling is a fixed distance coupling.

4.1 Heat kernel functional inequalities on Euclidean space

Let P_t be the heat semigroup for the Laplace operator Δ in $\mathbb{R}^n$. Recall that the Euclidean heat kernel is given by

$$p_t(x, y) = \frac{1}{(4\pi t)^{n/2}} e^{-\|y-x\|^2/(4t)},$$

so that

$$P_t f(x) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} f(y) e^{-\|y-x\|^2/(4t)} dy.$$

The following result easily follows from properties of convolutions and Jensen's inequality.

Proposition 4.1.1. *Let P_t be the heat semigroup on $\mathbb{R}^n$. Suppose $f \in C^1(\mathbb{R}^n)$ is*

bounded and has bounded first derivatives then

$$\|\nabla P_t f\|^q \leq P_t\left(\|\nabla f\|^q\right),$$

for all $q \geq 1$.

Proof. We know that

$$
\begin{aligned}
P_t f(x) &= \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^m} f(y) e^{-\|y-x\|^2/(4t)} dy \\
&= \int_{\mathbb{R}^m} f(y) K_t(x-y)\, dy \\
&= (f \star K_t)(x)
\end{aligned}
$$

where

$$K_t(x) = \frac{1}{(4\pi t)^{m/2}} e^{-\|x\|^2/(4t)}.$$

By a property of convolutions we have that

$$\frac{\partial}{\partial x_i}(f \star K_t) = \frac{\partial f}{\partial x_i} \star K_t,$$

so that

$$\nabla P_t f = P_t\left(\nabla f\right).$$

Hence

$$
\begin{aligned}
|\nabla P_t f|^p &= |P_t(\nabla f)|^p \\
&\leqslant (P_t(|\nabla f|))^p \\
&= \left(\int |\nabla f(x)| \, K_t(x-y) \, dx \right)^p \\
&\leqslant \int |\nabla f(x)|^p \, K_t(x-y) \, dx = P_t(|\nabla f|^p)
\end{aligned}
$$

where we used Jensen's inequality. $\qquad \square$

We now show that the same result can be proven using synchronous coupling under slightly different assumptions. The technique shown in the next proof will be the motivation for results in later chapters.

Proposition 4.1.2. *Let P_t be the heat semigroup on $\mathbb{R}^n$. Synchronous coupling implies that for all $f \in C^2(\mathbb{R}^n)$ with bounded seconded derivatives we have*

$$
\|\nabla P_t f\|^q \leqslant P_t(\|\nabla f\|^q),
$$

for all $q \geq 1$.

Proof. Consider two copies of synchronously coupled Brownian motions

$$
\begin{aligned}
X_t &= B_t^p = p + B_t, \\
\widetilde{X}_t &= \widetilde{B_t^p} = \widetilde{p} + B_t,
\end{aligned}
$$

where B_t is standard Brownian motion. Note that

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

for all $t \geqslant 0$. Recall that for a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(\mathbf{x}) \approx L(\mathbf{x}) = f(\tilde{\mathbf{x}}) + \nabla f(\tilde{\mathbf{x}}) \cdot (\mathbf{x} - \tilde{\mathbf{x}}).$$

More precisely

$$|f(\mathbf{x}) - f(\tilde{\mathbf{x}})| \leqslant |\nabla f(\tilde{\mathbf{x}}) \cdot (\mathbf{x} - \tilde{\mathbf{x}})| + |R|.$$

where

$$|R| \leqslant \frac{1}{2} C_f \left(\sum_{i=1}^{n} |x_i - \tilde{x}_i|^2 \right),$$

and C_f is a bounded on the Hessian of f.

By using an estimate on the remainder R of Taylor's approximation to f and the assumption that $f \in C^2(\mathbb{R}^d)$ has bounded second derivatives, there exists a $C_f \geqslant 0$ such that

$$
\begin{aligned}
&\left| f(X_t) - f\left(\widetilde{X_t}\right) \right| \\
&= \left| \sum_{i=1}^{d} \partial_{p_i} f\left(\widetilde{X_t}\right) (p_i - \widetilde{p}_i) + R\left(\widetilde{X_t}\right) \right| \\
&\leqslant \sum_{i=1}^{d} \left| \partial_{p_i} f\left(\widetilde{X_t}\right) \right| |p - \widetilde{p}| + \frac{C_f}{2} n \, |p - \widetilde{p}|^2.
\end{aligned}
$$

Using this estimate and Jensen's inequality we see that

$$
\begin{aligned}
|P_t f(p) - P_t f(\widetilde{p})| &= \left| \mathbb{E}^{(p,\widetilde{p})} \left[f(X_t) - f\left(\widetilde{X_t}\right) \right] \right| \\
&\leqslant \mathbb{E}^{(p,\widetilde{p})} \left[\left| f(X_t) - f\left(\widetilde{X_t}\right) \right| \right] \\
&\leqslant \sum_{i=1}^{d} \mathbb{E}^{(p,\widetilde{p})} \left[\left| \partial_{p_i} f\left(\widetilde{X_t}\right) \right|^q \right]^{\frac{1}{q}} |p - \widetilde{p}| + \frac{C_f}{2} n |p - \widetilde{p}|^2 \\
&= \sum_{i=1}^{d} \left(P_t \left(|\partial_{p_i} f|^q \right)(\widetilde{p}) \right)^{\frac{1}{q}} |p - \widetilde{p}| + \frac{C_f}{2} n |p - \widetilde{p}|^2 .
\end{aligned}
$$

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

$$
\begin{aligned}
\|\nabla_p P_t f(p)\| &= \limsup_{\widetilde{p} \to p} \frac{|P_t f(p) - P_t f(\widetilde{p})|}{|p - \widetilde{p}|} \\
&\leqslant \sum_{i=1}^{d} \left(P_t \left(|\partial_{p_i} f|^q \right)(\widetilde{p}) \right)^{\frac{1}{q}} ,
\end{aligned}
$$

which proves the statement. $\qquad\square$

Proposition 4.1.3 (Local reverse Poincaré inequality). Let P_t be the heat semigroup on $\mathbb{R}^n$. Synchronous coupling implies that if $f \in C^2(\mathbb{R}^n)$ with bounded second derivatives then

$$
2t \nabla |P_t f|^2 \leqslant P_t f^2 - (P_t f)^2 .
$$

Proof. We synchronously couple $\left(B_t^x, B_t^{\widetilde{x}} \right)$. Let C be a bounded on the second deriva-

tives of f. First we compute

$$
\begin{aligned}
|P_t f(x) - P_t f(\tilde{x})|^2 &= |P_s P_{t-s} f(x) - P_s P_{t-s} f(\tilde{x})|^2 \\
&= \left| \mathbb{E}^{(x,\tilde{x})} \left[P_{t-s} f(B_s^x) - P_{t-s} f(B_s^{\tilde{x}}) \right] \right|^2 \\
&\leq \mathbb{E}^{(x,\tilde{x})} \left[\left| P_{t-s} f(B_s^x) - P_{t-s} f(B_s^{\tilde{x}}) \right|^2 \right] \\
&\leq \mathbb{E}^{(x,\tilde{x})} \left[\left| \left| \nabla P_{t-s} f(B_s^{\tilde{x}}) \right| \left| B_s - \tilde{B}_s \right| + C \left| B_s - \tilde{B}_s \right|^2 \right|^2 \right] \\
&= \mathbb{E}^{(x,\tilde{x})} \left[\left| \nabla P_{t-s}(B_s^{\tilde{x}}) \right|^2 \left| B_s - \tilde{B}_s \right|^2 + 2 \left| \nabla P_{t-s}(B_s^{\tilde{x}}) \right| \left| B_s - \tilde{B}_s \right|^3 \right] \\
&\quad + C \mathbb{E}^{(x,\tilde{x})} \left[\left| B_s - \tilde{B}_s \right|^4 \right] .
\end{aligned}
$$

Integrating both sides with respect to s from $s = 0$ and $s = t$ we have

$$
\begin{aligned}
&|P_t f(x) - P_t f(\tilde{x})|^2 t \\
&\leqslant \int_0^t \mathbb{E}^{(x,\tilde{x})} \left[\left| \nabla P_{t-s}(B_s^{\tilde{x}}) \right|^2 \left| B_s - \tilde{B}_s \right|^2 + 2 \left| \nabla P_{t-s}(B_s^{\tilde{x}}) \right| \left| B_s - \tilde{B}_s \right|^3 + C \left| B_s - \tilde{B}_s \right|^4 \right] ds \\
&= \int_0^t \mathbb{E}^{(x,\tilde{x})} \left[\left| \nabla P_{t-s}(B_s^{\tilde{x}}) \right|^2 |x - \tilde{x}|^2 + 2 \left| \nabla P_{t-s}(B_s^{\tilde{x}}) \right| |x - \tilde{x}|^3 + C |x - \tilde{x}|^4 \right] ds \\
&= \int_0^t \left[P_s \left(|\nabla P_{t-s}(\tilde{x})|^2 \right) |x - \tilde{x}|^2 + 2 P_s \left(|\nabla P_{t-s}(B_s^{\tilde{x}})| \right) |x - \tilde{x}|^3 + C |x - \tilde{x}|^4 \right] ds \\
&= \frac{1}{2} \left(P_t f^2(\tilde{x}) - (P_t f(\tilde{x}))^2 \right) |x - \tilde{x}|^2 + 2 \int_0^t \left[P_s \left(|\nabla P_{t-s}(\tilde{x})| \right) \right] ds \, |x - \tilde{x}|^3 + C t \, |x - \tilde{x}|^4
\end{aligned}
$$

Dividing out by $|x - \tilde{x}|^2$ we have

$$
\begin{aligned}
\frac{|P_t f(x) - P_t f(\tilde{x})|^2}{|x - \tilde{x}|^2} t \quad \leqslant \quad & \frac{1}{2} \left(P_t f^2(\tilde{x}) - (P_t f(\tilde{x}))^2 \right) \\
& + 2 \int_0^t \left[P_s \left(|\nabla P_{t-s}(\tilde{x})| \right) \right] ds \, |x - \tilde{x}| + C t \, |x - \tilde{x}|^2 .
\end{aligned}
$$

Taking $\tilde{x} \to x$ we have the desired result. $\qquad\qquad\qquad\qquad\square$

4.2 Heat kernel gradient estimates for the Kolmogorov diffusion

We now consider the Kolmogorov operator

$$L = \frac{1}{2}\frac{\partial^2}{\partial x^2} + x\frac{\partial}{\partial y}.$$

This operator satisfies the weak *Hörmander's condition* condition since the vector fields $\left\{\frac{\partial}{\partial x}, x\frac{\partial}{\partial y}\right\}$ and its lie brackets span $\mathbb{R}^2$. Thus L is a hypoelliptic operator. Its corresponding *carré du champ* operator is $\Gamma(f, f) = \left(\frac{\partial f}{\partial x}\right)^2$. Its corresponding diffusion process stated at (x, y) is

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

where B_t is a standard Brownian motion. Our goal in this section is to illustrate the use of the coupling technique in proving gradient estimates on the heat semigroup of a hypoelliptic diffusion. In particular we prove a sharp Driver-Melcher type inequality. We refer the reader to Section 1.1.6 for an introduction on the Kolmogorov diffusion and the relevant functional inequalities.

Theorem 4.2.1. Let P_t be the heat semigroup for the Kolmogorov diffusion in $\mathbb{R}^2$. Suppose $f \in C^\infty(\mathbb{R}^2)$ such that $\left|\frac{\partial^2 f}{\partial x^2}\right| \leqslant M$ for some $M > 0$. We have that for all $1 \leqslant p < \infty$

$$\sqrt{\Gamma(P_t f, P_t f)} \leqslant \left(P_t\left(\Gamma(f, f)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}} + t\left|\frac{\partial}{\partial y}P_t f\right|. \qquad (4.2.1)$$

Proof. Recall that $f(x) - f(a) = f'(a)(x - a) + E$ where there exists a ξ between x and a such that

$$|E| \leqslant \frac{f''(\xi_l)}{2!} |x - a|^2 .$$

We use use the following estimate for a multivariable function $f(x_1, x_2) = f_{x_2}(x_1)$ so that

$$\left| f_{x_2}(x_1) - f_{x_2}(\tilde{x}_1) \right| \; \leqslant \; \left| f'_{x_2}(\tilde{x}_1) \right| |x_1 - \tilde{x}_1| + \frac{1}{2} \left| f''_{x_2}(\xi_{x_2}) \right| |x_1 - \tilde{x}_1|^2$$

where ξ_{x_2} is between x_1 and $\tilde{x}_1$. We consider the following coupling of Kolmogorov diffusions. Let $\mathbf{X}_t$ be a Kolmogorov diffusion started at (x_1, x_2) so that

$$\begin{aligned}
\mathbf{X}_t &= \left(X_t^1, X_t^2 \right) \\
&= \left(x_1 + B_t, x_2 + tx_1 + \int_0^t B_t ds \right)
\end{aligned}$$

where B_t is a brownian motion started at 0. Let $t_0 > 0$ be fixed time. Define

$$\begin{aligned}
\tilde{\mathbf{X}}_t &= \left(\tilde{X}_t^1, \tilde{X}_t^2 \right) \\
&= \left(\tilde{x}_1 + B_t, x_2 + t_0 (x_1 - \tilde{x}_1) + t\tilde{x}_1 + \int_0^t B_t ds \right) .
\end{aligned}$$

Now $\mathbf{X}_t$ is a Kolmogorov diffusion started at the point $(\tilde{x}_1, x_2 + t_0 (x_1 - \tilde{x}_1))$. Note that $\left| X_t^1 - \tilde{X}_t^1 \right| = |x_1 - \tilde{x}_1|$ for all time t, and at $t = t_0$ we have that $X_{t_0}^2 = \tilde{X}_{t_0}^2$. We will use the latter to apply the one dimensional Taylor formula to a two-variable

79

function as discussed above. Thus at a fixed time t_0, we have that

$$
\begin{aligned}
\left| f\left(\mathbf{X}_{t_0}^1\right) - f\left(\tilde{\mathbf{X}}_{t_0}\right) \right| &\leqslant \left| \partial_x f\left(\tilde{\mathbf{X}}_{t_0}\right) \right| \left| \tilde{X}_{t_0}^1 - \tilde{X}_{t_0}^2 \right| + \frac{M}{2} \left| \tilde{X}_t^1 - \tilde{X}_t^2 \right|^2 \\
&= \left| \partial_x f\left(\tilde{\mathbf{X}}_{t_0}\right) \right| \left| x_1 - \tilde{x}_1 \right| + \frac{M}{2} \left| x_1 - \tilde{x}_1 \right|^2
\end{aligned}
$$

Now

$$
\begin{aligned}
|P_t f(x_1, x_2) - P_t f\left(\tilde{x}_1, x_2 + t_0\left(x_1 - \tilde{x}_1\right)\right)| &= \left| \mathbb{E}^{((x_1,x_2),(\tilde{x}_1,x_2+t_0(x_1-\tilde{x}_1)))} \left[f\left(\mathbf{X}_t\right) - f\left(\tilde{\mathbf{X}}_t\right) \right] \right| \\
&\leqslant \mathbb{E}^{((x_1,x_2),(\tilde{x}_1,x_2+t_0(x_1-\tilde{x}_1)))} \left[\left| f\left(\mathbf{X}_t\right) - f\left(\tilde{\mathbf{X}}_t\right) \right| \right].
\end{aligned}
$$

At $t = t_0$ we have that for $p \geq 1$

$$
\begin{aligned}
&|P_{t_0} f(x_1, x_2) - P_{t_0} f\left(\tilde{x}_1, x_2 + t_0\left(x_1 - \tilde{x}_1\right)\right)| \\
&\leqslant \mathbb{E}^{((x_1,x_2),(\tilde{x}_1,x_2+t_0(x_1-\tilde{x}_1)))} \left[\left| f\left(\mathbf{X}_{t_0}\right) - f\left(\tilde{\mathbf{X}}_{t_0}\right) \right| \right] \\
&\leqslant \mathbb{E}^{((\tilde{x}_1,x_2+t_0(x_1-\tilde{x}_1)))} \left[\left| \partial_x f\left(\tilde{\mathbf{X}}_{t_0}\right) \right| \right] |x_1 - \tilde{x}_1| + \frac{M}{2} |x_1 - \tilde{x}_1|^2 \\
&\leqslant \left(\mathbb{E}^{((\tilde{x}_1,x_2+t_0(x_1-\tilde{x}_1)))} \left[\left| \partial_x f\left(\tilde{\mathbf{X}}_{t_0}\right) \right|^p \right] \right)^{\frac{1}{p}} |x_1 - \tilde{x}_1| + \frac{M}{2} |x_1 - \tilde{x}_1|^2 \\
&= (P_{t_0} \left(|\partial_x f\left(\tilde{x}_1, x_2 + t_0\left(x_1 - \tilde{x}_1\right)\right)|^p\right))^{\frac{1}{p}} |x_1 - \tilde{x}_1| + \frac{M}{2} |x_1 - \tilde{x}_1|^2 ,
\end{aligned}
$$

so that

$$
\begin{aligned}
\frac{|P_{t_0} f(x_1, x_2) - P_{t_0} f\left(\tilde{x}_1, x_2 + t_0\left(x_1 - \tilde{x}_1\right)\right)|}{|x_1 - \tilde{x}_1|} &\leqslant \\
\leqslant (P_{t_0} \left(|\partial_x f\left(\tilde{x}_1, x_2 + t_0\left(x_1 - \tilde{x}_1\right)\right)|^p\right))^{\frac{1}{p}} + \frac{M}{2} |x_1 - \tilde{x}_1| .
\end{aligned}
$$

Now taking $\tilde{x}_1 \to x_1$ we have that

$$\limsup_{\tilde{x}_1 \to x_1} \frac{|P_{t_0}f(x_1, x_2) - P_{t_0}f(\tilde{x}_1, x_2 + t_0(x_1 - \tilde{x}_1))|}{|x_1 - \tilde{x}_1|} \tag{4.2.2}$$

$$\leqslant \limsup_{\tilde{x}_1 \to x_1} (P_{t_0}(|\partial_x f(\tilde{x}_1, x_2 + t_0(x_1 - \tilde{x}_1))|^p))^{\frac{1}{p}} + 0$$

$$= (P_{t_0}(|\partial_x f(x_1, x_2)|^p))^{\frac{1}{p}}$$

$$= P_{t_0}\left(\Gamma(f(x_1, x_2), f(x_1, x_2))^{\frac{p}{2}}\right)^{\frac{1}{p}} \tag{4.2.3}$$

Looking at the left side of (4.2.3) we rewrite it as

$$\limsup_{\tilde{x}_1 \to x_1} \frac{|P_{t_0}f(x_1, x_2) - P_{t_0}f(\tilde{x}_1, x_2 + t_0(x_1 - \tilde{x}_1))|}{|x_1 - \tilde{x}_1|}$$

$$= \limsup_{h \to 0} \frac{|P_{t_0}f((x_1, x_2) + h(-1, t_0)) - P_{t_0}f(x_1, x_2)|}{h}$$

$$= \limsup_{\tilde{x}_1 \to x_1} \frac{|P_{t_0}f(\tilde{x}_1, x_2 + t_0(x_1 - \tilde{x}_1)) - P_{t_0}f(x_1, x_2)|}{|x_1 - \tilde{x}_1|}$$

$$= |\nabla P_{t_0}f(x_1, x_2) \cdot \langle 1, -t_0 \rangle|$$

$$= \left| \frac{\partial}{\partial x} P_{t_0}f(x_1, x_2) - t_0 \frac{\partial}{\partial y} P_{t_0}f(x_1, x_2) \right|.$$

We have the desired result after a rearranging of terms. $\qquad \square$

Remark 4.2.2. Note that this inequality is sharp. Let $f(x, y) = y$ so that $P_t f = y + tx$. Thus $\Gamma(P_t f, P_t f) = t^2$, $\Gamma(f, f) = 0$ and $\frac{\partial f}{\partial y} = 1$. Thus the left hand side of (4.2.1) is

$$\sqrt{\Gamma(P_t f, P_t f)} = t$$

81

while the right hand side is

$$\left(P_t\left(\Gamma\left(f,f\right)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}} + t\left|\frac{\partial}{\partial y}P_t f\right| = 0 + t.$$

4.3 Heat kernel gradient estimates on the Heisenberg group

We define $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ as the unique left-invariant vector fields with $\mathcal{X}_e = \partial_x$, $\mathcal{Y}_e = \partial_y$, and $\mathcal{Z}_e = \partial_z$, so that

$$\mathcal{X} = \partial_x - y\partial_z,$$
$$\mathcal{Y} = \partial_y + x\partial_z,$$
$$\mathcal{Z} = \partial_z.$$

As pointed out in [GL16, Example 6.1], the (sum of squares) operator

$$\Delta_{\mathcal{H}} = \mathcal{X}^2 + \mathcal{Y}^2$$

is a natural sub-Laplacian for the Heisenberg group.

Let $\mathbf{X}_t$ be the Markov processes associated to $\Delta_{\mathcal{H}}$. We call this process the Brownian motion on the Heisenberg group. The semigroup associated to the Heisenberg sub-laplacian is the following

$$P_t f(\mathbf{x}) = \mathbb{E}^{\mathbf{x}}\left[f\left(\mathbf{X}_t\right)\right].$$

In [Kuw10], Kuwada showed that (1.1.10) for $f \in C_c^\infty(X)$ where X is a sub-Riemannian space implies an upper gradient estimate. By the Kuwada duality, the upper gradient estimate formulated in [Kuw10] implies the existence of a coupling $\left(\mathbf{X}_t, \tilde{\mathbf{X}}_t\right)$ of Brownian motions on the Heisenberg groups started at $(\mathbf{x}, \tilde{\mathbf{x}})$, for each $t > 0$, satisfying

$$d_{CC}\left(\mathbf{X}_t, \tilde{\mathbf{X}}_t\right) \leqslant K d\left(\mathbf{x}, \tilde{\mathbf{x}}\right). \tag{4.3.1}$$

In the following, we give a direct proof of the converse of this. That is, given a coupling satisfying (4.3.1) we can prove the Driver-Melcher gradient estimate (1.1.10). We note that the following argument also works for any Carnot group.

Theorem 4.3.1. Suppose there exists a coupling of two Brownian motions on the Heisenberg group $\left(\mathbf{X}_t, \tilde{\mathbf{X}}_t\right)$ started at $(\mathbf{x}, \tilde{\mathbf{x}})$ that satisfies

$$d_{CC}\left(\mathbf{X}_1, \tilde{\mathbf{X}}_1\right) \leqslant K d_{CC}\left(\mathbf{x}, \tilde{\mathbf{x}}\right), \text{ a.s,}$$

for some constant $K \geq 1$. This coupling implies

$$|\nabla_{\mathbb{H}} P_t f|^p \leqslant K^p P_t\left(|\nabla_{\mathbb{H}} f|^p\right),$$

for all $p \geq 1$.

Proof. By the dilations in the group $\mathbb{H}$ it suffices to prove

$$|\nabla_{\mathbb{H}} P_t f|^p \leqslant K_p P_t\left(|\nabla_{\mathbb{H}} f|^p\right),$$

for $t = 1$ (See [DM05] and [BBBC08]). Let $\left(\mathbf{X}_t, \tilde{\mathbf{X}}_t\right)$ be a coupling of two Brownian

motions on the Heisenberg group started at $(\mathbf{x}, \tilde{\mathbf{x}})$ satisfying

$$d_{CC}\left(\mathbf{X}_1, \tilde{\mathbf{X}}_1\right) \leqslant K d_{CC}\left(\mathbf{x}, \tilde{\mathbf{x}}\right), \text{ a.s.} \tag{4.3.2}$$

If $y = (y_1, y_2, y_3)$ and $y = (\tilde{y}_1, \tilde{y}_2, \tilde{y}_3)$ then $P_1\left(f, \tilde{y}\right)(y) = f(\tilde{y}) + \nabla_{\mathbb{H}} f(\tilde{y}) \cdot \begin{pmatrix} y_1 - \tilde{y}_1 \\ y_2 - \tilde{y}_2 \end{pmatrix}$
is the first order $\mathbb{H}-$Taylor polynomial. The following estimate is given in (Theorem 20.3.2, [BLU07]) and is an analogue to Taylor's inequality for Carnot groups:

$$|f(y) - P_1(f, \tilde{y})(y)| \leqslant c_1 d_{CC}(y, \tilde{y})$$
$$\times \sup_{d(z,e) \leqslant b^2 d(\tilde{y}^{-1} \star y, e)} \left\{ \left|\mathcal{X}^2 u\right|(\tilde{y} \star z), \left|\mathcal{Y}^2 u\right|(\tilde{y} \star z), |\mathcal{X}\mathcal{Y} u|(\tilde{y} \star z) \right\}.$$

Thus we have that

$$|f(y) - f(\tilde{y})| \leqslant \left| \nabla_{\mathbb{H}} f(\tilde{y}) \cdot \begin{pmatrix} y_1 - \tilde{y}_1 \\ y_2 - \tilde{y}_2 \end{pmatrix} \right| + \left| f(y) - f(\tilde{y}) - \nabla_{\mathbb{H}} f(\tilde{y}) \cdot \begin{pmatrix} y_1 - \tilde{y}_1 \\ y_2 - \tilde{y}_2 \end{pmatrix} \right|$$

$$\leqslant |\nabla_{\mathbb{H}} f(\tilde{y})| \left| \begin{pmatrix} y_1 - \tilde{y}_1 \\ y_2 - \tilde{y}_2 \end{pmatrix} \right|$$

$$+ c_1 d_{CC}(y, \tilde{y})^2 \sup_{d(z,e) \leqslant b^2 d(\tilde{y}^{-1} \star y, e)} \left\{ \left|\mathcal{X}^2 u\right|(\tilde{y} \star z), \left|\mathcal{Y}^2 u\right|(\tilde{y} \star z), |\mathcal{X}\mathcal{Y} u|(\tilde{y} \star z) \right\}$$

$$\tag{4.3.3}$$

Using estimate (4.3.3) with $y = \mathbf{X}_1$ and $\tilde{y} = \tilde{\mathbf{X}}_1$ we have that

$$|P_1 f(\mathbf{x}) - P_1 f(\tilde{\mathbf{x}})|$$

$$= \left| \mathbb{E}^{\mathbf{x}} f\left(\mathbf{X}_1\right) - \mathbb{E}^{\tilde{\mathbf{x}}} f\left(\tilde{\mathbf{X}}_1\right) \right|$$

$$\leqslant \mathbb{E}^{(x,\tilde{x})} \left[\left| f\left(\mathbf{X}_1\right) - f\left(\tilde{\mathbf{X}}_1\right) \right| \right]$$

$$\leqslant \mathbb{E} \left[\left| \nabla_{\mathbb{H}} f(\tilde{\mathbf{X}}_1) \right| \cdot \left| \begin{pmatrix} b_1 - \tilde{b}_1 \\ b_2 - \tilde{b}_2 \end{pmatrix} \right| \right]$$

$$+ c_1 \mathbb{E} \left[d_{CC}\left(\mathbf{X}_1, \tilde{\mathbf{X}}_1\right)^2 \right. \tag{4.3.4}$$

$$\times \sup_{d(z,e) \leqslant b^2 d_{CC}(\mathbf{X}_1, \tilde{\mathbf{X}}_1)} \left\{ \left| \mathcal{X}^2 f \right| \left(\tilde{\mathbf{X}}_1 \star z\right), \left| \mathcal{Y}^2 f \right| \left(\tilde{\mathbf{X}}_1 \star z\right), \left| \mathcal{X}\mathcal{Y} f \right| \left(\tilde{\mathbf{X}}_1 \star z\right) \right\} \left. \right] \tag{4.3.5}$$

Recall that $|(x, y, z)|_{\mathbb{H}} = \left((x^2 + y^2)^2 + z^2 \right)^{\frac{1}{4}}$ is a homogeneous norm on $\mathbb{H}$ that is equivalent to $d_{CC}\left((x, y, z), e\right)$. By Lemma A.0.8 it is easy to see that

$$\left| \begin{pmatrix} b_1 - \tilde{b}_1 \\ b_2 - \tilde{b}_2 \end{pmatrix} \right| \leqslant d_{CC}\left(\mathbf{x}, \tilde{\mathbf{x}}\right). \tag{4.3.6}$$

Plugging (4.3.2) and (4.3.6) into (4.3.5) we have that

$$|P_1 f(\mathbf{x}) - P_1 f(\tilde{\mathbf{x}})|$$

$$\leqslant K d_{CC}(\mathbf{x}, \tilde{\mathbf{x}}) \mathbb{E}\left[\left\|\nabla_{\mathbb{H}} f(\tilde{\mathbf{X}}_1)\right\|\right]$$

$$+ c_1 d_{CC}(\mathbf{x}, \tilde{\mathbf{x}})^2 \mathbb{E}\left[\sup_{d(z,e)\leqslant b^2 d_{CC}(\mathbf{X}_1,\tilde{\mathbf{X}}_1)} \left\{\left|\mathcal{X}^2 f\right|\left(\tilde{\mathbf{X}}_1 \star z\right), \left|\mathcal{Y}^2 f\right|\left(\tilde{\mathbf{X}}_1 \star z\right), \left|\mathcal{X}\mathcal{Y} f\right|\left(\tilde{\mathbf{X}}_1 \star z\right)\right\}\right]$$

$$\leqslant K d_{CC}(\mathbf{x}, \tilde{\mathbf{x}}) \left(P_1\left(|\nabla_{\mathbb{H}} f(\tilde{\mathbf{x}})|^p\right)\right)^{\frac{1}{p}}$$

$$+ c_1 d_{CC}(\mathbf{x}, \tilde{\mathbf{x}})^2 \mathbb{E}\left[\sup_{d(z,e)\leqslant b^2 d_{CC}(\mathbf{X}_1,\tilde{\mathbf{X}}_1)} \left\{\left|\mathcal{X}^2 f\right|\left(\tilde{\mathbf{X}}_1 \star z\right), \left|\mathcal{Y}^2 f\right|\left(\tilde{\mathbf{X}}_1 \star z\right), \left|\mathcal{X}\mathcal{Y} f\right|\left(\tilde{\mathbf{X}}_1 \star z\right)\right\}\right],$$

$$(4.3.7)$$

where we used Jensen's inequality in the last line for $p \geq 1$. Now since $f \in C^\infty(\mathbb{H})$ and has bounded first and second derivatives then we can bound the expectation on the last term by a constant $K(f) < \infty$. Thus we have

$$|P_1 f(\mathbf{x}) - P_1 f(\tilde{\mathbf{x}})| \quad \leqslant \quad K d_{CC}(\mathbf{x}, \tilde{\mathbf{x}}) \left(P_1\left(|\nabla_{\mathbb{H}} f(\tilde{\mathbf{x}})|^p\right)\right)^{\frac{1}{p}}$$

$$+ c_1 \cdot K(f) d_{CC}(\mathbf{x}, \tilde{\mathbf{x}})^2.$$

Dividing out by $d_{CC}(\mathbf{x}, \tilde{\mathbf{x}})$ and taking $\tilde{\mathbf{x}} \to \mathbf{x}$ we have that

$$|\nabla_{\mathbb{H}} P_1 f(\mathbf{x})| \quad = \quad \limsup_{\tilde{\mathbf{x}} \to \mathbf{x}} \frac{|P_1 f(\mathbf{x}) - P_1 f(\tilde{\mathbf{x}})|}{d_{CC}(\mathbf{x}, \tilde{\mathbf{x}})}$$

$$\leqslant \quad \limsup_{\tilde{\mathbf{x}} \to \mathbf{x}} \left[K\left(P_1\left(|\nabla_{\mathbb{H}} f(\tilde{\mathbf{x}})|^p\right)\right)^{\frac{1}{p}} + c_1 \cdot K(f) d_{CC}(\mathbf{x}, \tilde{\mathbf{x}})\right]$$

$$= \quad K\left(P_1\left(|\nabla_{\mathbb{H}} f(\mathbf{x})|^p\right)\right)^{\frac{1}{p}},$$

which gives us

$$|\nabla_{\mathbb{H}} P_1 f(\mathbf{x})|^p \leqslant K^p P_1\left(|\nabla_{\mathbb{H}} f(\mathbf{x})|^p\right).$$

86

$\square$

Remark 4.3.2. In [DM05], the authors showed

$$|\nabla_{\mathbb{H}} P_t f|^p \leqslant C_p(t) P_t \left(|\nabla_{\mathbb{H}} f|^p \right)$$

on the Heisenberg group where $C_p(t)$ is the best constant. They showed that $C_p(t) = C > 1$. Specifically they showed that $C_2(t) > 2$. Note the Theorem 4.3.1 gives a characterization of this constant as $C_p(t) = K^p$. Here K is the best constant that one can obtain by finding a coupling that satisfies $d_{CC}\left(\mathbf{X}_1, \tilde{\mathbf{X}}_1\right) \leqslant K d_{CC}\left(x, y\right)$.

Chapter 5

Heat semigroup functional Inequalities for the Kolmogorov diffusion

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. [BG17, BBG14]), quasi-invariance of heat kernel measures in infinite dimensions (e.g. [BGM13, Gor17]), functional inequalities such as Poincaré and log-Sobolev type inequalities (e.g. [DM05, BB12, Kuw10, Wan16]), and the study of convergence to equilibrium for hypocoercive diffusions (e.g. [Bau17a, BT18]). In particular, the gradient bounds we present in this paper might be used to prove a spectral gap existence similarly to [BW14] 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 [Kol34], where he obtained an explicit expression for the transition density of the diffusion process whose generator is the operator L. Later L. Hörmander in [Hör67] 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 [Ham11], 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. [Bau17a, BB12, BBG14, BG17]) in the last years, the coupling techniques have seen recent progress for such degenerate operators. In [BACK95], 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 [BK16], S. Banerjee and W.Kendall used a non-Markovian strategy to couple the iterated Kolmogorov diffusion. The most relevant to our results is [BGM16], where coupling techniques have been used to prove gradient estimates on the Heisenberg group considered as a sub-Riemannian manifold.

The second part of the dissertation is organized as follows. We start by considering Kolmogorov diffusions in Section 5.1, where we use both generalized Γ-calculus and coupling techniques to prove gradient estimates such as in Proposition 5.1.5 and

Proposition 5.1.10. 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 Chapter 6. The relativistic diffusion has been introduced by R. Dudley and studied extensively in [Bai08, Ang15, Dud66, Dud67, DH09, FLJ07, FLJ12, IM15, JM07, McK63], while the history of related objects both in mathematics and physics can be found in [Dun08]. Observe that generalized Γ-calculus gives relatively simple proofs of functional inequalities for the relativistic diffusion compared to previous results.

In Section 7.0.3 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 [vR04]. 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 [GL17]. If such a coupling can be constructed, then our results and techniques would be valid for even a larger class of hypoelliptic diffusions.

5.1 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 5.1.1 ([BGM18]). 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 [Hör67], 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.

5.1.1 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 [Wan14, pp. 300-303].

However, our methods are different and yield optimal results as we explain in Remark 5.1.6.

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

$$\Gamma(f) = \frac{1}{2}\sigma^2 \|\nabla_p f\|^2, \tag{5.1.1}$$

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 5.1.2 ([BGM18]). 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

$$
\begin{aligned}
\Gamma^{\alpha,\beta}(f,g) &:= \sum_{i=1}^d \left(\frac{\partial f}{\partial p_i} - \alpha\frac{\partial f}{\partial \xi_i}\right)\left(\frac{\partial g}{\partial p_i} - \alpha\frac{\partial g}{\partial \xi_i}\right) + \beta\sum_{i=1}^d \frac{\partial f}{\partial \xi_i}\frac{\partial g}{\partial \xi_i} \\
&= \langle\nabla_p f, \nabla_p g\rangle - \alpha\langle\nabla_p f, \nabla_\xi g\rangle - \alpha\langle\nabla_\xi f, \nabla_p g\rangle + \left(\alpha^2 + \beta\right)\langle\nabla_\xi f, \nabla_\xi g\rangle, \tag{5.1.2}
\end{aligned}
$$

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 5.1.3 ([BGM18])**.** *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 5.1.4 ([BGM18])**.** We will repeatedly use the following simple computation. Suppose $\alpha\left(s\right), \beta\left(s\right) \in C^1 \left([0, \infty) \right)$. Then for $f \in C^\infty \left(\mathbb{R}^d \times \mathbb{R}^d \right)$

$$\phi'\left(s\right) = 2 P_s \left(\Gamma_2^{\alpha(s),\beta(s)} \left(P_{t-s} f \right) \right) \tag{5.1.3}$$

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

where ϕ is the functional

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

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

Proposition 5.1.5 (Bakry-Émery type estimate, [BGM18])**.** 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 [CCFI11, 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 (5.1.3) and Lemma 5.1.3

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

$$\geqslant 2P_s \left(\alpha(s)\|\nabla_\xi(P_{t-s}f)\|^2 - \langle \nabla_\xi(P_{t-s}f), \nabla_p(P_{t-s}f) \rangle \right)$$

$$+ \langle \nabla_\xi(P_{t-s}f), \nabla_p(P_{t-s}f) \rangle + s\|\nabla_\xi P_{t-s}f\|^2 = 0.$$

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 $s \to 0$. This follows since

$$
\begin{aligned}
|(P_s f)(p, \xi) - f(p, \xi)| &\leq \mathbb{E}^{(p,\xi)} \left[|f(X_s) - f(p, \xi)| \right] \\
&\leqslant C \mathbb{E}^{(p,\xi)} \left[|X_s - (p, \xi)| \right] \\
&= C \mathbb{E}^{(0,0)} \left[\left| \left(B_s, ps + \int_0^s B_u du \right) \right| \right] \\
&\leqslant C \mathbb{E}^0 \left[|B_s| \right] + |p| \, s + C \int_0^s \mathbb{E}^0 |B_u| \, du \\
&= C \sqrt{\frac{2s}{\pi}} + |p| \, s + C \frac{2}{3} \sqrt{\frac{2}{\pi}} s^{\frac{3}{2}},
\end{aligned}
$$

where C is the Lipschitz constant.

Similarly 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 5.1.6 (Bakry-Émery type estimate is sharp, [BGM18])**.** 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

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

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 5.1.5 are sharp.

Proposition 5.1.7 (Reverse Poincaré inequality, [BGM18]). 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 (5.1.1), (5.1.2), (5.1.3) and Lemma 5.1.3 we have

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

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 5.1.8 (Reverse log-Sobolev inequality, [BGM18]). 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

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

97

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

$$
\begin{aligned}
\phi'(s) = &- P_s((P_{t-s}f)\Gamma^{\alpha(s),\beta(s)}(\ln P_{t-s}f)) + 2(t-s)P_s((P_{t-s}f)\Gamma_2^{\alpha(s),\beta(s)}(\ln P_{t-s}f)) \\
&- 2(t-s)\alpha'(s)\sum_{i=1}^{d} P_s\left((P_{t-s}f)\frac{\partial \ln P_{t-s}f}{\partial \xi_i}\frac{\partial \ln P_{t-s}f}{\partial p_i}\right) \\
&+ 2(t-s)\alpha(s)\alpha'(s)P_s[(P_{t-s}f)\|\nabla_\xi \ln P_{t-s}f\|^2] \\
&+ (t-s)\beta'(s)P_s[(P_{t-s}f)\|\nabla_\xi \ln P_{t-s}f\|^2] \\
\geqslant &- P_s((P_{t-s}f)\|\nabla_p \ln P_{t-s}f\|^2) = -\frac{2}{\sigma^2}P_s((P_{t-s}f)\Gamma(\ln P_{t-s}f)).
\end{aligned}
$$

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 [BB12, Proposition 3.4]). We deduce therefore the following functional inequality.

Theorem 5.1.9 (Wang-Harnack inequality, [BGM18])**.** 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 5.1.8 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 [BB12, 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 (5.1.2).

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. ☐

5.1.2 Coupling

In this section, we use coupling techniques to prove Proposition 5.1.5 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 5.1.10, 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 5.1.10 ([BGM18]). Let $f \in C^2 \left(\mathbb{R}^d \times \mathbb{R}^d \right)$ with bounded second deriva-

tives. 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

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

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

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

101

Using this estimate and Jensen's inequality we see that

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

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

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

which proves the statement. $\qquad\square$

Remark 5.1.11 ([BGM18]). When $q = 2$, this coincides with the conclusion of Proposition 5.1.5. 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.

Chapter 6

Functional inequalities for the relativistic diffusion

In this chapter 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 [Dud66] and studied by J. Franchi and Y. Le Jan in [FLJ07]. 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 (6.0.3) in Corollary 6.0.4 is sharper than the ones given in Theorems 7.0.3 and 7.0.8. 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 [FLJ07]. 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 [FLJ12, 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 [Dud66], 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

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

104

By [FLJ12, 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

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

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

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 ∇_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

$$\Gamma^Z(f) := \|\nabla_\xi f\|^2, \tag{6.0.1}$$

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

Theorem 6.0.1 (Curvature-dimension condition, [BGM18])**.** 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

106

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

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 (6.0.1). 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

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

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

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

$$W(p, \xi) := 1 + \xi_0^2 + \|\vec{\xi}\|^2 + d_R(p_0, p)^2, \quad p \in \mathbb{H}^d, \xi \in \mathbb{R}^{1,d}, \qquad (6.0.2)$$

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

$$W \geqslant 1,$$

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

$$LW \leqslant CW \text{ for some constant } C > 0,$$

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

We shall make use of the Lyapunov function W defined by (6.0.2) to prove the following result.

Theorem 6.0.2 (Gradient estimate, [BGM18]). 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

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

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

Then

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

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

108

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 6.0.1 we have

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

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

$$a'(s)\,\Gamma\,(P_{t-s}f) + b'(s)\,\Gamma^Z\,(P_{t-s}f) + 2a(s)\left(-\frac{d}{2}\sigma^2\Gamma\,(P_{t-s}f) - \frac{1}{4}\Gamma^Z\,(P_{t-s}f)\right) =$$

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

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 (6.0.2) and a cutoff argument as in [Bau16, Theorem 7.3], we deduce from a parabolic comparison principle

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

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 6.0.3 (Poincaré type inequality, [BGM18]). 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 6.0.2 we have that

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

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

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

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

This implies

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

$\square$

The next corollary gives us an equivalent estimate to the one in Theorem 6.0.2. The estimate (6.0.3) will be similar to the one we will obtain in Theorem 7.0.8 in a more general setting.

Corollary 6.0.4 ([BGM18]). For any $f \in C_0^\infty \left(\mathbb{H}^d \times \mathbb{R}^{1,d}\right)$, the gradient estimate

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

is equivalent to

$$\Gamma(P_t f) \leqslant e^{d\sigma^2 t} P_t \left(\Gamma\left(f\right)\right) + \frac{e^{d\sigma^2 t} - 1}{2d\sigma^2} P_t \left(\Gamma^Z \left(f\right)\right). \tag{6.0.3}$$

Moreover, one has

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

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 $\Gamma^Z (f) \geqslant -2d\sigma^2 \Gamma(f) - 4\Gamma_2 (f)$ 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

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

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

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

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

Putting these together we have

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

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

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

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

Chapter 7

Heat semigroup functional inequalities 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 6.0.1. In section 7.0.1, we prove gradient bounds for a Kolmogorov type diffusions on $\mathbb{R}^k \times \mathbb{R}^k$ using a Γ-calculus approach. In section 7.0.2, we show that the results in section 7.0.1 are applicable to a large class of diffusions. In section 7.0.3, we prove gradient bounds when X_t is assumed to live on a Riemannian manifold using coupling techniques. In section 7.0.4, we generalize the results in section 7.0.3 to iterated Kolmogorov diffusions. Finally in section 7.0.5, we prove gradient bounds when X_t is assumed to live in the Heisenberg group.

7.0.1 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 that L is elliptic and that $\sigma : \mathbb{R}^k \to \mathbb{R}^k$ is a C^1 map such that

$$C_\sigma := \left(\sum_{i,j=1}^{d} \left(V_i \sigma_j \right)^2 \right)^{\frac{1}{2}} < \infty. \tag{7.0.1}$$

We consider functions on $\mathbb{R}^k \times \mathbb{R}^k$ with $f(p, \xi)$, $p, \xi \in \mathbb{R}^k$. By Proposition B.0.1 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

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

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

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

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

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

In the next lemma we compute the iterated carré du champ

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

and

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

Lemma 7.0.1. *If $f \in C^\infty\left(\mathbb{R}^k \times \mathbb{R}^k\right)$ then*

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

and

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

Proof. Take

$$\mathcal{L} = \sum_{i=1}^{k}V_i^2 + V_0 + \sum_{i=1}^{N}\sigma_i(p)\frac{\partial}{\partial \xi_i}.$$

First it is not too hard to see that $\Gamma(f) = \sum_j (V_j f)^2$. Then

$$\begin{aligned}
\mathcal{L}\Gamma &= \sum_{i=1}^{k}V_i^2\Gamma + V_0\Gamma + \sum_{i=1}^{k}\sigma_i(p)\frac{\partial \Gamma}{\partial \xi_i}\\
&= I + II,
\end{aligned}$$

116

where $I = \sum_{i=1}^{k} V_i^2 \Gamma$. Then

$$
\begin{aligned}
I &= \sum_{i=1}^{k} V_i V_i \left(\sum_{j=1}^{k} (V_j f)^2 \right) \\
&= \sum_{i=1}^{k} V_i \left(\sum_{j=1}^{k} 2 \, (V_j f) \, (V_{ij} f) \right) \\
&= 2 \sum_{i=1}^{k} \left[\sum_{j=1}^{k} (V_{ij} f)^2 + \sum_{j=1}^{k} (V_j f) \, (V_{iij} f) \right]
\end{aligned}
$$

and

$$
\begin{aligned}
II &= \sum_{i=1}^{k} \sigma_i(p) \frac{\partial}{\partial \xi_i} \left(\sum_{j=1}^{k} (V_j f)^2 \right) + V_0 \left(\sum_{j=1}^{k} (V_j f)^2 \right) \\
&= \sum_{i=1}^{k} \sigma_i(p) \left(\sum_{j=1}^{k} 2 \, (V_j f) \left(\frac{\partial}{\partial \xi_i} V_j f \right) \right) + \sum_{j=1}^{k} 2 \, (V_j f) \, (V_0 V_j f)
\end{aligned}
$$

Let

$$
\Gamma(f, Lf) = \sum_{i=1}^{k} (V_i f) \, (V_i L f) = III.
$$

Then

$$
\begin{aligned}
III &= \sum_{i=1}^{k} V_i f \left(V_i \left(\sum_{j=1}^{k} V_j^2 f + V_0 f + \sum_{j} \sigma_j(p) \frac{\partial f}{\partial \xi_j} \right) \right) \\
&= \sum_{i} (V_i f) \left(\sum_{j=1}^{k} V_{ijj}^3 f + V_i V_0 f + \sum_{j=1}^{k} V_i \, (\sigma_j(p)) \frac{\partial f}{\partial \xi_j} + \sum_{j=1}^{k} \sigma_j(p) V_i \frac{\partial f}{\partial \xi_j} \right)
\end{aligned}
$$

Combining everything we have

$$\Gamma_2(f) = \frac{1}{2}I + \frac{1}{2}II - III$$

$$= \sum_i \sum_j (V_{ij}f)^2 + \sum_i \sum_j (V_j f)(V_{iij}f)$$

$$+ \sum_i \sum_j \sigma_i(p)(V_j f)\left(\frac{\partial}{\partial \xi_i}V_j f\right) + \sum_j (V_j f)(V_0 V_j f)$$

$$- \sum_i \sum_j (V_i f)(V_{ijj}^3 f) - \sum_i (V_i f)(V_i V_0 f)$$

$$- \sum_i \sum_j (V_i f) V_i(\sigma_j(p))\frac{\partial f}{\partial \xi_j} - \sum_i \sum_j (V_i f)\sigma_j(p)V_i\frac{\partial f}{\partial \xi_j}$$

$$= \sum_i \sum_j (V_{ij}f)^2 + \sum_i \sum_j (V_i f)(V_{jji}f - V_{ijj}f)$$

$$+ \sum_i (V_i f)(V_0 V_i f - V_i V_0 f) - \sum_i \sum_j (V_i f) V_i(\sigma_j(p))\frac{\partial f}{\partial \xi_j}$$

Hence

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

A similar computation will arrive at

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

$\square$

Theorem 7.0.2 (Curvature-dimension inequality, [BGM18]). 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. Recall that by Lemma 7.0.1 we have

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

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

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

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

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

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

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

Using this inequality with the previous one give us the desired first curvature-dimension inequality. The second inequality follows again follows by Lemma 7.0.1

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

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

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 Chapter 6, 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 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 7.0.3 (Gradient estimate, [BGM18]). Suppose Assumption 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 $x \in \mathbb{R}^k \times \mathbb{R}^k$

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

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

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

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 7.0.2 we have

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

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

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

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

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 [Bau16, Theorem 7.3], we deduce from a parabolic comparison principle,

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

Observe that

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

while

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

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

□

Corollary 7.0.4 (Poincaré type inequality, [BGM18]). If $C_\sigma > 2\rho$ then for any $f \in C_0^\infty \left(\mathbb{R}^k \times \mathbb{R}^k \right)$ and $t \geqslant 0$

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

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

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

By Theorem 7.0.3 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

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

$\square$

7.0.2 Examples

To illustrate the results in Section 7.0.1 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 (7.0.1) and

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

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 [Hsu02, p. 77]). The generator of X_t is

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

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

To apply Theorem 7.0.3 we first need to construct an appropriate Lyapunov function W for the operator $\mathcal{L}$ satisfying Assumption 2. Once we construct W, we will spend the rest of the section verifying Assumption 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 [LY86] 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

$$W(p, \xi) := K + \|\xi\|^2 - \ln p(p_0, p, \tau), p \in M, \xi \in \mathbb{R}^k, \tag{7.0.2}$$

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

Lemma 7.0.5 ([BGM18]). *The function W defined in (7.0.2) 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 [Ham93,LY86],

one has for some constants $C_1, C_2 > 0$

$$\|\nabla_p \ln p (p_0, p, \tau)\|^2 \leqslant C_1 + C_2 d_M (p_0, p)^2 \,, \tag{7.0.3}$$

$$\Delta_p (- \ln p (p_0, p, \tau)) \leqslant C_1 + C_2 d_M (p_0, p)^2 \,. \tag{7.0.4}$$

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

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

$$\leqslant C (K - \ln (p (p_0, p, \tau))) \tag{7.0.5}$$

for a fixed $\tau \geqslant 0$ and a constant $C > 0$. Hence, $\|\nabla_\xi W\| + \|\nabla_p W\| \leqslant CW$ can be shown using (7.0.3), (7.0.5) 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 (7.0.4), (7.0.5), 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 7.0.5 proves that W defined by (7.0.2) is a Lyapunov function satisfying Assumption 2. As a consequence, Theorem 7.0.3 can be applied to complete Riemannian manifolds with $\text{Ric} \geqslant \rho$ since the condition $\text{Ric} \geqslant \rho$ is equivalent to

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

7.0.3 Coupling

Let (M, g) be a complete connected d-dimensional Riemannian manifold. It is not necessarily embedded in $\mathbb{R}^k$. We consider the process

$$X_t = \left(B_t, \int_0^t \sigma(B_s)\, ds \right), \tag{7.0.6}$$

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

$$|\sigma(p) - \sigma(\widetilde{p})| \leqslant C_\sigma d_M(p, \widetilde{p}), \tag{7.0.7}$$

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 $f \in C_0^\infty\left(M \times \mathbb{R}^k\right)$.

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 [vRS05, Wan97, vR04] 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

$$d_M\left(B_t, \widetilde{B}_t\right) \leqslant e^{-Kt/2} d_M\left(p, \widetilde{p}\right) \tag{7.0.8}$$

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 [Wan97, Cra91], or by a central limit theorem argument for the geodesic random walks as in [vR04]. It turns out that the existence of the coupling satisfying (7.0.8) is equivalent to

$$\|\nabla P_t f\| \leqslant e^{-Kt} P_t\left(\|\nabla f\|\right), \tag{7.0.9}$$

for all $f \in C_0^\infty(M)$ and all $t > 0$. We also point out that in [PP16], M. Pascu and I. Popescu constructed explicit Markovian couplings where equality in (7.0.8) is attained for $t \geqslant 0$ given some extra geometric assumptions.

The coupling by parallel transport that gives (7.0.8) is in the elliptic setting. In this section, we will use the coupling by parallel transport to induce a coupling for (7.0.6) 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 7.0.6 ([BGM18])**.** Let (M, g) be a Riemannian manifold. If $f \in C^1(M)$ then

$$\lim_{r\to 0} \sup_{\widetilde{p}:0<d_M(p,\widetilde{p})\leqslant r} \frac{|f(p) - f(\widetilde{p})|}{d_M(p,\widetilde{p})} = \|\nabla f(p)\|. \tag{7.0.10}$$

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 $F'(s) = g\left(\nabla f(\gamma(s)), \gamma'(s)\right)$, 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 (7.0.10) must be at least $\|\nabla f(p)\|$. This proves (7.0.10).

$\square$

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

Lemma 7.0.7 ([BGM18]). *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*

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

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

Proof. Let $p, \widetilde{p} \in M$ with $T_1 = d_M(p, \widetilde{p})$ 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

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

$$\leqslant \left\|\nabla_p f\left(\gamma\left(0\right),\beta(0)\right)\right\| = \left\|\nabla_p f\left(\widetilde{p},\widetilde{\xi}\right)\right\|.$$

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

$$\left|f(p,\xi) - f(\widetilde{p},\widetilde{\xi})\right| = |F(T_1,T_2) - F\left(0,0\right)|$$

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

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 $\left(B_t,\widetilde{B}_t\right)$ starting at $(p,\widetilde{p})$ by parallel transport satisfying (7.0.8), as introduced in [vRS05, vR04] . 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

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

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

Theorem 7.0.8 (Bakry-Émery type estimate, [BGM18])**.** Let M be a complete connected Riemannian manifold such that $\mathrm{Ric}\left(M\right) \geqslant K$ for some $K \in \mathbb{R}$. Let σ be a $C_\sigma-$Lipschitz map as in (7.0.7) and $f \in C^2\left(M\times\mathbb{R}^k\right)$ with a bounded Hessian. Then for every $q \geqslant 1$ and $t \geqslant 0$,

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

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

$$d_M\left(B_t, \widetilde{B}_t\right) \leqslant e^{-Kt/2} d_M\left(p, \widetilde{p}\right), \tag{7.0.11}$$

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

$$d_E\left(Y_t, \widetilde{Y}_t\right) \leqslant \int_0^t \left|\sigma(B_s) - \sigma\left(\widetilde{B}_s\right)\right| ds \leqslant C_\sigma \int_0^t d_M\left(B_s, \widetilde{B}_s\right) ds$$
$$\leqslant C_\sigma d_M\left(p, \widetilde{p}\right) \int_0^t e^{-Ks/2} ds = C_\sigma \left(\frac{1 - e^{-Kt/2}}{K/2}\right) d_M\left(p, \widetilde{p}\right), \tag{7.0.12}$$

where we used (7.0.7) and (7.0.8) . 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

$$d_M\left(B_t, \widetilde{B}_t\right) \leqslant d_M\left(p, \widetilde{p}\right), \tag{7.0.13}$$

Âfor all $t \geqslant 0$. A similar computation as in (7.0.12) gets us the estimate

$$d_E\left(Y_t, \widetilde{Y}_t\right) \leqslant C_\sigma t d_M\left(p, \widetilde{p}\right), \tag{7.0.14}$$

from (7.0.13). Combining (7.0.11) and (7.0.13) we get

$$d_M\left(B_t, \widetilde{B}_t\right) \leqslant K_1(t) d_M\left(p, \widetilde{p}\right), \qquad (7.0.15)$$

while combining (7.0.12) and (7.0.14) we have

$$d_E\left(Y_t, \widetilde{Y}_t\right) \leqslant K_2(t) d_M\left(p, \widetilde{p}\right), \qquad (7.0.16)$$

for all $t \geqslant 0$, where all of these inequalities hold $\mathbb{P}^{(\mathbf{x}, \widetilde{\mathbf{x}})}$−almost surely. By Lemma 7.0.7, 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

$$\left| f\left(B_t, Y_t\right) - f\left(\widetilde{B}_t, \widetilde{Y}_t\right) \right| \leqslant \left\| \nabla_p f\left(\widetilde{B}_t, \widetilde{Y}_t\right) \right\| d_M\left(B_t, \widetilde{B}_t\right) + \left\| \nabla_\xi f\left(\widetilde{B}_t, \widetilde{Y}_t\right) \right\| d_E\left(Y_t, \widetilde{Y}_t\right)$$
$$+ C_f \left(d_M\left(B_t, \widetilde{B}_t\right) + d_E\left(Y_t, \widetilde{Y}_t\right) \right)^2, \qquad (7.0.17)$$

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

Using inequalities (7.0.15), (7.0.16) and (7.0.17), we have that for $f \in C_0^2\left(M \times \mathbb{R}^k\right)$

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

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

$$|P_t f (p, \xi) - P_t f (\widetilde{p}, \xi)|$$

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

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

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

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

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

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 7.0.6, we have the desired result. □

Remark 7.0.9 ([BGM18])**.** The constants obtained in Theorem 7.0.8 using the coupling technique are sharper than the constants in Theorem 7.0.3 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 7.0.10 ([BGM18])**.** We note that when applying the triangle inequality to the right hand sides of the inequalities in Propositions 5.1.5, 5.1.10, we recover Theorem 7.0.8 when the manifold is $M = \mathbb{R}^d$. Here we have $k = d$, $\sigma(\mathbf{x}) = \mathbf{x}$ and $C_\sigma = 1$.

Example 7.0.1 (Velocity spherical Brownian motion, [BGM18])**.** 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 [ABT15] and further studied in [BT18]. When $\mathcal{M} = \mathbb{R}^{d+1}$ and $\sigma = \kappa = 1$ the diffusion is of the form $X_t = \left(B_t, \int_0^t B_s ds \right)$ 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

$$d_E(\mathbf{x}, \mathbf{y}) \leqslant d_{\mathbb{S}^d}(\mathbf{x}, \mathbf{y}), \tag{7.0.18}$$

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 (7.0.18) shows that $\sigma : \mathbb{S}^d \to \mathbb{R}^{d+1}$ is a $C_\sigma = 1$-Lipschtiz map. Thus we can apply Theorem 7.0.8 to the manifold $M = \mathbb{S}^d$, since $Ric = (d-1)g$ where g is the Riemannian metric.

Example 7.0.2. Suppose $k = 1$. Fix a $p_0 \in M$. We consider the map $\sigma : M \to \mathbb{R}$ defined by

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

Note that this map satisfies

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

135

for all $p, \tilde{p} \in M$. Thus we can always apply Theorem 7.0.8 to the process

$$X_t = \left(B_t, \int_0^t d_M \left(B_s, p_0 \right) ds \right),$$

where B_t is Brownian motion on M.

7.0.4 Iterated Kolmogorov diffusions

Our technique can also be applied in studying *iterated Kolmogorov diffusions* similar to those studied by Banerjee and Kendall in [BK16]. 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 7.0.8, we get the following result.

Theorem 7.0.11 ([BGM18])**.** Let M be a complete connected Riemannian manifold such that $\text{Ric}(M) \geqslant K$ 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

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

for $q \geqslant 1$, where

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

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

7.0.5 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

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

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

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

The horizontal distribution is defined by $\mathcal{H} = \text{span}\{\mathcal{X}, \mathcal{Y}\}$, 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_{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 $X \in \mathcal{H}$

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

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 [GL16, 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

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

138

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 [BBBC08, Li06, DM05, Eld10]. In particular, the L^1-gradient bounds for the heat semigroup have been proven in [BBBC08] and [Li06]. As pointed out in [Kuw10], 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

$$d_{CC}\left(B_t^p, \widetilde{B}_t^p\right) \leqslant K d_{CC}\left(p, \widetilde{p}\right), \tag{7.0.19}$$

almost surely for some constant $K \geqslant 1$ that does not depend on $p, \widetilde{p}, t$. We remark that in [BJ18], the authors show that any coupling that satisfy (7.0.19) on $\mathbb{G}$ must be non-Markovian. This further highlights the need for more non-Markovian coupling techniques as in [BK16, BGM16].

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 7.0.7 and let P_t be the heat semigroup associated with X_t. Using a similar argument as in Proposition 7.0.7 with the sub-Riemannian metric g and the horizontal gradient $\nabla_\mathcal{H}$, we can get an estimate

$$|f(p, \xi) - f(\widetilde{p}, \xi)| \leqslant \|\nabla_\mathcal{H} f(\widetilde{p}, \xi)\| d_{CC}(p, \widetilde{p}) + \|\nabla_\xi f(\widetilde{p}, \xi)\| d_E\left(\xi, \widetilde{\xi}\right)$$
$$+ C_f \left(d_{CC}(p, \widetilde{p}) + d_E\left(\xi, \widetilde{\xi}\right)\right)^2, \tag{7.0.20}$$

for functions $f \in C_0^\infty(\mathbb{G} \times \mathbb{R}^3)$, where $C_f \geqslant 0$. The argument in Theorem 7.0.8 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 (7.0.20). Thus using (7.0.19) and (7.0.20) for the Heisenberg group we obtain the following result.

Theorem 7.0.12 ([BGM18]). For all $q \geqslant 1$ and $f \in C_0^\infty \left(\mathbb{G} \times \mathbb{R}^3 \right)$,

$$\| \nabla_{\mathcal{H}} P_t f \|^q \leqslant K^q P_t \left(\left(\| \nabla_{\mathcal{H}} f \| + C_\sigma t \, \| \nabla_\xi f \| \right)^q \right). \tag{7.0.21}$$

The best constant K in (7.0.21) is not known. The best known estimate for K as of this writing is $K \geqslant \sqrt{2}$ (see [DM05, Proposition 2.7]).

Example 7.0.3 ([BGM18]). 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 $X_t = \left(B_t^p, \xi + \int_0^t \sigma \left(B_s^p \right) ds \right)$. A straightforward computation (see Lemma A.0.8) 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 (7.0.7). Thus Theorem 7.0.12 can be applied to X_t.

Appendix A

Heisenberg group geodesics and the Carnot-Carathéodory distance

Let $(M, \mathcal{H}, g)$ be a sub-Riemannian manifold.

Definition A.0.1. With $q \in M$ and $p \in T_q^\star M$ define $H(q, p) = \frac{1}{2} \langle \beta(p), \beta(p) \rangle_q$ where $\langle \cdot, \cdot \rangle$ is the cometric (fiber-wise). H is called the *Hamiltonian*. Elements in TM are *velocity* vectors and $T^\star M$ are *momentum* vectors.

Definition A.0.2. A curve $(x(t), p(t))$ in $T^\star M$ where $x(t) \in M$ and $p(t) \in T_{x(t)}^\star M$ is said to satisfy the *Hamilton-Jacobi Equations* (HJEs) if

$$
\begin{cases}
\dot{x}^i = \frac{\partial H}{\partial p_i} (x(t), p(t)) \\
\dot{p}_i = -\frac{\partial H}{\partial x^i} (x(t), p(t)).
\end{cases}
$$

We can write the Hamilton-Jacobi equations in local coordinates in the following way. Let $\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \right\}$ be a basis in $T_x M$ and $\{dx^1, \ldots, dx^n\}$ a basis in $T_x^\star M$. If

$p = \sum p_i dx^i$ is a momentum vector then we can rewrite the Hamiltonian as

$$H(x, p) = \frac{1}{2} \sum_i p_i p_j \beta^{ij}(x),$$

where

$$\beta^{ij} = \left\langle \beta\left(dx^i\right), \beta\left(dx^j\right) \right\rangle.$$

It is a fact that the Hamilton Jacobi Equations have a unique solution for a given intial $x(0) = x \in M$ and $p(0) = p \in T_x^\star M$, for a short time. See [Mon02] for a proof.

Theorem A.0.3. *Let $\xi(t) = (\gamma(t), p(t))$ be a solution to the HJE on $T^\star M$ for a sub-Riemannian Hamiltonian H. Then $\gamma(t)$ is a (locally) minimizing sub-Riemannian geodesic.*

Definition A.0.4. The curve γ as given in Theorem A.0.3 is called a *normal* (sub-Riemannian) geodesic.

We recall the definition of the Heisenberg group. Let $\mathbb{H}^3$ be identified with $\mathbb{R}^3$ with the law

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

which makes $\mathbb{R}^3$ into a non-commutative group. We call $\mathbb{H}^3 = (\mathbb{R}^3, \star)$ the *Heisenberg group*. The left invariant vector fields are

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

Proposition A.0.5. The curve $\gamma(t)$ is horizontal on the $\mathbb{H}^3$ if and only if $\dot{z} = -\frac{1}{2}(\dot{x}y - \dot{y}x)$ for $t \in [a, b]$.

Proof. Recall that horizontal means that $\dot{\gamma}(t) \in \mathcal{H}_{\gamma(t)}$ for all t. The goal is to write $\dot{\gamma}(t)$ in terms of X, Y, Z. We have

$$
\begin{aligned}
\dot{\gamma}(t) &= \dot{x}\partial_x + \dot{y}\partial_y + \dot{z}\partial_z \\
&= \dot{x}\left(\partial_x - \frac{y}{2}\partial_z\right) + \dot{y}\left(\partial_y + \frac{x}{2}\partial_z\right) \\
&\quad + \dot{x}\frac{y}{2}\partial_z - \dot{y}\frac{x}{2}\partial_z + \dot{z}\partial_z \\
&= \dot{x}X + \dot{y}Y + \dot{x}\frac{y}{2}\partial_z - \dot{y}\frac{x}{2}\partial_z + \dot{z}\partial_z,
\end{aligned}
$$

but $\dot{x}X + \dot{y}Y \in \mathcal{H}_{\gamma(t)}$ and $\left(\dot{x}\frac{y}{2} - \dot{y}\frac{x}{2} + \dot{z}\right)\partial_z \notin \mathcal{H}_{\dot{\gamma}(t)}$. Thus we must have that $\dot{\gamma}(t) \in \Delta_{\gamma(t)}$ if and only $\dot{x}\frac{y}{2} - \dot{y}\frac{x}{2} + \dot{z} = 0$. $\square$

Assume $\{X, Y\}$ are orthonormal, which means $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{H}$. Let's take any co-vector $p = p_x dx + p_y dy + p_z dz$. We will compute sub-Riemannian Hamiltonian

$$
H(q, p) = \frac{1}{2}\left(\langle X(q), p \rangle^2 + \langle Y(q), p \rangle^2\right).
$$

Recall that for $p \in T_q^\star M$ we have $\langle X(q), p \rangle = p(X(q))$ where $X(q) \in T_q M$. Define $P_X : T^\star M \to \mathbb{R}$ by $P_X = p(X(q))$ where $(q, p) \in T^\star M$. Also $\langle v, \alpha \rangle = \alpha(v)$ so that $\langle e_i, \alpha^j \rangle = \delta_i^j$. For $X = \partial_x - \frac{y}{2}\partial_z$ and $Y = \partial_y + \frac{x}{2}\partial_z$ we have

$$
\begin{aligned}
P_X &= \left\langle \partial_x - \frac{y}{2}\partial_z, p_x dx + p_y dy + p_z dz \right\rangle \\
&= p_x - \frac{y}{2}p_z
\end{aligned}
$$

143

so

$$P_X\left(x, y, z, p_x, p_y, p_z\right) = p_x - \frac{y}{2}p_z$$

where $q = (x, y, z)$ and $p = (p_x, p_y, p_z)$. Now

$$\begin{aligned}
P_Y &= \left\langle \partial_y + \frac{x}{2}\partial_z, p_x dx + p_y dy + p_z dz \right\rangle \\
&= p_y + \frac{x}{2}p_z
\end{aligned}$$

so

$$P_Y\left(x, y, z, p_x, p_y, p_z\right) = p_y + \frac{x}{2}p_z.$$

Now since $Z = \partial_z$ then

$$P_Z\left(x, y, z, p_x, p_y, p_z\right) = p_z.$$

Thus the Hamiltonian is

$$\begin{aligned}
H\left(q, p\right) &= \frac{1}{2}\left(\langle X(q), p\rangle^2 + \langle Y(q), p\rangle^2\right) \\
&= \frac{1}{2}\left(\left(p_x - \frac{y}{2}p_z\right)^2 + \left(p + \frac{x}{2}p_z\right)^2\right).
\end{aligned}$$

The Hamilton Jacobi Equations become

$$\begin{aligned}
\dot{q} &= \frac{\partial H}{\partial p} \\
\dot{p} &= -\frac{\partial H}{\partial q}.
\end{aligned}$$

which breaks down to by plugging in what we know

$$
\begin{aligned}
\dot{x} &= \frac{\partial H}{\partial p_x} = p_x - \frac{y}{2}p_z \\
\dot{y} &= \frac{\partial H}{\partial p_y} = p_y + \frac{x}{2}p_z \\
\dot{z} &= \frac{\partial H}{\partial p_z} = -\frac{y}{2}\left(p_x - \frac{y}{2}p_z\right) + \frac{x}{2}\left(p_y + \frac{x}{2}p_z\right) \\
&= \frac{1}{4}\left(x^2 + y^2\right)p_z + \frac{1}{2}\left(xp_y - yp_x\right) \\
\dot{p}_x &= -\frac{\partial H}{\partial x} = -\frac{1}{2}p_z\left(p_y + \frac{x}{2}p_z\right) \\
\dot{p}_y &= \frac{\partial H}{\partial y} = \frac{1}{2}p_z\left(p_x - \frac{y}{2}p_z\right) \\
\dot{p}_z &= -\frac{\partial H}{\partial z} = 0.
\end{aligned}
$$

Thus we can rewrite with the momentum functions and obtain

$$
\begin{aligned}
\dot{x} &= P_X & \text{(A.0.1)} \\
\dot{y} &= P_Y & \text{(A.0.2)} \\
\dot{z} &= \frac{1}{2}xP_Y - \frac{1}{2}yP_X & \text{(A.0.3)} \\
\dot{P}_X &= -P_Z P_Y & \text{(A.0.4)} \\
\dot{P}_Y &= P_Z P_X & \text{(A.0.5)} \\
\dot{P}_Z &= 0. & \text{(A.0.6)}
\end{aligned}
$$

Let $\gamma(t) = (x(t), y(t), z(t))$ with initial values $q_0 = (x, y, z) = (0, 0, 0)$ and $p_0 =$

(ξ, η, θ). Differentiating equations (A.0.1)-(A.0.3) we obtain

$$\ddot{x} = \dot{P}_X$$

$$\ddot{y} = \dot{P}_Y$$

$$\begin{aligned}
\ddot{z} &= \frac{1}{2}\dot{x}P_Y + \frac{1}{2}x\dot{P}_Y - \frac{1}{2}\dot{y}P_X - \frac{1}{2}y\dot{P}_X \\
&= \frac{1}{2}\dot{x}\dot{y} + \frac{1}{2}x\ddot{y} - \frac{1}{2}\dot{y}\dot{x} - \frac{1}{2}y\ddot{x} \\
&= \frac{1}{2}x\ddot{y} - \frac{1}{2}y\ddot{x}.
\end{aligned}$$

Also since $P_z = p_z$ then $P_z(t) = 0$. Plugging this back into the equation (A.0.4) and (A.0.5) we get

$$\dot{P}_X = -\theta P_Y$$

$$\dot{P}_Y = \theta P_X.$$

Combining we obtain the equations

$$\begin{cases}
\ddot{x} = -\theta\dot{y} \\
\ddot{y} = \theta\dot{x} \\
\ddot{z} = \frac{1}{2}\left(x\ddot{y} - y\ddot{x}\right)
\end{cases}$$

Solving this we obtain the following solutions:

$$x(t) = \frac{\xi}{|\theta|} \sin\left(|\theta|\, t\right) - \frac{\eta}{|\theta|} \left(\cos\left(|\theta|\, t\right) - 1\right),$$

$$y(t) = -\frac{\xi}{|\theta|} \left(\cos\left(|\theta|\, t\right) - 1\right) - \frac{\eta}{|\theta|} \sin\left(|\theta|\, t\right)$$

$$z(t) = \frac{\xi^2 + \eta^2}{2\,|\theta^2|} \left(|\theta|\, t - \sin\left(|\theta|\, t\right)\right).$$

We summarize this in the following theorem.

Theorem A.0.6 (Heisenberg group geodesics). Let $\gamma(t) = (x(t), y(t), z(t))$ be a curve on $\mathbb{H}^3$ with initial position $q_0 = (x, y, z) = (0, 0, 0)$ and initial momentum $p_0 = (\xi, \eta, \theta)$. The curve given by

$$x(t) = \frac{\xi}{|\theta|} \sin\left(|\theta|\, t\right) - \frac{\eta}{|\theta|} \left(\cos\left(|\theta|\, t\right) - 1\right), \tag{A.0.7}$$

$$y(t) = -\frac{\xi}{|\theta|} \left(\cos\left(|\theta|\, t\right) - 1\right) - \frac{\eta}{|\theta|} \sin\left(|\theta|\, t\right) \tag{A.0.8}$$

$$z(t) = \frac{\xi^2 + \eta^2}{2\,|\theta^2|} \left(|\theta|\, t - \sin\left(|\theta|\, t\right)\right). \tag{A.0.9}$$

satisfies the initial conditions above and is a normal geodesic.

Let $d_{CC}(e, g)$ be Carnot-Caratheodory distance. Then for a given $g = (a, b, c) \in \mathbb{H}^3 \cong \mathbb{R}^3$ we find (ξ, η, θ) to such that $x(t_0) = a, y(t_0) = b, z(t_0) = c$. Then if

$\gamma_0(t) = (x_0(t), y_0(t), z_0(t))$ satisfies A.0.7-A.0.9 then

$$
\begin{aligned}
d_{CC}(e, g) &= \inf \left\{ \int_0^{t_0} \sqrt{\dot{x}^2 + \dot{y}^2} ds, \gamma = (x, y, z) \text{ horizontal } \gamma(0) = e, \gamma(t_0) = (a, b, c) \right\} \\
&= \int_0^{t_0} \sqrt{(\dot{x}_0)^2 + (\dot{y}_0)^2} ds \\
&= \int_0^{t_0} \sqrt{\xi^2 + \eta^2} ds \\
&= t_0 \sqrt{\xi^2 + \eta^2},
\end{aligned}
$$

since geodesics are globally length minimizing on the Heisenberg group.

Lemma A.0.7. *Consider the point $g = (0, 0, c)$ $c \neq 0$. Then $d_{CC}(e, g) \sim c^{\frac{1}{2}}$.*

Proof. Recall that by letting $x_0(t_0) = 0$ and $y_0(t_0) = 0$ and we obtain $\sin|\theta| t_0 = 0$ and $\cos|\theta| t_0 = 1$, say $|\theta| t_0 = 2\pi$. Also $z_0(t_0) = c$ which we obtain

$$
\frac{\xi^2 + \eta^2}{2|\theta^2|} |\theta| t_0 = c.
$$

Thus $|\theta| t_0 = 2\pi$ and $\xi^2 + \eta^2 = \frac{c}{\pi} \theta^2$. Hence

$$
t_0 \sqrt{\xi^2 + \eta^2} = \sqrt{\frac{c}{\pi}} |\theta| t_0 = 2\sqrt{\pi}\sqrt{c}.
$$

$\square$

Lemma A.0.8. *Suppose $(a, b, c) \in \mathbb{H}$. Then*

$$
\sqrt{a^2 + b^2} \leqslant d_{CC}(e, (a, b, c)).
$$

Proof. By equations A.0.7-A.0.9 the normal geodesic from the identity to the point

148

(a, b, c) has parametrization

$$
\begin{aligned}
x_0(t) &= \frac{\xi}{|\theta|} \sin(|\theta| t) - \frac{\eta}{|\theta|} \left(\cos(|\theta| t) - 1 \right), \\
y_0(t) &= -\frac{\eta}{|\theta|} \sin(|\theta| t) - \frac{\xi}{|\theta|} \left(\cos(|\theta| t) - 1 \right) \\
z_0(t) &= \frac{\xi^2 + \eta^2}{2 |\theta^2|} \left(|\theta| t - \sin(|\theta| t) \right).
\end{aligned}
$$

Given $g = (a, b, c) \in \mathbb{H}_3 \cong \mathbb{R}^3$ we can find (ξ, η, θ) such that $x(t_0) = a, y(t_0) = b, z(t_0) = c$ so that

$$
d_{CC}(e, g) = t_0 \sqrt{\xi^2 + \eta^2}.
$$

Now

$$
\begin{aligned}
a = x_0(t_0) &= \frac{\xi}{|\theta|} \sin(|\theta| t_0) - \frac{\eta}{|\theta|} \left(\cos(|\theta| t_0) - 1 \right), \\
b = y_0(t_0) &= -\frac{\eta}{|\theta|} \sin(|\theta| t_0) - \frac{\xi}{|\theta|} \left(\cos(|\theta| t_0) - 1 \right) \\
c = z_0(t_0) &= \frac{\xi^2 + \eta^2}{2 |\theta^2|} \left(|\theta| t_0 - \sin(|\theta| t_0) \right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
a^2 &= \frac{\xi^2}{|\theta|^2} \sin^2(|\theta| t_0) - 2 \frac{\xi \eta}{|\theta|^2} \sin(|\theta| t_0) \left(\cos(|\theta| t_0) - 1 \right) + \frac{\eta^2}{|\theta|^2} \left(\cos(|\theta| t_0) - 1 \right)^2, \\
b^2 &= \frac{\eta^2}{|\theta|^2} \sin^2(|\theta| t_0) + 2 \frac{\eta \xi}{|\theta|^2} \sin(|\theta| t_0) \left(\cos(|\theta| t_0) - 1 \right) + \frac{\xi^2}{|\theta|^2} \left(\cos(|\theta| t_0) - 1 \right)^2
\end{aligned}
$$

So that

$$
\begin{aligned}
a^2 + b^2 \;&=\; \frac{\xi^2 + \eta^2}{|\theta|^2}\sin^2\left(|\theta|\,t_0\right) + \frac{\xi^2 + \eta^2}{|\theta|^2}\left(\cos\left(|\theta|\,t_0\right) - 1\right)^2 \\
&\quad\; \frac{\xi^2 + \eta^2}{|\theta|^2}\left[\sin^2\left(|\theta|\,t_0\right) + \left(\cos\left(|\theta|\,t_0\right) - 1\right)^2\right] \\
&=\; \frac{\xi^2 + \eta^2}{|\theta|^2}\left[\sin^2\left(|\theta|\,t_0\right) + \cos^2\left(|\theta|\,t_0\right) - 2\cos\left(|\theta|\,t_0\right) + 1\right] \\
&=\; 2\frac{\xi^2 + \eta^2}{|\theta|^2}\left[1 - \cos\left(|\theta|\,t_0\right)\right] \\
&=\; 4\frac{\xi^2 + \eta^2}{|\theta|^2}\sin^2\left(\frac{|\theta|\,t_0}{2}\right)
\end{aligned}
$$

So that

$$
\begin{aligned}
\sqrt{a^2 + b^2} \;&=\; 2\sqrt{\frac{\xi^2 + \eta^2}{|\theta|^2}}\left|\sin\left(\frac{|\theta|\,t_0}{2}\right)\right| \\
&\leq\; \frac{2}{|\theta|}\sqrt{\xi^2 + \eta^2}\,\frac{|\theta|\,t_0}{2} \\
&=\; t_0\sqrt{\xi^2 + \eta^2} \\
&=\; d_{CC}(e, g),
\end{aligned}
$$

where I used the elementary inequality $|\sin x| \leq |x|$ for all x. $\qquad\square$

Appendix B

The generator of Kolmogorov-type diffusions

Consider the process

$$\boldsymbol{X}_t = \left(X_t, \int_0^t \sigma\left(X_s\right) ds \right)$$

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, bounded vector fields. We assume L is elliptic. We also assume $\sigma : \mathbb{R}^k \to \mathbb{R}^k$ is a C^1 map such that

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

and that σ is not zero almost everywhere. We let $\mathcal{L}$ be the generator for $\boldsymbol{X}_t$.

Proposition B.0.1. *The generator $\mathcal{L}$ of the process $\boldsymbol{X}_t = \left(X_t, \xi + \int_0^t \sigma\left(X_s\right) ds \right)$ is given by*

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

for function $f \in C_0^\infty \left(\mathbb{R}^k \times \mathbb{R}^k \right)$.

Proof. Take $f(p,\xi) \in C_0^\infty \left(\mathbb{R}^k \times \mathbb{R}^k \right)$. We can suppose that X_t is a diffusion of the form

$$dX_t = a\left(X_t\right) dB_t + b\left(X_t\right) dt, \quad X_0 = p$$

where the a, b are smooth and bounded. We wish to compute

$$\mathcal{L}f(p,\xi) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}^{(p,\xi)} \left[f\left(X_s, \xi + \int_0^t \sigma\left(X_s\right) ds \right) - f(p,\xi) \right].$$

By Taylor's theorem we have

$$f\left(X_t, \xi + \int_0^t \sigma(X_s) ds \right) - f(p,\xi)$$

$$= \nabla_p f(p,\xi) \cdot (X_t - p) + \nabla_\xi f(p,\xi) \cdot \left(\int_0^t \sigma\left(X_s\right) ds \right)$$

$$+ \frac{1}{2} \sum_{i \neq j} \frac{\partial^2 f(p,\xi)}{\partial p_i \partial p_j} \left(X_s^i - p_i \right) \left(X_s^j - p_j \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2 f(p,\xi)}{\partial \xi_i^2} \left(\int_0^t \sigma_i\left(X_s\right) ds \right)^2 + \frac{1}{2} \sum_{i \neq j} \frac{\partial^2 f(p,\xi)}{\partial \xi_i \partial \xi_j} \left(\int_0^t \sigma_i\left(X_s\right) ds \right) \left(\int_0^t \sigma_j\left(X_s\right) ds \right)$$

$$+ \frac{1}{2} \sum_{i \neq j} \frac{\partial^2 f(p,\xi)}{\partial p_i \partial \xi_j} \left(X_s^i - p_i \right) \left(\int_0^t \sigma_j\left(X_s\right) ds \right) + R_t$$

$$= I + II + III + IV + V + VI + R_t,$$

where R_t is the remainder term. We start by computing the limit involving term I,

$$\lim_{t\to 0} \frac{1}{t} \mathbb{E}^{(p,\xi)} \left[\nabla_p f(p,\xi) \cdot (X_t - p)\right] = \lim_{t\to 0} \frac{1}{t} \sum_{i=1}^{k} \frac{\partial f(p,\xi)}{\partial p_i} \mathbb{E}^{(p,\xi)} \left[(X_t^i - p_i)\right]$$

$$= \sum_{i=1}^{k} b_i(p) \frac{\partial f(p,\xi)}{\partial p_i},$$

since by Lebesgue's differentiation theorem we have

$$\lim_{t\to 0} \frac{1}{t} \left[\int_0^t \mathbb{E}^{(p,\xi)} \left[b_i(X_s)\right] ds\right] = b_i(p).$$

Computing the limit involving term II,

$$\lim_{t\to 0} \frac{1}{t} \mathbb{E}^{(p,\xi)} \left[\nabla_\xi f(p,\xi) \cdot \left(\int_0^t \sigma(X_s) ds\right)\right] = \lim_{t\to 0} \frac{1}{t} \sum_{i=1}^{k} \frac{\partial f(p,\xi)}{\partial \xi_i} \left(\int_0^t \mathbb{E}^{(p,\xi)} \left[\sigma_i(X_s)\right] ds\right)$$

$$= \sum_{i=1}^{k} \frac{\partial f(p,\xi)}{\partial \xi_i} \left(\lim_{t\to 0} \frac{1}{t} \int_0^t \mathbb{E}^{(p,\xi)} \left[\sigma_i(X_s)\right] ds\right)$$

$$= \sum_{i=1}^{k} \sigma_i(p) \frac{\partial f(p,\xi)}{\partial \xi_i},$$

where the last equality is due to Lebesgue's differentiation theorem and the fact that σ is $C-$Lipschitz. This can be seen by

$$\lim_{t\to 0} \left|\frac{1}{t} \int_0^t \mathbb{E}^{(p,\xi)} \left[\sigma_i(X_s)\right] ds - \sigma_i(p)\right| = \lim_{t\to 0} \left|\frac{1}{t} \int_0^t \mathbb{E}^{(p,\xi)} \left[\sigma_i(X_s) - \sigma_i(p)\right] ds\right|$$

$$\leqslant \lim_{t\to 0} \frac{C}{t} \int_0^t \mathbb{E}^{(p,\xi)} |X_s - p| ds$$

$$= C \cdot \mathbb{E}^{(p,\xi)} |p - p| = 0.$$

Computing the limit involving term III,

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}^{(p,\xi)}\left[\frac{1}{2}\sum_{i,j}\frac{\partial^2 f\left(p,\xi\right)}{\partial p_i \partial p_j}\left(X_s^i-p_i\right)\left(X_s^j-p_j\right)\right]=\frac{1}{2}\sum_{i,j}\left(a(p)a(p)^T\right)_{i,j}\frac{\partial^2 f\left(p,\xi\right)}{\partial p_i \partial p_j}.$$

Computing the limit involving term IV,

$$\limsup_{t\to 0}\frac{1}{t}\mathbb{E}^{(p,\xi)}\left[\frac{1}{2}\sum_{i=1}^{k}\frac{\partial^2 f\left(p,\xi\right)}{\partial \xi_i^2}\left(\int_0^t\sigma_i\left(X_s\right)ds\right)^2\right]=$$
$$\limsup_{t\to 0}\frac{1}{t}\frac{1}{2}\sum_{i=1}^{k}\frac{\partial^2 f\left(p,\xi\right)}{\partial \xi_i^2}\mathbb{E}^{(p,\xi)}\left[\left(\int_0^t\sigma_i\left(X_s\right)ds\right)^2\right].$$

We prove this limit is zero. To do so, it suffices to show $\limsup_{t\to 0}Y_t \leqslant C$ where $Y_t=\mathbb{E}^{(p,\xi)}\left[\left(\int_0^t\sigma_i\left(X_s\right)\frac{ds}{t}\right)^2\right]$. For if this was true then

$$\limsup_{t\to 0}t\cdot\mathbb{E}^{(p,\xi)}\left[\left(\int_0^t\sigma_i\left(X_s\right)\frac{ds}{t}\right)^2\right]=\limsup_{t\to 0}t\cdot Y_t\leqslant 0\cdot C=0.$$

To see this we use the fact that σ_i must have linear growth and get

$$Y_t=\mathbb{E}^{(p,\xi)}\left[\left(\frac{1}{t}\int_0^t\sigma_i\left(X_s\right)ds\right)^2\right]$$
$$\leqslant\mathbb{E}^{(p,\xi)}\left[\frac{1}{t}\int_0^t\sigma_i\left(X_s\right)^2ds\right]$$
$$\leqslant\mathbb{E}^{(p,\xi)}\left[\frac{1}{t}\int_0^t\left(C_1\left|X_s\right|+C_2\right)^2ds\right]$$
$$=\frac{1}{t}\int_0^t\mathbb{E}^{(p,\xi)}\left[\left(C_1\left|X_s\right|+C_2\right)^2\right]ds$$

then taking limits and using Lebesgue's differentiation theorem we have

$$\limsup_{t\to 0} Y_t \leqslant \mathbb{E}^{(p,\xi)}\left[(C_1\left|X_0\right|+C_2)^2\right] = (C_1\left|p\right|+C_2)^2.$$

Computing the limit involving term V we have

$$\limsup_{t\to 0} \frac{1}{t}\mathbb{E}^{(p,\xi)}\left[\frac{1}{2}\sum_{i\neq j}\frac{\partial^2 f\left(p,\xi\right)}{\partial\xi_i\partial\xi_j}\left(\int_0^t \sigma_i\left(X_s\right)ds\right)\left(\int_0^t \sigma_j\left(X_s\right)ds\right)\right] =$$
$$\limsup_{t\to 0}\frac{1}{t}\frac{1}{2}\sum_{i\neq j}\frac{\partial^2 f\left(p,\xi\right)}{\partial\xi_i\partial\xi_j}\mathbb{E}^{(p,\xi)}\left[\left(\int_0^t\int_0^t \sigma_i\left(X_s\right)\sigma_j\left(X_u\right)dsdu\right)\right] = 0,$$

which follows similarly to the previous cases.

Estimating the limit involving term VI, we have

$$\limsup_{t\to 0}\frac{1}{t}\mathbb{E}^{(p,\xi)}\left[\frac{1}{2}\sum_{i\neq j}\frac{\partial^2 f\left(p,\xi\right)}{\partial p_i\partial\xi_j}\left(X_t^i - p_i\right)\left(\int_0^t \sigma_j\left(X_s\right)ds\right)\right] \leq$$
$$\frac{1}{2}\sum_{i\neq j}\frac{\partial^2 f\left(p,\xi\right)}{\partial p_i\partial\xi_j}\left(\limsup_{t\to 0}\mathbb{E}^{(p,\xi)}\left[\left(X_t^i - p_i\right)\left(\frac{1}{t}\int_0^t \sigma_j\left(X_s\right)ds\right)\right]\right).$$

Using the Cauchy-Schwarz inequality and a previous estimate we have,

$$\limsup_{t\to 0}\mathbb{E}^{(p,\xi)}\left[\left(X_t^i - p_i\right)\left(\frac{1}{t}\int_0^t \sigma_j\left(X_s\right)ds\right)\right]$$
$$\leqslant \limsup_{t\to 0}\left(\mathbb{E}^{(p,\xi)}\left[\left(X_t^i - p_i\right)^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}^{(p,\xi)}\left[\left(\frac{1}{t}\int_0^t \sigma_j\left(X_s\right)ds\right)^2\right]\right)^{\frac{1}{2}}$$
$$\leqslant \limsup_{t\to 0}\left(\mathbb{E}^{p_i}\left[\left(X_t^i - p_i\right)^2\right]\right)^{\frac{1}{2}}(C_1\left|p\right|+C_2).$$

Since X_t is hypoelliptic then there exists a smooth transition kernel $p_t^i(x,y)$ for X_t^i, so that by Dominated Convergence Theorem and the fact that a,b are smooth and

bounded we have that

$$\lim_{t \to 0} \mathbb{E}^{p_i} \left[\left(X_t^i - p_i \right)^2 \right] = \lim_{t \to 0} \int (y - p_i)^2 \, p_t^i \, (p_i, y) \, dy = \mathbb{E}^{p_i} \left[\left(X_0^i - p_i \right)^2 \right] = 0.$$

Let C be a bound on the third derivatives of f. By Taylor's theorem the error term R_t can be bounded by

$$|R_t| \leqslant \frac{C}{6} \sum_{i,j,k} \left| \mathbb{E}^{(p,\xi)} \left[A_i B_j C_k \right] \right|,$$

where the A_i, B_j, C_k are all either of the form $(X_s^{\cdot} - p_{\cdot})$ or $\int_0^t \sigma_{\cdot} (X_s) \, ds$. A similar analysis can be done as above to show that

$$\limsup_{t \to 0} \frac{R_t}{t} = 0.$$

To summarize we showed that

$$
\begin{aligned}
\mathcal{L} f(p, \xi) &= \sum_{i=1}^d b_i(p) \frac{\partial f(p, \xi)}{\partial p_i} + \frac{1}{2} \sum_{i,j} \left(a(p) a(p)^T \right)_{i,j} \frac{\partial^2 f(p, \xi)}{\partial p_i \partial p_j} \\
&\quad + \sum_{i=1}^k \sigma_i(p) \frac{\partial f(p, \xi)}{\partial \xi_i} \\
&= L_p f(p, \xi) + \sum_{i=1}^k \sigma_i(p) \frac{\partial f(p, \xi)}{\partial \xi_i},
\end{aligned}
$$

as needed. $\qquad \square$

Bibliography

[ABT15]   Jürgen Angst, Ismaël Bailleul, and Camille Tardif, *Kinetic Brownian motion on Riemannian manifolds*, Electron. J. Probab. **20** (2015), no. 110, 40. MR 3418542

[Ang15]   Jürgen Angst, *Poisson boundary of a relativistic diffusion in curved spacetimes: an example*, ESAIM Probab. Stat. **19** (2015), 502–514. MR 3423304

[BACK95] 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

[Bai08]   Ismael Bailleul, *Poisson boundary of a relativistic diffusion*, Probab. Theory Related Fields **141** (2008), no. 1-2, 283–329. MR 2372972 (2008m:60155)

[Bak94]   Dominique Bakry, *L'hypercontractivité et son utilisation en théorie des semigroupes*, Lectures on probability theory (Saint-Flour, 1992), Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, pp. 1–114. MR 1307413

[Bak06]   ———, *Functional inequalities for Markov semigroups*, Probability measures on groups: recent directions and trends, Tata Inst. Fund. Res., Mumbai, 2006, pp. 91–147. MR 2213477

[Bau16]   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

[Bau17a]   ———, *Bakry-Émery meet Villani*, J. Funct. Anal. **273** (2017), no. 7, 2275–2291. MR 3677826

[Bau17b]   ———, *Stochastic analysis on sub-Riemannian manifolds with transverse symmetries*, Ann. Probab. **45** (2017), no. 1, 56–81. MR 3601645

[BB12]   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

[BB16]   ———, *Reverse Poincaré inequalities, isoperimetry, and Riesz transforms in Carnot groups*, Nonlinear Anal. **131** (2016), 48–59. MR 3427969

[BBBC08]   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)

[BBG14]   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

[BE85]   D. Bakry and Michel Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206. MR 889476

[BG17]   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

[BGL14]   Dominique Bakry, Ivan Gentil, and Michel Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014. MR 3155209

[BGM13]   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

[BGM16]   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, 2016.

[BGM18]   Fabrice Baudoin, Maria Gordina, and Phanuel Mariano, *Gradient bounds for kolmogorov type diffusions*, ArXiv:1803.01436 Preprint, 2018.

[BJ18]   Michel Bonnefont and Nicolas Juillet, *Couplings in L^p distance of two Brownian motions and their Lévy area*, Arxiv preprint, 2018.

[BK16]   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

[BK17]   _____, *Rigidity for Markovian maximal couplings of elliptic diffusions*, Probab. Theory Related Fields **168** (2017), no. 1-2, 55–112. MR 3651049

[BL06]   Dominique Bakry and Michel Ledoux, *A logarithmic Sobolev form of the Li-Yau parabolic inequality*, Rev. Mat. Iberoam. **22** (2006), no. 2, 683–702. MR 2294794

[BLU07]   A. Bonfiglioli, E. Lanconelli, and F. Uguzzoni, *Stratified Lie groups and potential theory for their sub-Laplacians*, Springer Monographs in Mathematics, Springer, Berlin, 2007. MR 2363343

[BT18]   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

[BW14]   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

[CCFI11]   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)

[CCG07]   Ovidiu Calin, Der-Chen Chang, and Peter Greiner, *Geometric analysis on the Heisenberg group and its generalizations*, AMS/IP Studies in Advanced Mathematics, vol. 40, American Mathematical Society, Providence, RI; International Press, Somerville, MA, 2007. MR 2310372

[CGL93]   Giovanna Citti, Nicola Garofalo, and Ermanno Lanconelli, *Harnack's inequality for sum of squares of vector fields plus a potential*, Amer. J. Math. **115** (1993), no. 3, 699–734. MR 1221840

[Cra91]   M. Cranston, *Gradient estimates on manifolds using coupling*, J. Funct. Anal. **99** (1991), no. 1, 110–124. MR 1120916 (93a:58175)

[Cra92]   _____, *A probabilistic approach to gradient estimates*, Canad. Math. Bull. **35** (1992), no. 1, 46–55. MR 1157463 (93e:35011)

[CTW10]   Shu-Cheng Chang, Jingzhi Tie, and Chin-Tung Wu, *Subgradient estimate and Liouville-type theorem for the CR heat equation on Heisenberg groups*, Asian J. Math. **14** (2010), no. 1, 41–72. MR 2726594

[CY75]   S. Y. Cheng and S. T. Yau, *Differential equations on Riemannian manifolds and their geometric applications*, Comm. Pure Appl. Math. **28** (1975), no. 3, 333–354. MR 0385749

[DH09]   Jörn Dunkel and Peter Hänggi, *Relativistic Brownian motion*, Phys. Rep. **471** (2009), no. 1, 1–73. MR 2503143

[DM05]   Bruce K. Driver and Tai Melcher, *Hypoelliptic heat kernel inequalities on the Heisenberg group*, J. Funct. Anal. **221** (2005), 340–365.

[Dud66]   R. M. Dudley, *Lorentz-invariant Markov processes in relativistic phase space*, Ark. Mat. **6** (1966), 241–268. MR 0198540 (33 #6695)

[Dud67]   ———, *A note on Lorentz-invariant Markov processes*, Ark. Mat. **6** (1967), 575–581 (1967). MR 0216567 (35 #7398)

[Dun08]   Jörn Dunkel, *Relativistic Brownian motion and diffusion processes*, Ph.D. thesis, Universität Augsburg, 2008.

[Eld10]   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)

[FLJ07]   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)

[FLJ12]   ———, *Hyperbolic dynamics and Brownian motion*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2012, An introduction. MR 2977883

[GL16]   Maria Gordina and Thomas Laetsch, *Sub-Laplacians on Sub-Riemannian Manifolds*, Potential Anal. **44** (2016), no. 4, 811–837. MR 3490551

[GL17]   ———, *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

[Gol79]   Sheldon Goldstein, *Maximal coupling*, Z. Wahrsch. Verw. Gebiete **46** (1978/79), no. 2, 193–204. MR 516740

[Gor17]   Maria Gordina, *An application of a functional inequality to quasi-invariance in infinite dimensions*, pp. 251–266, Springer New York, New York, NY, 2017.

[Gri75]   David Griffeath, *A maximal coupling for Markov chains*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **31** (1974/75), 95–106. MR 0370771

[Ham93]   Richard S. Hamilton, *A matrix Harnack estimate for the heat equation*, Comm. Anal. Geom. **1** (1993), no. 1, 113–126. MR MR1230276 (94g:58215)

[Ham11]   _____, *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

[HK00]   Piotr Hajłasz and Pekka Koskela, *Sobolev met Poincaré*, Mem. Amer. Math. Soc. **145** (2000), no. 688, x+101. MR 1683160 (2000j:46063)

[Hör67]   Lars Hörmander, *Hypoelliptic second order differential equations*, Acta Math. **119** (1967), 147–171. MR 0222474 (36 #5526)

[Hsu02]   Elton P. Hsu, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics, vol. 38, American Mathematical Society, Providence, RI, 2002. MR 1882015 (2003c:58026)

[IM15]   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

[JM07]   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

[Kar47]   Kari Karhunen, *über lineare Methoden in der Wahrscheinlichkeitsrechnung*, Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys. **1947** (1947), no. 37, 79. MR 0023013

[Ken86]   Wilfrid S. Kendall, *Nonnegative Ricci curvature and the Brownian coupling property*, Stochastics **19** (1986), no. 1-2, 111–129. MR 864339

[Ken89]   _____, *Coupled Brownian motions and partial domain monotonicity for the Neumann heat kernel*, J. Funct. Anal. **86** (1989), no. 2, 226–236. MR 1021137

[Ken07]   _____, *Coupling all the Lévy stochastic areas of multidimensional Brownian motion*, Ann. Probab. **35** (2007), no. 3, 935–953. MR 2319712 (2008h:60339)

[Ken10]   _____, *Coupling time distribution asymptotics for some couplings of the Lévy stochastic area*, Probability and mathematical genetics, London Math. Soc. Lecture Note Ser., vol. 378, Cambridge Univ. Press, Cambridge, 2010, pp. 446–463. MR 2744251

[Kol34]   A. Kolmogoroff, *Zufällige Bewegungen (zur Theorie der Brownschen Bewegung)*, Ann. of Math. (2) **35** (1934), no. 1, 116–117. MR 1503147

[KS91]   Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR 1121940 (92h:60127)

[Kuw10]   Kazumasa Kuwada, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), no. 11, 3758–3774. MR 2606871 (2011d:35109)

[Li06]   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)

[Loe48]   Michel Loeve, *Sur l'équivalence asymptotique des lois*, C. R. Acad. Sci. Paris **227** (1948), 1335–1337. MR 0027975

[LPW09]   David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson. MR 2466937

[LY86]   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)

[McK63]   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

[Mon02]   Richard Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, vol. 91, American Mathematical Society, Providence, RI, 2002. MR 1867362 (2002m:53045)

[Neu96]   Daniel Neuenschwander, *Probabilities on the Heisenberg group*, Lecture Notes in Mathematics, vol. 1630, Springer-Verlag, Berlin, 1996, Limit theorems and Brownian motion. MR 1439509

[Pit76]   J. W. Pitman, *On coupling of Markov chains*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **35** (1976), no. 4, 315–322. MR 0415775

[PP16]   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.

[RS94]   S. T. Yau R. Schoen, *Lectures on differential geometry*, International Press, Boston, 1994.

[SS90]   M. Yu. Sverchkov and S. N. Smirnov, *Maximal coupling for processes in $D[0, \infty]$*, Dokl. Akad. Nauk SSSR **311** (1990), no. 5, 1059–1061. MR 1072656

[TL86]   L. Rogers T. Lindvall, *Coupling of multi-dimensional diffusions by reflection*, Ann. Probab. **14** (1986), 860–872.

[vR04]   Max-K. von Renesse, *Intrinsic coupling on Riemannian manifolds and polyhedra*, Electron. J. Probab. **9** (2004), no. 14, 411–435 (electronic). MR 2080605

[vRS05]   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)

[Wan97]   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)

[Wan08]   Limin Wang, *Karhunen-Loeve expansions and their applications*, ProQuest LLC, Ann Arbor, MI, 2008, Thesis (Ph.D.)–London School of Economics and Political Science (United Kingdom). MR 3301633

[Wan14]   Feng-Yu Wang, *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

[Wan16]   _____, *Derivative formulas and Poincaré inequality for Kohn-Laplacian type semigroups*, Sci. China Math. **59** (2016), no. 2, 261–280. MR 3454046

[Yau75]   S.T. Yau, *Harmonic functions on complete riemannian manifolds*, Comm. Pure Appl. Math. **28** (1975), 201–228.

[Yor91]   Marc Yor, *The laws of some Brownian functionals*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), Math. Soc. Japan, Tokyo, 1991, pp. 1105–1112. MR 1159295